

Perceptual Quality Model for Enhancing
Telemedical Video Under Restricted
Communication Channels

PERCEPTUAL QUALITY MODEL FOR ENHANCING
TELEMEDICAL VIDEO UNDER RESTRICTED
COMMUNICATION CHANNELS

BY
TARALYN SCHWERING, B.Eng.

A THESIS
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING
AND THE SCHOOL OF GRADUATE STUDIES
OF MCMASTER UNIVERSITY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF APPLIED SCIENCE

© Copyright by Taralyn Schwering, March 2015

All Rights Reserved

Master of Applied Science (2015)
(Electrical & Computer Engineering)

McMaster University
Hamilton, Ontario, Canada

TITLE: Perceptual Quality Model for Enhancing Telemedical
Video Under Restricted Communication Channels

AUTHOR: Taralyn Schwering
B.Eng., (Electrical and Biomedical Engineering)
McMaster University, Hamilton, Canada

SUPERVISOR: Dr. Thomas E. Doyle, Ph.D., P.Eng. and Dr. David M.
Musson, Ph.D., M.D. (Co-supervisor)

NUMBER OF PAGES: xv, 123

Abstract

Telemedicine has the potential to provide greatly increased access to medical training, consultation and care for individuals and populations located in remote regions of the Earth. In order for this to happen, telemedicine relies on delivery of high quality video over bandwidth limited internet connections. Methods of video encoding to accommodate these connections reduce the perceptual quality of telemedical video, reducing its clinical or educational impact.

This thesis aims to understand how changes in bitrate, frame rate and frame size affect the perceptual quality of telemedical video through studies of objective and perceptual quality of two types of medical simulation video. H.264/AVC encoding was used to encode two medical simulation videos with varying bitrates, frame rates and frame sizes. Objective frame image quality tests and subjective video quality tests were performed on the resulting videos.

It was observed that the objective frame image quality, as measured by the Structural Similarity (SSIM), is linearly related to the number of pixels per bit of the encoded video. It was also determined that perceptual quality of these videos was dependent primarily on the frame rate and frame image quality of the encoded video.

Models are proposed from the results of each type of video showing how perceptual quality can be determined by the parameters chosen to encode a video (maximum

bitrate, frame rate and frame size). The models proposed are unique for the type and purpose of video under investigation as well as the encoding method used. The models created will be useful to encode video for transmission over limited bandwidth networks while maintaining diagnostic and education quality.

Acknowledgements

I would like to thank my supervisors, Dr. Doyle and Dr. Musson for thier knowledge and guidance throughout this process. Recognition is due to Dr. Bingxian Wang and Lisa Bonney at the Centre for Simulation Based Learning for their help with all things simulation. I would also like to acknowledge Ben Kinsella for assisting me with parts of this thesis and fellow graduate students J.J. Booth, Samantha Chan and Omar Boursalie for their comaraderie and friendship over the last two years.

Finally, I would like to thank my family, my parents and my sister Monika. I am grateful for all the support they have given me while working on this thesis.

Notation and abbreviations

σ	standard deviation
ACR	absolute category rating
AVC	advanced video coding
b	bit rate
CABAC	context-adaptive binary arithmetic coding
CAVLC	context-adaptive variable length coding
crf	constant rate factor
df	degrees of freedom
DSCQS	double stimulus continuous quality scale
DSCS	double stimulus comparison scale
DSIS	double stimulus impairment scale
ECG	electrocardiogram
exp-golomb	Exponential-Golomb coding
fps	frames per second
H	Kruskal Wallis test statistic
H.264	H.264/MPEG-4 AVC video compression format
HVS	human visual system
IDR	instantaneous decoding refresh

JVT	joint video team
kbit/s	kilobit per second
MOVIE	MOtion-based Video Integrity Evaluation
MPEG	moving pictures expert group
MPEG-2	MPEG-2/H.262 video compression format
MS	mean square
NAL	network abstraction layer
p-value	significance level
QP (or qp)	quantization parameter
r	frame rate
$R[]$	predicted perceptual video quality rating
R-square	coefficient of determination
RMSE	root mean square error
s	frame size
SNR	signal to noise ratio
SS	sum of squares
SSCQE	single stimulus continuous quality evaluation
SSIM	structural similarity index
SSIM[x]	predicted average frame image quality (SSIM)
SS-HR	single stimulus with hidden reference removal
SVC	scalable video coding
VCEG	video coding experts group
VCL	video coding layer
VQM	video quality metric
x	encoding ratio

Contents

Abstract	iii
Acknowledgements	v
Notation and abbreviations	vi
1 Introduction	1
1.1 Telemedicine	1
1.1.1 Telemedicine for Medical Education or Consultation	2
1.1.2 Delivery of Health to Remote Locations	4
1.2 Video Transmission in Telemedicine	6
1.2.1 Video Compression	7
1.2.2 Video Quality Assessment	9
1.3 Thesis Statement	10
1.4 Organization of Thesis	11
2 Background	13
2.1 Video Encoding	13
2.1.1 H.264/AVC Encoding Format	14

2.1.2	Encoding Parameters	22
2.2	Video Quality Analysis	24
2.2.1	Objective Image Quality Tests	25
2.2.2	Subjective Video Quality Tests	28
3	Experimental Design	33
3.1	Source Video	34
3.2	Test Video Encoding	35
3.3	Subjective Video Quality Test	39
3.3.1	Selected Design	43
3.3.2	Procedure	44
3.3.3	Participants	45
3.3.4	Viewing Conditions	46
3.3.5	Data Collection	46
3.4	Objective Frame Quality Test	46
3.5	Statistical Methods	47
3.5.1	Spearman’s Rank Correlation Coefficient	47
3.5.2	Kruskal-Wallis Analysis of Variance of Ranks	48
3.5.3	Box Plot	49
4	Results	50
4.1	Simulation Room View Results	51
4.1.1	Set 1: Frame Rate	51
4.1.2	Set 2: Bitrate	54
4.1.3	Set 3: Frame Size	57

4.1.4	Set 4: Bitrate and Frame Rate	60
4.1.5	Set 5: Frame Rate and Frame Size	63
4.1.6	Set 6: Bitrate and Frame Size	66
4.2	Glide Scope Video Results	69
4.2.1	Set 1: Frame Rate	69
4.2.2	Set 2: Bitrate	72
4.2.3	Set 3: Frame Size	75
4.2.4	Set 4: Bitrate and Frame Size	78
4.2.5	Set 5: Frame Rate and Frame Size	81
4.2.6	Set 6: Bitrate and Frame Size	84
5	Model Development	88
5.1	Simulation Room View Video	88
5.1.1	Interpretation of Results	88
5.1.2	Encoding Ratio	89
5.1.3	Model 1: Frame Image Quality	90
5.1.4	Model 2: Human Video Quality Rating	93
5.1.5	Composite Model	96
5.2	Glide Scope Video	96
5.2.1	Observations from Previous Data	98
5.2.2	Encoding Ratio	99
5.2.3	Model 1: Frame Image Quality	99
5.2.4	Model 2: Human Video Quality Rating	99
5.2.5	Composite Model	103

6 Discussion	106
6.1 Limitations	106
6.2 Benefits	107
6.3 Uses	107
6.3.1 Example	108
6.4 Further Directions	109
7 Conclusion	110
A Subjective Video Quality Test Instructions	112

List of Figures

2.1	Division of frames into macroblocks and slices [1]	16
2.2	Functions Performed by the Video Coding Layer (VCL)	17
2.3	Relationship between i- p- and b-frames	17
2.4	Transform Coding of the Prediction Residual	19
2.5	Quantization of Transform Coefficients	20
2.6	Entropy Coding of Quantized Transform Coefficients	21
2.7	Network Abstraction Layer	22
2.8	Reference and test images and resulting SSIM index map	27
2.9	Continuous Absolute Quality Scale	31
2.10	Continuous Comparison Scale	31
3.1	Flow chart of experiment processes	34
3.2	Frames from original video of the simulation room. Camera position is stationary while figures in the video move and interact.	36
3.3	Frames from original GlideScope video.	37
3.4	Relative Frame Sizes	38
3.5	Video frame at several levels of degradation due to encoding parameters	41
3.5	Video frame at several levels of degradation due to encoding parameters	42
3.6	Quality Rating Scale Used for Experiment	44

4.1	Room View Perceptual Video Quality Ratings for Set 1	52
4.2	Correlation between average frame image quality and perceptual quality rating for Set 1	53
4.3	Room View Perceptual Video Quality Ratings for Set 2	55
4.4	Correlation between average frame image quality and perceptual quality rating for Set 2	56
4.5	Room View Perceptual Video Quality Ratings for Set 3	58
4.6	Correlation between average frame image quality and perceptual quality rating for Set 3	59
4.7	Room View Perceptual Video Quality Ratings for Set 4	61
4.8	Correlation between average frame image quality and perceptual quality rating for Set 4	62
4.9	Room View Perceptual Video Quality Ratings for Set 5	64
4.10	Correlation between average frame image quality and perceptual quality rating for Set 5	65
4.11	Room View Perceptual Video Quality Ratings for Set 6	67
4.12	Correlation between average frame image quality and perceptual quality rating for Set 6	68
4.13	Glide Scope Perceptual Video Quality Ratings for Set 1	70
4.14	Correlation between average frame image quality and perceptual quality rating for Set 1	71
4.15	Glide Scope Perceptual Video Quality Ratings for Set 2	73
4.16	Correlation between average frame image quality and perceptual quality rating for Set 2	74

4.17	Glide Scope Perceptual Video Quality Ratings for Set 3	76
4.18	Correlation between average frame image quality and perceptual quality rating for Set 3	77
4.19	Glide Scope Perceptual Video Quality Ratings for Set 4	79
4.20	Correlation between average frame image quality and perceptual quality rating for Set 4	80
4.21	Glide Scope Perceptual Video Quality Ratings for Set 5	82
4.22	Correlation between average frame image quality and perceptual quality rating for Set 5	83
4.23	Glide Scope Perceptual Video Quality Ratings for Set 6	86
4.24	Correlation between average frame image quality and perceptual quality rating for Set 6	87
5.1	Alternative Models for Frame Image Quality based off Objective Frame Quality Results	91
5.2	Model for Frame Image Quality based off Objective Frame Quality Results	92
5.3	Plot of perceptual video quality ratings; minimum, maximum and mean scores.	94
5.4	Surface plot of perceptual video quality model along with plot of actual minimum, maximum and mean ratings.	95
5.5	Surface plot of perceptual video quality model as a function of bit rate, frame rate and frame size along with plot of actual minimum, maximum and mean ratings.	97

5.6	Alternative Models for Frame Image Quality based off Objective Frame Quality Results	100
5.7	Model for Frame Image Quality based off Objective Frame Quality Results	101
5.8	Plot of perceptual video quality ratings; minimum, maximum and mean scores.	102
5.9	Surface plot of perceptual video quality model along with plot of actual minimum, maximum and mean ratings.	103
5.10	Surface plot of perceptual video quality model as a function of bit rate, frame rate and frame size along with plot of actual minimum, maximum and mean ratings.	105
A.1	114
A.2	115
A.3	115

Chapter 1

Introduction

1.1 Telemedicine

Telemedicine is defined by many as the use of telecommunication technology to deliver healthcare services or share medical knowledge, often over long distances [2–5]. While the proliferation of the internet in recent decades has widely expanded the field of telemedicine, examples exist as far back as 1959 when two-way video was used for group therapy sessions [5]. In 1965, satellite communication was used for the first time to broadcast an open-heart surgery from North America to Europe. It included a 2-way television link to allow participants to discuss the procedure as it occurred [6]. Since then, early telemedicine has developed through the use of telephones and fax machines and later email and video conferencing for remote clinical consultations [3].

The motivations behind the use of telemedicine, which have remained the same throughout the development of the field, are based on removing physical barriers and increasing availability of information required for the delivery of medical care [5].

Examples include:

- Assistance to local healthcare providers in cases of increased and unplanned need of medical services due to disaster [4];
- Supplement health care services in understaffed areas such as rural health centres [3, 7];
- Provide health care services to individuals with no access due to remote location such as oil rig workers, travellers on ships and planes, remote research locations [2, 8];
- Increase efficiency of health care processes through better data communication in the form of home monitoring and emergency ambulance services [2, 3].

The field of telemedicine has grown to become very diverse, however this thesis will only cover topics in telemedicine which directly relate to it. Of particular interest are the following topics: delivering healthcare to remote areas; medical education and consultation between healthcare providers; and medical video and data transmission. Topics that are not covered include: chronic disease; remote patient monitoring; telemedicine applications for therapy or psychology; virtual reality for rehabilitation; and others.

1.1.1 Telemedicine for Medical Education or Consultation

There are various ways in which the use of telemedicine can be used for different aspects of the education of health care providers including; consultation and collaboration between healthcare providers, formal lectures or courses delivered to students through telemedicine and skills mentoring or remote simulation based learning.

The use of telemedicine for consultation between medical providers has been shown to be beneficial to the health care teams and the patients involved [9]. Multidisciplinary teams dispersed over large areas are able to meet to discuss patients and cases with the use of teleconferencing [10]. In cases where health care is being provided at an isolated centre or by nurse practitioners, telemedicine allows for consultation and collaboration with specialist teams located at central hospitals [11] and the use of video for this consultation helps to improve the communication between the sites and to foster relationships [9].

Traditional lectures presented in an electronic format either one way-video or an interactive session with real-time feedback and collaboration between participants can allow healthcare providers access to training that would otherwise not be available [9]. In [12], a pediatric resuscitation course was provided remotely to doctors in Iraq by trained instructors in Florida. The course, which was delivered using teleconferencing links between the instructors and participants, was implemented to address the need for improved emergency medical care in Iraq. Tele-education provided a safe and inexpensive means of addressing the needs of the region without compromising the safety of the instructors involved.

Less traditional teaching scenarios such as one-on-one mentoring can benefit from telemedicine, especially in the case of skills learning. Often used for teaching laparoscopic surgical skills, telementoring involves real-time feedback from a mentor in a different physical location than the surgeon performing the operation [13]. While research on the effects of telementoring is still limited, it has been shown to have similar success rates to in-person mentoring for laparoscopic surgery and is predicted to be useful for meeting increased demand for surgical education [10].

Of particular interest to this thesis is the use of telemedicine for providing simulation based medical education involving either remotely located instructors or facilitators, simulation scenarios which take place away from traditional simulation or both. The use of telemedicine for providing portable or remote simulation is motivated by a desire for increased access to simulation based learning [14–18]. Examples of use include instruction using low fidelity simulation, interactive lessons involving simulated patients and full scale medical simulation (such as emergency, anesthesia or surgery) located outside a simulation centre or with the instructor connected to the scenario remotely [13, 15–17, 19, 20].

Similar to other types of telemedicine, the widespread adoption of teleconsultation and telementoring is hindered by technical limitations such as limited bandwidth [13], reduced sound and video quality [12], and human limitations such as required training for new technologies [9]. Despite the benefits of telemedicine for education and consultation between healthcare providers, due to its relative inconvenience compared with traditional methods, teleconsultation is only used in cases where it is necessary due to physical barriers [9].

1.1.2 Delivery of Health to Remote Locations

A big benefit of telemedicine is the ability to deliver healthcare to remote populations with little or no healthcare providers. The previous section on consultation between healthcare providers discusses the uses of telemedicine for assisting regional healthcare providers, such as a nurse practitioner in a rural area, with dealing with difficult medical cases. This section describes scenarios in which telemedicine is used for the benefit of remote populations and the technology used in those situations.

Many remote locations are restricted by the internet connection types available. Remote research stations, off shore oil and gas operations and travellers on ships are all restricted to the use of satellite communication for access to the internet. A signal is required to travel from a location on earth, to the satellite in orbit and back down to earth in another location. Because of this, signals are highly dependent on location and weather and in ideal cases speeds of 1 Mbit/s and delays of 250 ms or more can be expected [21, 22]. Conditions measured in Eureka Nunavut found that the available bandwidth and latency was highly variable with typical speeds of 340 Kbit/s upload and 120 Kbit/s download and a latency of 680 ms [23].

There are many examples of telemedicine being used to provide medical services to support offshore workers in the oil and gas industry. Traditionally, the medical needs of workers of offshore operations were taken care of by medics with limited medical training. These medics could range from non-medical staff with some emergency medical training [24] to paramedics or nurses with emergency medical training [8].

These personnel are required to deal with a wide range of medical injuries and illnesses; common examples include dental problems, sprains and strains, fractures, digestive issues and respiratory illness [8]. Due to the infrequency and variety of the cases encountered by these individuals, even trained personnel undergo skills degradation [8, 24]. Telemedicine is used in these situations to provide supervision and support to onboard medical personnel by trained doctors onshore.

In the early 1990s, [25] tested the provision of telemedical services via satellite for travellers on aircraft and ships. Information sent included colour images, audio, 3 channel ECG and blood pressure and special consideration was taken to limit the size of the data being sent via satellite.

The MERMAID project as described in [26], is a global 24-h multilingual telemedicine surveillance and emergency service for both maritime and remote land-based services. The project was very ambitious in that it hoped to operate over different types of maritime telecommunication systems to provide an autonomous guidance system for paramedics, teleconsultation capability including interactive sound and video as well as a medical records system.

More recently, full telemedical service is provided to full-time maritime operations such as offshore oil rigs by a number of different companies [27]. These setups, available at a cost to the oil company, include diagnostic tools (such as ECG or endoscope), video conferencing equipment, medical supplies and medications as well as the services of medical personnel [27]. Medical care is still provided by an onboard medic under the supervision of an onshore medical advisor who is able to assess the patient remotely [8]. The onshore doctor is able to provide both direct medical interpretations of symptoms as well as shared clinical management of the patient with the onboard provider [8]. The use of telemedicine systems described here has enhanced patient care and peace of mind by providing direct access to medical doctors as well as reduced costs and risks through improved patient triage and reduce the instance of unnecessary evacuations [27].

1.2 Video Transmission in Telemedicine

All telemedicine systems involve transmitting medical data over telecommunication networks. This data could be raw signals (such as an ECG or blood pressure signal), audio, video or medical images. Video transmission has advantages over audio and still images for transmitting medically relevant information in real time.

As mentioned in the previous sections, telemedicine has great potential for removing barriers of medical treatment for individuals in remote or underpopulated area. A major obstacle to the successful implementation of telemedicine in these areas however is the lack of high speed internet connections required to accommodate high quality video. Additionally, the conditions required for medical video differs depending on the content and context of the video.

Current research in developing telemedicine focusses on video compression technology for medical video and to meet the specific needs of telemedicine users. Additionally, new standards for assessing the quality of medical video through diagnostic validation are required to test the effectiveness of telemedical solutions.

1.2.1 Video Compression

Several video compression techniques exist that are used for the compression of medical video in order to accommodate high quality medical video over a variety of network connections including low bandwidth connections.

The H.264/AVC video compression standard has been around since 2003 and was developed jointly by the Moving Pictures Expert Group (MPEG) and the ITU-T Video Coding Experts Group (VCEG) to improve video compression performance for video to be sent over networks such as conversational services and video on demand [1]. It provides improvements over previous standards such as MPEG-2 due to improved prediction methods, better encoding efficiency, robustness to data errors and flexibility over a variety of networks [1]. Most modern telemedical solutions make use of these benefits of the H.264 encoding standard in order to provide high quality medical video [28].

Additional encoding techniques can be applied in order to increase encoding efficiency and diagnostic quality of medical video. Region of interest encoding allows for different levels of compression to be applied to regions of the same image or video. A region of interest selected by a medical expert or through an algorithm is encoded with a higher allocation of bits to pixels than background regions of the image [29,30]. The goal is to achieve diagnostically lossless compression, a term applied to compressed video which retains necessary medical content [29]. For example, medical images or videos from ultrasound or CT can be encoded such that different anatomical features are encoded to different quantization (or compression) levels depending on their level of importance to the medical expert or viewer [29,30].

Another method for medical video compression is Scalable Video Coding or SVC. SVC allows efficient coding and sending of video streams for multiple viewers with different resources and viewing capabilities. Video streams can be scaled up or down to adapt to network conditions, user preferences or terminal capabilities [31]. Scalability can refer to: temporal scalability or changes in frame rate; spatial scalability or changes in resolution; quality resolution or changes in quantization; or in some cases region of interest scalability where regions of the image are scaled differently [31,32]. Generally a base layer is encoded to H.264 standards with additional enhancement layers available to increase the video quality when sufficient bandwidth is available [32].

1.2.2 Video Quality Assessment

Objective Assessment

Several techniques exist for the evaluation of video quality without the costly involvement of human rating processing. The most basic means of video quality evaluation is through Peak Signal to Noise Ratio, which measures the mean square error between an original and a distorted video [33]. The SSIM method for video quality assessment attempts to quantify the perceived visual distortion of images and videos through measures of the structural similarity between the reference and distorted samples [34].

Furthermore, the Video Quality Metric (VQM) and Motion-based Video Integrity Evaluation (MOVIE) were developed to further attempt to approximate human perceptual video quality [33].

Beyond these metrics, there exists a strong need for video-quality metrics that are clinically driven in order to evaluate the quality of video in telemedicine solutions [30]. Existing objective methods fail to assess the diagnostic content of telemedical video [35]. Additionally, different medical modalities have different diagnostic requirements for telemedical video [35].

Subjective Assessment

In order to better assess the diagnostic content of telemedicine videos, subjective video assessment tools have been used and modified to apply to medical videos. In particular these methods aim to classify medical videos as diagnostically lossless, meaning that all information needed for medical assessment is retained in a compressed or degraded video [29].

Traditional subjective video assessment tools (which will be further discussed in Chapter 2) typically involve subjects rating the perceived quality of video samples on a scale from 1 to 5 where 1 represents bad quality and 5 represents excellent quality. When applied to medical videos, subjective assessment is typically performed by medical experts and rating is done on the diagnostic content of the video rather than the perceived quality [36, 37]. Because of the different diagnostic features required by different medical modalities, some studies involve rating video for each diagnostic feature or criteria separately [29, 30].

This method of video assessment is much more costly and takes significantly more time than objective methods, especially for a high number of videos. Hybrid methods involving the use of subjective assessment to validate objective assessment methods are useful in reducing the time and cost of subjective assessment while retaining the focus on diagnostic video quality.

Typically hybrid methods of video quality assessment involve obtaining the statistical correlation between objective and subjective methods in order to decide which objective method give the best diagnostic quality evaluation [36]. Panayides et. al. employed this method for Atherosclerotic Plaque ultrasound video and found that weighted SNR for the diagnostic region of interest had the highest correlation to diagnostic scores for the video [30].

1.3 Thesis Statement

Telemedicine has the potential to provide greatly increased access to medical training, consultation and care for individuals and populations located in remote regions of the Earth. In order for this to happen, telemedicine relies on delivery of high quality

video over bandwidth limited internet connections. Methods of encoding videos to accommodate these connections reduce the resolution, frame rate, image quality of the transmitted image or some combination of the three in order to reduce the video file size or bitrate. This process reduces the quality arbitrarily in order to accommodate these connections. A more effective encoding method would be able to take into account the unique requirements for telemedicine to better retain the diagnostic or educational quality of the video at lower bitrates. This thesis aims to understand how changes in bitrate, frame rate and frame size affect the perceptual quality of telemedical video through studies of the perceptual quality of two types of medical simulation video. Results of these studies lead to the proposal that the perceptual quality of telemedical video is dependent on a combination of the frame image quality and frame rate of encoded video. Furthermore, models are proposed for each of the medical simulation videos which illustrate how the quality of these videos is directly influenced by their encoding. The proposed models are unique to the exact type of video they were created for and have the potential improve telemedicine through encoding processes which retain the diagnostic quality of videos at reduced bitrates.

1.4 Organization of Thesis

This thesis is organized as follows.

Chapter 2 presents details on the H.264/AVC encoding standard which is used often in telemedicine and will be employed in the experiment contained in this thesis. Also in chapter 2 is information on several established methods for assessing video quality; both objective methods to measure frame image quality and subjective methods which will be used to quantify the perceptual quality of a video.

Chapter 3 contains the particulars of the experimental design of the video quality assessment undertaken for this thesis including the selected videos for study; methods and variations in encoding of these videos; chosen procedures for both the subjective and objective video quality tests; and statistical methods used for analysis.

Chapter 4 comprises the results of these tests for both sample videos; the average frame image quality and the perceptual video quality ratings are presented along with the results of statistical analysis of the data.

Chapter 5 includes both interpretation of the results contained in chapter 4 and several models which are based off these interpretations. Separate but related models are developed for each of the sample videos under investigation.

Chapter 6 presents a discussion of the results and models found in this thesis including: the limitations and advantages of the models created; how the models could be used to improve telemedicine; and the implications for future research.

Chapter 7 summarizes the contents and concludes the thesis.

Chapter 2

Background

This thesis aims to understand how encoding and file size (or bitrate) affect the diagnostic and educational quality of medical simulation video. To this end, it proposes that quality depends primarily on frame image quality and frame rate.

This chapter covers background information on the encoding and video quality assessment processes employed in the thesis. Of the encoding process it is to be understood what happens to the video as it is encoded and how the process can be controlled by encoding parameters. Two video quality assessment methods are covered: objective assessment of frame image quality and subjective assessment of perceptual video quality.

2.1 Video Encoding

Video encoding methods exist to reduce the file size of videos for the purpose of storage or transmission. The H.264/AVC encoding format was developed in 2003 by the Joint Video Team (JVT) made up of both the Video Coding Experts Group (VCEG) and

the Moving Pictures Expert Group (MPEG) [1]. The purpose of H.264/AVC was to improve encoding efficiency (smaller bit rate for the same level of fidelity) over existing formats including the MPEG-2 video standard which was commonly used for the transmission of video over cable and satellite and for storage on DVD.

When the encoding efficiency of H.264/AVC was compared to previous video coding standards (MPEG-2, H.263, and MPEG-4), it was found to have greatly increased efficiency [38]. PSNR measurements show an average bitrate savings of 37.44% for video streaming and 27.37% for low delay video streaming over MPEG-4 coding [38]. These results were found to be conservative when they were validated using informal subjective tests [38].

Transmission of high-definition video over internet connections and storage on Blu-Ray Discs requires higher encoding efficiency than provided by earlier encoding standards. H.264/AVC has been shown to achieve the required reduction in bitrate for comparable video quality and is also compatible with many transport layers or storage media types. H.264/AVC is standardized through restrictions on the encoded format and the decoding process, allowing flexibility in the implementation and optimization of encoders [1].

2.1.1 H.264/AVC Encoding Format

H.264/AVC encoding consist of 2 parts: the video coding layer (VCL) responsible for encoding and compressing a video; and the network abstraction layer (NAL) which packages the encoded video for transmission or storage [1].

Video Coding Layer

The video coding layer includes several steps to convert a video input signal into an encoded bitstream. These steps are prediction, transform, quantization and bitstream encoding as shown in Figure 2.2 [39].

Input video signals are split into 16x16 pixel blocks of the frame called macroblocks to be processed by the video coding layer [1]. Continuous groups or sequences of macroblocks within a picture are referred to as slices. Each slice is self-contained, meaning it can be decoded independently of other slices [1]. Flexible macroblock ordering is a concept that refers to alternative arrangements of slices within a picture based on the desired encoding (rather than continuous blocks). Slice groups (composed of smaller continuous slices) are used to group areas of the picture when encoding as shown in Figure 2.1 [1].

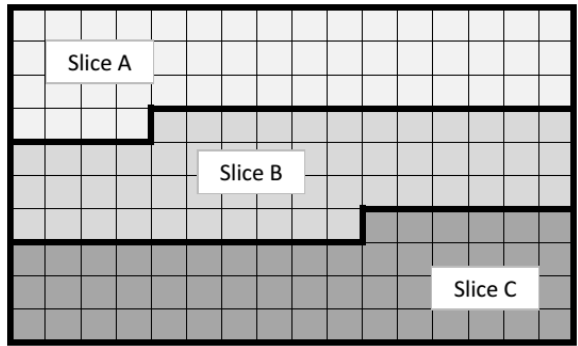
Slices can be classified by the type of prediction used for encoding of that frame [1]:

i-slice Intra-slice prediction is used to predict the luma and chroma of each pixel in a macroblock based only on previously encoded macroblocks in the same slice or frame.

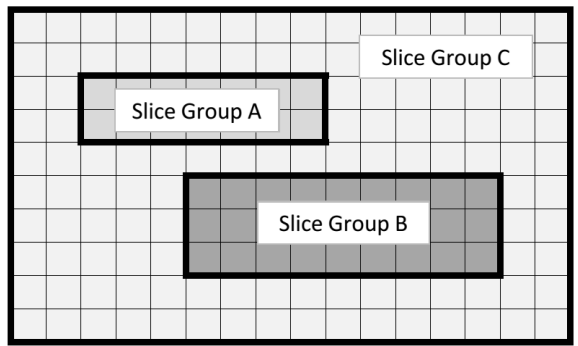
p-slice Inter-slice prediction is used to predict each pixel based on macroblocks in a previous slice or frame.

b-slice Bi-directional inter-slice prediction is used to predict each pixel based on macroblocks in a previous and a future slice or frame.

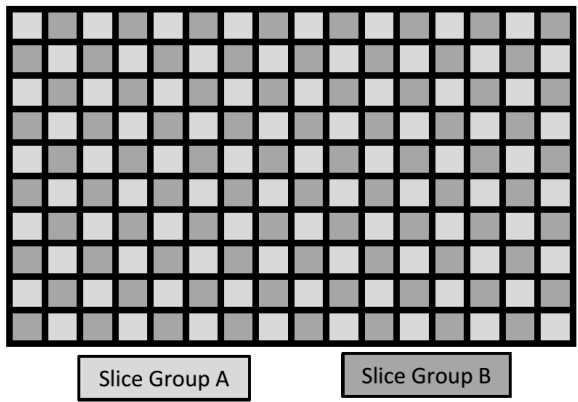
Figure 2.3 shows the relationship between i-, p- and b- frames and off of which frames they are predicted. P-frames are predicted using only the preceding i-frame b-frames are predicted using both the preceding i- or p- frame and the subsequent p- frame.



(a) Continuous slices



(b) Region-of-interest slice groups



(c) Checkerboard slice groups

Figure 2.1: Division of frames into macroblocks and slices [1]

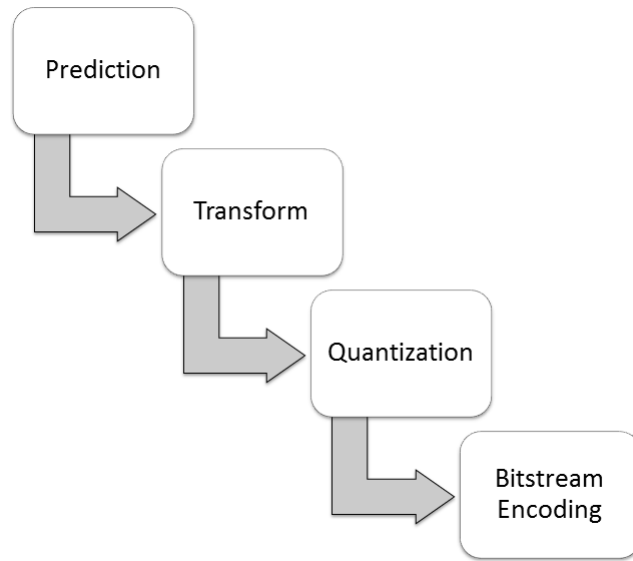


Figure 2.2: Functions Performed by the Video Coding Layer (VCL)

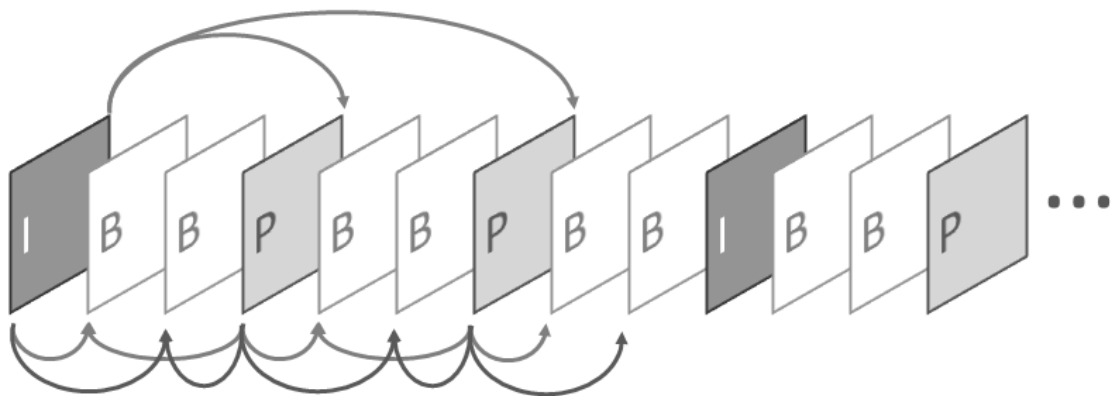


Figure 2.3: Relationship between i- p- and b-frames

Prediction

The first step performed by the VCL is prediction. Two types of prediction are used for different slice types.

Intra-slice prediction estimates pixel values using prediction modes in the spatial domain. Several prediction modes exist suited to different structures and details in the frame. 4x4 modes predict each 4x4 pixel block separately and are suited to areas with high amounts of detail while 16x16 modes are suitable for smooth areas with less detail [1].

Inter-slice prediction occurs within the time domain and involves using motion vectors to predict the content of frames based off previous or future frames [1].

The result of this process is a predicted frame that may differ significantly from the actual frame. This difference, known as the prediction residual, can also be viewed as the error of the prediction process. Since the prediction process can be recreated at the decoder, only this error needs to be retained in order to reconstruct the original frame.

Transform Coding

In order to encode the prediction residual of each macroblock, it is transformed into the (spatial) frequency domain using a 4x4 integer transform with similar properties to the discrete cosine transform [1].

As shown in Figure 2.4, the result is a set of transform coefficients that can be used to recreate the prediction residual through the inverse transform.

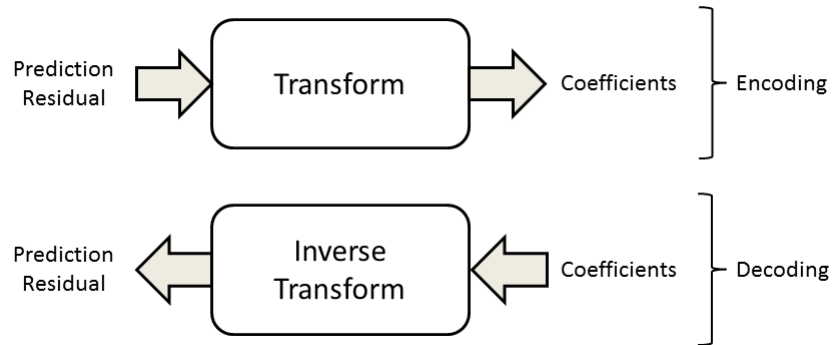


Figure 2.4: Transform Coding of the Prediction Residual

Quantization

Quantization refers to the process of mapping the full set of transform to a smaller discrete set of values. This process reduces the precision of the coefficients and allows users to control the video quality to file size ratio via the quantization parameter. The quantization parameter (QP) can take on values of 0-51 where higher values lead to higher amounts of quantization [1]. For example a high QP allows coefficients to be mapped to a smaller set of discrete values than a small QP and usually results in many of the coefficients with lower values going to zero after quantization [39]. While this may negatively affect the video quality it also reduces the final file size, leading to a trade- off of quality for size,

A quantization parameter of 0 results in no quantization and effectively no loss in video quality. While the video is still encoded with H.264/AVC, it can be considered lossless since no precision is lost and it can be decoded to the same original quality.

The inverse process of quantization occurring at the decoder is scaling. The coefficients are scaled back up to their original values but the loss of precision through quantization cannot be reclaimed.

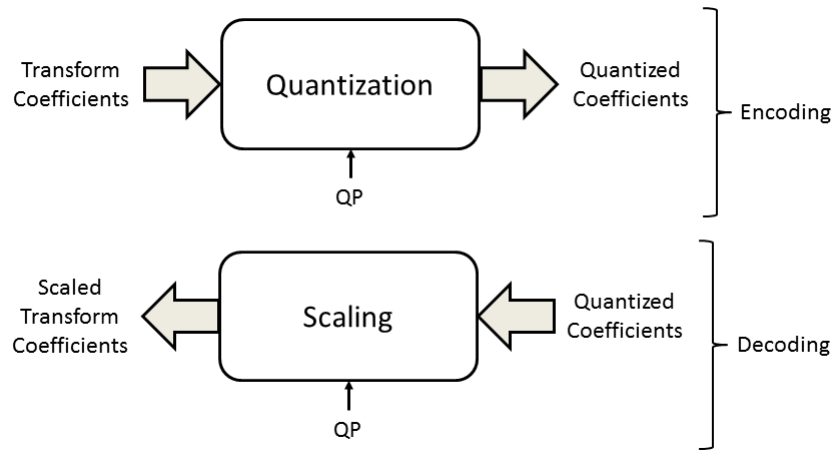


Figure 2.5: Quantization of Transform Coefficients

Entropy Coding

Quantized coefficients undergo entropy encoding to compress the data further in a lossless manner so it is suitable for storage or transmission. Two types of entropy coding are used, one for encoding of quantized coefficients and another for encoding all other data including syntax elements. Exponential-Golomb coding (exp-golomb) is used for encoding syntax elements [1]. This encoding method is a single infinite extent code word method capable of universally encoding all syntax elements.

Context-Adaptive Variable Length Coding (CAVLC) and Context- Adaptive Binary Arithmetic Coding (CABAC) are more efficient encoding methods for encoding the quantized coefficients. These methods take advantage of the high number of trailing zeroes to encode the data more efficiently than previous encoding methods [1]. CABAC has higher encoding efficiency than CAVLC but also requires more processing and is not universally supported by decoders [1].

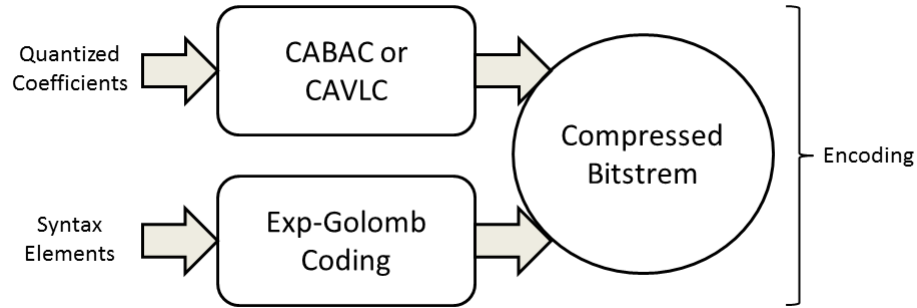


Figure 2.6: Entropy Coding of Quantized Transform Coefficients

Network Abstraction Layer

As mentioned previously, the job of formatting and packaging the above bitstream for transport or storage falls under the domain of the network abstraction layer. Figure 2.7 shows the units and sequences that the encoded bitstream consists of.

The smallest packet of information dealt with by the network abstraction layer is called a NAL unit. A NAL unit consists of an integer number of bytes and contains both a header to identify the contents of the unit and a payload of the contents [1]. NAL units can be classified into two categories:

- VCL NAL units which contain coded data samples of video pictures and
- Non-VCL NAL units which contain additional information on top of the coded data including parameter sets (which will be discussed later).

NAL units can be arranged in either byte-stream format or packet format depending on the requirements for the transport or storage system chosen by the user.

A series of NAL units which together make up one coded picture is referred to as an *access unit*. The access unit will contain the coded data of the frame as well as supplementary information such as timing information and redundant coded information

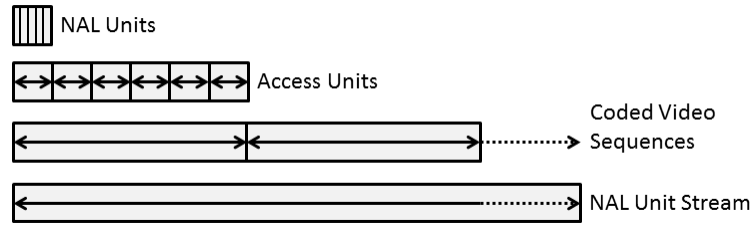


Figure 2.7: Network Abstraction Layer

for that single coded picture [1].

A set of access units that can be decoded independently into a video sequence is collectively known as a *coded video sequence*. Each coded video sequence begins with an *instantaneous decoding refresh* (IDR) access unit containing a single i-frame. This frame and all subsequent frames do not rely on any frames in earlier video sequences to decode hence each sequence is independently coded [1].

Parameter sets are sets of additional encoding information required for decoding a coded picture or video sequence. Sets which apply to an entire video sequence are termed *sequence parameter sets* while *picture parameter sets* apply to the decoding of one or more pictures within a sequence [1].

2.1.2 Encoding Parameters

All H.264/AVC encoding and transcoding of videos for this project was done using the FFmpeg software tool and the x264 video codec included with libavcodec library [40]. FFmpeg and all codecs used for this project are freely available under the GNU General Public License.

The x264 video codec allows the user to specify several encoding settings or parameters to customize properties of the encoded video. This section will outline some

of these parameters which were used for this thesis.

Frame Rate

The frame rate at which a video is recorded affects the raw file size of the video. Reducing the frame rate by a factor of 2 (for example from 30 to 15 frames per second) results in half of the frames being omitted from the encoded video and a significant reduction in file size. This is an effective way of reducing file size or video bitrate without affecting other video encoding options.

Frame Size

Similar to frame rate, changing the frame size will also affect the raw file size of the video. Specifying a frame size smaller than the original causes fewer pixels in each frame to be included in the encoded video and results in a smaller file size.

While frame size and resolution are closely related, resolution refers to the number of pixels per square inch when represented on screen or in print. For the experiment described in this thesis all videos were viewed at the same resolution but at different sizes so the term frame size more accurately describes what is being investigated in this thesis. If all the images were scaled up to be viewed as full screen on the same size monitor, then resolution would be a more accurate description.

Bitrate

The encoding efficiency is controlled by the amount of quantization in an encoded video which can be directly or indirectly controlled by users through the x264 settings for constant rate factor (crf) and quantization parameter (qp) [41]. By doing this, the

encoder targets the same quality level for all frames and the video bitrate is adjusted to meet these needs.

Encoding efficiency can also be controlled through bitrate settings which will target a specific video bitrate and apply the best possible quantization to achieve the desired file size [41]. While quantization of the encoded video is not directly known, the video bitrate is known which can be much more useful for deciding the required bandwidth for video streaming applications.

Preset

The preset option allows users to specify a ratio of encoding speed to file size. Some applications such as live streaming or video conferencing may require fast encoding speed at the expense of lower quality or larger file size. Applications with less stringent time restraints can afford to spend more time on encoding resulting in more efficient encoding (higher quality for the same file size or smaller file size for the same quality level).

2.2 Video Quality Analysis

This section reviews the required information on the chosen method for assessing frame image quality, the Structural Similarity Index, as well as background information pertaining to perceptual video quality assessment and the requirements for diagnostic and educational quality in medical simulation video.

2.2.1 Objective Image Quality Tests

This thesis proposes that perceptual video quality depends on both frame image quality and frame rate. In order to accomplish this, an objective method is needed for assessing the image quality of the individual frames of encoded test videos. Several full-reference methods exist for measuring the quality of an image by comparing it with a reference image of perfect quality. The two most common methods are PSNR (peak signal to noise ratio) and SSIM (structural similarity). The PSNR is calculated based on the amount of error present in an image. While it is simple to calculate it can be a poor reflection of human opinion of image quality [42]. For this reason, SSIM, which is better than PSNR at approximating the image quality as perceived by humans, was selected as the method for assessing frame image quality in this thesis.

Structural Similarity Index

The Structural Similarity (SSIM) index was proposed in 2004 by Wang as a full-reference method to assess image quality based off the idea that perceived image quality is related to the amount of structural distortion present in the image [34]. Alternative methods of image quality assessment, deal with measures of error and either the magnitude of that error or an approximation of how that error is perceived by the Human Visual System (HVS) based off the results of psychovisual experiments [34].

In contrast to these approaches, which deal with the individual functions of the HVS, the SSIM metric assumes that the overall function of the HVS is to recognise structures in the visual field [34]. To this end, the SSIM metric is a measure of the structural distortions of an image in contrast to an original (hence it is a full reference method).

The equation for the structural similarity between 2 signals x and y is given by:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (2.1)$$

where

- μ_x is the mean of signal x ,
- μ_y is the mean of signal y ,
- σ_x^2 is the variance of x ,
- σ_y^2 is the variances y and
- σ_{xy} is the covariance of x and y .

Constants c_1 and c_2 are added to stabilize the SSIM value when $(\mu_x^2 + \mu_y^2)$ or $(\sigma_x^2 + \sigma_y^2)$ are close to zero [34].

The index value in equation 2.1 is calculated for 8x8 pixel windows over the entire image. Each window is represented by the resulting index value on an SSIM index map of the image. Figure 2.8c shows an example of this SSIM map for the reference and test images in Figures 2.8b and 2.8b. Areas of the test image with high structural similarity to the reference are white and the darker areas represent areas of the image with higher structural distortion. The values can also be averaged over the entire image to give the mean SSIM for the image [34]. The SSIM value for the image in Figure 2.8 was calculated to be 0.9105.

This thesis makes use of the SSIM function included in the Image Processing Toolbox of the 2014a release of Matlab [43].



(a) Reference image



(b) Test image



(c) SSIM index map

Figure 2.8: Reference and test images and resulting SSIM index map

2.2.2 Subjective Video Quality Tests

The International Telecommunications Union has published several recommendations on the methods for subjective video quality assessment for broadcast television and multimedia applications [44,45]. The purpose of the recommended tests is to validate the overall video quality for multimedia applications, although they are also used for the validation of objective video quality assessment methods [45]. Within these recommendations, several of the testing methods include [44,45]:

- Double Stimulus Impairment Scale (DSIS)
- Double Stimulus Comparison Scale (DSCS)
- Double Stimulus Continuous Quality Scale(DSCQS)
- Absolute Category Rating (ACR)
- Single Stimulus Continuous Quality Evaluation (SSCQE)
- Single Stimulus with Hidden Reference Removal (SS-HR).

These methods differ from each other based on three distinct elements: single vs double stimulus design; scale type used; and scale descriptors used.

Double stimulus methods involve either direct or indirect quality comparisons between two video samples. In contrast to this are single stimulus methods which involve ratings of individual video samples. Single stimulus with hidden reference removal refers to the inclusion of a reference video within the video samples during a single stimulus test which is used for indirect comparison between the reference and the test videos.

Two different scale types are used for subjective quality tests: *ordinal* and *continuous* scales. An ordinal (or ordered categorical) scale consists of a discrete number of ordered responses as shown in Tables 2.1, 2.2 and 2.3 [44,46]. The values assigned to each category have no mathematical relation to the ratings except to indicate the order of the categories.

Continuous scales employ the same category descriptions used in ordinal scales but allow for continuous ratings along a line as shown in Figures 2.9 and 2.10 [44,46]. The selection along the line is converted to a value between 0 and 100 based on the distance along the line. Similar to the ordinal scale, the actual value of the rating on a continuous scale has little mathematical relation to the quality of the video; for example, a video with twice the quality rating is not necessarily twice as good [46]. Svensson found that the results from ordinal and continuous scales are consistent in order and that they may be used interchangeably when considering the order or rank of the results [46]. It was also noted that continuous scales provide greater freedom of choice to the rater [46].

The third way that these scales can be defined is by the descriptors or category used. Table 2.1 depicts an impairment scale used to rate the degradation discernable between two video samples [44, 45]. Comparison scales as shown in Table 2.2 and Figure 2.10 require the rater to distinguish the preferred video clip between two samples. Absolute scales can be used for both single or double stimulus tests as they involve only the rating the quality of a video clip in absolute terms as shown in Table 2.3 and Figure 2.9 [44, 45].

The subjective video quality assessment methods mentioned above all have strengths and weaknesses that make some more suited for a study than others. For example,

Score	Assessment
5	Imperceptible
4	Perceptible but not annoying
3	Slightly Annoying
2	Annoying
1	Bad

Table 2.1: Impairment Scale

Score	Assessment
-3	Much worse
-2	Worse
-1	Slightly worse
0	Same
1	Slightly better
2	Better
3	Much better

Table 2.2: Comparison Scale

Score	Assessment
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

Table 2.3: Absolute Category Scale

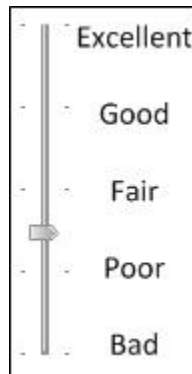


Figure 2.9: Continuous Absolute Quality Scale

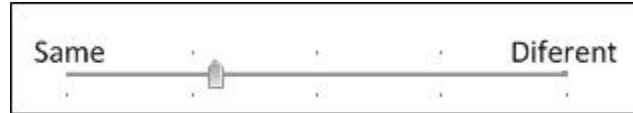


Figure 2.10: Continuous Comparison Scale

impairment methods are useful for testing high quality systems where detection of minor impairments by viewers is important [45]. Comparison methods are also useful for fine discrimination between the quality of videos [45]. In general, double stimulus methods give more accurate results than single stimulus methods, but do so with the added drawback that they take much longer to perform [47].

One reason for the better performance of double stimulus methods is the reduction of context effects when viewers are shown the reference sequence along with the test sequence. Context effects refer to effects in the results of an experiment due to the ordering or intensity of test sequence impairments during an experiment [47].

Hidden reference removal may be an effective way of reducing context effects in a single stimulus design, since it removes any bias in test scores due to the content or quality of the reference video or bias of the viewer [45]. Pinson and Wolf found that the results from a single stimulus method (SSCQE) with hidden reference removal

were consistent with the results from double stimulus methods of quality assessment (DSCQS and DSCS) [47]. This establishes that single stimulus methods can be as effective as double stimulus methods of subjective video quality assessment, especially when time is a limiting factor in the experimental design.

Pinson and Wolf also found that the effect of past test sequences on current quality ratings is minimized after 8-15 seconds [47].

Seshadrinathan et al. performed a large scale subjective study of video quality using a single stimulus method adapted from the above guidelines. The study employed single stimulus assessment with hidden reference removal and a continuous absolute quality scale [33]. The use of hidden reference removal converted the results from absolute quality ratings to difference quality ratings between the reference and the test sequences. As well, the use of a continuous scale (including the same descriptors used in a categorical scale) allowed for subtle differences in the quality ratings of similar test sequences [33].

Chapter 3

Experimental Design

The purpose of this study is to determine the effects of changes in encoding frame size and frame rate on the image quality and perceptual quality of simulation video. The aim is to understand the ideal parameters for H.264/AVC encoding of simulation video in bandwidth limited situations. This was accomplished by measuring the frame image quality and perceptual video quality of simulation videos encoded at various H.264/AVC settings.

The flow chart in Figure 3.1 shows the steps that were involved in this experiment. Section 3.1 discusses the simulation videos that were chosen as the source video for this experiment. Section 3.2 discusses the creation of encoded test sets from the original videos through the various H.264/AVC encoding settings. Section 3.3 introduces the subjective video quality test that was used for this experiment as well as particulars of the test participants. Section 3.4 discusses how objective image quality assessment was used for this experiment. Finally, Section 3.5 presents the statistical methods used for this experiment.

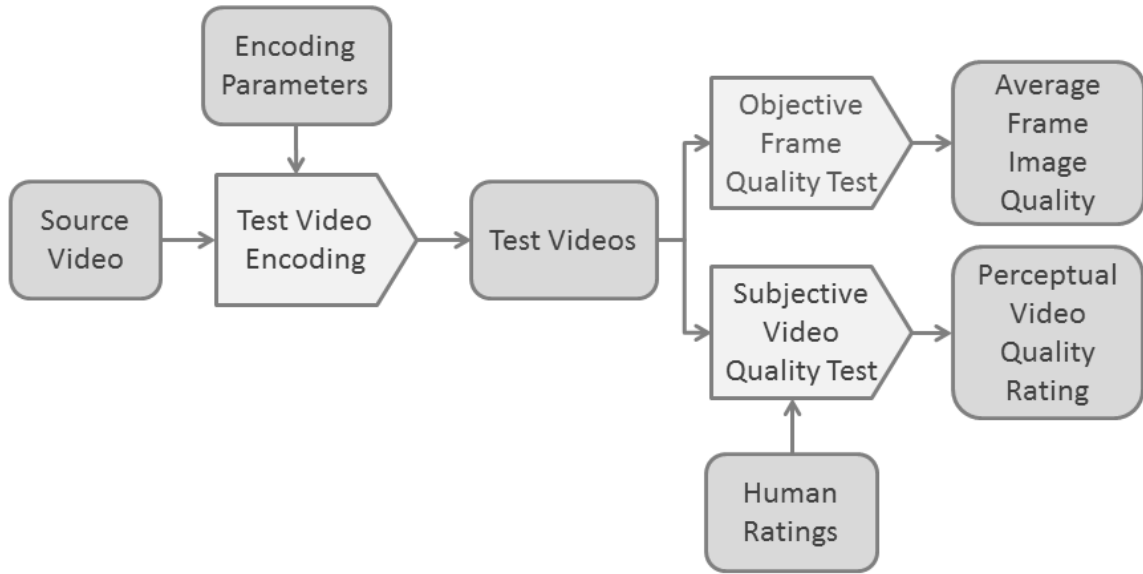


Figure 3.1: Flow chart of experiment processes

3.1 Source Video

The two original videos off of which all these tests are performed were recorded at the Centre for Simulation Based Learning at McMaster University in Hamilton Ontario. Both were chosen as examples of telemedical video to exemplify the type of videos used by an instructor to assess a simulated medical scenario.

The first recording contains footage of the simulation room with individuals interacting with the manikin. This type of video would enable an instructor to visualize the events happening in the simulation room and to understand the context of the simulation. The recording was made using IP Cameras used for local simulations and were encoded to HD standards using H.264/AVC encoding. Due to the technology available, raw or unencoded video of the subject was not available. While not perfect, the quality of this video is near perfect and without any visible coding artifacts. This agrees with the requirements put forth by the Video Quality Experts Group for

quality of experience analysis [44]. Three frames taken from the original video of the simulation room are given in Figure 3.2. The camera position and simulation manikin are stationary while 2 figures move within the frame and interact with the manikin.

The second recording is the footage from a GlideScope during an attempt at intubation of the manikin. This video was chosen to represent a potential task to be completed during a simulation that an instructor would be evaluating. The recording was obtained through the video out port of the GlideScope connected through a video capture card (Aver Media EZmaker USB Gold) and was recorded to a raw video format at a resolution of 640x480 and a raw bitrate of 640 Kbps. Figure 3.3 shows three frames taken from the original Glide Scope video. Since the camera is located on a device inserted into the mouth and throat of a simulation manikin, the camera view moves and changes with the operator's movements and an intubation tube can be seen being inserted in the trachea of the simulation manikin.

3.2 Test Video Encoding

As mentioned previously, both original videos were encoded using H264/AVC encoding via the x264 codec included in the FFmpeg codec library. This experiment took advantage of the ability to control encoding bitrate, frame size and frame rate when using this encoding method.

Five encoding bitrates were chosen to accommodate a range of low bandwidth conditions: 50, 100, 150, 200 and 250 kilobits per second (kbit/s). By setting the encoding bitrate to these values, the video size is constrained through reduced video quality. When encoding, the maximum bitrate was set using the *-maxrate* setting in ffmpeg in addition to the *-bufsize* setting which specifies the size of the buffer used



Figure 3.2: Frames from original video of the simulation room. Camera position is stationary while figures in the video move and interact.

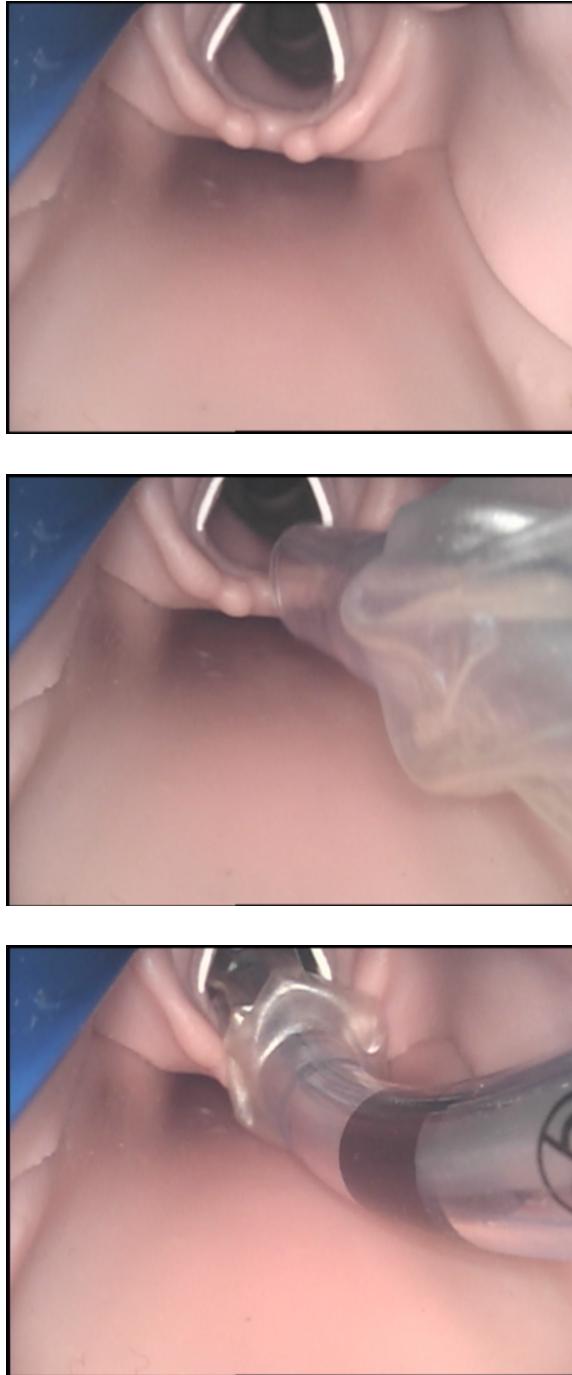


Figure 3.3: Frames from original GlideScope video.

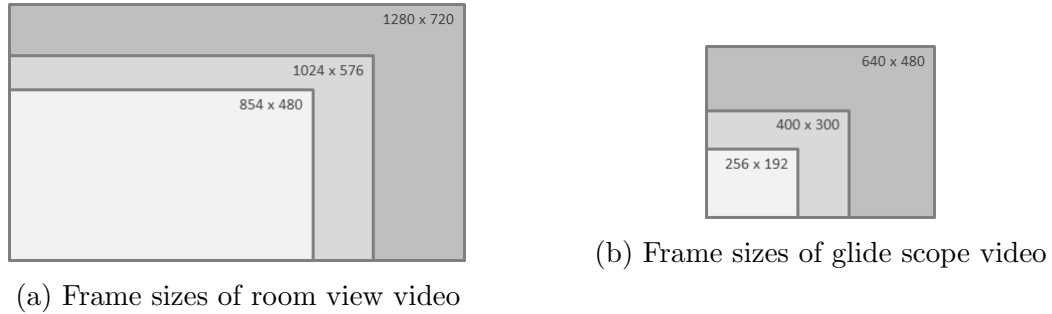


Figure 3.4: Relative Frame Sizes

to calculate bitrate. For consistency, the buffer size was set to twice the maximum bit rate for all encoding cases.

Each video was encoded to its original frame size and to two smaller frame sizes which were chosen because they were suitable for viewing the encoded videos on a computer monitor without significantly sacrificing the resolution of the original. The room view video, which was originally encoded with an aspect ratio of 16:9, was encoded to 854x480, 1024x576 and 1280x720. The glide scope video was originally recorded with an aspect ratio of 4:3 and was encoded to the frame sizes 256x192, 400x300 and 648x480. Figure 3.4 and Table 3.1 show the relative sizes of the frames dimensions chosen.

Both source videos were recorded at a frame rate of 30 frames per second (fps) which is the standard for video recording in North America. The reduced frame rates chosen for this study were selected by dividing the previous frame rate in half. This ensures that the encoded frames are selected evenly from the available frames. The resulting frames rates for the encoded videos are 30, 15, 7.5, 3.75 and 1.875 fps. While it seems counterintuitive to select a non-integer frame rate, it is not a rigid constraint of the number of frames displayed in a second but rather an average rate of frames over a longer period of time.

All videos encoded for this study were encoded with the *-preset* parameter set to *fast* in order to prioritize encoding time over file size, which is favourable for live streaming of video with little or no delay.

The objective frame quality tests discussed below required reference frames with the same size and position in the video as the frames tested. This required several videos encoded to these frame rates and frame sizes with lossless encoding. This was done by setting the *-qp* parameter to 0 rather than setting the maximum bitrate.

Videos were encoded in sets with either one or two encoding parameters varying while others were held constant. Several videos are included in multiple sets resulting in a total of 19 unique test videos for each of the source videos. A complete list of the encoding parameters for each video is included in Table 3.2.

Figure 3.5 shows the same frame of several different test videos each exhibiting different levels of degradation due to the parameters used for encoding.

3.3 Subjective Video Quality Test

The purpose of the perceptual video quality test was to measure the effect of encoding conditions on the perceived quality of the video with respect to its intended purpose. The quality was determined by the viewers ability to understand the context of events in the simulation room view video and to evaluate task performance in the glide scope video rather than evaluating the presence of encoding artifacts as would be the case with non-telemedical video streaming.

	Room View Video		Glide Scope Video	
	Dimensions	Pixels	Dimensions	Pixels
large	1280x720	921,600	640x480	307,600
medium	1024x576	589,824	400x300	120,000
small	854x480	409,920	256x192	49,152

Table 3.1: Frame Dimensions

	Video	Bitrate	Frame Rate	Frame Size	Varying Parameter(s)
Set 1	1	250 kbit/s	30	large	frame rate
	2	250 kbit/s	15	large	
	3	250 kbit/s	7.5	large	
	4	250 kbit/s	3.75	large	
	5	250 kbit/s	1.875	large	
Set 2	1	250 kbit/s	30	large	bitrate
	6	200 kbit/s	30	large	
	7	150 kbit/s	30	large	
	8	100 kbit/s	30	large	
	9	50 kbit/s	30	large	
Set 3	1	250 kbit/s	30	large	frame size
	10	250 kbit/s	30	medium	
	11	250 kbit/s	30	small	
Set 4	1	250 kbit/s	30	large	bitrate and frame rate
	12	200 kbit/s	15	large	
	13	150 kbit/s	7.5	large	
	14	100 kbit/s	3.75	large	
	15	50 kbit/s	1.875	large	
Set 5	1	250 kbit/s	30	large	frame rate and frame size
	16	250 kbit/s	15	medium	
	17	250 kbit/s	7.5	small	
Set 6	6	200 kbit/s	30	large	bitrate and frame size
	18	150 kbit/s	30	medium	
	19	100 kbit/s	30	small	

Table 3.2: Encoding Parameters of Test Videos



(a) Reference video frame (encoded at 1280x720, 30 fps and quantization parameter of 0)



(b) Frame of test video 5 with minimal degradation (encoded at 1280x720, 1.875 fps and 250 kbit/s)

Figure 3.5: Video frame at several levels of degradation due to encoding parameters



(c) Frame of test video 6 with moderate degradation (encoded at 1280x720, 30 fps and 200 kbit/s)



(d) Frame of test video 9 with extreme degradation (encoded at 1280x720, 30 fps and 50 kbit/s)

Figure 3.5: Video frame at several levels of degradation due to encoding parameters

3.3.1 Selected Design

A single-stimulus continuous-quality scale with hidden-reference removal was chosen for the design of this test. The continuous absolute quality rating scale used is shown in Figure 3.6.

The continuous scale was selected because it does not restrain ratings to discrete values and therefore allows a greater range of ratings. As previously mentioned, results for a continuous scale are consistent with the results for a categorical scale when considering rank or order of the data [46].

A *single stimulus* design was chosen over the alternative *double stimulus* design because it requires less time and results in much faster tests.

The original video clip was included along with encoded clips during the test as a reference. The rating each participant provides for the reference clip was subtracted from their scores for all other videos to provide difference ratings as shown in the following equation

$$\text{Test Video Rating} = \text{test clip score} - \text{reference clip score} + 100.$$

Possible scores can range between 200 and 0 where ratings greater than 100 represent a higher quality than the reference video. This hidden-reference removal process eliminated issues with participant bias in the scores that is introduced when a single stimulus design is used [47].

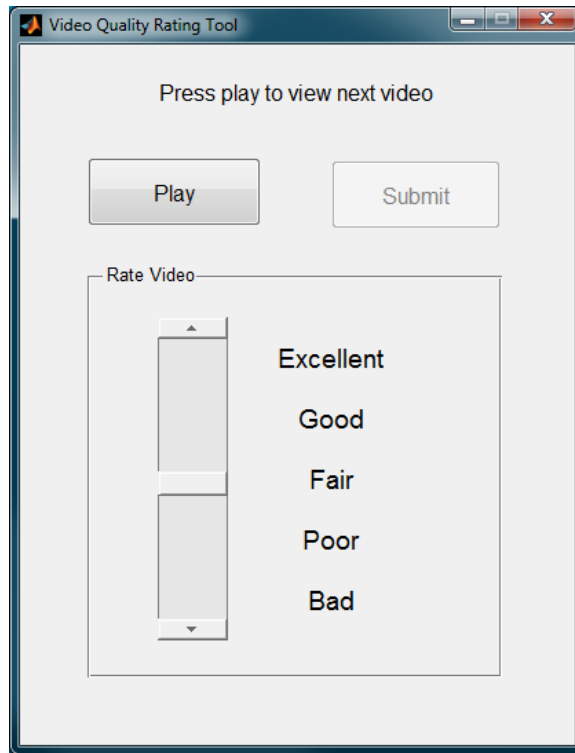


Figure 3.6: Quality Rating Scale Used for Experiment

3.3.2 Procedure

The tests were automated to run using Matlab with minimal intervention from the tester. Participants were told to interact with the on screen dialog and that all instructions should be self-explanatory.

In order to familiarize the participants with the testing process they were presented with instructions at the beginning of their test session. Additional instructions were included for the test involving the glide scope video so that participants with no medical background could understand the task they were viewing. The full instructions provided to the participants are included in Appendix A. Participants had the ability to review the instructions for as long as they needed to before proceeding with the

test. They were also encouraged to ask questions for further clarification.

Prior to the rest of the test, 5 sample clips were presented to the participants to rate in order to become accustomed to the rating process. Again, they were able to repeat the process as many times as they preferred and were allowed to ask questions during this time.

The actual test composed of 20 clips (19 test clips and 1 reference clip) which were presented to each participant in a random order. In order to ensure ratings were based off the entire length of the clip, participants were required to wait until the complete clip was viewed before submitting their quality judgement.

During the test, each video clip was shown and rated independently as per the single stimulus convention. The same 10-12 second clip was used for each test video, ensuring the content was the same across all videos.

Each test ran for approximately 10 minutes and most participants chose to sit both tests (simulation room view and glide scope videos) consecutively.

3.3.3 Participants

Twelve participants were recruited to participate in rating of the test videos. Participants were all considered to be non-experts in the area of video picture quality and had normal vision as is recommended by the International Telecommunications Union [45].

Participants were asked to judge video quality based on the purpose and features in the video. The video clips used for this study were easy to understand by individuals with no medical background. Because of this it was decided that non-expert participants would be sufficient to judge the quality of the videos with respect to the

intended purpose.

3.3.4 Viewing Conditions

All tests took place at a desk in an office with daytime lighting and minimal distractions. The test setup included 2 monitors (1 laptop and 1 external monitor). Videos were played on the external monitor while the rating dialog appeared on the laptop monitor. All videos had frame sizes smaller than the monitor they were viewed on and were played with a black background.

3.3.5 Data Collection

The video ratings provided by participants through the rating dialog in Figure 3.6 were converted to a value between 0 and 100 based on the position on the scale. These values were saved in order of clip number (given in Table 3.2) for each individual participant.

3.4 Objective Frame Quality Test

The SSIM Matlab function was used to measure the image quality of the frames of the encoded test videos. As introduced in Section 2.2.1, SSIM is a full reference image quality metric to measure the structural similarity of a degraded image against a reference image of perfect quality. Reference frames were created by encoding the original video to the same frame rate and frame size as the test video using lossless encoding, ensuring that the resulting videos have the same number and size of frames. Figure 3.5a shows a frame of the reference video that was used for the calculation of

SSIM for test videos encoded at 1280x720 and 30 fps.

The experiment involved measuring the SSIM of every individual frame of the test videos listed in Table 3.2. The SSIM values were then averaged over the entire length of the test video to give a single average SSIM value for each video representing the *average frame image quality*.

3.5 Statistical Methods

Nonparametric statistics are statistical methods that can be used for analysis of data which does not conform to a normal or Gaussian distribution [48]. Specifically, these methods concentrate on the rank or order of the data [49]. Because nonparametric methods make fewer assumptions about the data, they can be considered more robust than the alternative parametric methods if also less statistically powerful [48].

Nonparametric statistical methods are recommended for the analysis of ordinal data such as the data collected from the perceptual video quality tests since the numerical value of the ratings do not absolutely represent the video quality but are rather a relative quality rating between the videos suitable for ordering of the videos [46,50].

The tests discussed below all use a level of risk of $\alpha = 0.05$ to test for significance. This indicates a less than 5% probability that the result occurred due to chance.

3.5.1 Spearman's Rank Correlation Coefficient

The Spearman's rank correlation coefficient is the nonparametric equivalent of Pearson's correlation, used to measure the relationship between two variables which include ordinal data [49]. It is used in this thesis to measure the correlation between

frame image quality and perceptual video quality scores.

The test results give a ρ value ranging between -1 and 1 representing the correlation between the variables and a p -value representing the significance of the result. A positive or negative ρ represents a positive or negative correlation between the two variables while a ρ of zero indicates no correlation. A p -value of less than 0.05 indicates a statistically significant result that the variables correlated [51].

3.5.2 Kruskal-Wallis Analysis of Variance of Ranks

The Kruskal-Wallis test is the nonparametric equivalent of the ANOVA test to check if more than 2 unrelated samples originate from the same distribution [49]. It will be employed in this thesis to check if there is a significant difference in the median perceptual quality ratings of videos within the same test set.

The null hypothesis of the Kruskal Wallis test is that the mean ranks of the groups are the same [49]. A p -value given as the result can reject this null hypothesis if $p < 0.05$, indicating that there is a significant difference between at least 2 of the groups [51]. The results also give the degrees of freedom between the groups (represented by df) and the H -statistic for the test, which represents the variance of ranks between the groups [51].

In the event that the null hypothesis is rejected, followup tests are required to determine which groups originate from a different distribution [51]. The *multcompare* method in Matlab was used for this additional test.

3.5.3 Box Plot

Boxplots are used to display the distribution of perceptual quality ratings for videos within a set. These plots are interpreted in the following manner:

- Each group is represented in a column on the plot;
- The central red line is the median value for each group;
- The edges of the box are the 25th and 75th percentiles within each group;
- The whiskers extend to the most extreme data points within 2.7σ ; and
- Outliers beyond this range are plotted with a + [51].

Chapter 4

Results

In order to present the results of the above experiments in a meaningful way, data will be presented in sets of three or five videos with only one or two of the encoding parameters varying within the test set. Test sets 1-3 have one encoding parameter changing while the other two stay constant and sets 4-6 have two parameters changing while one stays constant. Comparison of videos within the sets and between the sets will allow us to make assumptions about the nature of the data in order to build a model. Results from the simulation room view video are presented in Section 4.1 while the glide scope results are presented in Section 4.2.

The following results are given for each video test set:

- Average frame image quality as measured by the average SSIM,
- Distribution of perceptual video quality ratings given as a boxplot graph,
- Spearman's rank correlation coefficient and scatterplot depicting correlation between average frame image quality and perceptual quality ratings,

Video	Bitrate	Frame Rate	Frame Size	Average SSIM
1	250 kbit/s	30	1280x720	0.9445
2	250 kbit/s	15	1280x720	0.9543
3	250 kbit/s	7.5	1280x720	0.9634
4	250 kbit/s	3.75	1280x720	0.9693
5	250 kbit/s	1.875	1280x720	0.9707

Table 4.1: Simulation Room View Average SSIM Results for Set 1

- Kruskal-Wallis one-way analysis of variance by ranks testing of perceptual quality ratings of videos within a set originate from the same distribution, and
- Results of followup tests indicating which data samples come from different distributions.

As mentioned in Section 3.5, non-parametric statistical methods (Kruskal Wallis test and Spearman's rank correlation coefficient) have been employed due to the ordinal nature of the perceptual video quality ratings.

4.1 Simulation Room View Results

4.1.1 Set 1: Frame Rate

Test set 1 consists of five videos encoded at 250 kbit/s and a frame size of 1280x720 with frame rates ranging from 1.875 to 30 fps. The encoding parameters and average frame image quality values for this set are presented in Table 4.1. When the frame rate of the encoded video is decreased, the average frame image quality (as measured by the SSIM) increases.

The distribution of perceptual quality ratings for the videos in set 1 can be found in Figure 4.1 while Figure 4.2 shows these values plotted against the average SSIM.

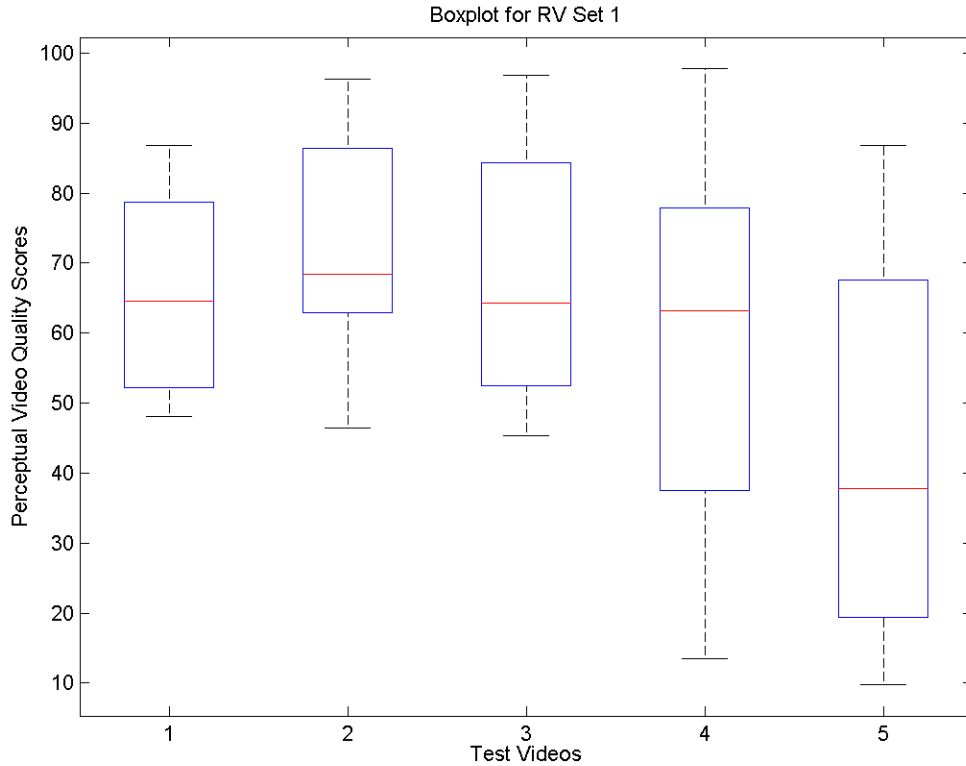


Figure 4.1: Room View Perceptual Video Quality Ratings for Set 1

The spearman's rank correlation coefficient between SSIM and ratings for set 1 is $\rho = -0.2933$ ($p = 0.0229$). Kruskal-Wallis one-way analysis of variance of ranks results for set 1 is given in Table 4.2.

While none of the videos in set 1 show a statistically significant difference in perceptual quality rating, the general trend shows that more participants rated the videos poorly when encoded at low and very low frame rates (3.75 and 1.875 fps). It can also be seen that there is a negative correlation between average frame image quality and perceptual video quality for this set. These results show that decreasing video frame rate while maintaining bitrate and frame size will both increase the

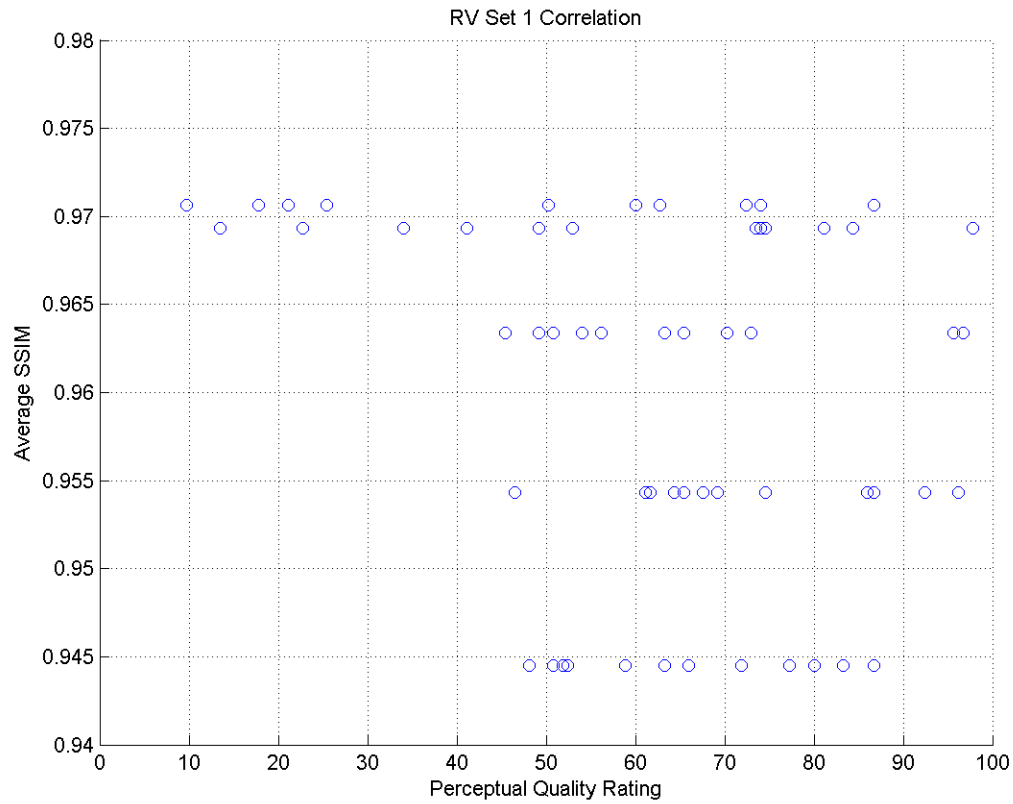


Figure 4.2: Correlation between average frame image quality and perceptual quality rating for Set 1

Source	SS	df	MS	H	p-value
Columns	2448.75	4	612.1875	8.0309	0.090452
Error	15541.25	55	282.5682		
Total	17990	59			

Table 4.2: Kruskal-Wallis one-way analysis of variance of ranks results for Set 1

Video	Bitrate	Frame Rate	Frame Size	Average SSIM
1	250 kbit/s	30	1280x720	0.9445
6	200 kbit/s	30	1280x720	0.9372
7	150 kbit/s	30	1280x720	0.9260
8	100 kbit/s	30	1280x720	0.9045
9	50 kbit/s	30	1280x720	0.7912

Table 4.3: Simulation Room View Average SSIM Results for Set 2

average frame image quality and diminish the perceptual quality of the video.

4.1.2 Set 2: Bitrate

The videos in test set 2 were all encoded at 30 fps and a frame size of 1280x720 with bitrates ranging from 50 to 250 kbit/s. As seen by the results in Table 4.3, the average frame image quality decreases along with the encoding bitrate of the videos in set 2.

Figures 4.3 and 4.4 show a boxplot of the distribution of perceptual quality ratings for set 2 and these ratings plotted against average SSIM, respectively. The Spearman's correlation between SSIM and perceptual ratings was calculated to be $\rho = 0.7076$ ($p = 0.0000$).

Kruskal-Wallis one-way analysis of variance of ranks results for set 2 is given in Table 4.4. Significant differences were found between the perceptual quality ratings for the following videos:

- Videos 1 and 8 ($p = 0.0012$);
- Videos 1 and 9 ($p = 0.0000$);
- Videos 6 and 8 ($p = 0.0433$); and

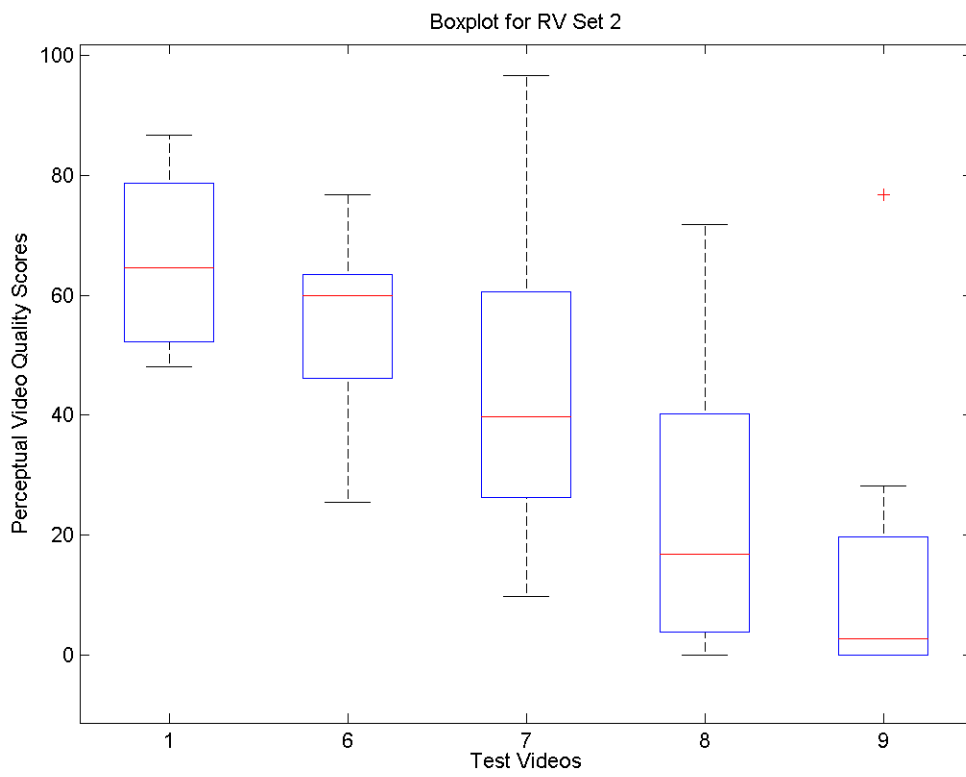


Figure 4.3: Room View Perceptual Video Quality Ratings for Set 2

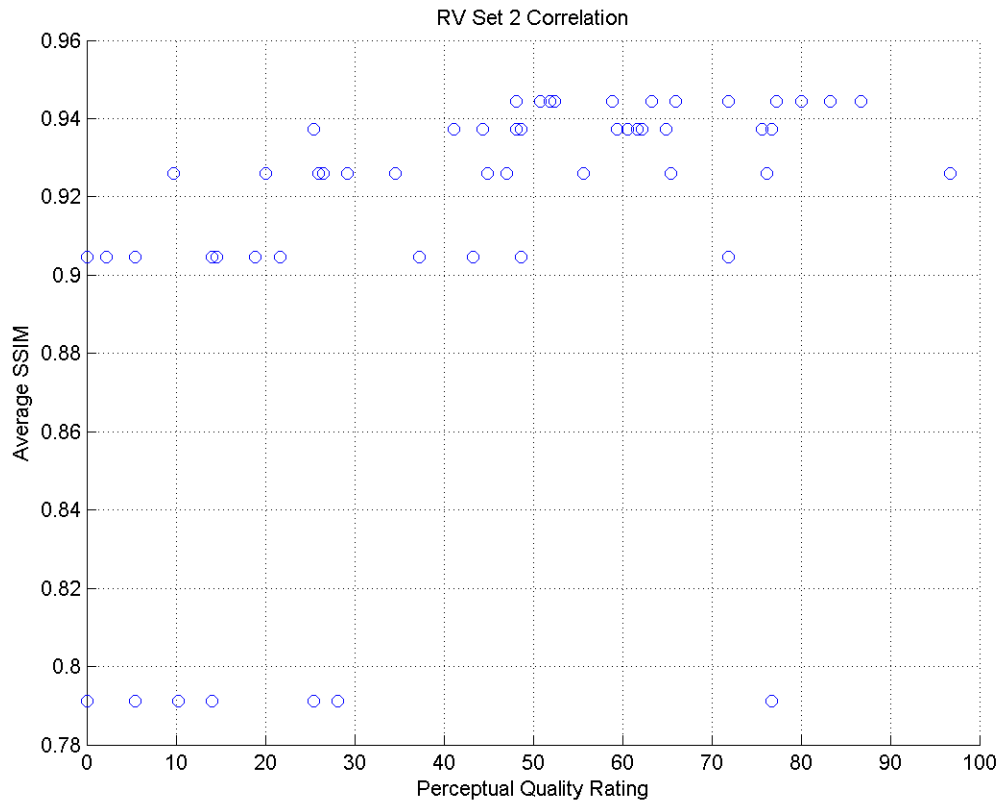


Figure 4.4: Correlation between average frame image quality and perceptual quality rating for Set 2

Source	SS	df	MS	H	p-value
Columns	9099.625	4	2274.9063	29.9088	5.1082e-06
Error	8850.875	55	160.925		
Total	17950.5	59			

Table 4.4: Kruskal-Wallis one-way analysis of variance for Set 2

Video	Bitrate	Frame Rate	Frame Size	Average SSIM
1	250 kbit/s	30	1280x720	0.9445
10	250 kbit/s	30	1024x576	0.9569
11	250 kbit/s	30	854x480	0.9601

Table 4.5: Simulation Room View Average SSIM Results for Set 3

- Videos 6 and 9 ($p = 0.0026$).

These results show a significant decrease in perceptual video quality ratings for videos encoded at low bitrates (50 and 100 kbit/s) as compared to bitrates at the higher end used in this study (200 and 250 bit/s). A strong correlation between frame image quality and perceptual video quality is also apparent when frame rate and frame size are constant.

4.1.3 Set 3: Frame Size

The videos in test set 3 are encoded at 250 kbit/s and 30 fps at all 3 frame sizes. Table 4.5 shows the encoding parameters and average frame image quality values for the videos in set 3. Set 3 exhibits a similar trend to set 1, in that when fewer pixels are represented by the same number of bits (either through fewer frames as in set 1 or smaller frames as in set 3) the frame image quality (as measured by SSIM) increases.

The distribution of perceptual quality ratings is presented in the boxplot in Figure 4.5. Figure 4.6 shows the perceptual ratings plotted against the average frame image quality for set 3. The Spearman's rank correlation coefficient between the two was

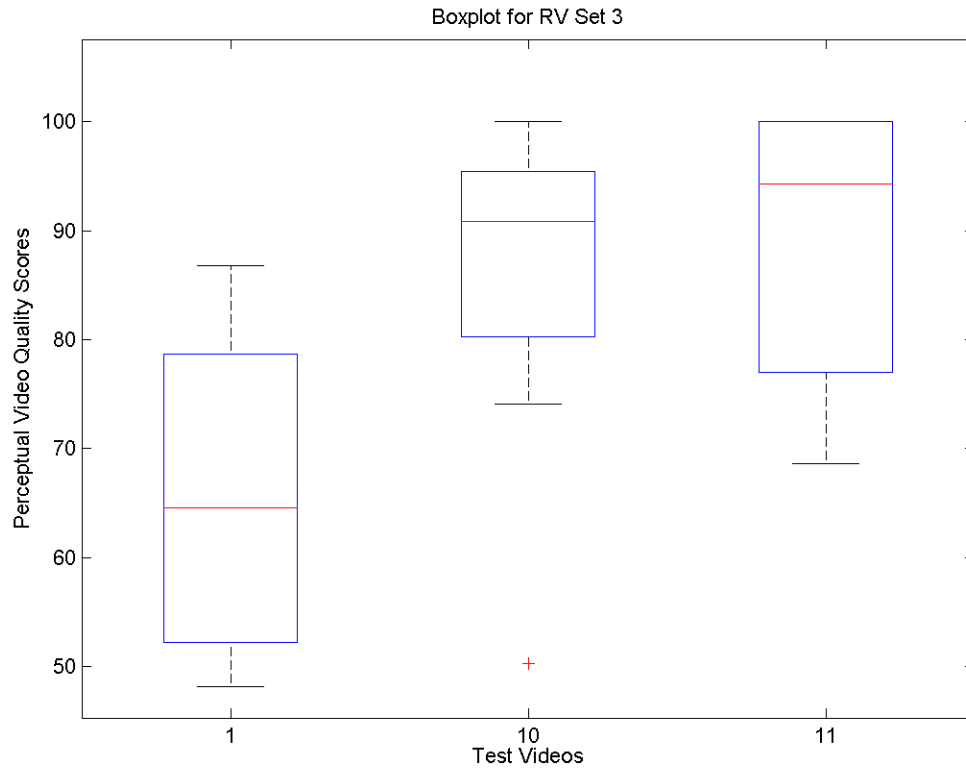


Figure 4.5: Room View Perceptual Video Quality Ratings for Set 3

calculated as $\rho = 0.5632$ ($p = 0.0003$).

Kruskal-Wallis one-way analysis of variance of ranks results for set 3 is given in Table 4.6. Significant differences were found between the following videos:

- Videos 1 and 10 ($p = 0.0113$); and
- Videos 1 and 11 ($p = 0.0025$).

The video encoded at 1280x720 was rated significantly lower than the videos encoded at 1024x5760 and 854x480. Set 3 also exhibits a strong correlation between average frame image quality and perceptual video quality despite the variation in

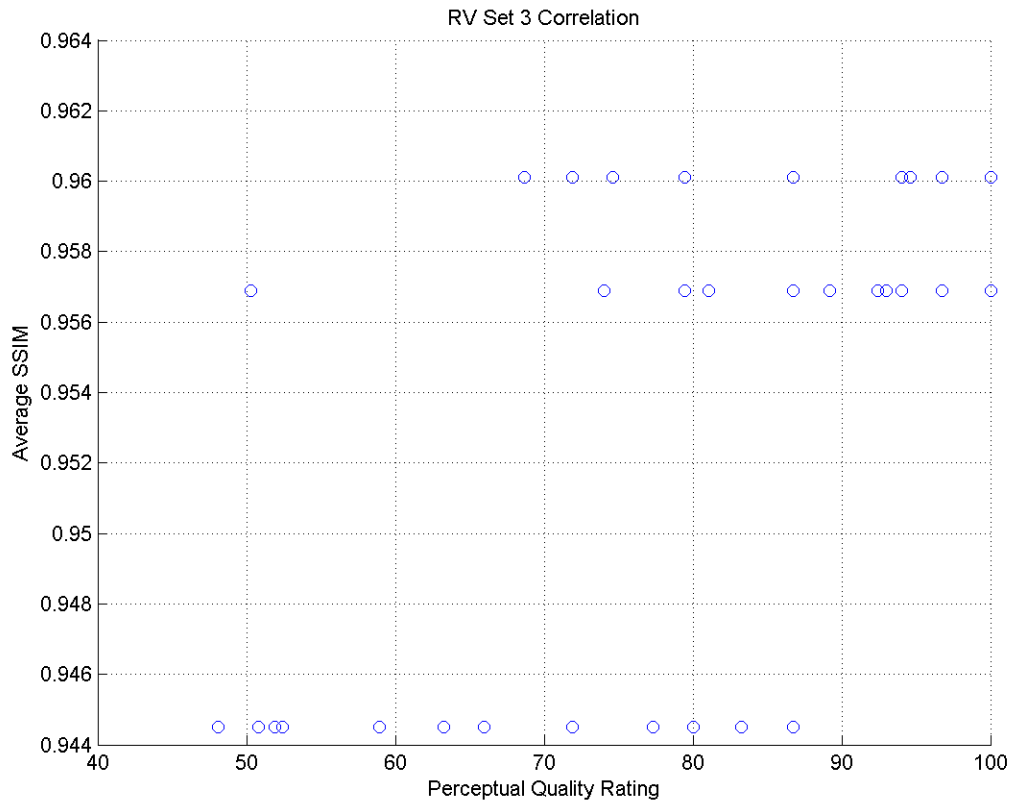


Figure 4.6: Correlation between average frame image quality and perceptual quality rating for Set 3

Source	SS	df	MS	H	p-value
Columns	1440.7917	2	720.3958	13.0523	0.0014646
Error	2422.7083	33	73.4154		
Total	3863.5	35			

Table 4.6: Kruskal-Wallis one-way analysis of variance for Set 3

Video	Bitrate	Frame Rate	Frame Size	Average SSIM
1	250 kbit/s	30	1280x720	0.9445
12	200 kbit/s	15	1280x720	0.9484
13	150 kbit/s	7.5	1280x720	0.9512
14	100 kbit/s	3.75	1280x720	0.9499
15	50 kbit/s	1.875	1280x720	0.9292

Table 4.7: Simulation Room View Average SSIM Results for Set 4

Source	SS	df	MS	H	p-value
Columns	5851.7917	4	1462.9479	19.1953	0.00071946
Error	12134.7083	55	220.6311		
Total	17986.5	59			

Table 4.8: Kruskal-Wallis one-way analysis of variance for Set 4

frame size.

4.1.4 Set 4: Bitrate and Frame Rate

Test set 4 consists of videos which were encoded at bitrates ranging from 50 to 250 kbit/s and frame rates of 1.875 to 30 fps, all at a frame size of 1280x720. Encoding parameters and average SSIM results for each video are included in table 4.7.

The distribution of perceptual quality ratings for set 4 is given in the boxplot in Figure 4.7. Figure 4.8 is a plot of the perceptual quality ratings against the average SSIM values. The correlation between the two is $\rho = 0.1344$ ($p = 0.3058$).

Kruskal-Wallis one-way analysis of variance results for set 4 is given in Table 4.8. Significant differences were found between the following videos:

- Videos 1 and 15 ($p = 0.0047$); and
- Videos 12 and 15 ($p = 0.0012$).

Sets 2 and 4 were both encoded with variations in bitrate while set 4 also has

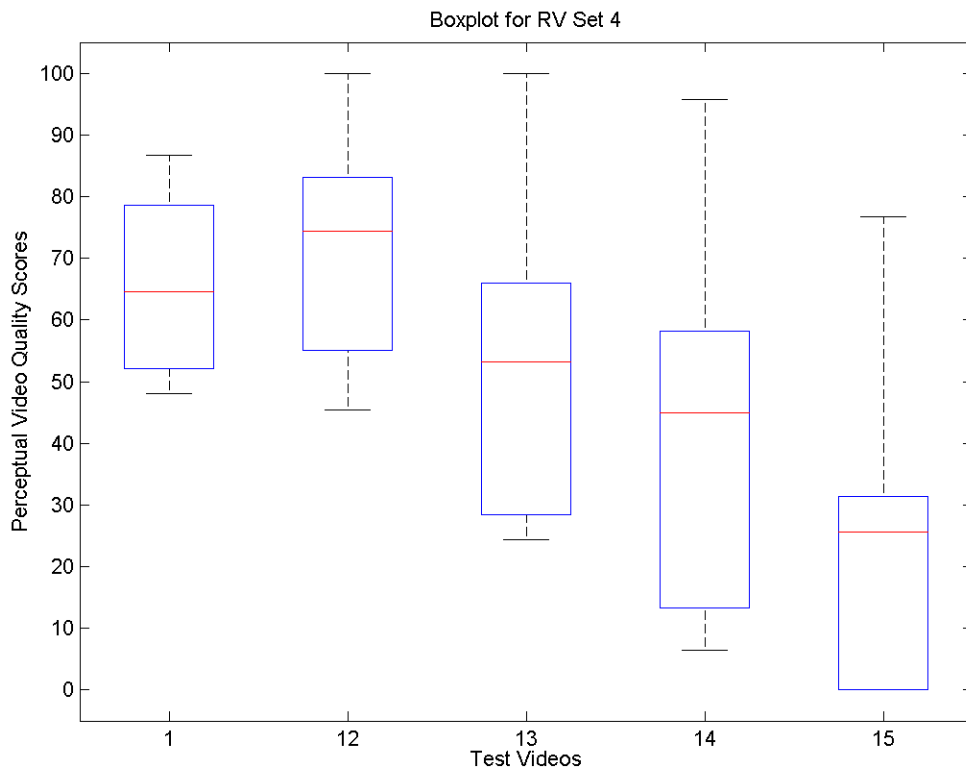


Figure 4.7: Room View Perceptual Video Quality Ratings for Set 4

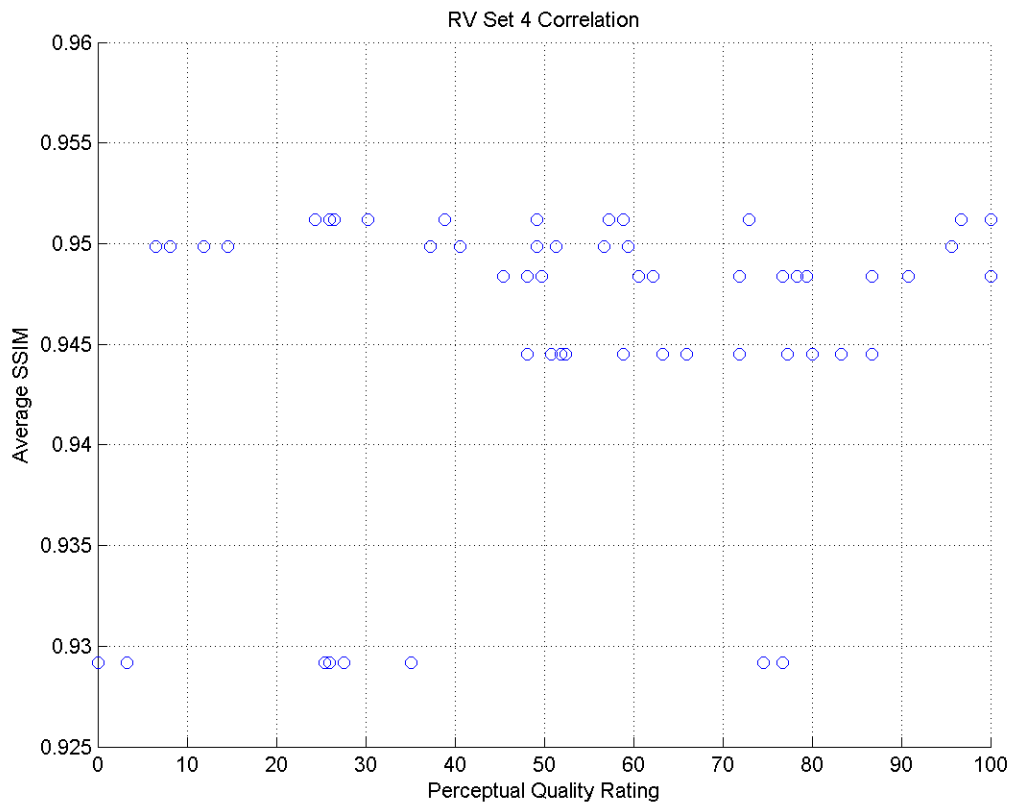


Figure 4.8: Correlation between average frame image quality and perceptual quality rating for Set 4

		Frame rate				
		30	15	7.5	3.75	1.875
Bitrate	250	0.9445	0.9543	0.9634	0.9693	0.9707
	200	0.9372	0.9484			
	150	0.9260		0.9512		
	100	0.9045			0.9499	
	50	0.7912				0.9292

Table 4.9: Average SSIM results for videos in sets 1, 2 and 4

Video	Bitrate	Frame Rate	Frame Size	Average SSIM
1	250 kbit/s	30	1280x720	0.9445
16	250 kbit/s	15	1024x576	0.9646
17	250 kbit/s	7.5	854x480	0.9718

Table 4.10: Simulation Room View Average SSIM Results for Set 5

variation in the frame rate used for encoding. Sets 1 and 4 both possess a variation in frame rate, while set 4 also varies in the encoding bitrate used. Direct comparison between the results of these sets allows for greater insight into the results. Table 4.9 shows the average SSIM values for the videos in these sets along with their respective frame rates and bitrates. It can be seen that the average frame image quality increases across all bitrates when the frame rate is lowered while it decreases across all frame rates for decreases in bitrate.

4.1.5 Set 5: Frame Rate and Frame Size

Test set 5 consists of videos which were encoded at 250 kbit/s and a range of frame rates and frame sizes. The encoding parameters for set 5 as well as the average SSIM values are given in table 4.10.

A boxplot of the perceptual quality ratings for set 5 is given in 4.9. The correlation between average frame image quality and perceptual quality rating as shown in Figure

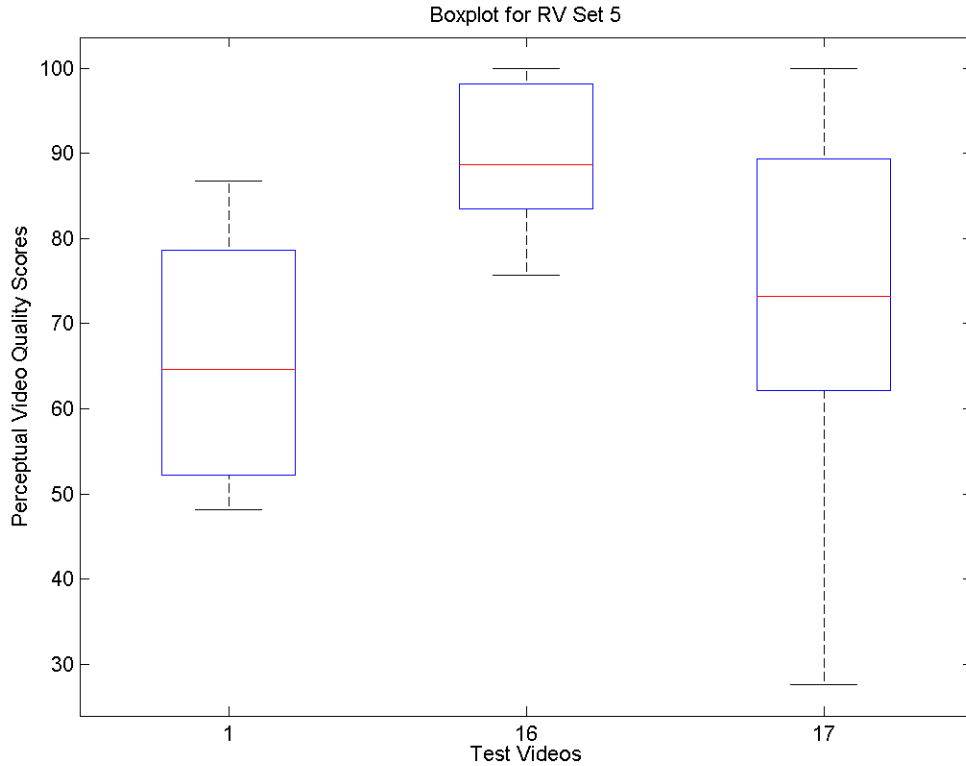


Figure 4.9: Room View Perceptual Video Quality Ratings for Set 5

4.10 was calculated as $\rho = 0.2001$ ($p = 0.2414$).

Kruskal-Wallis one-way analysis of variance of ranks results for set 5 is given in Table 4.11. Significant differences were found between the following videos:

- Videos 1 and 16 ($p = 0.0017$).

Test set 5 (encoded with varying frame size and frame rate) combines the variations in encoding parameters of test set 1 (varied frame rate) and test set 3 (varied frame size). Table 4.12 shows the average SSIM results for the videos in these sets. It can be seen that the average frame image quality increases across all frame sizes when frame rate is reduced and that it increases when the frame size is reduced.

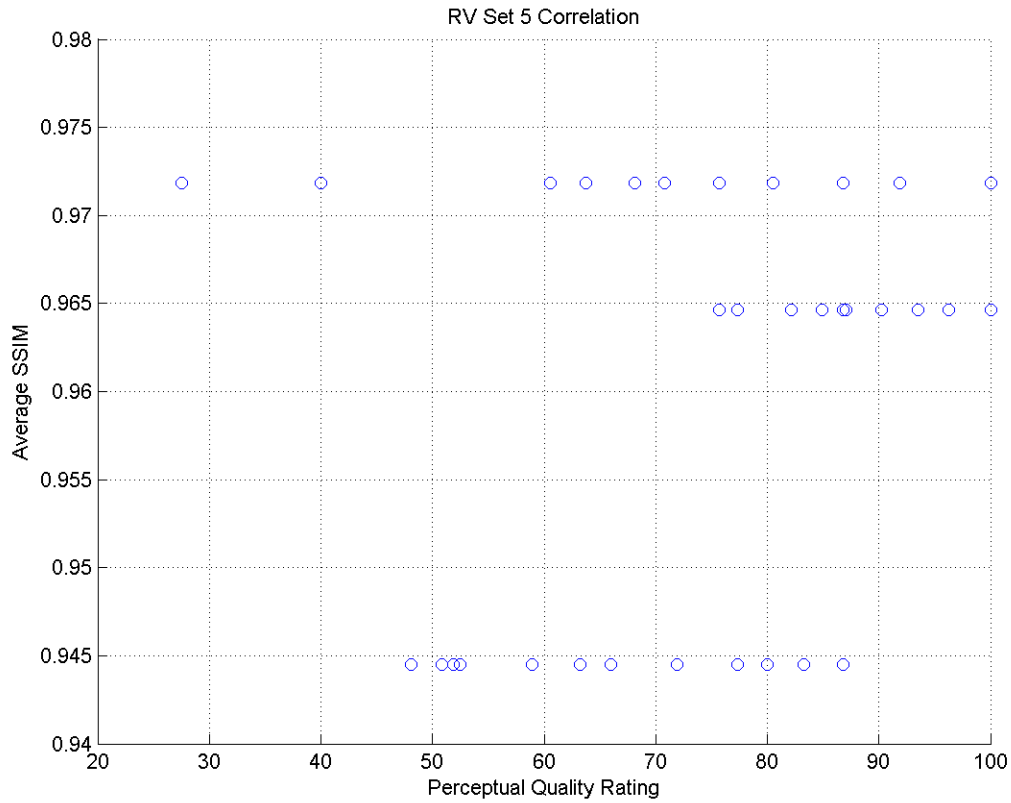


Figure 4.10: Correlation between average frame image quality and perceptual quality rating for Set 5

Source	SS	df	MS	H	p-value
Columns	1355.5417	2	677.7708	12.2531	0.0021841
Error	2516.4583	33	76.2563		
Total	3872	35			

Table 4.11: Kruskal-Wallis one-way analysis of variance for Set 5

		Frame rate				
		30	15	7.5	3.75	1.875
Frame size	1280x720	0.9445	0.9543	0.9634	0.9693	0.9707
	1024x576	0.9569	0.9646			
	854x480	0.9601		0.9718		

Table 4.12: Average SSIM results for videos in sets 1, 3 and 5

Video	Bitrate	Frame Rate	Frame Size	Average SSIM
6	200 kbit/s	30	1280x720	0.9372
18	150 kbit/s	30	1024x576	0.9428
19	100 kbit/s	30	854x480	0.9332

Table 4.13: Simulation Room View Average SSIM Results for Set 6

Source	SS	df	MS	H	p-value
Columns	249.875	2	124.9375	2.252	0.32433
Error	3633.625	33	110.1098		
Total	3883.5	35			

Table 4.14: Kruskal-Wallis one-way analysis of variance for Set 6

4.1.6 Set 6: Bitrate and Frame Size

The videos in test set 6 vary in the frame size and bitrate at which they were encoded but were all encoded at 30 fps. Table 4.13 shows the encoding parameters and the average frame image quality for the videos included in the set.

The perceptual quality ratings for set 6 are presented in the boxplot in Figure 4.11. Figure 4.12 shows the correlation between average frame image quality and perceptual quality rating which was calculated using Spearman's rank correlation as $\rho = 0.2309$ ($p = 0.1754$).

The Kruskal-Wallis one-way analysis of variance of ranks results for set 6 given in Table 4.14 show that there were no significant differences between the ratings of all the videos in set 6.

The parameters varied in set 6 (bitrate and frame size) are a combination of the parameters varied in sets 2 (bitrate) and 3 (frame size). Table 4.15 shows the average SSIM values for these sets and their respective encoding parameters. It can be seen from this table that the average frame image quality decreases with bitrate and increases with smaller frame size.

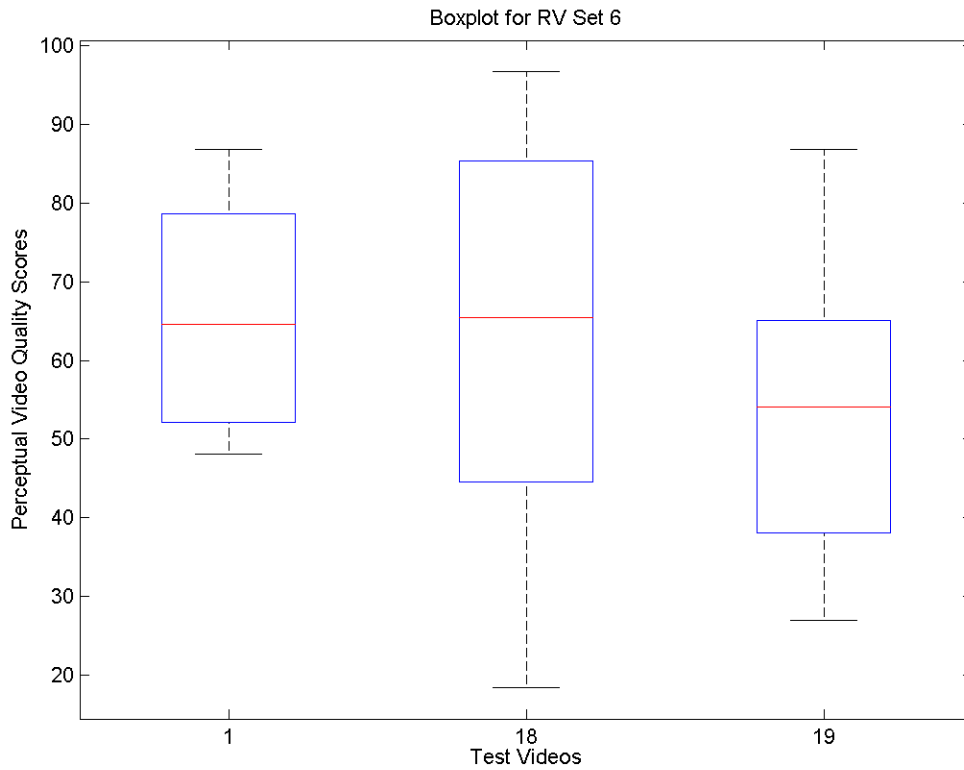


Figure 4.11: Room View Perceptual Video Quality Ratings for Set 6

		Bitrate				
		250	200	150	100	50
Frame size	1280x720	0.9445	0.9372	0.9260	0.9045	0.7912
	1024x576	0.9569	0.9428			
	854x480	0.9601	0.9332			

Table 4.15: Average SSIM results for videos in sets 2, 3 and 6

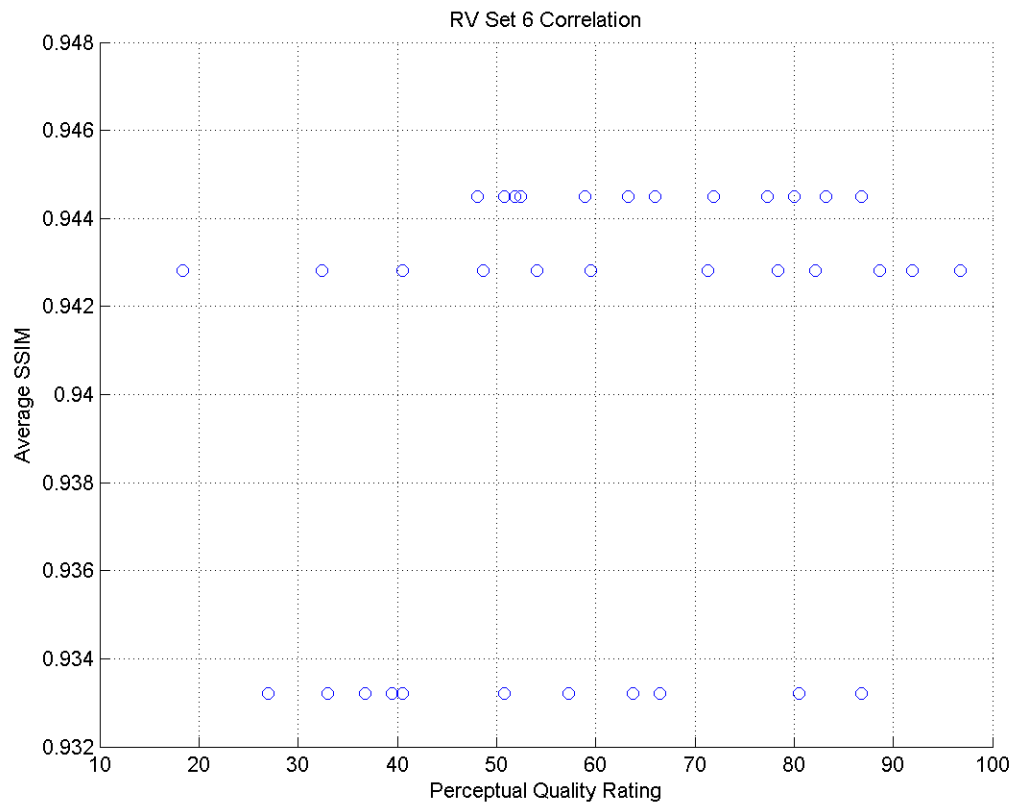


Figure 4.12: Correlation between average frame image quality and perceptual quality rating for Set 6

Video	Bitrate	Frame Rate	Frame Size	Average SSIM
1	250 kbit/s	30	640x480	0.9875
2	250 kbit/s	15	640x480	0.9915
3	250 kbit/s	7.5	640x480	0.9942
4	250 kbit/s	3.75	640x480	0.9960
5	250 kbit/s	1.875	640x480	0.9965

Table 4.16: Glide Scope Average SSIM Results for Set 1

4.2 Glide Scope Video Results

4.2.1 Set 1: Frame Rate

Set 1 consists of videos encoded at 250 kbit/s and a frame size of 640x480 and at a range of frame rates from 1.875 to 30 frames per second. Table 4.16 includes the average frame image quality of the videos in set 1 as measured by average SSIM. Similar to the results from the simulation room view video, the average frame image quality increases when a lower encoding frame rate is used.

Perceptual video quality ratings for set 1 are presented in the boxplot in Figure 4.13. Figure 4.14 shows the perceptual ratings plotted against the averageSSIM values and the Spearman's rank correlation coefficient between the two is 4.14 was calculated as $\rho = -0.5128$ ($p = 0.0001$).

Kruskal-Wallis one-way analysis of variance results for set 1 is given in Table 4.17. Significant differences were found between the perceptual quality ratings of the following videos:

- Videos 2 and 4 ($p = 0.0435$); and
- Videos 2 and 5 ($p = 0.0183$).

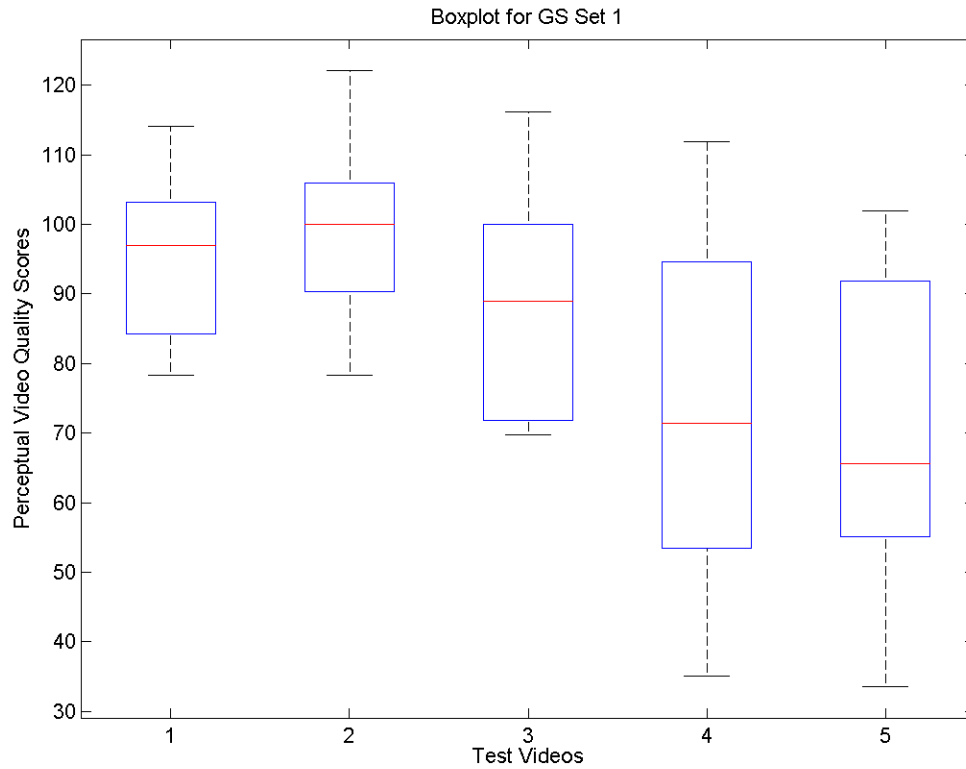


Figure 4.13: Glide Scope Perceptual Video Quality Ratings for Set 1

Source	SS	df	MS	H	p-value
Columns	3182.2	4	795.55	14.9895	0.0047231
Error	7220.3	45	160.4511		
Total	10402.5	49			

Table 4.17: Kruskal-Wallis one-way analysis of variance for Set 1

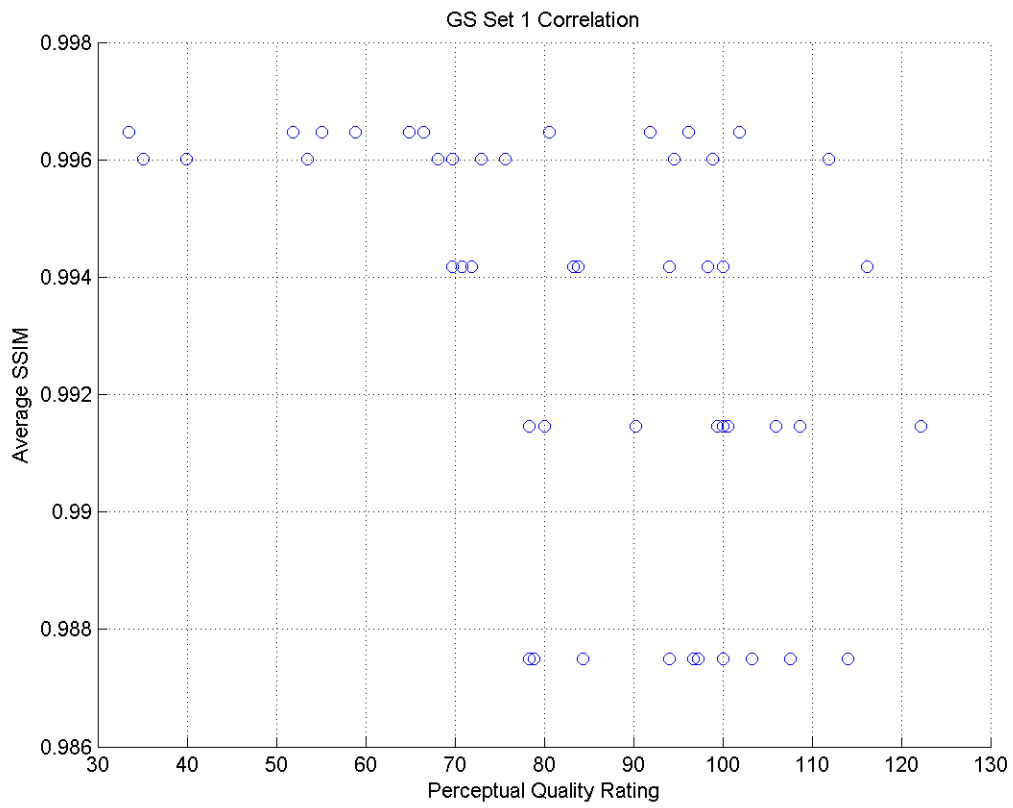


Figure 4.14: Correlation between average frame image quality and perceptual quality rating for Set 1

Video	Bitrate	Frame Rate	Frame Size	Average SSIM
1	250 kbit/s	30	640x480	0.9875
6	200 kbit/s	30	640x480	0.9851
7	150 kbit/s	30	640x480	0.9807
8	100 kbit/s	30	640x480	0.9722
9	50 kbit/s	30	640x480	0.8221

Table 4.18: Glide Scope Average SSIM Results for Set 2

The results from glide scope set 1 show significantly higher perceptual video quality ratings for the video encoded at 15 fps compared to the videos encoded at 3.75 and 1.875 fps. A negative correlation between perceptual video quality ratings and average SSIM is also apparent.

4.2.2 Set 2: Bitrate

The videos in set 2 were encoded at 30 fps and a frame size of 1280x720 with bitrates ranging from 50 to 250 kbit/s. The average frame image quality for these videos as measured by average SSIM is presented in Table 4.18 along with the encoding parameters of each video. It is apparent from these results that average frame image quality decreases at lower bitrates.

The boxplot in Figure 4.15 depicts the distribution of perceptual video quality ratings for the videos in set 2. These values plotted against average SSIM are presented in the scatterplot in Figure 4.16. The Spearman's rank correlation coefficient between these values is $\rho = 0.6723$ ($p = 0.0000$).

Kruskal-Wallis one-way analysis of variance results for set 2 is given in Table 4.19. Significant differences were found between the perceptual quality ratings of the following videos:

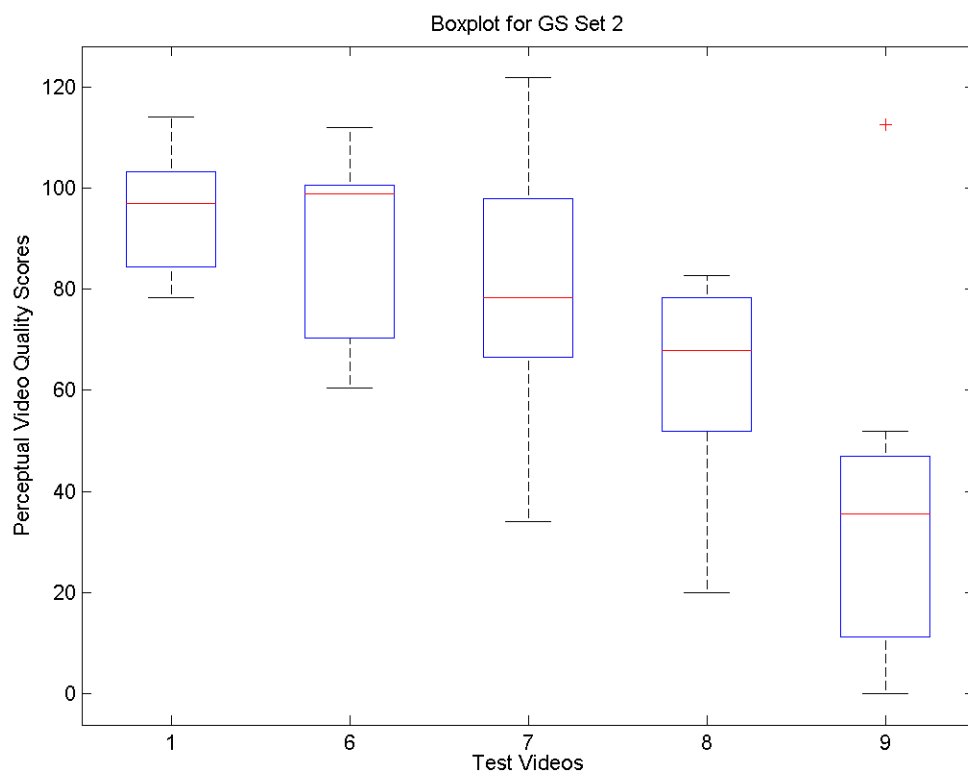


Figure 4.15: Glide Scope Perceptual Video Quality Ratings for Set 2

Source	SS	df	MS	H	p-value
Columns	4812.85	4	1203.2125	22.6509	0.00014867
Error	5598.65	45	124.4144		
Total	10411.5	49			

Table 4.19: Kruskal-Wallis one-way analysis of variance for Set 2

Video	Bitrate	Frame Rate	Frame Size	Average SSIM
1	250 kbit/s	30	640x480	0.9875
10	250 kbit/s	30	400x300	0.9914
11	250 kbit/s	30	256x192	0.9886

Table 4.20: Glide Scope Average SSIM Results for Set 3

- Videos 1 and 8 ($p = 0.0409$);
- Videos 1 and 9 ($p = 0.0004$); and
- Videos 6 and 9 ($p = 0.0030$).

Set 2 results from show significant differences in perceptual video quality ratings for videos with high and low bitrates (250 kbit/s vs 100 and 50 kbit/s, 200 kbit/s vs 50 kbit/s). The average SSIM values are highly correlated with the perceptual ratings.

4.2.3 Set 3: Frame Size

All the videos in set 3 were encoded at a frame rate of 30 fps and a bitrate of 250 kbit/s with frame sizes varying from 256x192 to 640x480. As seen in Table 4.20 the average frame image quality does not show a clear increase for smaller frame sizes, unlike the results from the simulation room view video.

The perceptual quality ratings for set 3 are displayed in the boxplot in Figure 4.17 and plotted against average SSIM in Figure 4.18. The Spearman's rank correlation

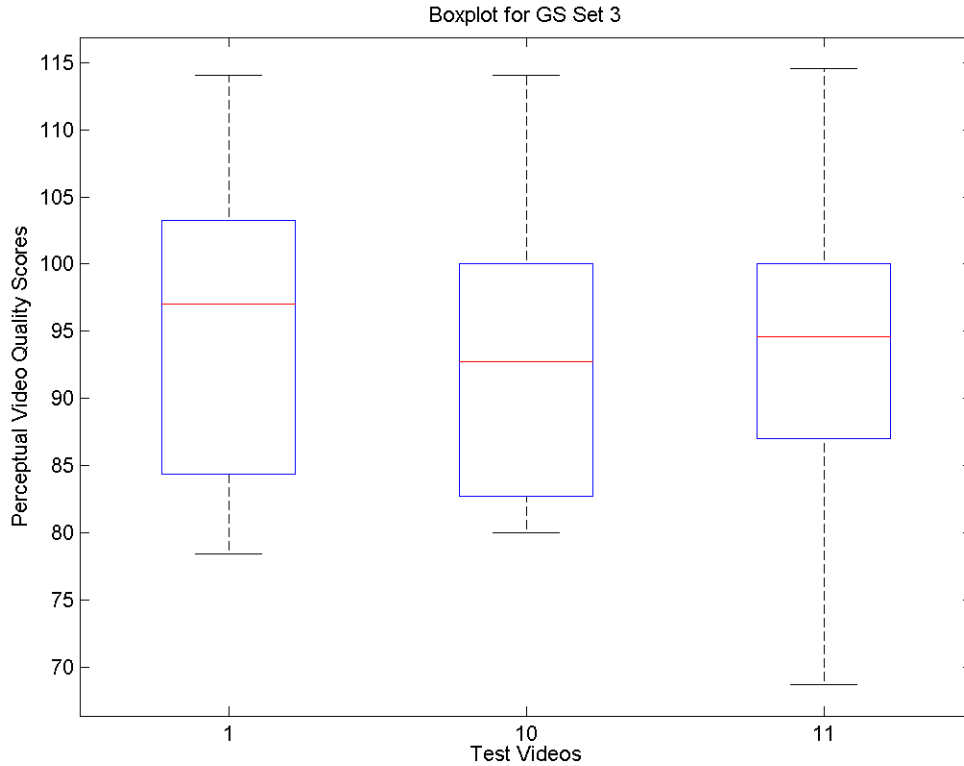


Figure 4.17: Glide Scope Perceptual Video Quality Ratings for Set 3

coefficient between average frame image quality and perceptual quality rating was calculated to be $\rho = -0.1134$ ($p = 0.5509$).

Kruskal-Wallis one-way analysis of variance results for set 3 is given in Table 4.21.

The results of set 3 do not show any significant difference in the perceptual quality ratings of videos encoded at different frame sizes. This set also shows a weak negative correlation between average SSIM and perceptual ratings which differs from the results seen in the simulation room view video.

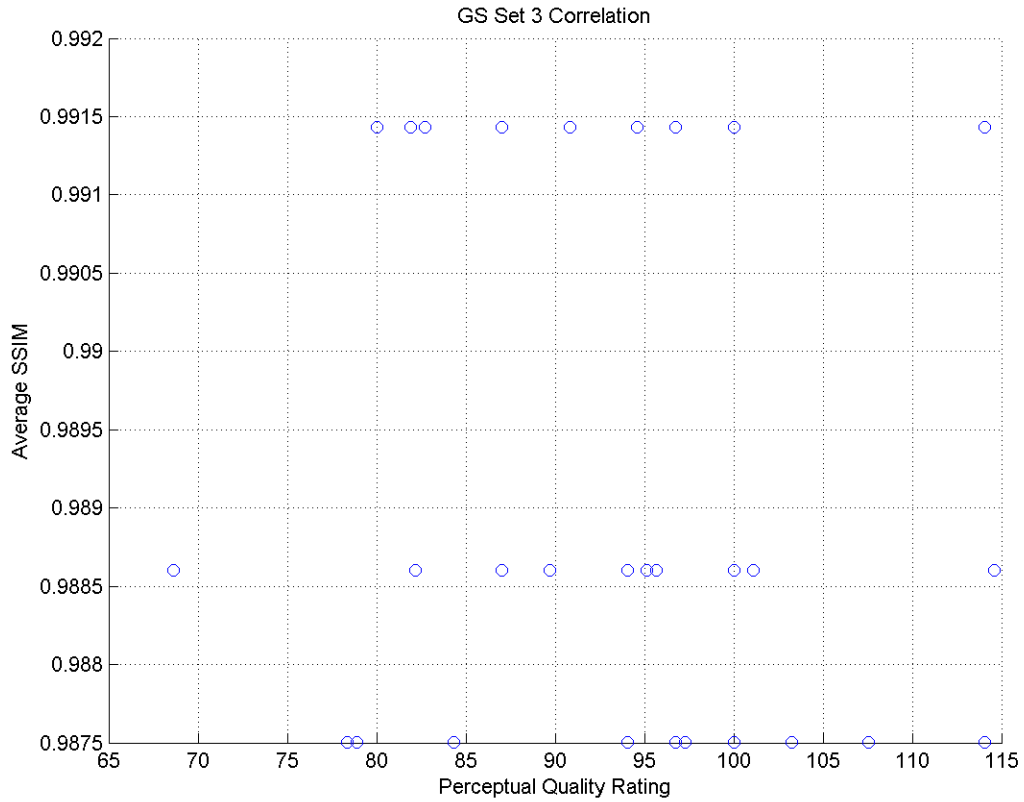


Figure 4.18: Correlation between average frame image quality and perceptual quality rating for Set 3

Source	SS	df	MS	H	p-value
Columns	30.15	2	15.075	0.39016	0.82277
Error	2210.85	27	81.8833		
Total	2241	29			

Table 4.21: Kruskal-Wallis one-way analysis of variance for Set 3

Video	Bitrate	Frame Rate	Frame Size	Average SSIM
1	250 kbit/s	30	640x480	0.9875
12	200 kbit/s	15	640x480	0.9897
13	150 kbit/s	7.5	640x480	0.9916
14	100 kbit/s	3.75	640x480	0.9934
15	50 kbit/s	1.875	640x480	0.9918

Table 4.22: Glide Scope Average SSIM Results for Set 4

4.2.4 Set 4: Bitrate and Frame Size

Set 4 includes videos encoded at varying bitrates and frame rates from 50 kbit/s and 1.875 fps to 250 kbit/s and 30 fps, all with a frame size of 640x480. The average frame image quality results for this set are presented in Table 4.22 along with the encoding parameters of each video.

The distribution of perceptual quality ratings for set 4 is given in the boxplot in Figure 4.19. The Spearman's rank correlation coefficient between average frame image quality and perceptual quality rating as shown by the scatterplot in figure 4.20 was calculated as $\rho = 0.1344$ ($p = 0.3058$).

Kruskal-Wallis one-way analysis of variance results for set 4 is given in Table 4.23. Significant differences were found between the following videos:

- Videos 1 and 15 ($p = 0.0003$);
- Videos 12 and 15 ($p = 0.0002$); and
- Videos 12 and 15 ($p = 0.0068$).

The results for the videos in set 4, which vary in frame rate and bitrate, can be directly compared to the results for the videos in set 1, which vary in frame rate, and set 2, which vary in bitrate. The average SSIM results for these sets are given in table

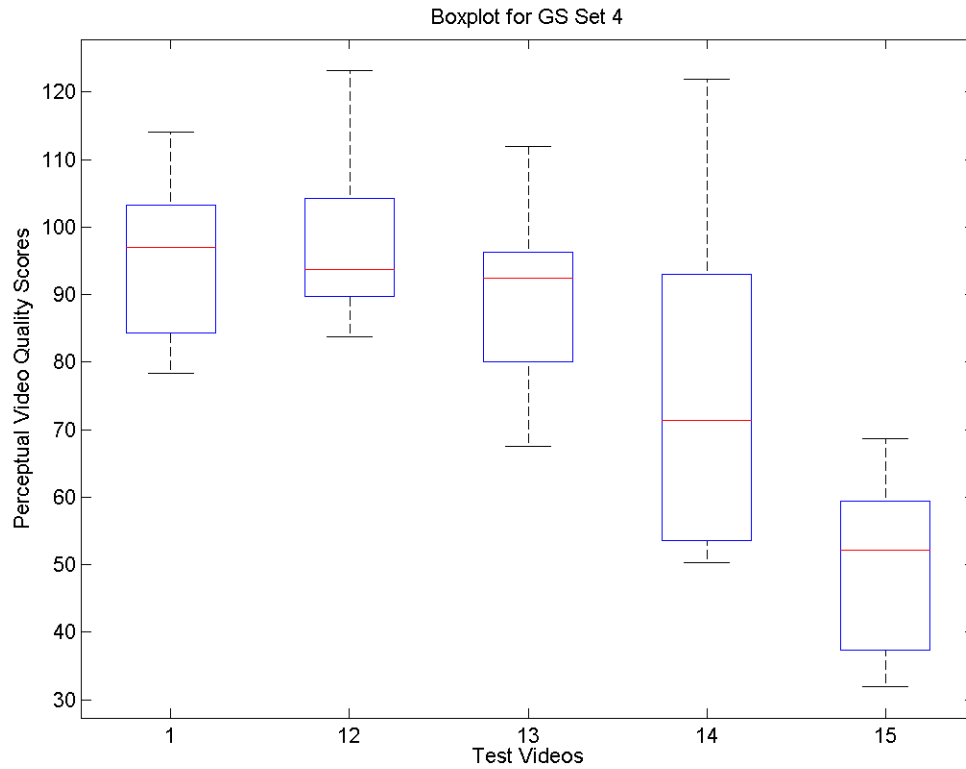


Figure 4.19: Glide Scope Perceptual Video Quality Ratings for Set 4

Source	SS	df	MS	H	p-value
Columns	5419.75	4	1354.9375	25.5084	3.9752e-05
Error	4991.25	45	110.9167		
Total	10411	49			

Table 4.23: Kruskal-Wallis one-way analysis of variance for Set 4

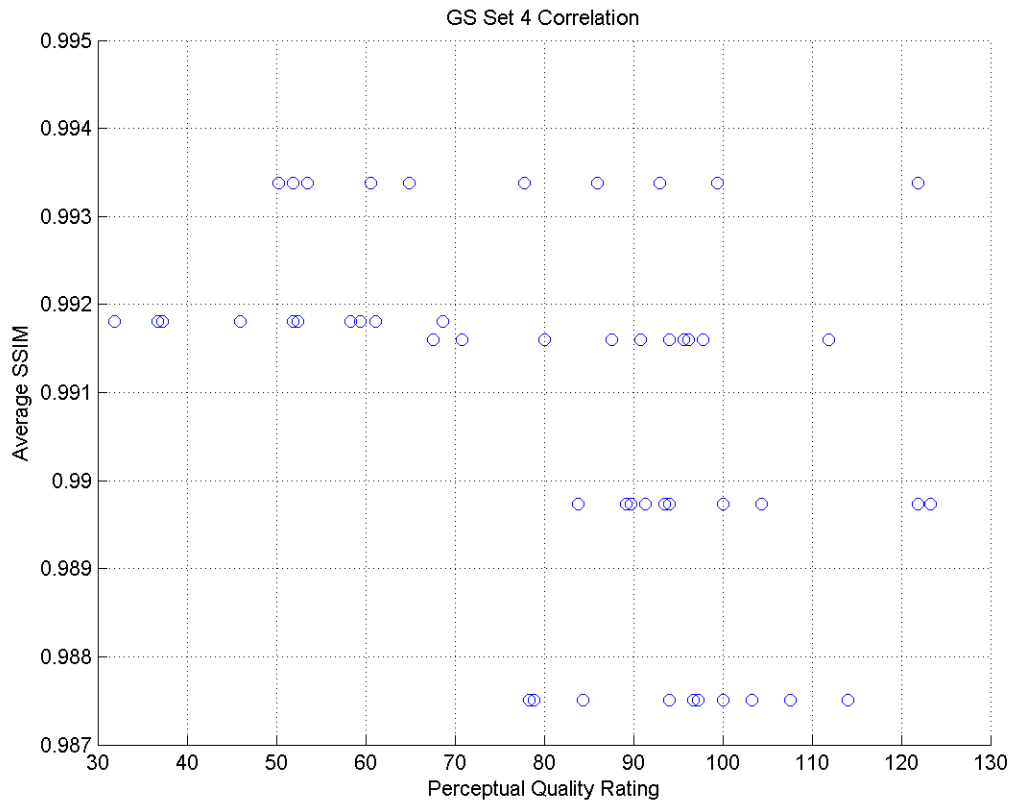


Figure 4.20: Correlation between average frame image quality and perceptual quality rating for Set 4

		Frame rate				
		30	15	7.5	3.75	1.875
Bitrate	250	0.9875	0.9915	0.9942	0.9960	0.9965
	200	0.9851	0.9897			
	150	0.9807		0.9916		
	100	0.9722			0.9934	
	50	0.8221				0.9918

Table 4.24: Average SSIM results for videos in sets 1, 2 and 4

Video	Bitrate	Frame Rate	Frame Size	Average SSIM
1	250 kbit/s	30	640x480	0.9875
16	250 kbit/s	15	400x300	0.9930
17	250 kbit/s	7.5	256x192	0.9935

Table 4.25: Glide Scope Average SSIM Results for Set 5

4.24. These results show that the average frame quality increases across all bitrates when the frame rate is lowered and decrease with bitrate across all frame rates. These results show that average frame image quality decreases with bitrate across all frame rates and that it increases for lower frame rates across all bitrates.

4.2.5 Set 5: Frame Rate and Frame Size

The average SSIM results for test set 5, which consists of videos encoded at 250 kbit/s and from rates form 7.5 to 30 fps and frame sizes from 256x192 to 640x480, is presented in table 4.25.

Figure 4.21 shows the distribution of perceptual video quality ratings for set 5 while Figure 4.22 shows these ratings plotted against the calculates average SSIM vales for the videos in this set. The Spearman's rank correlation coefficient between average frame image quality and perceptual quality rating was calculated to be $\rho = -0.0543$ ($p = 0.7758$).

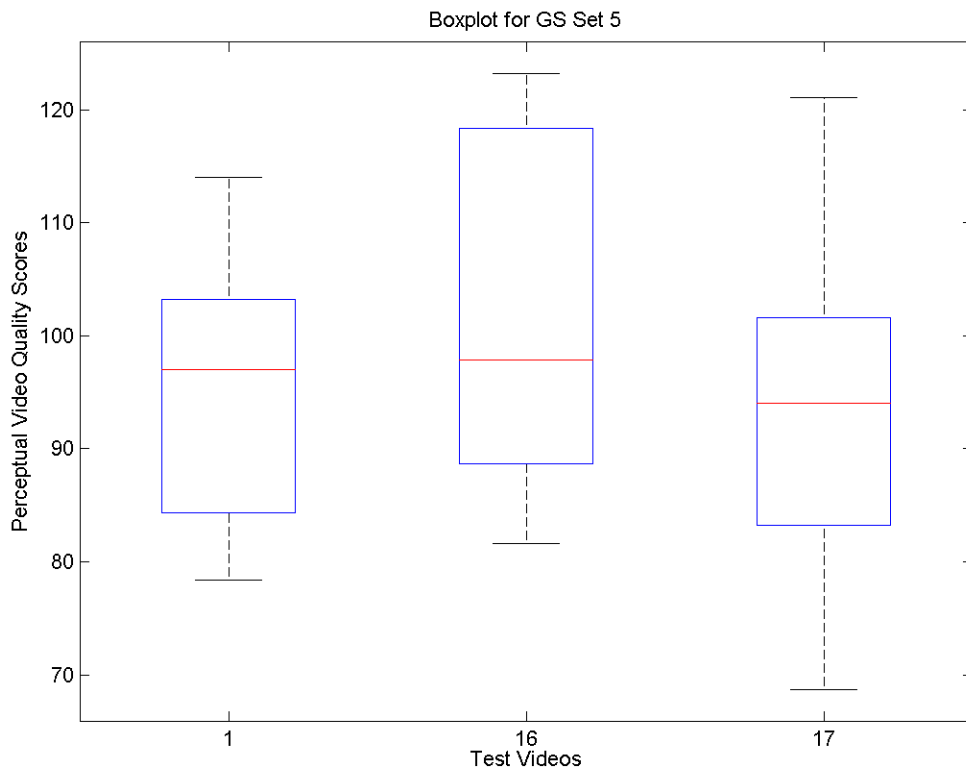


Figure 4.21: Glide Scope Perceptual Video Quality Ratings for Set 5

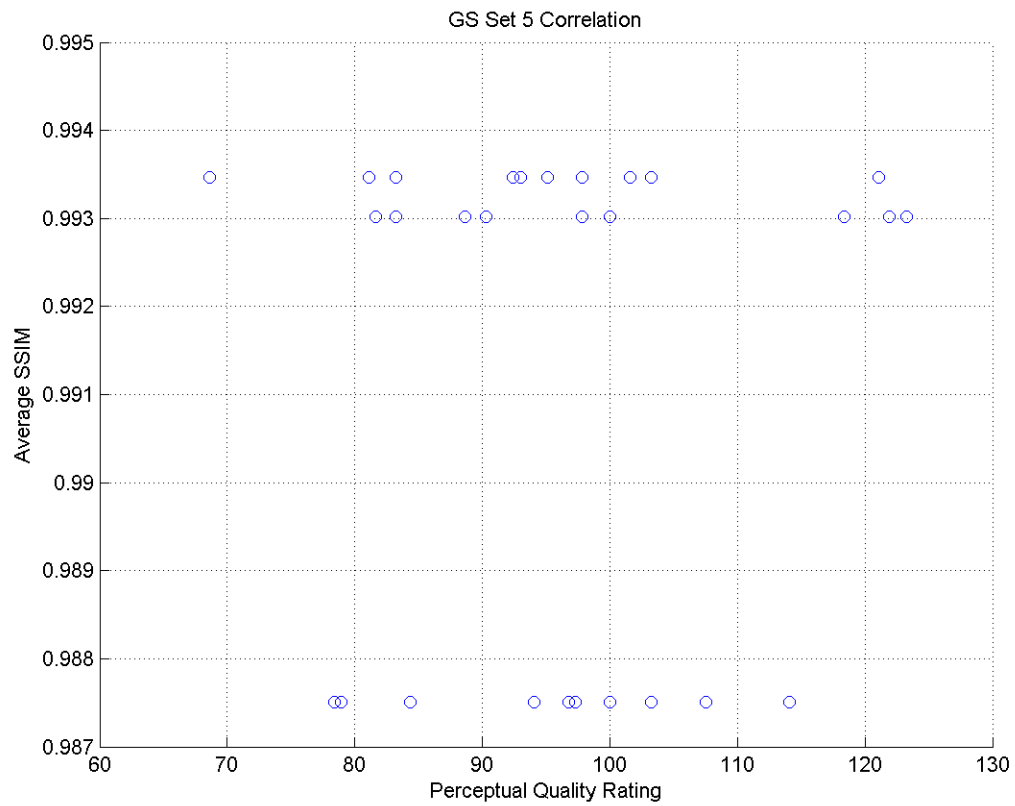


Figure 4.22: Correlation between average frame image quality and perceptual quality rating for Set 5

Source	SS	df	MS	H	p-value
Columns	52.55	2	26.275	0.67867	0.71224
Error	2192.95	27	81.2204		
Total	2245.5	29			

Table 4.26: Kruskal-Wallis one-way analysis of variance for Set 5

		Frame rate				
		30	15	7.5	3.75	1.875
Frame size	640x480	0.9875	0.9915	0.9942	0.9960	0.9965
	400x300	0.9914	0.9930			
	256x192	0.9886		0.9935		

Table 4.27: Average SSIM results for videos in sets 1, 3 and 5

Kruskal-Wallis one-way analysis of variance results for set 5 is given in Table 4.26. It was found that there was no significant difference in the perceptual quality ratings for any of the videos in the set.

Set 5 can be directly compared to set 1 (which includes videos with varying frame rates) and set 3 (which includes videos with varying frame sizes). The average SSIM values for these sets can be seen in Table 4.27. The frame image quality increases for smaller frame rates across all frame sizes similar to the trend observed in the simulation room view video. With the exception of one video, the average SSIM increases with smaller frame sizes across the three frame rates investigated.

4.2.6 Set 6: Bitrate and Frame Size

Set 6 consists of videos encoded at 30 fps, bitrates ranging from 100 kbit/s to 200 kbit/s and frame sizes from 256x192 to 640x480. The encoding parameters and average frame image quality for these videos are given in Table 4.28.

The perceptual quality ratings for this set are shown in the boxplot in Figure

Video	Bitrate	Frame Rate	Frame Size	Average SSIM
6	200 kbit/s	30	640x480	0.9851
18	150 kbit/s	30	400x300	0.9873
19	100 kbit/s	30	256x192	0.9859

Table 4.28: Glide Scope Average SSIM Results for Set 6

Source	SS	df	MS	H	p-value
Columns	1.95	2	0.975	0.025178	0.98749
Error	2244.05	27	83.113		
Total	2246	29			

Table 4.29: Kruskal-Wallis one-way analysis of variance for Set 6

4.23 and again plotted against average SSIM in the scatterplot in Figure 4.24. A Spearman's rank correlation coefficient of $\rho = -0.0071$ ($p = 0.9704$) was calculated between the average frame image quality and the perceptual quality ratings.

Kruskal-Wallis one-way analysis of variance of ranks results for set 6 is given in Table 4.29. There is no statistical difference in the distribution of ranks of the perceptual ratings for any of the videos in the set.

Test set 6, which was encoded with varying frame size and bitrate, combines the varying encoding parameters of sets 2 (bitrate) and 3 (frame size). The average SSIM of these sets is directly compared in Table 4.30. As with the room view video results, frame image quality decreases across all frame sizes with lower encoding frame rate. Additionally smaller encoding frame sizes lead to higher SSIM values with one exception.

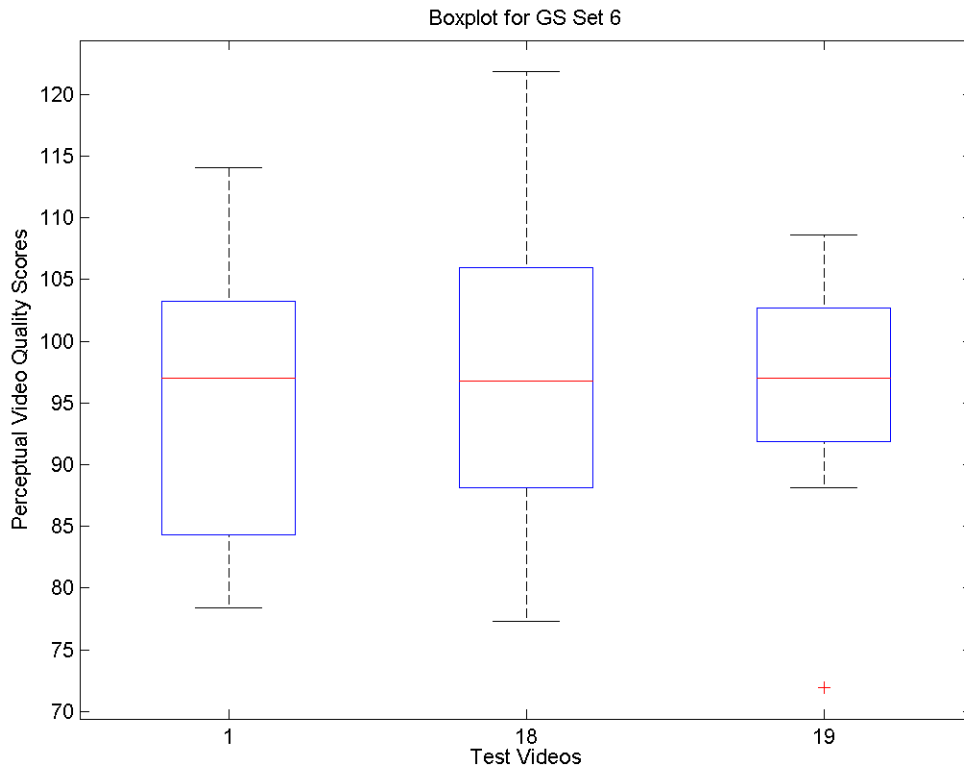


Figure 4.23: Glide Scope Perceptual Video Quality Ratings for Set 6

		Bitrate				
		250	200	150	100	50
Frame size	640x480	0.9875	0.9851	0.9807	0.9722	0.8221
	400x300	0.9914		0.9873		
	256x192	0.9886			0.9859	

Table 4.30: Average SSIM results for videos in sets 2, 3 and 6

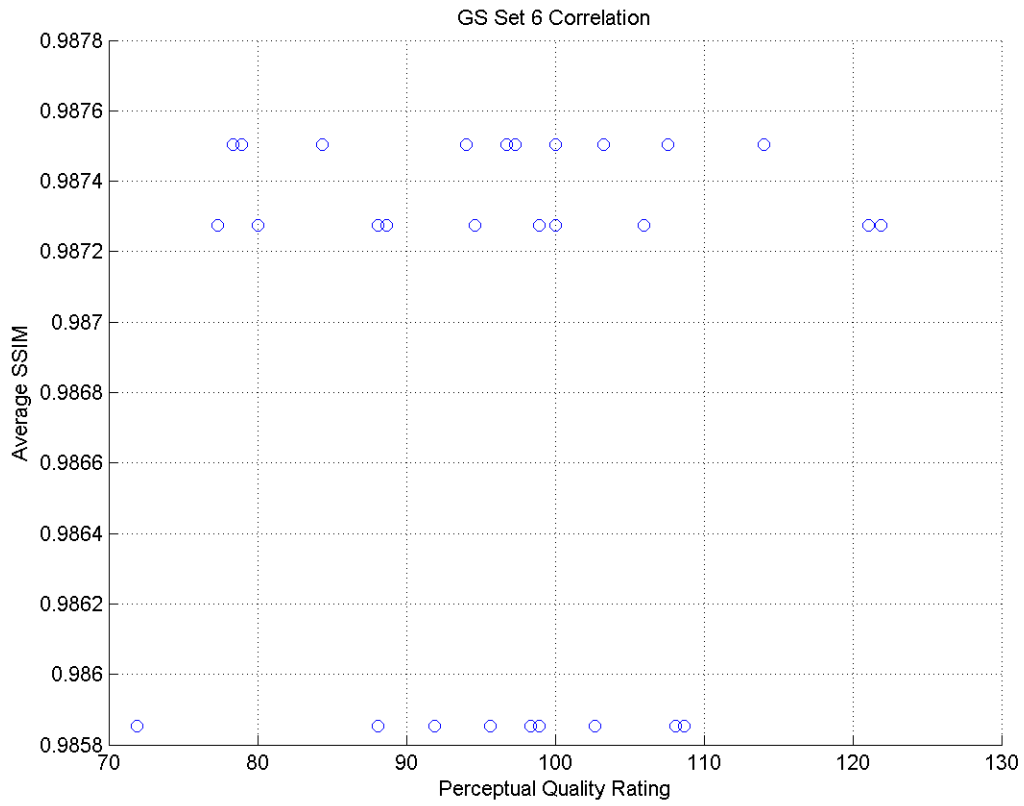


Figure 4.24: Correlation between average frame image quality and perceptual quality rating for Set 6

Chapter 5

Model Development

5.1 Simulation Room View Video

The following observations and models are based on the average frame image quality and perceptual video quality results of the simulation room view video presented in Chapter 4. Test video 9 was classified as an outlier due to the fact that the encoding ratio, frame image quality and perceptual rating for this video were all far removed from the rest of the data. The results for this video were removed from the data to ensure that the single video did not impact the models created.

5.1.1 Interpretation of Results

The following observations can be made from the results in Chapter 4 about the relationships between encoding parameters (frame rate, frame size and bitrate), average frame image quality (as measured by average SSIM) and the perceptual video quality rating of the simulation room view video.

1. Decreases in frame rate results in higher average frame image quality (Sections 4.1.1, 4.1.4 and 4.1.5).
2. Decreases in bitrate results in lower average frame image quality (Sections 4.1.2, 4.1.4 and 4.1.6).
3. Smaller frame size results in higher average frame image quality (Sections 4.1.3, 4.1.5 and 4.1.6).
4. When videos encoded at the same frame rate are compared, a strong positive correlation exists between the average frame image quality and the perceptual video quality rating (Sections 4.1.2 and 4.1.3).
5. Sets encoded with varying frame rate show a negative correlation between the average frame image quality and the perceptual quality ratings (Section 4.1.1).

The first three observations describe how frame rate, bit rate and frame size each influence the average frame image quality. These observations contributed to the equation and model for average frame image quality described in Sections 5.1.2 and 5.1.3.

From the last two observations, it can be ascertained that the perceptual video quality rating depends primarily on frame rate and average frame image quality. These observations lead to the model for perceptual video quality rating described in Section 5.1.4.

5.1.2 Encoding Ratio

In order to simplify the process of analyzing the results, a new value was introduced to combine the effects of the three encoding parameters under consideration. The

encoding ratio, denoted by x is the ratio of pixels to bits in encoded videos. The formula is as follows

$$x = \frac{r \times s}{b} \quad (5.1)$$

where r represents *frame rate*, s represents *frame size* and b represents *bit rate* of the encoded video.

A similar concept was introduced by Rao et al. called *bpp* or *bits per pixel* and is calculated as the inverse of equation 5.1 [29]. The *bpp* was used to designate the quality level of a video. For example, a video encoded at 30 fps and 360x240 pixels using MPEG-2 was found to be diagnostically lossless at a *bpp* of 0.42 [29]. Implicit in this conclusion is the concept that diagnostic video quality can be accurately quantified by the *bpp* (or the encoding ratio) of the encoded video. This thesis tests this assumption using the results collected and leading to the model created for average frame image quality.

5.1.3 Model 1: Frame Image Quality

The average SSIM was plotted against the encoding ratio for all the test videos (with the exception of video 9 as discussed above). Three alternative models were fit to the data using the least squares method and the goodness of fit of the first, second and third order polynomial models is provided in Table 5.1. In this table, R-square refers to the square of the correlation between the actual and predicted values [52]. The adjusted R-square adjusts this value based on the degrees of freedom and is be used to compare the goodness of fit of multiple models [52]. Models with a R-squared or adjusted R-squared value closer to 1 demonstrate a better fit [52]. Root mean square error (RMSE) is also used to determine the godnss of fit [52]. Models with a lower

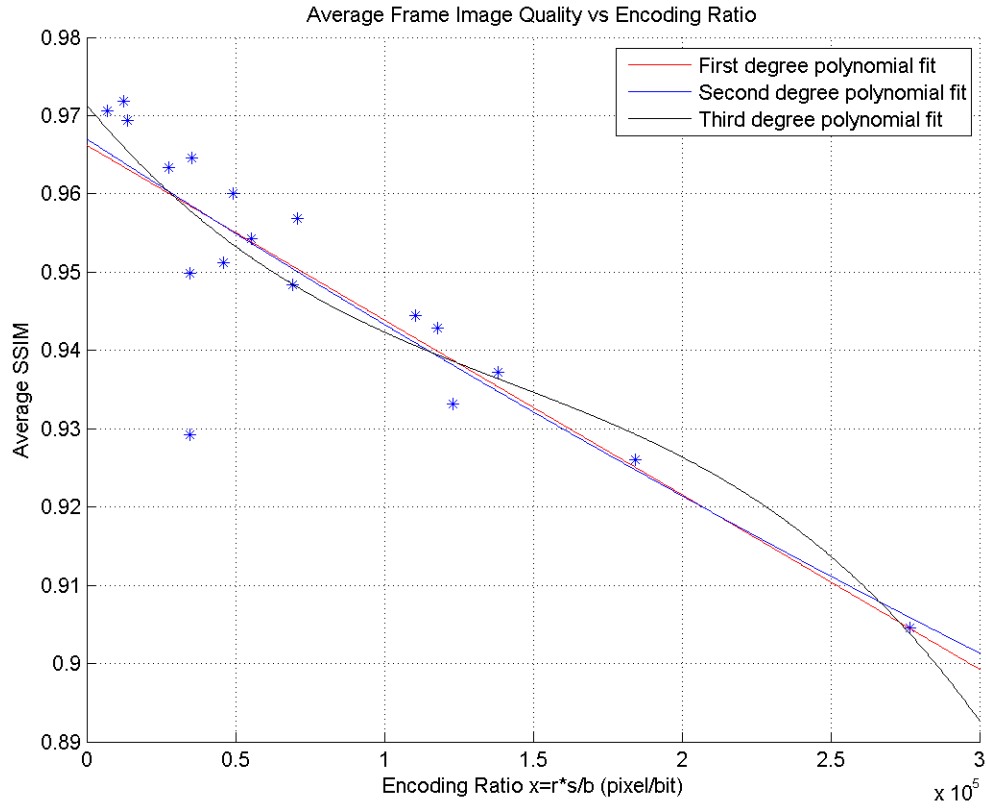


Figure 5.1: Alternative Models for Frame Image Quality based off Objective Frame Quality Results

RMSE have a better fit than models with a higher RMSE value.

As shown in Figure 5.1, the higher order models were close to linear. The linear model (shown in Figure 5.2) also had the best fit of the three (lowest RMSE and highest adjusted R-square) and so it was selected over the higher order models.

The equation for the function relating the frame image quality to encoding ratio is

$$SSIM[x] = -2.23 \times 10^{-7}x + 0.966. \quad (5.2)$$

The negative slope shown in Figure 5.2 can be interpreted as the frame image

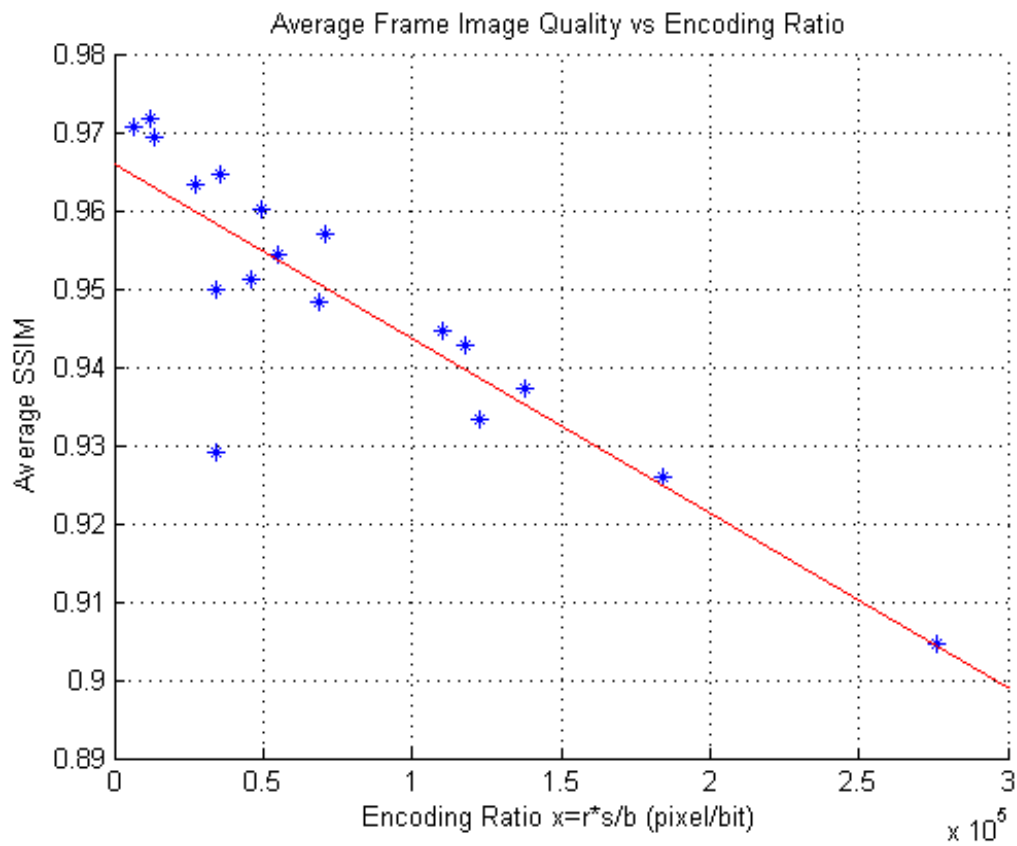


Figure 5.2: Model for Frame Image Quality based off Objective Frame Quality Results

	Poly 1	Poly 2	Poly 3
R-square	0.7645	0.7654	0.7768
Adjusted R-square	0.7498	0.7341	0.729
RMSE	0.008929	0.009203	0.009292

Table 5.1: Goodnes of Fit of alternative models for Frame Image Quality

quality decreasing as the number of pixels per bit increases. For example, increasing the frame rate or frame size while maintaining the same bit rate will increase the number of pixels for each bit (encoding ratio) and result in a lower average frame image quality. Conversely, when the bit rate is increased for the same frame rate and size, the encoding ratio is decreased due to a larger number of bits and the resultant average frame image quality goes up.

5.1.4 Model 2: Human Video Quality Rating

It was observed that the perceptual video quality rating is dependent on both frame rate and frame image quality. The relationship between these three values can be understood by plotting them as shown in Figure 5.3.

Using the least squares method, first, second and third order polynomial models were fit to the data. The goodness of fit of these models is presented in Table 5.2. The second degree model was chosen because it had the best fit of the three alternatives (lowest RMSE and highest ajusted R-square).

	Poly 1	Poly 2	Poly 3
R-square	0.8011	0.9877	0.9912
Adjusted R-square	0.7746	0.9825	0.9812
RMSE	9.243	2.574	2.667

Table 5.2: Goodness of Fit of alternative models for Pereptual Video Quality

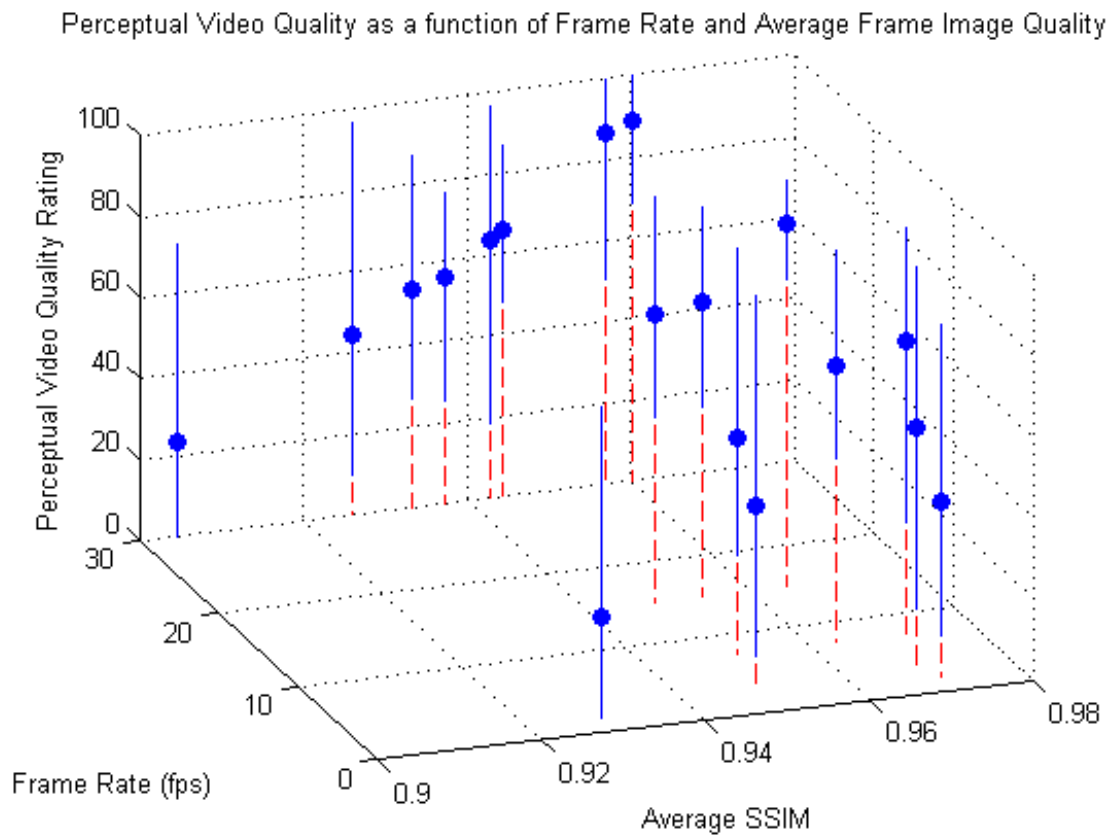


Figure 5.3: Plot of perceptual video quality ratings; minimum, maximum and mean scores.

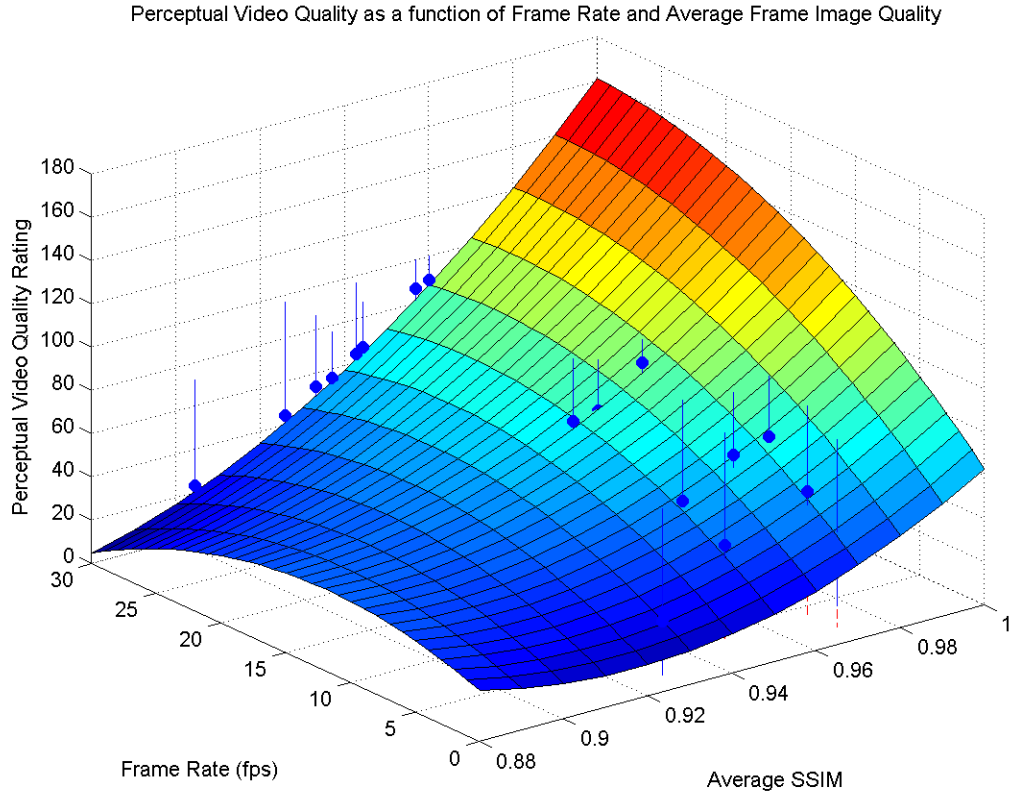


Figure 5.4: Surface plot of perceptual video quality model along with plot of actual minimum, maximum and mean ratings.

Figure 5.4 contains a surface plot of the selected model which can be described by the following equation:

$$R[SSIM, r] = 6993SSIM^2 - 9.968 \times 10^{-2}r^2 + 32.52SSIM \times r - 1.282 \times 10^4SSIM - 26.26r + 5890 \quad (5.3)$$

where R represents the predicted perceptual video quality rating, $SSIM$ represents the frame image quality and r represents the frame rate of the encoded

video. Values of perceptual quality rating less than zeros are not meaningful so where $R[SSIM, r] < 0$ as defined by equation 5.3, it can be assumed that $R[SSIM, r] = 0$ as is shown in Figure 5.4.

5.1.5 Composite Model

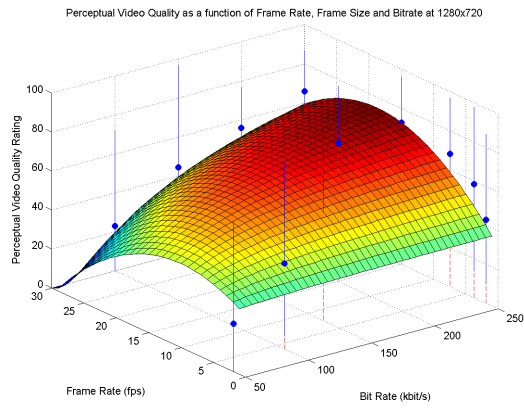
By combining equations 5.1, 5.2 and 5.3, a composite model can be created to determine how the three encoding parameters interact to influence perceptual video quality. The equation for this model is

$$R[b, r, s] = 3.48 \times 10^{-10} \frac{r^2 s^2}{b^2} - 7.25 \times 10^{-6} \frac{r^2 s}{b} - 1.54 \times 10^{-4} \frac{r s}{b} - 9.97 \times 10^{-2} r^2 + 5.14r + 31.4 \quad (5.4)$$

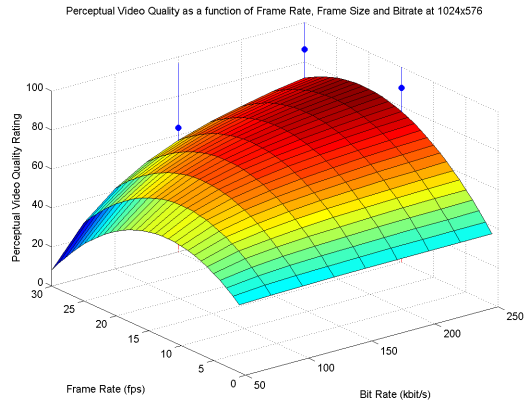
where R represents the predicted perceptual video quality rating and b , s and r represent the bit rate, frame size and frame rate of the encoded video. Again values of quality ratings less than zero are not meaningful so it should be assumed $R[b, r, s] = 0$ where $R[b, r, s] < 0$ as defined by equation 5.4. Figure 5.5 shows the model plotted as a function of frame rate and bit rate for various frame sizes along with the actual perceptual video quality ratings.

5.2 Glide Scope Video

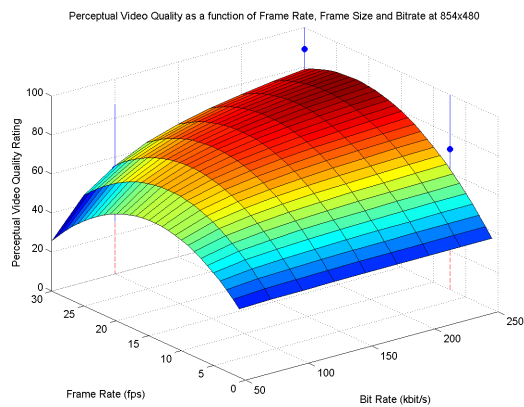
The following observations and models are based off the average frame image quality and perceptual video quality results of the glide scope video presented in Chapter 4. Due to being identified as an outlier, the results for test video 9 were removed from



(e) Frame size = 1280x720



(f) Frame size = 1024x576



(g) Frame size = 854x480

Figure 5.5: Surface plot of perceptual video quality model as a function of bit rate, frame rate and frame size along with plot of actual minimum, maximum and mean ratings.

the data again.

5.2.1 Observations from Previous Data

The following observations were made from the results of the glide scope video included in Chapter 4 and used in the development of the models included later in this section.

1. Decreases in frame rate results in higher average frame image quality (Sections 4.2.1, 4.2.4 and 4.2.5).
2. Decreases in bitrate results in lower average frame image quality (Sections 4.2.2, 4.2.4 and 4.2.6).
3. Smaller frame size results in higher average frame image quality (Sections 4.2.3, 4.2.5 and 4.2.6).
4. When videos encoded at the same frame rate and frame size are compared, a strong positive correlation exists between the average frame image quality and the perceptual video quality rating (Sections 4.2.2).
5. Sets encoded with varying frame rate show a negative correlation between the average frame image quality and the perceptual quality ratings (Section 4.1.1).

These observations can again be separated into factors affecting the average frame image quality (1-3) which will contribute to the models in Sections 5.2.2 and 5.2.3 and those affecting the perceptual video quality (4-5) which will contribute to the model in Sections 5.2.4.

	Poly 1	Poly 2	Poly 3
R-square	0.8699	0.8714	0.877
Adjusted R-square	0.8618	0.8542	0.8506
RMSE	0.002197	0.002256	RMSE: 0.002284

Table 5.3: Goodness of fit of alternative models for Frame Image Quality

5.2.2 Encoding Ratio

The same variable *encoding ratio* as used to build the models for simulation room view video will be used for the models for the glide scope video. Again, the encoding ratio denoted by x in equation 5.1 represents the ratio of pixels to bits in the encoded video.

5.2.3 Model 1: Frame Image Quality

The average frame image quality for all the Glide Scope videos was plotted as a function of the encoding ratio and three models of first second and third degree polynomials were fit to the data as shown in Figure 5.6. Table 5.3 shows the goodness of fit of these three models. The linear model shown in Figure 5.7 was chosen because it had the best fit (lowest RMES, highest adjusted R-square).

The equation for the function relating the frame image quality to encoding ratio is

$$SSIM[x] = -2.34 \times 10^{-7}x + 0.995. \quad (5.5)$$

5.2.4 Model 2: Human Video Quality Rating

Similar to the simulation room view video, the observations of the glide scope video results point to the fact that perceptual video quality depends on the frame rate a

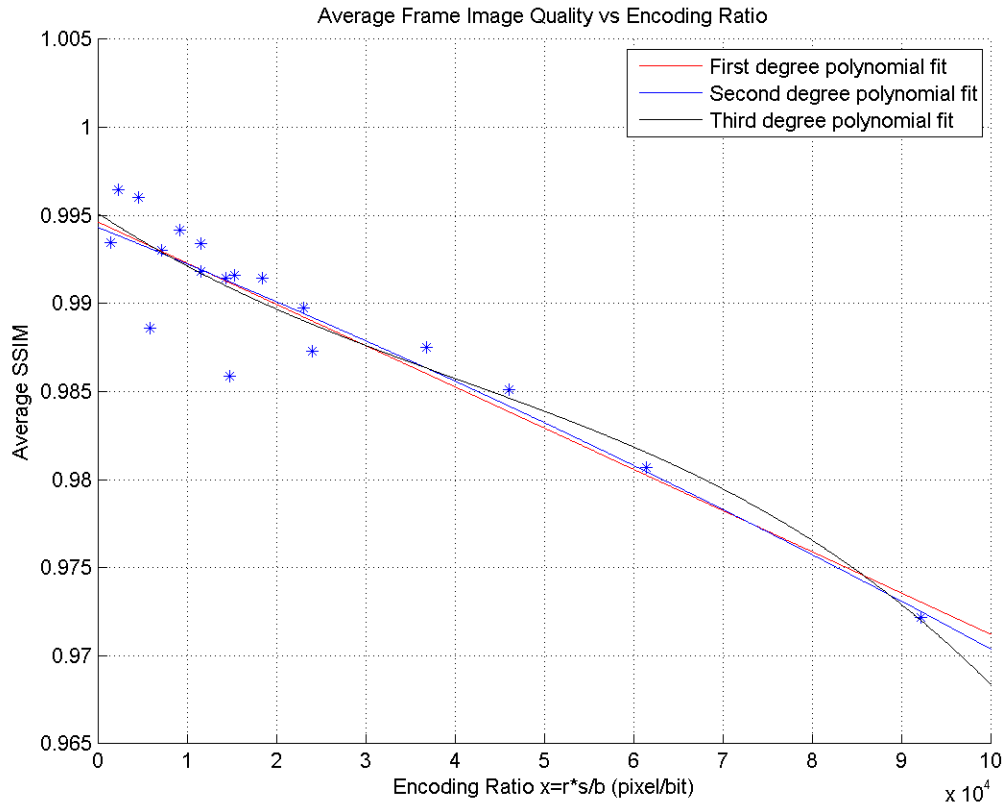


Figure 5.6: Alternative Models for Frame Image Quality based off Objective Frame Quality Results

video is encoded at and the average frame image quality of the video. Figure 5.8 shows a plot of the minimum, maximum and mean perceptual video quality ratings for the glide scope videos as a function of both average frame image quality and encoding frame rate. For this fitting, all test videos were included.

Three models of first, second and third degree polynomials were fit to the data using the least-squares method. Table 5.4 shows the goodness of fit for the three models attempted. The third degree polynomial model has the best fit (lowest RMSE and highest R-square), however the second degree polynomial model was chosen instead

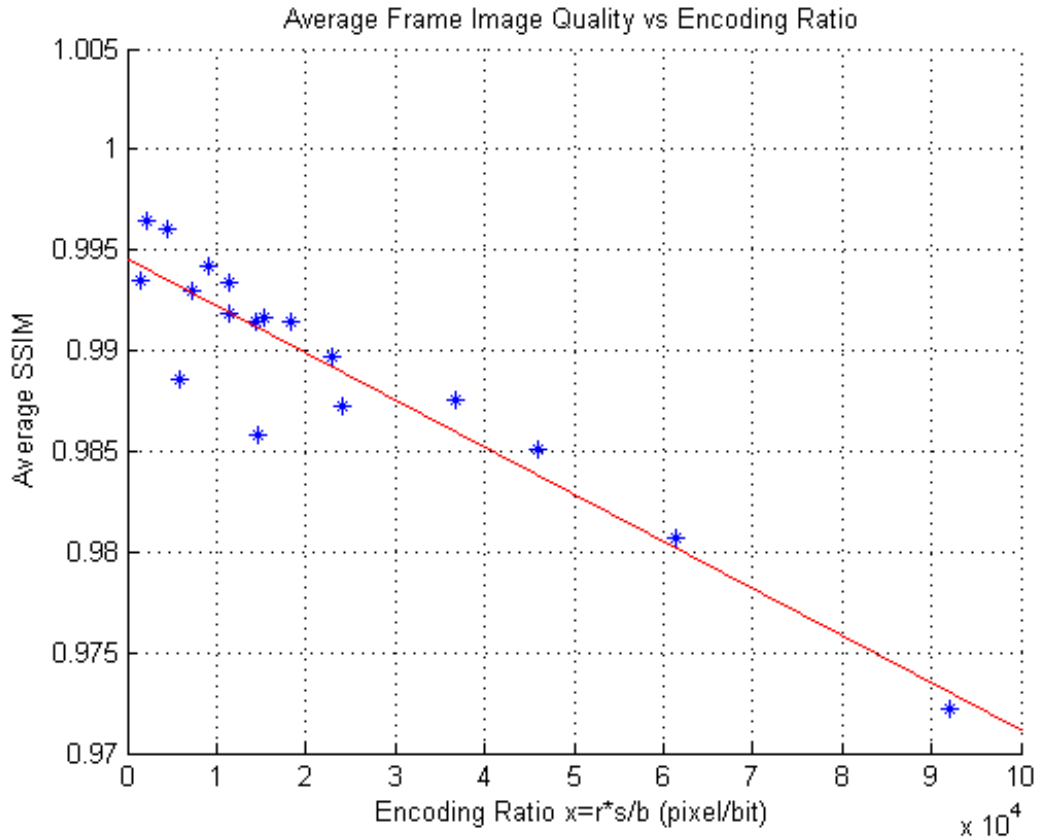


Figure 5.7: Model for Frame Image Quality based off Objective Frame Quality Results

because it offered a simpler equation with fewer coefficients. In a real-time application, the computational effort added due to the use of a third order equation is not justified by the marginal improvement in fit.

	Poly 1	Poly 2	Poly 3
R-square	0.5039	0.9134	0.9777
Adjusted R-square	0.4377	0.8773	0.9527
RMSE	10.61	4.958	3.078

Table 5.4: Goodness of Fit of alternative models for Perceptual Video Quality

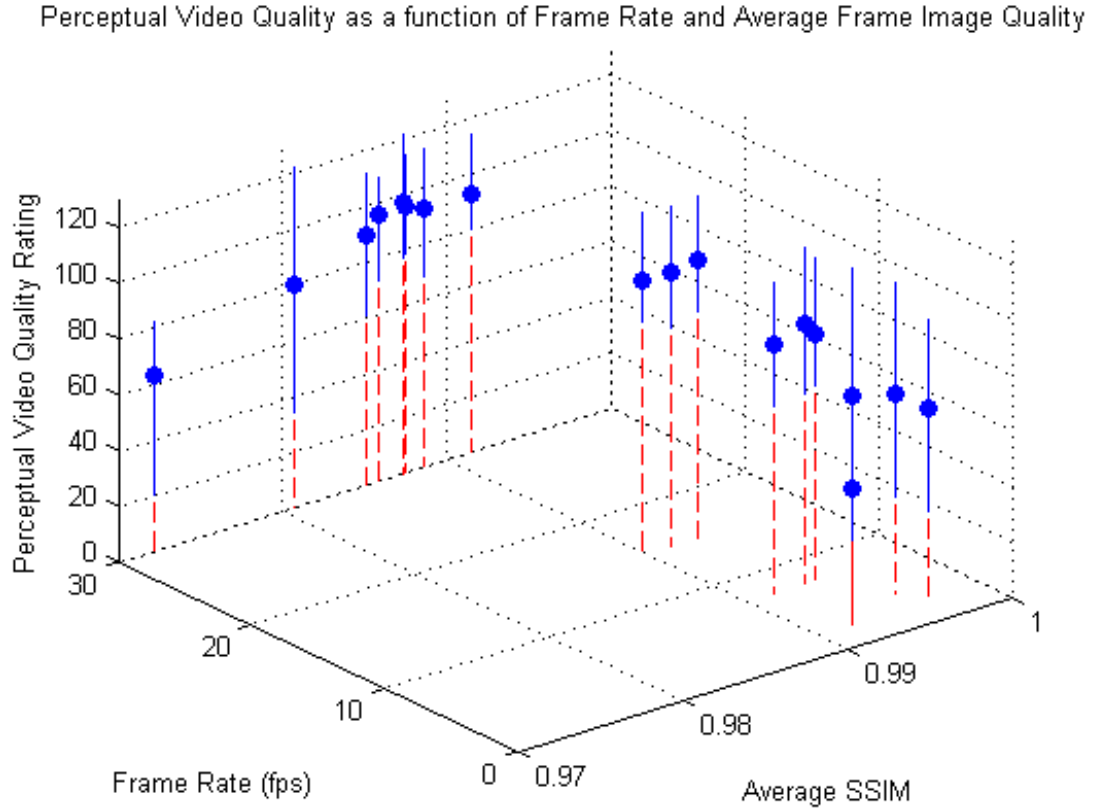


Figure 5.8: Plot of perceptual video quality ratings; minimum, maximum and mean scores.

This model is shown in Figure 5.9 and is described by the equation:

$$R[SSIM, r] = -9.918 \times 10^4 SSIM^2 - 0.1449 \times 10^{-2} r^2 - 116.4 SSIM \times r + 1.999 \times 10^5 SSIM + 121.4r - 1.007 \times 10^5 \quad (5.6)$$

where R represents the predicted perceptual video quality rating, $SSIM$ represents the frame image quality and r represents the frame rate of the encoded video. Values for the perceptual quality rating less than zero are not meaningful so it can

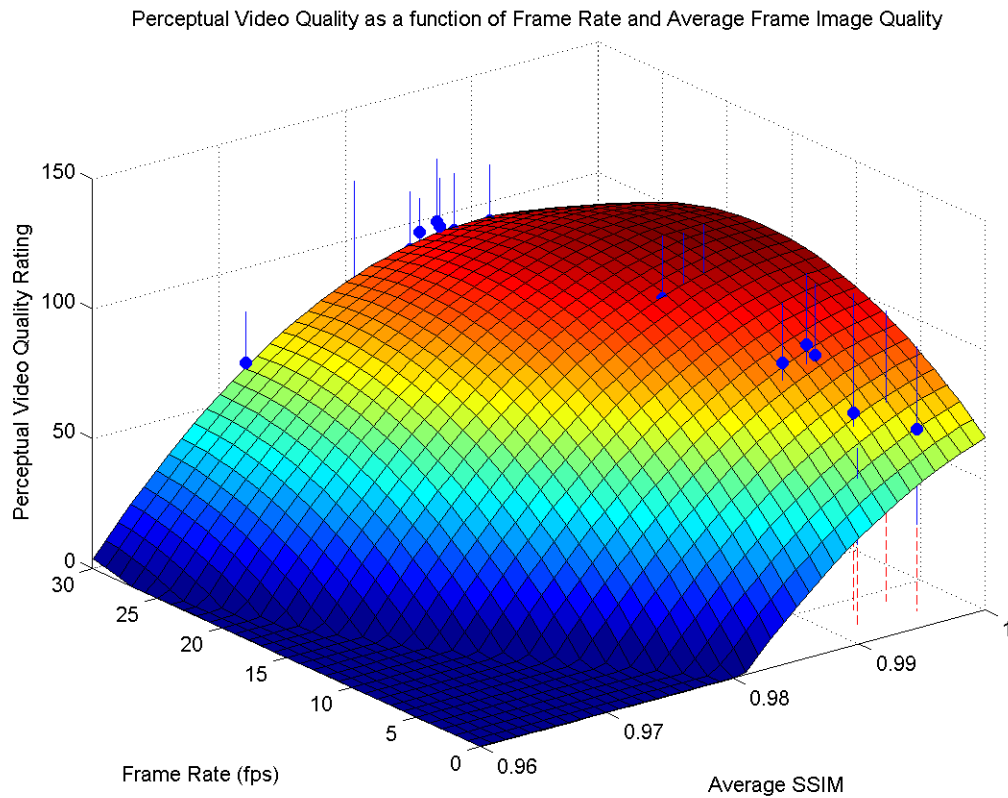


Figure 5.9: Surface plot of perceptual video quality model along with plot of actual minimum, maximum and mean ratings.

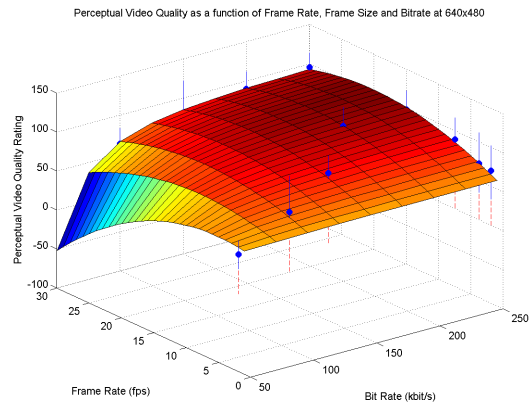
be assumed that $R = 0$ where $R < 0$ as defined by equation 5.6. This is reflected in Figure 5.9.

5.2.5 Composite Model

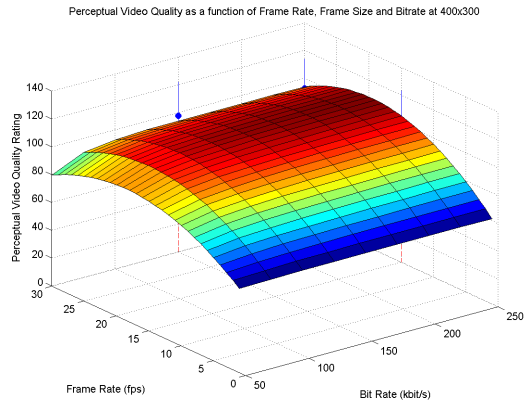
In order to describe how the perceptual quality of a video is described solely based on its encoding parameters (frame rate, bitrate and frame size), equations 5.1, 5.5 and 5.6 above were combined to create a single composite model as shown in equation 5.7,

$$R[b, r, s] = -5.43 \times 10^{-9} \frac{r^2 s^2}{b^2} + 2.72 \times 10^{-5} \frac{r^2 s}{b} - 5.95 \times 10^{-4} \frac{rs}{b} - 0.145r^2 + 5.54r + 56.2 \quad (5.7)$$

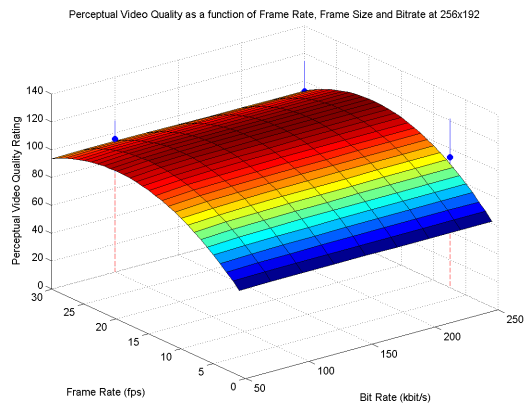
where R represents the predicted perceptual video quality rating and b , s and r represent the bit rate, frame size and frame rate of the encoded video. Figure 5.10 shows the model plotted as a function of frame rate and bit rate for various frame sizes along with the actual perceptual video quality ratings. Although not covered by any points on these plots, it can be assumed again that $R = 0$ where $R < 0$ as defined by equation 5.7.



(a) Frame size = 640x480



(b) Frame size = 400x300



(c) Frame size = 256x192

Figure 5.10: Surface plot of perceptual video quality model as a function of bit rate, frame rate and frame size along with plot of actual minimum, maximum and mean ratings.

Chapter 6

Discussion

6.1 Limitations

The models created in the previous chapter have several limitations. As can be seen when the results for the simulation room view video and the glide scope video are compared, the models differ significantly between the two video types. The actual mathematical models created from the data collected are dependent on both the video itself and the encoding used. For example viewer perception and encoding efficiency are likely to differ dramatically between videos with different content. The complexity of motion, colours and shapes in a video contributes to differences in perceived quality as well as encoding efficiency. Differences in encoding methods (codecs, encoding standards, settings) will also introduce variation in the resulting models. For example, MPEG-2 encoding results in a much lower quality for the same pixel per bit ratio than H.264 encoding and even among codecs adhering to the H.264/AVC standard, there can be disparity in the encoding efficiency.

It is also important to keep the purpose of the video in mind when looking at

the results from this and other studies. These tests were designed to measure the effectiveness of the video to convey room and task awareness. Because of this these models describe specifically the diagnostic and educational quality of these videos. Again, it is important to note that this will vary between medical modality, further emphasizing the dependence of these models on the video content and purpose.

6.2 Benefits

Despite these limitations, the models created demonstrate several properties that can lead to a better understanding of diagnostic and educational video quality and how the quality is affected by interactions between encoding parameters. Specifically, the linear relationship between pixels per bit (referred to as encoding ratio) and frame image quality held true for both test videos despite a difference in the slope of the relationship. Additionally, the perceptual video quality ratings of both videos appeared to be dependent most significantly on frame rate and frame image quality, although each video did show a difference weighting of importance of the two factors.

6.3 Uses

The observations and models discussed in this thesis can be useful to inform the encoding of telemedicine videos over slow or inconsistent internet connections. For streaming techniques that adapt to the available network conditions, the models created can provide a guide or algorithm to determine the encoding frame rate and size to provide the highest diagnostic or educational video quality with little or no user interaction.

6.3.1 Example

Consider the scenario of training a medic in a remote location (such as a remote research station) to perform a procedure to stabilize a patient for transport, such as an intubation. The high travel time and cost required for in person instruction necessitates the use of remote instruction and telesimulation to accomplish this training. (Another benefit is just in time training in which the medic could receive the required training at any time as the need for it arises.)

Available internet access at remote research stations is likely limited to satellite internet connections. These connections require a signal to be sent from earth to a satellite in orbit and back down to a different location on earth and are often hampered with high delay times and inconsistency depending on the location and conditions of the remote site.

An instructor mentoring a medic to complete an intubation of a simulated patient would require 2 video feeds accomplish proper instruction: a glide scope video to view the procedure in detail and a room view video to understand the context of the medical event (patient, location, equipment, personnel, etc.).

With the help of the available models, encoding of each of the 2 video feeds could be accomplished according to the appropriate model to meet the available network conditions and to optimize the educational quality of the video. Changes in the network conditions due to instability of satellite internet connections would be accompanied by changes in encoding in order to comply with the models.

The use of these models allows for streaming telesimulation video over poor internet connections with the quality required for the specific purpose of the video.

Automated processes to optimize quality will require no user interaction while ensuring a high quality of service to the user.

6.4 Further Directions

While the models as they currently exist are capable of improving the decision making process of existing scalable video coding, even more potential lies in the frequent or continual collection of perceptual quality data in order to further refine the existing models. Collection of this data would allow for the use of machine learning in order to continually refine and improve the models of perceptual video quality to provide better quality of service for various medical modalities, users and video content.

Chapter 7

Conclusion

The average frame image quality of a video clip encoded with H.264/AVC video encoding was found by this thesis to have a linear relationship to the ratio of pixels to bits of the encoded video. Practically, this finding implies that the frame image quality of a video decreases as the number of pixels per bit increases. The equations defining this relationship were found to be different for each of the videos tested in the thesis (Figures 5.2, 5.7 and equations 5.2, 5.5) suggesting that factors beyond the pixel to bit ratio also influence the frame image quality such as aspect ratio and video content.

It was also found that the perceptual quality of telemedical video depends primarily on the frame rate and the frame image quality of the video. How these factors combine to affect perceptual quality is unique for videos with different content and purpose, such as the room view and glide scope videos discussed in this thesis.

The following models were created for each of the video types discussed in this thesis:

- Perceptual quality as a function of frame rate and frame image quality

- Perceptual quality as a function of encoding parameters (frame rate, frame size and encoded bitrate).

The models created in this thesis have the potential to greatly improve the quality of experience of telemedicine systems through smarter allocation of bits when encoding video for limited bitrates. The existence of multiple models for different types of videos allows for video encoding specific to the particular purpose of that video. Overall, this has the potential to improve the convenience, quality and usefulness of telemedicine especially in situations requiring low bitrate videos (such as locations employing satellite or dial up internet connections) and enhancing telemedicine in these situations is a critical element of increasing the delivery of healthcare to remote populations.

Appendix A

Subjective Video Quality Test

Instructions

The following instructions were given to the participants of the subjective video quality test at the start of their test. These instructions were included for tests with both the simulation room view video and the glide scope video.

Welcome and thank you for participating in this experiment. The following introduction will present you with instructions for completing this experiment.

In this experiment you will see a number of short video clips each approximately 10 seconds on the screen in front of you.

Each time a clip is played, you should judge its quality by using the sliding scale available and press submit when the rating reflects your judgement.

Your quality rating should reflect the extent to which you are able to understand the contents of the video (not the level of distortion present

in the video).

Observe carefully the entire video clip before submitting your judgement.

Press play to begin the experiment and to continue onto the next clip after submitting a judgement.

The test begins with 5 training clips to allow you to become accustomed to the process. If you have further questions, please ask for clarification during these training clips.

Additional instructions were provided to participants during the glide scope video test to familiarize them with the contents of the video. These instructions are as follows:

The following slides will help you to understand the context of the videos to be rated in this test.

A tracheal intubation is the placement of a flexible tube into a patient's trachea in order to maintain an open airway while anesthetized (among other purposes). A GlideScope video laryngoscope can be used during intubation to visualize the larynx and assist in difficult procedures.

The videos to be rated in this test are taken using a GlideScope video laryngoscope during a tracheal intubation of a medical manikin. The next few slides will provide you with a brief introduction to the procedure and anatomy to better understand the contents of the video clips.

This diagram [Figure A.1] shows how the breathing tube is positioned in a patient's airway during a tracheal intubation. The tube must be inserted into the patient's trachea (C) while the esophagus (D) is avoided.

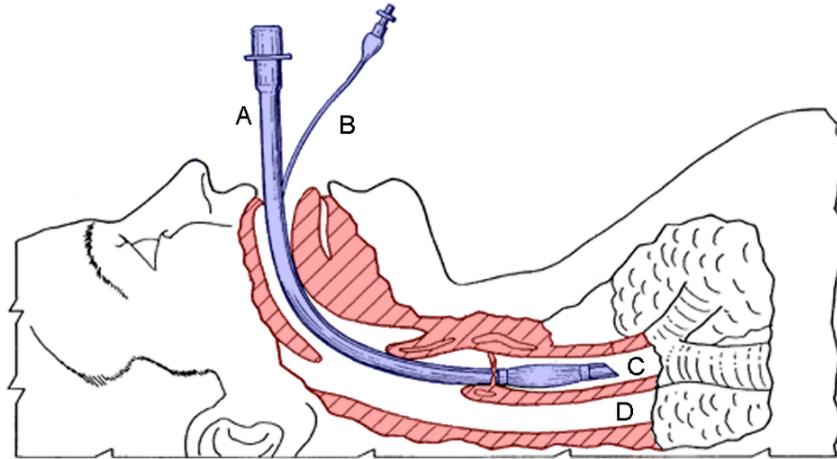


Figure A.1

This image [Figure A.2] depicts how the glide scope is inserted into the patient's airway during the procedure. The illuminated area represents the area visible in the video.

This diagram [Figure A.3] depicts the anatomical features visible during the intubation in the video clips. Notice the opening for the trachea located above the opening for the esophagus.



Figure A.2

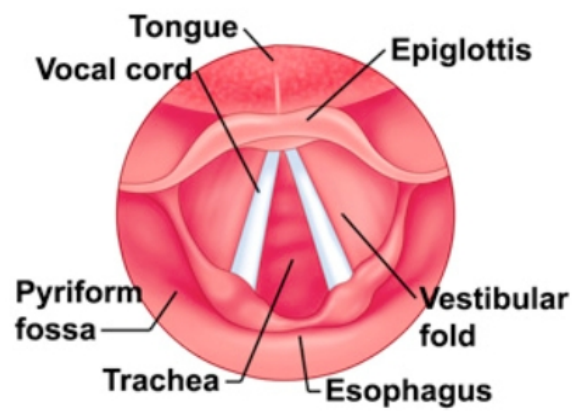


Figure A.3

Bibliography

- [1] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the h.264/avc video coding standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 7, pp. 560–576, 2003.
- [2] C.-F. Lin, "Mobile telemedicine: A survey study," *Journal of medical systems*, vol. 36, no. 2, pp. 511–520, 2012.
- [3] C. Pattichis, E. Kyriacou, S. Voskarides, M. Pattichis, R. Istepanian, and C. N. Schizas, "Wireless telemedicine systems: an overview," *Antennas and Propagation Magazine, IEEE*, vol. 44, no. 2, pp. 143–153, 2002.
- [4] V. Garshnek and F. M. B. Jr, "Applications of telemedicine and telecommunications to disaster medicine: historical and future perspectives," *Journal of the American Medical Informatics Association : JAMIA*, vol. 6, pp. 26–37, Jan-Feb 1999.
- [5] D. A. Perednia and A. Allen, "Telemedicine technology and clinical applications," *Journal of the American Medical Association*, vol. 273, no. 6, pp. 483–488, 1995.
- [6] M. E. DeBakey, "Telemedicine has now come of age," *Telemedicine Journal*, vol. 1, no. 1, pp. 3–4, 1995.

- [7] O. S. Adewale, "An internet-based telemedicine system in nigeria," *International Journal of Information Management*, vol. 24, no. 3, pp. 221–234, 2004.
- [8] W. Ponsonby, F. Mika, and G. Irons, "Offshore industry: medical emergency response in the offshore oil and gas industry," *Occupational medicine (Oxford, England)*, vol. 59, pp. 298–303, Aug 2009.
- [9] S. Binks and J. Bengler, "Tele-education in emergency care," *Emergency medicine journal : EMJ*, vol. 24, pp. 782–784, Nov 2007.
- [10] K. M. Augestad, J. G. Bellika, A. Budrionis, T. Chomutare, R. O. Lindsetmo, H. Patel, C. Delaney, and M. M. M. M. Project, "Surgical telementoring in knowledge translation—clinical outcomes and educational benefits: a comprehensive review," *Surgical innovation*, vol. 20, pp. 273–281, Jun 2013.
- [11] S. G. Mouggiakakou, E. Kyriacou, K. Perakis, H. Papadopoulos, A. Androulidakis, G. Konnis, R. Tranfaglia, L. Pecchia, U. Bracale, and C. Pattichis, "A feasibility study for the provision of electronic healthcare tools and services in areas of greece, cyprus and italy," *Biomedical engineering online*, vol. 10, 2011.
- [12] R. I. Donaldson, D. A. Mulligan, K. Nugent, M. Cabral, E. R. Saleeby, W. Ansari, S. Tajer, and M. Gausche-Hill, "Using tele-education to train civilian physicians in an area of active conflict: Certifying iraqi physicians in pediatric advanced life support from the united states," *The Journal of pediatrics*, vol. 159, no. 3, pp. 507–509. e1, 2011.

- [13] I. Choy, A. Fecso, J. Kwong, T. Jackson, and A. Okraïneç, "Remote evaluation of laparoscopic performance using the global operative assessment of laparoscopic skills," *Surgical endoscopy*, vol. 27, no. 2, pp. 378–383, 2013.
- [14] E. M. Hayden, D. D. Navedo, and J. A. Gordon, "Web-conferenced simulation sessions: A satisfaction survey of clinical simulation encounters via remote supervision," *Telemedicine and e-Health*, vol. 18, no. 7, pp. 525–529, 2012.
- [15] T. Ikeyama, N. Shimizu, and K. Ohta, "Low-cost and ready-to-go remote-facilitated simulation-based learning," *Simulation in healthcare : journal of the Society for Simulation in Healthcare*, vol. 7, pp. 35–39, Feb 2012.
- [16] A. Mikrogianakis, A. Kam, S. Silver, B. Bakanisi, O. Henao, A. Okraïneç, and G. Azzie, "Telesimulation: an innovative and effective tool for teaching novel intraosseous insertion techniques in developing countries," *Academic Emergency Medicine*, vol. 18, no. 4, pp. 420–427, 2011.
- [17] A. Okraïneç, O. Henao, and G. Azzie, "Telesimulation: an effective method for teaching the fundamentals of laparoscopic surgery in resource-restricted countries," *Surgical endoscopy*, vol. 24, no. 2, pp. 417–422, 2010.
- [18] D. K. J. E. von Lubitz, B. Carrasco, F. Gabbrielli, T. Ludwig, H. Levine, F. Patricelli, C. Poirier, and S. Richir, "Transatlantic medical education: preliminary data on distance-based high-fidelity human patient simulation training," *Studies in Health Technology and Informatics*, vol. 94, p. 379, 2003.
- [19] D. K. Via, R. R. Kyle, J. D. Trask, C. H. Shields, and P. D. Mongan, "Using high-fidelity patient simulation and an advanced distance education network to teach

- pharmacology to second-year medical students,” *Journal of clinical anesthesia*, vol. 16, no. 2, pp. 144–151, 2004.
- [20] J. B. Cooper, D. Barron, R. Blum, J. K. Davison, D. Feinstein, J. Halasz, D. Raemer, and R. Russell, “Video teleconferencing with realistic simulation for medical education,” *Journal of clinical anesthesia*, vol. 12, no. 3, pp. 256–261, 2000.
- [21] CBC News, *New satellite to boost rural internet speeds*. Online: <http://www.cbc.ca/news/technology/new-satellite-to-boost-rural-internet-speeds-1.1010956>, Oct 2011.
- [22] G. Maral and M. Bousquet, *Satellite communications systems: systems, techniques and technology*. John Wiley & Sons, 2011.
- [23] T. E. Doyle and D. Musson, “A technological cybernetic analysis of simulation-based, remotely delivered medical skills training in an austere environment: Developing a test bed for spaceflight medicine.” unpublished, 2012.
- [24] F. Mair, S. Fraser, J. Ferguson, and K. Webster, “Telemedicine via satellite to support offshore oil platforms,” *Journal of telemedicine and telecare*, vol. 14, no. 3, pp. 129–131, 2008.
- [25] H. Murakami, K. Shimizu, K. Yamamoto, T. Mikami, N. Hoshimiya, and K. Kondo, “Telemedicine using mobile satellite communication,” *Biomedical Engineering, IEEE Transactions on*, vol. 41, no. 5, pp. 488–497, 1994.
- [26] G. Anogianakis, S. Maglavera, and A. Pomportsis, “Relief for maritime medical emergencies through telematics,” *Information Technology in Biomedicine, IEEE Transactions on*, vol. 2, no. 4, pp. 254–260, 1998.

- [27] D. L. Anscombe, "Healthcare delivery for oil rig workers: telemedicine plays a vital role," *Telemedicine journal and e-health : the official journal of the American Telemedicine Association*, vol. 16, pp. 659–663, Jul-Aug 2010.
- [28] M. Paul and G. Sorwar, "Encoding and decoding techniques for medical video signal transmission and viewing," in *Computer and Information Science, 2007. ICIS 2007. 6th IEEE/ACIS International Conference on*, pp. 750–756, IEEE, 2007.
- [29] S. P. Rao, N. S. Jayant, M. E. Stachura, E. Astapova, and A. Pearson-Shaver, "Delivering diagnostic quality video over mobile wireless networks for telemedicine," *International journal of telemedicine and applications*, vol. 2009, p. 1, 2009.
- [30] A. Panayides, M. S. Pattichis, C. S. Pattichis, C. Loizou, M. Pantziaris, and A. Pitsillides, "Atherosclerotic plaque ultrasound video encoding, wireless transmission, and quality assessment using h. 264," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 15, no. 3, pp. 387–397, 2011.
- [31] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the h. 264/avc standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 17, no. 9, pp. 1103–1120, 2007.
- [32] C. Doukas and I. Maglogiannis, "Adaptive transmission of medical image and video using scalable coding and context-aware wireless medical networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2008, p. 25, 2008.

- [33] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *Image Processing, IEEE transactions on*, vol. 19, no. 6, pp. 1427–1441, 2010.
- [34] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal processing: Image communication*, vol. 19, no. 2, pp. 121–132, 2004.
- [35] A. Panayides, M. S. Pattichis, C. S. Pattichis, and A. Pitsillides, "A tutorial for emerging wireless medical video transmission systems [wireless corner]," *IEEE Antennas and Propagation Magazine*, vol. 53, pp. 202–213, 2011 APRIL 2011.
- [36] M. Razaak and M. G. Martini, "Medical image and video quality assessment in e-health applications and services," *IEEE HEALTHCOM-The 1st Int. Work. on Service Science for e-Health*, pp. 6–10, 2013.
- [37] P. C. Pedersen, B. W. Dickson, and J. Chakareski, "Telemedicine applications of mobile ultrasound," in *Multimedia Signal Processing, 2009. MMSP'09. IEEE International Workshop on*, pp. 1–6, IEEE, 2009.
- [38] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan, "Rate-constrained coder control and comparison of video coding standards," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 7, pp. 688–703, 2003.
- [39] I. Richardson, "White paper: An overview of h.264 advanced video coding," *Vcodex/OneCodec*, 2007.
- [40] FFmpeg, *version N-55515-gbbbd959*. Online: <https://www.ffmpeg.org/>, 2013.

- [41] “FFmpeg and H.264 Encoding Guide.” Online: <https://trac.ffmpeg.org/wiki/Encode/H.264>. Accessed: August 2014.
- [42] “Image Processing Toolbox MATLAB R2014a Documentation.” Online: <http://www.mathworks.com/help/curvefit/>. Accessed: April 2015.
- [43] MATLAB, *version 8.3 (R2014a)*. Natick, Massachusetts: The MathWorks Inc., 2014.
- [44] Radiocommunication Sector of International Telecommunications Union (ITU-R BT.500-13), *Methodology for the subjective assessment of the quality of television pictures*, 01 2012.
- [45] Telecommunication Standardization Sector of International Telecommunications Union (ITU-T P.910), *Subjective video quality assessment methods for multimedia applications*, 04 2008.
- [46] E. Svensson, “Comparison of the quality of assessments using continuous and discrete ordinal rating scales,” *Biometrical Journal*, vol. 42, no. 4, pp. 417–434, 2000.
- [47] M. H. Pinson and S. Wolf, “Comparing subjective video quality testing methodologies,” in *Visual Communications and Image Processing 2003*, pp. 573–582, International Society for Optics and Photonics, 2003.
- [48] T. Hoskin, “Parametric and nonparametric: Demystifying the terms,” *Mayo Clinic CTSA BERD Resource*. Retrieved from <http://www.mayo.edu/mayo-edudocs/center-for-translational-science-activities-documents/berd-5-6.pdf>, 2012.

- [49] G. W. Corder and D. I. Foreman, *Nonparametric statistics: A step-by-step approach*. John Wiley & Sons, 2014.
- [50] M. Mu, A. Mauthe, G. Tyson, and E. Cerqueira, “Statistical analysis of ordinal user opinion scores,” in *Consumer Communications and Networking Conference (CCNC), 2012 IEEE*, pp. 331–336, IEEE, 2012.
- [51] “Statistics Toolbox MATLAB R2014a Documentation.” Online: <http://www.mathworks.com/help/stats/>. Accessed: February 2015.
- [52] “Curve Fitting Toolbox MATLAB R2014a Documentation.” Online: <http://www.mathworks.com/help/curvefit/>. Accessed: April 2015.