NTNU
Norwegian University of
Science and Technology

Video Quality Assessment in Broadcasting

Anders Prytz

Master of Science in Communication Technology

Submission date:	June 2010
Supervisor:	Andrew Perkis, IET
Co-supervisor:	Torkel Aamodt Thoresen, Telenor Satellite Broadcasting AS
	Fitri N. Rahayu, IET

Norwegian University of Science and Technology
Department of Electronics and Telecommunications

Problem Description

The assessment of video quality at Telenor Satellite Broadcasting AS (Telenor SBc) is done by a group of experts. This is time consuming and results are not always consistent with those of the user group.

To assess the quality, Telenor SBc use a video quality analysis (VQA) system from Video Clarity, for measuring objective visual quality. The VQA system includes three objective assessment methods for video quality.

VQEG has conducted several tests of objective quality assessment methods in an attempt to standardize methods for use in broadcasting [1], without recommending a specific method.

The task is to evaluate the use of the VQA system as a means for assessing the perceived quality of compressed content before delivery to the users. This is to be done by running a subjective evaluation [2] of video quality and compare the results to those from the VQA system.

[1] Video Quality Experts Group, Final report from the video quality experts group on the validation of objective models of video quality assessment, Phase II , 2003.

[2] ITU-T Recommendation BT.500, Methodology for the Subjective Assessment of the Quality of Television Pictures , 2000.

Assignment given: 22. January 2010
Supervisor: Andrew Perkis, IET

Abstract

In broadcasting, the assessment of video quality is mostly done by a group of highly experienced people. This is a time consuming task and demands lot of resources. In this thesis the goal is to investigate the possibility to assess perceived video quality with the use of objective quality assessment methods. The work is done in collaboration with Telenor Satellite Broadcasting AS, to improve their quality verification process from a broadcasting perspective.

The material used is from the SVT Fairytale tape [1] and a tape from the Norwegian cup final in football 2009. All material is in the native resolution of 1080i and is encoded in the H.264/AVC format. All chosen compression settings are more or less used in daily broadcasting.

A subjective video quality assessment been carried out to create a comparison basis of perceived quality. The subjective assessment sessions carried out by following ITU recommendations [2].

Telenor SBc provided a video quality analysing system, the Video Clarity Clearview system[3] that contains the objective PSNR, DMOS and JND. DMOS[4] and JND[5] are two pseudo-subjective assessment methods that use objective methods mapped to subjective results [3]. The methods hopefully predict the perceived quality and eases quality assessment in broadcasting.

The correlation between the subjective and objective results is tested with linear, exponential and polynomial fitting functions. The correlation for the different methods did not achieve a result that proved use of objective methods to assess perceived quality, independent of content. The best correlation result is 0.75 for the objective DMOS method. The analysis shows that there are possible dependencies in the relationship between subjective and objective results. By measuring spatial and temporal information[8] possible dependent correlation results are investigated.

The results for dependent relationships between subjective and objective results are good. There are some indications that the two pseudo-subjective methods, JND and DMOS, can be used to assess perceived video quality. This applies when the mapping functions are dependent on spatial and temporal information of the reference sequences. The correlation achieved for dependent fitting functions, that has a suitable progression, are in the range 0.9 – 0.98.

In the subjective tests, the subjects used were non-experts in quality evaluation. Some of the results indicate that subjects might have a problem with assessing sequences with high spatial information.

This thesis creates a basis for further research on the use of objective methods to assess the perceived quality.

Preface

This Master's thesis was written during the spring of 2010 at the Norwegian University of Science and Technology (NTNU). The thesis was carried out at the Department of Electronics and Telecommunication (IET) in collaboration with the company Telenor Satellite Broadcasting AS. The purpose of this thesis was to investigate the possibility to assess perceived quality by the use of a video quality analysing system.

I would like to thank Andrew Perkis and Fitri N. Rahayu for their support throughout the work on my thesis. They were good support on the scientific part and were of good help to solve problems.

Thanks to Torkel Aamodt Thoresen and Telenor Satellite Broadcasting AS for providing guidance, encoder, decoder and video quality analysing system. Torkel's help with ideas and facts about broadcasting was very useful.

Thanks to OB-TEAM AS and TV2 for providing football material for use in this test. Without their support the test would not contain football material.

A special thanks to my parents for their support over the years I have studied for my master's degree.

Anders Prytz

Anders Prytz
Trondheim, June 14, 2010

Contents

1 Introduction **1**
1.1 Background 1
1.2 Quality assessment 2
1.3 Scenario 3

2 Method **5**
2.1 Literature 5
2.2 Data gathering 6
2.2.1 How 6
Evaluation room 6
Subjects 6
Evaluation session 7
Video material 8
Compression 8
Objective methods 9
Hardware 11
2.2.2 Why 12
2.2.3 Alternatives 13
2.3 Possible sources of error 14

3 Results and Discussion **17**
3.1 Test Material 17
3.2 Subjective quality assessment results 19
3.2.1 Compression and content 21
3.2.2 Single stimulus 22

3.3 Correlation 23
3.3.1 Fitting function 24
3.3.2 DMOS calculation 25
3.3.3 Normalisation 27
3.4 Evaluation 29
3.4.1 PSNR 30
3.4.2 DMOS 32
3.4.3 JND 35

4 Conclusion 39
Observations 40
4.1 Future Work 40

Appendices 45

A Encoder configuration 45

B Equations 47
B.1 Screening equations 47
B.2 PSNR 48
B.3 Correlation 49

C Graphs 51

D Sequences 71

E Pictures 75
E.1 Training sequences 75
E.2 Stabilisation sequences 76
E.3 Test sequences 77
E.4 Evaluation area pictures 80
E.5 Screenshot of software 82
E.6 Equipment 84

F Subjective and objective testing sheets and results 85

List of Figures

2.1 A drawing of the area that was made to perform subjective evaluations. Curtains, floor and room surround the area. 7

2.2 Time schedule for the different playlists. Start screen, Vote and End screen are grey screens. 8

3.1 The spatial and temporal information relationship between the different reference contents used in testing. The lines indicate the mean value across test material. Only reference videos are measured. Original is found in figure C.7. 18

3.2 MOS score for each video. Distorted sequences, for the specific contents, are in the order A-M, with the reference as the first sequence. Colouring for each sequence can bee seen in Appendix D. 95% CI is included using the *errorbar* matlab function. Original is found in figure C.1. Each content is presented in figure C.3. 19

3.3 Example of the variation in objective results, for a specific subjective value, and vice versa. The black lines indicate the differences. It is the relationship between MOS and JND that are represented. Original is found in figure C.8. 21

3.4 MOS for each compression setting. The results presented are for each compressed sequences according to table 2.2. The colouring represents the setting changes that are not bitrate. The marks for the MOS value represent the different bitrates. 95% CI is included using the *errorbar* matlab function. Triangle = 8 Mbps, circle = 10 Mbps and square = 15 Mbps. Original is found in figure C.2. 22

3.5 Plot of the subjective relationship with DMOS and PSNR, with their fitting function included. The resulting correlation are found in table 3.3. Both fitting functions are made by the polynomial function. Original is found in figure C.5. 25

3.6 Plot of the subjective DMOS relationship with DMOS and PSNR, with their fitting function included. The resulting correlation are found in table 3.4. Both fitting functions are made by the polynomial function. Original is found in figure C.6. . . 26

3.7 Histogram of the MOS for every sequence used. The black line are a function of the normal distribution made by the MOS mean and standard deviation. Original is found in figure C.4. . 28

3.8 Relationship between PSNR and subjective DMOS. Original is found in figure C.10(a) 30

3.9 Fitting functions for the individual groups of objective PSNR. Original is found in figure C.12. 32

3.10 Relationship between objective and subjective DMOS. Original found in figure C.10(b) 32

3.11 Fitting functions for the individual groups of objective DMOS. Original is found in figure C.13. 34

3.12 Relationship between the results for JND and subjective DMOS. Original found in figure C.10(c) 35

3.13 Fitting functions for the individual groups for objective JND. Original is found in figure C.14. 36

C.1 MOS score for each video. Sequences for each content are in order A-M, with the reference as the first sequence. Colouring for each sequence can bee seen in section D. 95% Ci is included using the *errorbar* matlab function. Each content is presented in figure C.3. 51

C.2 MOS for each compression setting. The results presented are for each compressed sequences according to table 2.2. The colouring represent the setting changes that are not bitrate. The marks for the MOS value represents the different bitrates. 95% CI is included. Triangle = 8 Mbps, circle = 10 Mbps and square = 15 Mbps. 52
C.3 All contents MOS plotted with respect to the compression settings. The compressing code, showed in table 2.2 is in increasing order, with the reference first in the graphical presentation. 95% CI is included using the *errorbar* matlab function. 55
C.4 Histogram of the votes for every sequence used. The black line is a normal distribution made by the mean and standard deviation for all contents. 56
C.5 Plot of the subjective relationship with DMOS and PSNR, with their fitting function included. The resulting correlation are found in table 3.3. Both fitting functions are made by the polynomial function. 57
C.6 Plot of the subjective DMOS relationship with DMOS and PSNR, with their fitting function included. The resulting correlation are found in table 3.4. Both fitting functions are made by the polynomial function. 58
C.7 The spatial and temporal information relationship between the different reference contents used in testing. The lines indicate the mean value across test material. Only reference videos are measured. 59
C.8 Example of the variation in objective results, for a specific subjective value, and vice versa. The black lines indicate the differences. It is the relationship between MOS and JND that are represented. 59
C.9 MOS vs. objective results are plotted. Independent colour for each content described in Appendix D. 61
C.10 Subjective DMOS vs. objective results are plotted. Independent colour for each content described in Appendix D. 62
C.11 The normalised subjective DMOS vs. objective results are plotted. Independent colour for each content described in Appendix D. 64
C.12 Fitting functions for the individual groups of objective PSNR. . . 65
C.13 Fitting functions for the individual groups of objective DMOS. 67
C.14 Fitting functions for the individual groups of objective JND. . . 69

E.1 Picture taken from the training sequences 75

E.2 Picture taken from the stabilisation sequences 76
E.3 Picture taken from the test sequences 79
E.4 The evaluation area 81
E.5 Screenshot of the software used 83
E.6 The VQA system and HD-SDI to HDMI converter 84

List of Tables

2.1 Compression setting used without configuration. 9
2.2 Compression settings that are configured to produce the distorted material. The compression code is used in analysis of compression analysis. 10

3.1 Results from the measurement of spatial and temporal information. The results are measured according to [8] by the VQA system. The results are presented graphical in figure 3.1 18
3.2 $\mu_{content}$ and standard deviation for each content. 20
3.3 Correlation results between the objective measurements and the subjective MOS. Fitting is done with the fitting function seen in equations 3.1 and 3.2. 23
3.4 Correlation results between the objective measurements and the subjective DMOS (DMOS calculated using the VQEG test plan equation 3.3 [9]). Fitting is done with the fitting function seen in equations 3.1 and 3.2. 26
3.5 $\mu_{content}$ and standard deviation of the normalised score for each content 29
3.6 Correlation results between the objective measurements and the normalised subjective DMOS (DMOS calculated using the VQEG test plan equation 3.3 [9]). Fitting is done with the fitting function seen in equations 3.1 and 3.2. 29
3.7 Correlation results for the two groups of contents for PSNR. The spatial information is divided into low and high, with the mean value as divider. 31

3.8 Correlation results for the four individual groups of contents for objective DMOS. The groups are assigned with the spatial interval their are placed into. 33
3.9 Correlation results for the four groups of contents with JND measurement. The groups are assigned with Spatial-Temporal classification divided by the mean value. Low is the classification from mean and decreasing, while high is from mean and increasing. 36

D.1 Results of a measurement of the spatial and temporal information. The results are measured according to [8] by the VQA system 73

Abbreviations

CI	Confidence Interval
DMOS	Difference Mean Opinion Score
fps	frames per second
GOP	Group of pictures
HDTV	High-Definition Television
Hz	Hertz
....i	Interlaced video
ITU	International Telecommunication Union
JND	Just-Noticeable Differences
$\mu_{content}$	Average subjective score for each content
Mbps	Megabit per second
MOS	Mean Opinion Score
NTNU	Norwegian University of Science and Technology
....p	Progressive video
PLCC	Pearson Linear Correlation Coefficient
PSNR	Peak Signal-to-Noise Ratio
SROCC	Spearman Rank Order Correlation Coefficient
Telenor SBc	Telenor Satellite Broadcasting AS
VQA system	Video quality analysing system (Video Clarity Clearview [3])
VQEG	Video Quality Experts Group

CHAPTER 1

Introduction

1.1 Background

This thesis is produced in collaboration with Telenor SBc[1]. Telenor SBc broadcasts over 250 digital TV channels and 70 radio channels across the Nordic countries and throughout Europe by satellite. Telenor SBc also encodes numeral services for IPTV and terrestrial transmissions in Norway.

In this thesis the possible objective methods to make quality assessment of TV channels are evaluated

For satellite and terrestrial transmissions, several services are multiplexed together in a multiplex that occupies a specific bandwidth. Typically, a multiplex contains 8 – 20 services. Radio bandwidth, either by satellite or terrestrial transmissions, is a limited resource and therefore costly. To increase the number of services in a multiplex is obviously an advantage.

In the Nordic marked, there are several providers delivering essentially the same services. One of the big differentiators between them are quality. Bandwidth is a cost for the service providers and they are all trying to use the lowest possible bitrate in the transmissions. All broadcast services are compressed with either MPEG-2 or H.264/AVC (MPEG-4 Layer 10). The compression produces artefacts and general distortion in the image. To be able to provide the best quality, it is essential to find the optimal compression settings for the codecs. By optimising the encoding, the number of services in a multiplex can be increased without significant loss of quality.

[1] `http://telenorsbc.com/` Last retrieved: June 2010

Today, the video quality verification is performed by a group of engineers with high experience. This process is time consuming and therefore costly. Telenor SBc has today approximately 350 different encoders of several different models and software versions. In order to ease and improve the picture quality assessments, Telenor SBc has invested in a video quality analysing system. This system contains objective quality assessment methods that are supposed to correlate well with subjectively perceived quality.

The Video Quality Experts Group (VQEG) is working to standardise objective methods for quality analysis [6, 7], but no method have been standardised to date. To carry out subjective quality test, ITU have standardised test plans and methods [2, 9, 8] on how to perform these kind of test.

1.2 Quality assessment

The broadcast industry has been sceptical to the results produced by objective methods, with regard to subjective perceived quality. This scepticism resulted in limited use of objective method in quality assessment.

The objective behind this thesis is to find a relationship between subjective and objective quality assessment. The goal is to use objective methods to assess how people will perceive the quality of the content they are watching.

To evaluate the encoding quality, several short video sequences are selected. These have been chosen in order to represent difficult, but not unusual content. Some of the content are widely used as test material in the broadcast industry.

Test material is encoded with the same encoders as TV services are encoded in broadcast transmissions. The test material is assessed both subjectively and objectively. The results are analysed to investigate a possible relationship between subjective and objective quality assessments.

This thesis intends to find a mapping function for one or more objective method, which interprets objective results with regard to subjective perceived quality. This is to use the VQA system to assess how people would perceive different compression settings. The optimal compression setting for encoding systems can then be found a lot easier and more accurate.

An increasing number of broadcast services are delivered in HD resolution. HD comes in two different versions today, 720p and 1080i. In the future, it is believed that 1080p also will be transmitted. The most common HD version today is 1080i. In Europe, the only compression standard used for HD content is H.264/AVC.

In order to limit the scope of the work done in this thesis, only the combination of H.264/AVC and 1080i has been investigated.

1.3 Scenario

A typical broadcasting scenario is created: A viewer watching a TV service on a screen.

The only parameter varied in this thesis is the possible distortion through compression. It is difficult for a broadcaster to control the video quality at the end user, since little is known about the environment for the end user.

The output of the transmission head end is the last reliable point of evaluation when it comes to the quality of the encoding. This is due to the different transmission paths (fibre, xDSL, coax, terrestrial and satellite) and their impact on the perceived quality of service.

At the end user, there is no option to compare the picture quality with a reference signal or the service before compression. The viewed service is the only evaluation criteria regarding picture quality. A live transmission is only viewed once. It is therefore difficult to compare different transmissions.

The viewer is placed in a controlled environment with as few distortions as possible. In order to eliminate other distortion factors, the screen used is a high quality plasma TV. The distance between screen and viewer is five times picture height. There is no sound.

The material viewed is a combination of sports and different outdoor settings. Each clip is ten seconds long. Due to the limited length of each video clip, relative few scene changes are included in each clip.

CHAPTER 2

Method

2.1 Literature

The main literature study for this thesis is standardised methods on how to perform subjective video quality evaluations. The ITU recommendation BT. 500-11 [2], "Methodology for the subjective assessment of the quality of television pictures", is important due to the scenario chosen. Other articles/reports from VQEG and ITU were studied since the methods they present are used world wide in research on video quality assessment. The choice of literature that presents standardised methods are important to conduct test methodologies that are found good for the purpose.

The VQEG reports [7, 6] are used to understand the basics on how to investigate the relationship between subjective and objective assessment of video quality. There have been proposed and tested objective assessment methods over the years, and many of them have been evaluated by VQEG. VQEG works to find an objective quality assessment method that can be standardised.

The system under evaluation is based on objective assessment methods. How the methods works and previous evaluation are investigated through articles and reports [5, 4, 3]. A study of how other researchers have worked are performed to get input and inspiration to the task [10, 11].

2.2 Data gathering

This section describes the methodology of how the quality evaluation was performed, why the choices were made and alternative solutions that could be used.

2.2.1 How

Evaluation room

The room used to conduct subjective quality evaluation was modified to cope with the ITU recommendation [2] for a home environment. The room used for quality assessment was Café Media at NTNU, which is a multimedia laboratory. All windows in the Cafe Media were covered by corrugated cardboard in order to prevent the external light influencing the illumination of the room. An evaluation area was made, inside the room, by hanging up curtains. . Light brown laminate flooring, light-grey curtain and a white ceiling surrounded the evaluation area. A couch was placed at the wall, inside the area, as seating for the subjects. The height, length and width of the area was 290, 390 and 240 centimetres. A drawing of the area can be seen in figure 2.1. A 50 inch plasma monitor was placed at a approximate 5H distance from the subject, as recommended by [2] for the respective monitor size. Figure E.4 shows pictures of the room and the evaluation area.

Outside the area all possible light sources were turned off or covered. The test observer sat outside the evaluation area. The play out system was also placed outside the evaluation area. The only noise inside the room was fan noises from the play out server and a router. These noises were very low and constant.

The area was lightened with a 500 W halogen lamp places behind the monitor, lightning up towards the ceiling, as seen in figure E.4(e).

The level of room illumination was measured by “Meterman LM631 light-meter”, and it was in compliance with ITU recommendation [2] for home environment viewing condition.

Subjects

The subjects who participated in the quality evaluation were without any training or experience in picture evaluation. The majority were students and employees at NTNU. The average age was 25.7 years with 30 people participating. The youngest person was 20 years old, while the oldest was 53. 50% of the subject used some kind of vision correction, either glasses or contact lenses. No one of the subjects had any kind of colour blindness. Both vision

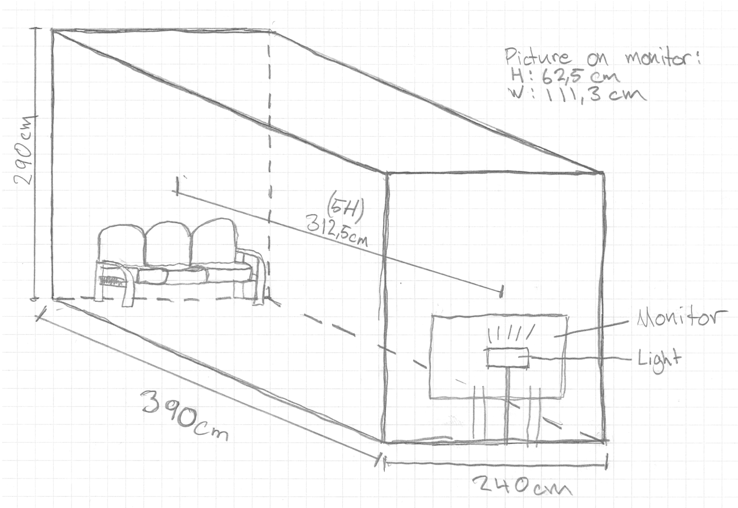


Figure 2.1: A drawing of the area that was made to perform subjective evaluations. Curtains, floor and room surround the area.

and colour blindness was checked before the subjects participated in the test. 40% of the subjects in this experiment were women.

Evaluation session

In advance of the session each subject received an instruction (seen in Appendix F) on how to perform the test. These instructions were repeated on arrival to the evaluation session. Each subject was checked on vision and colour vision with a Snellen chart and Ishihara test. The subjects went through a training session before the actual test was performed. After training the subject had opportunity to ask questions. When the subject had no further questions the test was performed.

The test was performed single stimulus with only one view of each sequence. When the test was started the subject could not stop or pause the session. The training playlist contained 2 video contents, which had 5 examples in total. The test playlist contained 5 stabilisation sequences and 84 test sequences. The actual test lasted for 19.4 minutes, in hand with [2].

Ten second sequences were shown continuous with a three second grey screen with letters in between. The letters on the grey screen showed which sequence to vote on. The voting was done on a sheet of paper with a letter for each sequence (seen in Appendix F). Both training and test were made with the set-up shown in figure 2.2.

Four different playlists were made. Each playlist was checked to not play

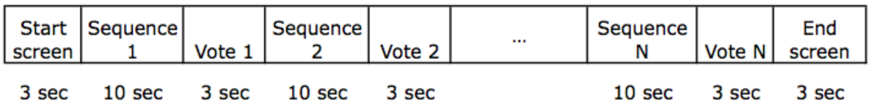


Figure 2.2: Time schedule for the different playlists. Start screen, Vote and End screen are grey screens.

sequences with the same content or compression setting following each other.

Video material

The sequences used in this thesis are captured from two different tapes of video material. The main part of the material used was captured from “The SVT high definition multi format test set” (hereafter described as the SVT Fairytale tape) [1]. Football material was captured from a tape borrowed from the OB-TEAM AS[1] and TV2.

The Fairytale material used in this thesis had the resolution 1920x1080. The material was interlaced with a frame rate at 50 Hz. The original material were captured at the VQA system from a HDCAM-SR tape, in *.yuv* 4:2:2.

The football tape is a recording from the camera source at the Norwegian cup final in football 2009. This tape was borrowed from OB-TEAM AS and TV2 with permission to use in this testing. The material had the resolution 1440x1080. The material was interlaced with a frame rate at 50 Hz. The original material were captured at the VQA system from an HDCAM tape, in *.yuv* 4:2:2.

It is important to note that all material used in this thesis is true interlaced i.e. there is a temporal difference between the two fields in one frame. Much of the material used for 1080i transmissions today, are derived from 1080p/24/25/30 Hz, and therefore without temporal difference between Field 1 and Field 2.

Detailed description of the sequences used can be found in Appendix D and pictures in Appendix E.

Compression

The coding process of video was done with a live coding system of the brand Tandberg[2]. The encoder was an EN5990 and the decoder was a RX1290. Both supported by Telenor SBc and previous used in broadcasting. The codec used was H.264/AVC.

[1] `http://www.obteam.no/` Last retrieved: June 2010.
[2] Later changed name to Ericsson.

All the video sequences were played out, and recorded on the VQA system. The VQA system has two broadcast modules with in/out put HDSDI interfaces. The software allowed one output and one input at the same time. This made it possible to play out the reference material and make it run through a compression process before capturing it. All editing and alignment of the sequences were done on the VQA system.

The encoder is previous used in broadcasting, and had many possible parameters to configure. In this thesis some of the parameters were picked out of interest. In Table 2.1 and 2.2 both the non-configured and configured settings are presented. An explanation of each setting is presented in appendix A. The configuration GUI can be seen in figure E.5(b), the settings listed in the GUI, that are not reviewed, are not relevant for the video quality.

Table 2.1: Compression setting used without configuration.

Name	Setting
Profile:	High @ level 4.0
Mode:	Standard (max 4.4s)
VBR:	off
Aspect Ratio:	16:9

Objective methods

A short description of the objective methods on the VQA system is presented below. All the measurements are full-reference methods, which means that they carry out the measurements by comparing a sequence with the reference sequence.

PSNR

PSNR is an quality assessment method and a direct measurement on how much the processed image/video is different from the original, pixel by pixel. PSNR is given in decibel and is a logarithmic function calculated by the possible number of colour representation and the mean squared error. Objective measurements settings are seen in figure E.5(c). The PSNR equation can be seen in Appendix B.2.

Table 2.2: Compression settings that are configured to produce the distorted material. The compression code is used in analysis of compression analysis.

Compression code	Bitrate	GOP length	GOP structure	Resolution	Bandwidth
A	10	12	IP	960	Sharp
B	15	12	IP	960	Sharp
C	10	12	IP	1920	Sharp
D	15	12	IP	1920	Sharp
E	8	12	IBBBP	960	Sharp
F	10	12	IBBBP	960	Sharp
G	15	12	IBBBP	960	Sharp
H	8	12	IBBBP	1440	Sharp
I	10	12	IBBBP	1440	Sharp
J	15	12	IBBBP	1440	Sharp
K	8	12	IBBBP	1920	Sharp
L	10	12	IBBBP	1920	Sharp
M	15	12	IBBBP	1920	Sharp

DMOS

DMOS is the name of one of the objective assessment method on the VQA system that give results "directly" comparable to subjective results. The measurement is based on MS-SSIM [4, 3], which gives a score that are mapped to DMOS through a polynomial fitting function. The mapping function is made, based on research and subjective testing, by the use of standard sequences that VQEG uses to conduct subjective studies. The objective results produced in this thesis are given in the range $0 - 4$, with zero for identical sequences.

Objective measurements settings are seen in figure E.5(d).

JND

The Sarnoff JND visual model [5] is a method, developed by Sarnoff/ Tektronix, which predicts the subjective rating for a group of human testers. The method analyse the image for makroblocks, blur, luminous variations etc. It predicts a score that is correlated to the JND scale using the VQEG database. JND result is a number, to use in equation 2.1, that tells how many expert viewers needed before one of them prefer the processed sequence instead of the original.

$$\text{NumberOfPeople} = 2^{(JND+1)} \tag{2.1}$$

Objective measurements settings are seen in figure E.5(e) and E.5(f). The JND settings that can be configured in the software are the viewing distance and luminance from the screen. The settings chosen were 5x viewing distance and 200 luminance, as recommended in [2] and used in the subjective assessment sessions.

Spatial and Temporal information

This is not an assessment method, but a measurement of the information in a video sequence. Spatial information is a measurement of the differences in each frame of a video sequence. Temporal information is a measurement of the differences between frames. The equations for spatial and temporal information can be found in [8].

Hardware

Essential equipment used to produce and show video content were the VQA system, encoder, decoder and monitor. These are explained briefly below.

Monitor

The monitor used to view the video sequences was Pioneer PDP-5000EX 50 inch plasma. All digital processing by the monitor were turned off. Only brightness of the screen was adjusted to meet recommendations [2]. The monitor can be seen in the evaluation area pictures in figure E.4

VQA system

To play out[3], capture, edit, organise and analyse the sequences, Video Clarity Clearview RTM [3] with its software was used. This system is made to handle HD material and the software contains objective measurement tools to measure video quality objectively. A screenshot of the software can be found in figure E.5(a) and the hardware can be seen in use in figure E.6(a).

Encoder/Decoder

To produce compressed material an encoder and decoder were used. This was a live coding system of the Tandberg brand that has been used in broadcasting. The encoder was an EN5990 and the decoder was a RX1290. The encoder used is operated through a GUI in a web browser, by connecting a computer through a network connection. A screenshot of the configuration

[3]To play out on an external screen, a HD-SDI to HDMI converter was needed. Pictured in figure E.6(b)

screen on the encoder can be seen in figure E.5(b) and the hardware in use can be seen in figure E.6(a).

2.2.2 Why

Evaluation room/area

The evaluation room/area was modified to follow the home environment recommendation, from [2], to cope with the broadcasting scenario for this thesis. The area was not furnished, such that there were no disturbing elements inside the area.

Subjects

The subjects were chosen to achieve a wide range in age and contain both sexes. In broadcasting both sexes are users and people from all ages are using services a broadcaster provides.

Evaluation session

The single stimulus method was chosen to achieve a viewing situation corresponding to a broadcasting situation. When people are using broadcast services the only evaluation criteria, regarding video quality, is the picture on the screen. It is not possible to compare the picture on the screen with other pictures or play a sequence in repeat[4]. The lettering of the grey screen is used to guide the subjects on which sequence to vote on.

Video material

The content chosen in the quality testing was from the SVT Fairytale tape [1] and football content provided by OB-TEAM AS[5]. The Fairytale material is material that is commonly used in video quality testing and encoding configuration. The football material was used since football is a large part of broadcasting in Norway.

Compression

To evaluate the picture quality of the broadcasted services, the picture that are sent for transmission, are the picture that easiest and most accurate can be evaluated. It is difficult to include transmission errors since this demands to occupy a transmission link on a satellite or be on another location to capture possible transmission errors.

[4]Provided that recording is not an option.

[5]`http://www.obteam.no/` Last retrieved: April 2010.

The settings are defined together with Telenor SBc to test configurations more or less used in HD broadcasting. The bitrates chosen range from the lowest (8Mbps) to the usual (15Mbps) bitrate for HD services. The GOP length at 12 is the length that is usually used. The GOP structure chosen is the ones with no and most B frames. This is to look at how B frames affects video quality. IBBBP is the structure usually used to encode HD-services in broadcasting. The resolution is a horizontal compression of the image, where the number of horizontal lines is limited in the encoding. Bandwidth is a bandpass filtering of the image; sharp is the setting for no filtering.

2.2.3 Alternatives

The standardised viewing methods from [2] have a couple of alternatives. The most common methods to use, beside single stimulus, are two of the double stimulus methods. The first one is that the reference video is played before each sequence to be compared against. The second is to have two similar screens and compare the sequences head-to-head. For these two methods it is easier to distinguish differences between the sequences. They are similar to a full reference objective method, but do not cope well with a broadcasting scenario since one do not have the possibility to compare the picture on screen with another.

The software has the possibility to play two different sequences on one screen with three different options. These are side-by-side, seamless-split and split-mirror. None of these will be used since the full picture is not shown.

To assess each sequence an automatic registration system can be used. This can ease the assessment for the subjects that do not need to check which sequence to assess. Buttons or a sliding bar can be made to communicate with a database to ease the registration and calculation of results. This is not chosen because of the time limit on this thesis. There were only 30 subjects in this evaluation, which do not take too long time to register and calculate manually.

The content used was interlaced. An alternative to this is progressive material. Today the HD material, in broadcasting, is usually broadcasted interlaced and it is to believe that this will continue for years to follow. The football material is a material not previously used in quality testing, but the lack of football material publicly available makes the material a good replacement. The material was not previously compressed.

MPEG-2 and -4 are the two compression standards that are used in broadcasting today. H.264/AVC, a part of the MPEG-4, is the standard that are usually used for HD material, and therefore used in this test. Other experimental standards, like JPEG-2000, could be tested but H.264/AVC is to believe to still last as the commonly used standard for the years to come.

How many content and different compression settings to use are difficult to be sure of, before the execution of the test. The results will indicate if the choice was good or bad. The choice for this thesis was to cope with the recommended time for an assessment session, and to not perform two assessment sessions with the subjects. More contents could be used to look at the effect of different contents in the subjective/objective relationship. This would though affect the number of compression settings. Another alternative is to increase the number of compression settings, but this would then affect the number of different contents to use. The chosen contents and settings are hopefully a golden mean, to cope with the subjective/objective relationship as a whole.

2.3 Possible sources of error

- **Evaluation area:** There are some noise sources in the room. The noises are low, but the subjects can be affected. Noises from the hallway outside the room can also affect the subjects, if the noises are "interesting". Multiple students use the room, and someone can go straight in. A poster will be hung on the door, but everyone who has access can enter.

- **Stimulus method:** The use of single stimulus might give results that are gathered on the top of the scale. The encoding settings chosen are settings used in broadcasting and can result in good quality in a subjective opinion. A double stimulus might give a result that is more related to the comparison between the sequences and the objective results, since the objective methods are full-reference methods. To find a relationship between the subjective assessment and the full-reference method, the material might be presented to the subjects in the same way. But because of the broadcasting scenario chosen for this thesis, the single stimulus method is used.

- **Voting method:** The method of voting on paper might be affected by the possibility that the subject is not fully concentrated for the whole 10 seconds. They might start to vote a little too early or use more time to vote than deposited. There is a possibility that they vote on wrong or miss a sequence and get their result rejected.

- **Content:** In some of the contents, there are chosen to have a scene change in the middle of the sequence. This might affect the evaluation of the subjects, because the changes in content. An example is the Gun

Run content, where it is first an image of a gun that do not have much movement. The content changes to an overview of a running crowd, which contains lot of details in movement. The content might influence which part of the sequence to assess. It may be natural for the subjects that they cast their votes to the last part of the sequence and the result might be different if the sub contents order were changed.

- **Number of contents:** The number of contents might be too small. Only six contents are used in the test. To find a relationship between subjective and objective results that is independent of the content, it is needed as many content as possible. The low number of content can make it difficult to find a reliable relationship.

- **Compression:** The compression settings chosen is based on possible settings used in broadcasting. These settings could possible result in acceptable quality for some contents, across all compression settings.

- **Subjects:** The group of people contributing in the test might be a little to young, regarding a broadcasting scenario. The target group for broadcasters are everyone. The subjective testing can also contain a questionnaire regarding the multimedia habits for each subject, which is not chosen.

- **Media habits:** Today the use of other sources is used to watch TV-services, like streaming over network. This applies especially for students that are known to not have the largest amount of cash and often resort to the use of more or less illegal sources. The widely use of Internet and computer, to look at multimedia sources, might give another expectation of video quality.

CHAPTER 3

Results and Discussion

3.1 Test Material

In a test of video quality it is important to have a good reference material, which have gone through as few processing steps as possible. The content used in this thesis were of high quality. Of the test content, 4 of the contents was from the SVT Fairytale tape [1]. The Fairytale content are commonly used in quality testing and are suitable to use in testing. The last two contents used are football content that are recorded from a HDCAM tape. The tape has captured images at the site of production. The content on a HDCAM tape is limited to the resolution 1440x1080, but still suitable to use in quality testing, since no processing have been performed on the content. The reason to use this content was to have content that are often used in broadcasting.

The contents spatial and temporal information was measured with use of the VQA system. The calculations are based on the scene characteristics equation in ITU-T rec P.910 [3, 8]. Figure 3.1 show a graphical representation of the relationship between the information for the different contents, given by the values in table 3.1.

The mean values are plotted to look at the variation of contents. Of the 6 sequences, two of them are close to the intersection between the two mean values, while the four others are spread in each category of high/low spatial and temporal information. The content used are of high variation and cover the different categories of contents. The fact that there are a wide spread in the content description is good, since there are a wide variety in the contents

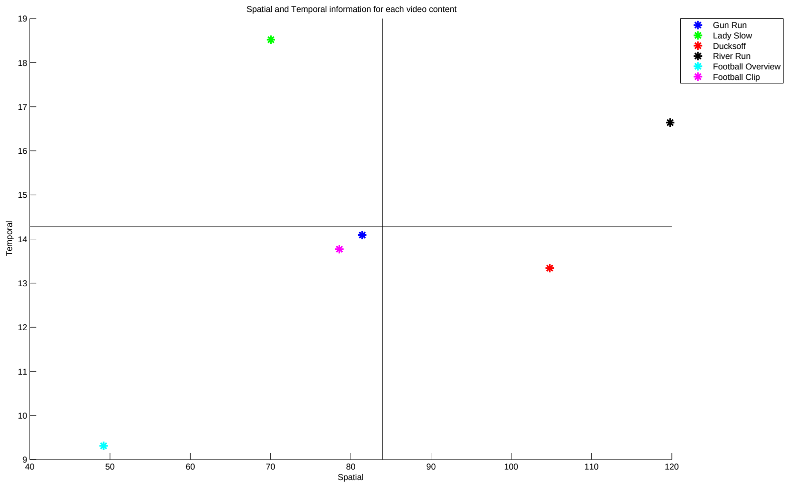


Figure 3.1: The spatial and temporal information relationship between the different reference contents used in testing. The lines indicate the mean value across test material. Only reference videos are measured. Original is found in figure C.7.

Table 3.1: Results from the measurement of spatial and temporal information. The results are measured according to [8] by the VQA system. The results are presented graphical in figure 3.1

	Gun Run	Lady Slow	Ducks Off	River Run	Football O	Football C
Spat	81.420	70.050	104.79	119.79	49.200	78.580
Temp	14.090	18.520	13.340	16.640	9.3100	13.770

broadcasted.

When content are plotted with different colour these following colour characteristics are used: **Blue** = Gun Run (E.3(d)), **green** = Lady Slow (E.3(e)), **red** = Ducks off (E.3(a)), **black** = River Run (E.3(f)), **cyan** = Football Overview (E.3(c)) and **magenta** = Football Clip (E.3(b)).

3.2 Subjective quality assessment results

The subjective results can be found in Appendix F. The total mean score for all sequences was 3.4377. All subject filled in the scoring sheet correct, hence no rejections. No subjects were rejected after a screening process, recommended by ITU (equations and details found in Appendix B.1) [2]. In figure 3.2 all sequences are shown with their MOS value given by the subjects in the test performed.

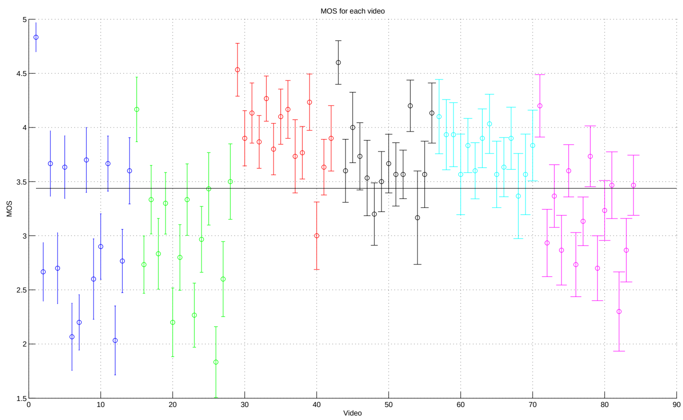


Figure 3.2: MOS score for each video. Distorted sequences, for the specific contents, are in the order A-M, with the reference as the first sequence. Colouring for each sequence can bee seen in Appendix D. 95% CI is included using the *errorbar* matlab function. Original is found in figure C.1. Each content is presented in figure C.3.

The reference sequence is always voted as the best sequence for all contents. The sequence with the lowest score is always the sequence with the lowest bitrate, highest resolution and the GOP structure IBBBP. This setting was expected to get the lowest votes, because of the low bitrate with minimum compression from the other settings. These facts tell that the subjects, as a group, have evaluated the expected best and worst sequences as the best and worst.

For each setting of the GOP structure and resolution, the MOS is always increasing for increasing bitrate. This fact applies for almost all contents in the test. Between the different contents, how much the MOS changes, for the increasing bitrate are different. This indicates that there are differences in how much the different contents distortion is perceived. These differences are reflected in the standard deviation for each content, seen in Table 3.2 and figure 3.2. There are three contents that are more varied than the others.

In figure 3.1 the spatial and temporal information for each content are plotted. The content that has the lowest perceived differences, Football Overview, is the content that have the lowest values for both spatial and temporal information. The relationship between how much the subjective score are varied and spatial and temporal information for the five other contents are not that obvious. The three contents that have the highest standard deviation are found inside the range of 65 – 85 on the spatial scale, with the other sequences on both sides on the scale.

Table 3.2: $\mu_{content}$ and standard deviation for each content.

Content	Gun Run	Lady Slow	Ducks Off	River Run	Football O	Football C
$\mu_{content}$	3.0738	2.9500	3.9310	3.7167	3.7690	3.1857
St.dev	0.7988	0.6149	0.3655	0.3924	0.2162	0.4952

For the temporal information, the contents do not group in the same way. The contents are scattered in the temporal information value. Football Overview with the lowest standard deviation has the smallest value, but the other contents are varied. In the range 13 – 14 are three contents collected. The content with highest standard deviation and the one with the second lowest are collected there with a difference of 0.4333 in the standard deviation value.

In Table 3.2 the mean value for all contents are listed. Three of the contents have a mean value around the mean of the voting scale (3), while the three others are above. The three contents with a mean around 3 are the three contents that have the highest variation between the different compression settings.

An explanation of the high mean value and varied standard deviation can be found in some subjects' comments after testing. Some subjects commented that there were difficult to assess the quality for a couple of contents. The contents mentioned were mainly Ducks Off and Football Overview. The distortion was difficult to notice especially for these two videos with single stimulus. The comments, made by the subjects, indicate that some of the contents and compression settings might have resulted in too good quality to achieve a total mean score at 3. The settings that were used are settings that are found useful in broadcasting, which might have given these results.

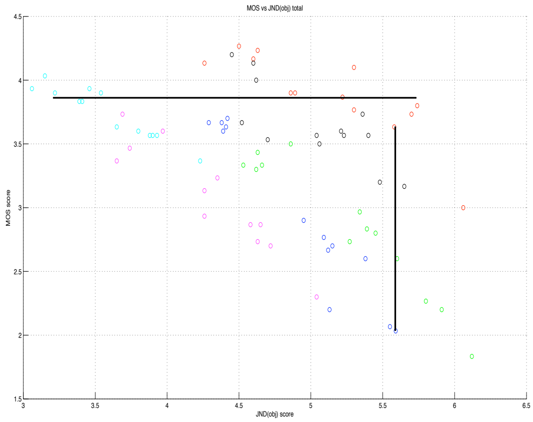


Figure 3.3: Example of the variation in objective results, for a specific subjective value, and vice versa. The black lines indicate the differences. It is the relationship between MOS and JND that are represented. Original is found in figure C.8.

Another issue with the results are that it is a wide variation of subjective results for one objective score, and vice versa. This is shown in figure 3.3 with the black lines. For the specific value on the subjective scale there are several different objective values. This applies also for a specific objective value with different subjective results. This fact indicates that it is difficult to clearly see a relationship between the objective and subjective scores. If the subject scores two different videos equal, the software assess those different. The inconsistency between the results might result in difficulties when the correlation is to be calculated.

3.2.1 Compression and content

In figure 3.4 the MOS for each compression setting is presented. It is noticed that bitrate is the setting that essentially changes the perception of quality. For each compression setting with changed GOP structure and resolution, the sequences scores are increased with increased bitrate. This indicates that the bitrate is the main setting that affect the perceived video quality.

For the 15 Mbps sequences the MOS is almost the same when the other settings are changed. For both GOP structures the MOS is higher for the

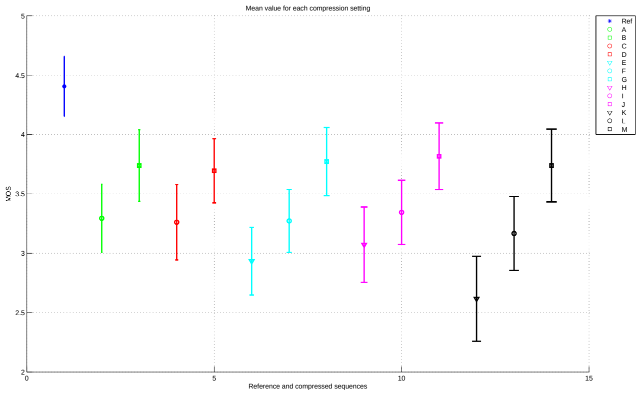


Figure 3.4: MOS for each compression setting. The results presented are for each compressed sequences according to table 2.2. The colouring represents the setting changes that are not bitrate. The marks for the MOS value represent the different bitrates. 95% CI is included using the *errorbar* matlab function. Triangle = 8 Mbps, circle = 10 Mbps and square = 15 Mbps. Original is found in figure C.2.

lower resolutions, with the 1440 as the best. The IP structure was not tested with 1440, but 960 gave a higher MOS than 1920. This indicates that the decrease in horizontal resolution might improve the perceived quality. The same changes in MOS are also seen for the other bitrates where the differences in MOS are larger.

Each content from figure 3.2 are presented individually in figure C.3. The different content follow the same relationship between MOS and bitrate as the overall result in figure 3.4, except the River Run content. Some of the contents do not change much between the different bitrate, but the MOS always increase for increasing bitrate. The River Run content do not follow the same relationship, especially seen comparing compression setting C and D. The result seen in figure C.3(d) indicate that it might be difficult for people to assess sequences with high spatial information. The temporal information might also have an effect due the fact that Ducks Off has high spatial information but lower temporal information and holds the relationship increasing MOS with increasing bitrate. It might be difficult for people to see the effects of the compression when it is much movement and details in the picture.

3.2.2 Single stimulus

The single stimulus method might have affected the test results in some way. The material and compression settings were chosen to cope with a broadcasting scenario. The low standard deviation and high mean scores for some of the contents might have been different if double stimulus had been used. The effect of the compression could be more evident with a comparison with the original. After watching the sequences a couple of times, the opinion is that there are much distortion in all contents, but this might not be that evident for inexperienced people who see each sequence only once.

3.3 Correlation

To investigate the relationship between the subjective and objective results, the correlation are calculated (equations in Appendix B.3). Correlation is a measure of the relationship between two variables, ranging from 0 (no correlation) to 1 (fully correlated) The correlation between the MOS and the different objective measurements are presented in Table 3.3. The fitting function will be reviewed in Section 3.3.1. The subjective and objective results can be found in Appendix F

Table 3.3: Correlation results between the objective measurements and the subjective MOS. Fitting is done with the fitting function seen in equations 3.1 and 3.2.

	Without fitting		Exp fitting		Pol fitting	
Model	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC
PSNR	0.14568	0.96201	0.14973	0.96201	0.37438	0.96201
DMOS	0.66270	0.96203	0.68911	0.96203	0.69452	0.96203
JND	0.52159	0.96193	0.54365	0.96193	0.55038	0.96193

The results show that there are a big difference between the pure objective PSNR and the pseudo subjective DMOS and JND. Previous research on PSNR [12] have shown that it is not the best objective metric, when using different content in the group of material evaluated. In [4] both the MS-SSIM (DMOS) and Sarnoff (JND) were tested on still image, with good results.

PSNR gave very poor result in this first calculation, with a correlation of 0.15. This result indicates, as with previous research, that PSNR is not accurate enough to use on different contents. A graphical view of the results can be seen in figure C.9. For PSNR the results are scattered and the results do not have an obvious gathering. Since SROCC is high, there is still a statistical relationship between the results.

In comparison with PSNR, the two pseudo subjective measurement, DMOS and JND, gave better correlation results. Both resulted in over a 0.5 correlation, with a linear dependency. The results are lower than compared to previous research on still images[4]. For the specific equipment used for this thesis, there are not found any previous results.

The graphical representation of DMOS and JND in figure C.9 indicate a better collection of the results, hence the improved correlation results. The graphical representation shows that the relationship is not optimal and that the better result for DMOS, compared to JND, is visible. The correlation results for the test are not great, but promising given that subjective quality tends to have a non-linear correlation with objective results. The possibility for a non-linear relationship is based on [13], and will be treated further in Section 3.3.1.

3.3.1 Fitting function

To investigate a non-linear relationship between objective and subjective results, fitting functions[1] were tried out. A possible non-linear relationship is found in [13], where VQEG have worked further with the idea. Two different VQEG proposed functions are presented below and used in this thesis. The fitting function can be used to find a non-linear relationship between objective and subjective results, and further be used as a mapping function.

The exponential fitting function from VQEG [7]

$$\text{MOSp} = \frac{\beta_1}{1 + exp(-\beta_2(VQR - \beta_3))} \tag{3.1}$$

The polynomial fitting function from the new VQEG test plan [9]

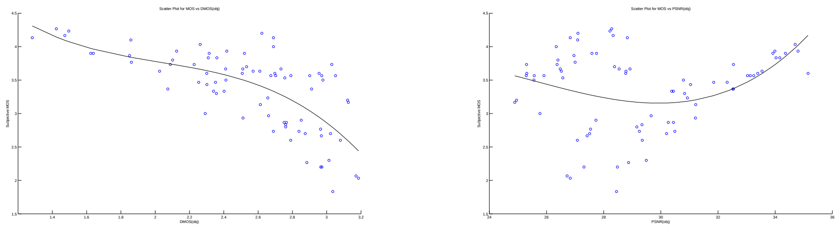
$$\text{MOSp} = \beta_1 \times VQR^3 + \beta_2 \times VQR^2 + \beta_3 \times VQR + \beta_4 \tag{3.2}$$

VQR is the video quality rating given by the objective methods.

The fitting functions try to find the parameters, β_i, that would describe the relationship between the objective and subjective results. These parameters can then be used in the equations to assess the subjective perceived quality from the objective results. In figure 3.5 the fitting function is shown with its respective objective/subjective results.

From the figure there is noticed that fitting functions that give a good correlation result, might not be suitable as a mapping function. DMOS obtain a fitting function that seems suitable, since it give an increasing relationship between the objective and subjective result. On the other hand, the fitting

[1] The fitting functions was made with the *nlinfit* function in MatLab



(a) MOS vs. DMOS(objective) (b) MOS vs. PSNR(objective)

Figure 3.5: Plot of the subjective relationship with DMOS and PSNR, with their fitting function included. The resulting correlation are found in table 3.3. Both fitting functions are made by the polynomial function. Original is found in figure C.5.

function for PSNR represents the same subjective value for two different objective results. From figure C.9(a), which give a content based plotting of the results, it is seen that the values the function tries to be fitted to, at the lower end of the PSNR scale, are from a different content than the others. These values are out of the main collection of results. Based on this, a good correlated fitting function has to be checked to be reliable, before used as a mapping function.

The correlation results of the fitting function are promising. The correlation value increase for the different objective results, but the gain is not that big. From figure 3.5(a) it is seen that the function is finding a non-linear relationship, but the results are still too scattered to find a suitable relationship.

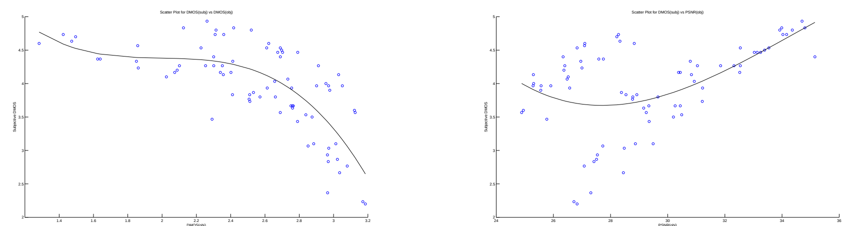
3.3.2 DMOS calculation

In the correlation results from previous section, the reference was not included in the calculation. The MOS was only for the distorted sequences. The lack of using the reference may have affected the results. From figure 3.2 it is seen that for some of the contents, the MOS for the distorted sequences are not assessed much above the reference. This is a fact that can be coped with by using DMOS. DMOS is a representation of the MOS, where the difference between the reference and the distorted sequence are taken into account. Below is the DMOS equation from [9], which is used in this thesis. The equation results the score 5 for a perfect result, and 1 for poor. PVS is the distorted sequences and SRC is the reference.

$$\text{DMOSp} = \text{MOS(PVS)} - \text{MOS(SRC)} + 5 \tag{3.3}$$

Table 3.4: Correlation results between the objective measurements and the subjective DMOS (DMOS calculated using the VQEG test plan equation 3.3 [9]). Fitting is done with the fitting function seen in equations 3.1 and 3.2.

	Without fitting		Exp fitting		Pol fitting	
Model	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC
PSNR	0.45129	0.96201	0.46112	0.96201	0.53883	0.96201
DMOS	0.64385	0.96203	0.74436	0.96203	0.74746	0.96203
JND	0.65246	0.96190	0.65237	0.96193	0.65282	0.96193



(a) DMOS(subjective) vs. DMOS(objective)

(b) DMOS(subjective) vs. PSNR(objective)

Figure 3.6: Plot of the subjective DMOS relationship with DMOS and PSNR, with their fitting function included. The resulting correlation are found in table 3.4. Both fitting functions are made by the polynomial function. Original is found in figure C.6.

The new correlation results are presented in Table 3.4. The results have an improved correlation compared to the previous from Section 3.3. The correlation with a linear relationship is improved for PSNR and JND, while the objective DMOS achieve about the same value. The improvement in the subjective results relationship to all the objective measurement are seen graphical in figure C.10. The previous scattering is now more clustered. This clustering of the relationships is to believe one of the main reasons to the improved linear correlation.

The reason for the improved results is probably the impact of the reference. The use of the difference between the reference and distorted sequences are probably a better comparison basis for video quality, since the subjects show differences in their assessment of the reference sequences. Also it is more comparable with the full-reference objective metrics. In a perfect world, the reference would get the top score from all subjects, but the reality is another thing. In this kind of study, with a hidden reference and different randomised playlists, the subjects are probably unconsciously influenced of the different sequences. The instruction is to not compare videos with each other, but people often do it unconsciously. The use of DMOS hopefully corrects some of these influences.

In figure 3.6 some of the fitting functions are presented for the subjective/objective relationship. The results from Table 3.4 are good, but a correlation at 0.75 is still not good enough to result in a mapping function to subjective quality. There are still a large variance between the function and many results. For PSNR the fitting function is not good, but DMOS achieve a good correlation with a function that look suitable. From the graphs in figure C.10 a good gathering of each content is observed, but between the contents there are a wide variation. This applies for all the objective measurements and from this it is still difficult to see a good relationship. The DMOS calculation influenced the correlation in a good way, but the original MOS might not be the best values to use in this calculation. In the next section a further improvement is attempted.

3.3.3 Normalisation

In an attempt to improve the correlation, a normalisation of the subjective results was conducted. This is based on the idea that the subjects may have used the voting scale differently, although they got the same instructions, as proposed in [10]. In this attempt the goal is to normalise the subjective votes within a 1–5 range, with the value 3 as a mean value. Normalisation of the

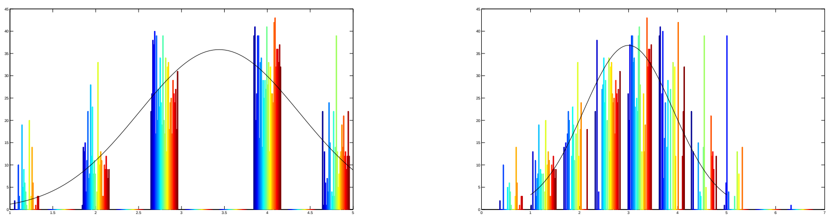
MOS results are done with the equation suggested by [10].

$$m'_{ij} = \frac{(m_{ij} - \bar{m_i} + \mu)}{\frac{4\sigma_i}{K}} \tag{3.4}$$

I represent the subject and J represents the video sequence. m'_{ij} is the normalised score for each subject. $\bar{m_i}$ is the mean score for each subject. μ is the mean value for all subjects and sequences. σ_i represents each subjects standard deviation. K is the maximum value in the test (5). When new results were calculated, all scores were divided by $\frac{\mu_{new}}{3}$ to obtain a mean value at the middle of the scale (3).

After normalisation, a screening of the results were conducted with the same procedure as in Section 3.2. No outliers were detected through this operation.

The new MOS results were plotted against a normal distribution with the same statistical values of μ and σ. The distribution of the original and normalised MOS results are presented in figure 3.7.



(a) Original scores. μ=3.4377, σ=0.9343 (b) Normalised scores. μ=3, σ=0.9103

Figure 3.7: Histogram of the MOS for every sequence used. The black line are a function of the normal distribution made by the MOS mean and standard deviation. Original is found in figure C.4.

The normalisation resulted in a smoother distribution of the results. While the original MOS were integers, the new MOS contain decimals. This is a reason why the new distribution seems more normalised. In the test there were only 30 subjects, which is only a small part of the users of broadcasting services at a daily basis. The new distribution might be a better result to use, provided that it should be used as a baseline on video quality, for all people using broadcasting services. This is based in the assumption that all distributions for a large amount of people is more or less normally distributed

The normalised means, in Table 3.5, have decreased compared to the originals. The process have not normalised each content towards a mean value of 3, but only adjusted the score for a total mean to 3.

Table 3.5: $\mu_{content}$ and standard deviation of the normalised score for each content

Content	Gun Run	Lady Slow	Ducks Off	River Run	Football O	Football C
$\mu_{content}$	2.6456	2.5342	3.4667	3.2737	3.3120	2.7677
St.dev	0.7741	0.5846	0.3633	0.3664	0.2040	0.4774

Table 3.6: Correlation results between the objective measurements and the normalised subjective DMOS (DMOS calculated using the VQEG test plan equation 3.3 [9]). Fitting is done with the fitting function seen in equations 3.1 and 3.2.

	Without fitting		Exp fitting		Pol fitting	
Model	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC
PSNR	0.46765	0.96201	0.47762	0.96201	0.55836	0.96201
DMOS	0.62213	0.96203	0.73122	0.96203	0.73381	0.96203
JND	0.66588	0.96193	0.66585	0.96193	0.66616	0.96193

After the normalisation, DMOS was calculated for the correlation calculation. This was done since DMOS show a better result than the MOS for the two first attempts of finding a relationship. As stated earlier, DMOS are more useful since the votes for the hidden reference sequences are taken into account.

The DMOS results from the normalised MOS do not improve the correlation significant. In fact, the correlation for objective DMOS decreases, while PSNR and JND is increased minimal. This can be seen, comparing the values in the tables 3.6 and 3.4. From the graphical presentation of the subjective/objective relationship in figure C.11, there is noticed that the contents retain the same relationship between each other. The normalisation procedure failed to improve the correlation between subjective and objective results.

3.4 Evaluation

Through analysis and modifications of the results, none of the objective measurements have proved to have a good enough correlation to the subjective results for the whole collection of contents. The best result for a fitting function achieved only a correlation of 0.75, which is not a good enough to be used as an overall mapping function. To obtain a mapping function that are useable, the correlation needs to be at least above 0.9 to achieve a reliable result.

Regarding the fitting functions, the new polynomial function from VQEG [9] seem to give a better fit between subjective and objective results, compared to an exponential or linear function.

A closer look at the results in figure C.9, C.10 and C.11 shows that the assessment of the different contents looks to be the biggest issue for each objective method. The treatment of the different contents also varies between the different objective methods. The fact that objective methods are dependent of content or compression type are found by earlier researchers [12]. While this research was done on PSNR, there is an indication in the results that similar dependencies may apply for the pseudo subjective methods as well.

With the use of figure C.7 and Table D.1, possible mapping function between the results, regarding the spatial/temporal information, will be investigated. In this analysis the subjective DMOS without normalisation, from Section 3.3.2, will be used.

3.4.1 PSNR

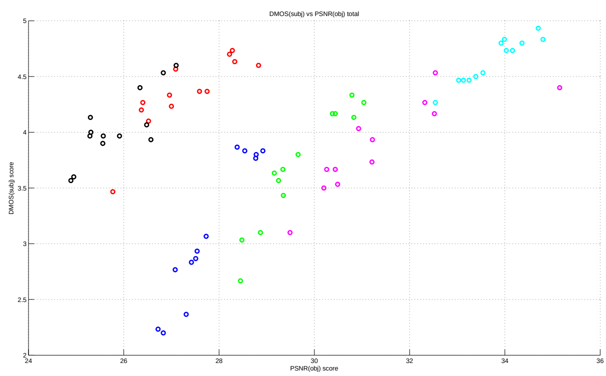


Figure 3.8: Relationship between PSNR and subjective DMOS. Original is found in figure C.10(a)

For PSNR the relationship between subjective and objective results is gathered in two different collections. This is seen in figure 3.8. The two groups of contents are in fact divided by the mean of the spatial information, from the figure 3.1. River Run and Ducks Off are sequences with high spatial information, and are separated from the other contents. The separation is the reason why the fitting function isn't useful and cannot be used as a mapping function.

In comparison with the spatial information, the temporal information seems to not have much influence on the relationship. The 4 contents that are gathered in one group contains the contents with the lowest and highest temporal information. The influence of temporal information seems to be low. From this fact the two sequences Gun Run and Football Clip, which are rated approximately equal in the information analysis, are expected to be assessed equal. This does not imply since the Lady Slow sequence are assessed in between the two contents. Although that the sequences are assessed a little different, a correlation/fitting test is done for the two different groups. The results are presented in Table 3.7 and their best fitting functions are shown in figure 3.9.

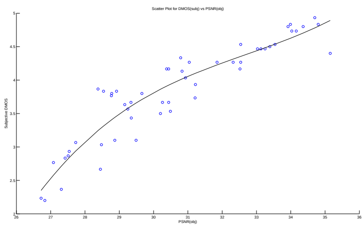
Table 3.7: Correlation results for the two groups of contents for PSNR. The spatial information is divided into low and high, with the mean value as divider.

		Without fitting		Exp fitting		Pol fitting	
Group	Model	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC
Low SI	PSNR	0.90911	0.94335	0.92721	0.94335	0.93104	0.94335
High SI	PSNR	0.84307	0.88889	0.85077	0.88889	0.85253	0.88889

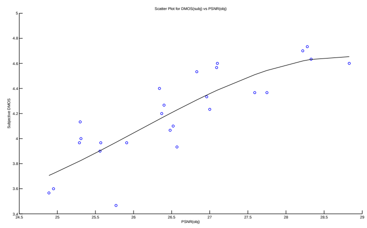
The results show a good relationship regarding the content with low spatial information. The other group have also a good result, but a correlation of at least 0.9 is wanted before the mapping function can be evaluated as useful. In a test of the correlation for each of the high spatial information contents, the numbers were 0.926 and 0.877 for Ducks Off and River Run respectively. The polynomial fitting function was used.

The results show that there is a possible dependency of spatial information, regarding the relationship between PSNR and subjective results. The contents with the low spatial information show a promising relationship, whereas it is uncertainties regarding the contents with high. The high ones had two contents in their group, while the low had four. If there were more contents in the high group the relationship might be better and the intervals that the analysis is based on might have to be different to find suitable mapping functions. There might also be a dependency of the temporal information, which is not seen with the contents in the test. The SROCC are lower for the high spatial information groups, which indicate that there are larger scatterings in the collection. This might also occur from the low number of samples in the analysis.

The fitting functions that achieve the high correlation is seen in figure 3.9. It is seen that the figure for low spatial information results in a relationship



(a) Low spatial information



(b) High spatial information

Figure 3.9: Fitting functions for the individual groups of objective PSNR. Original is found in figure C.12.

that are reasonable and useable. Figure 3.9(b) results in a function that stop increasing at high objective results. A correlation of 0.85 and a fitting function that not seems to be good enough to assess subjective quality cannot be used as a mapping function. The number of contents used in analysis may have affected the correlation and fitting function for high spatial information.

PSNR was not expected to achieve a good relationship with subjective results, although it shows an indication of a good correlation when applying dependencies to the analysis. The problem with the mapping occurred at high spatial information contents. The pseudo subjective measurement following are expected to give better results.

3.4.2 DMOS

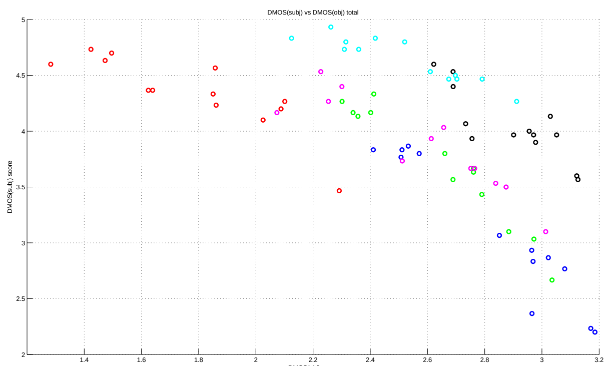


Figure 3.10: Relationship between objective and subjective DMOS. Original found in figure C.10(b)

DMOS proved to be the best objective measurement obtaining a correlation of approximately 0.75 with a polynomial fitting function. This is a good result compared to the other objective methods in the test, but not good enough to be used independently to assess the subjective quality.

For the objective DMOS the content look separated into two or four different collections. For a two group alternative Football Overview and River Run are collected together, while the four others are considered as a group. An interesting thing is that the two contents in the first group are those that have the most and least spatial information. Those are at the endpoints of the scale, but seem gathered in the relationship. Regarding the temporal information, it seems that DMOS assess independent of this. This can be seen when comparing the relationship for Football Clip and Lady Slow, which have an almost identical relationship between objective and subjective results. Their spatial information is almost the same, but the variation in temporal information is large. This is an indication that the temporal information does not have much influence in the assessment relationship.

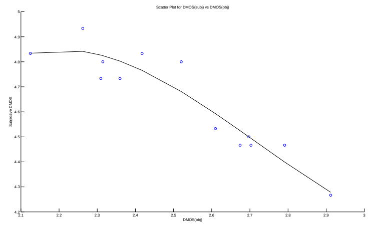
Based on the spatial dependency the contents are divided into four groups. This grouping puts Gun Run, Lady Slow and Football Clip in the same group, while the other contents are seen as separate groups. The groups, based on spatial information value, are seen in Table 3.8.

Table 3.8: Correlation results for the four individual groups of contents for objective DMOS. The groups are assigned with the spatial interval their are placed into.

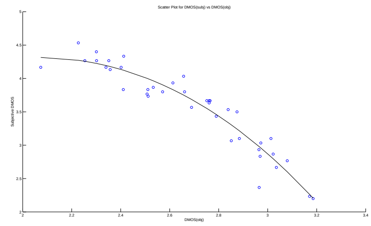
		Without fitting		Exp fitting		Pol fitting	
Group	Model	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC
<65	DMOS	0.91767	0.78571	0.93688	0.78571	0.94374	0.78571
65-85	DMOS	0.92648	0.92500	0.95519	0.92500	0.95529	0.92500
85-105	DMOS	0.83700	0.78571	0.91161	0.78571	0.91440	0.78571
>105	DMOS	0.83438	0.78571	0.83544	0.78571	0.92631	0.78571

The results show a great improvement in the correlation. With the polynomial fitting function, all groups obtain a correlation above 0.9. SROCC is high for the 65-85 group, but low for the other groups. This tells that the results are scattered in their relationship. This is most likely based on the fact that there are a small number of samples in the analysis. The result is promising, considering a relationship between the subjective and objective results. Figure

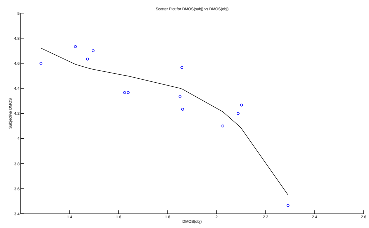
3.11 presents the fitting function compared to the results graphically for each group.



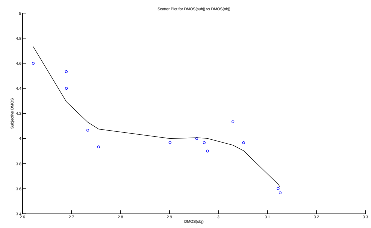
(a) <65 spatial information



(b) 65-85 spatial information



(c) 85-105 spatial information



(d) >105 spatial information

Figure 3.11: Fitting functions for the individual groups of objective DMOS. Original is found in figure C.13.

The fitting functions are not accurate for three of the collections. Those with a poor fitting function, are also those with a low SROCC. The function for the interval 65-85 is good and looks suitable to be used for assessment of subjective results. The others are not good, but might suffer from the low number of samples in analysis. Because of the low number of samples, the fitting function tries to be as close as possible to every sample and this result in a function that does not have a suitable progression. The 65-85 interval contain three times as many samples compared to the others, and might have a more suitable progression from this.

Another fact is that the content with spatial information below the mean seems to correlate better with subjective results. What this fact occurs from is unknown, but there seems to be a decrease in the correlation for increasing spatial information.

The relationship between objective and subjective DMOS looks to be independent of the temporal information. To use it as a mapping function there have to be made mapping functions based on spatial information.

3.4.3 JND

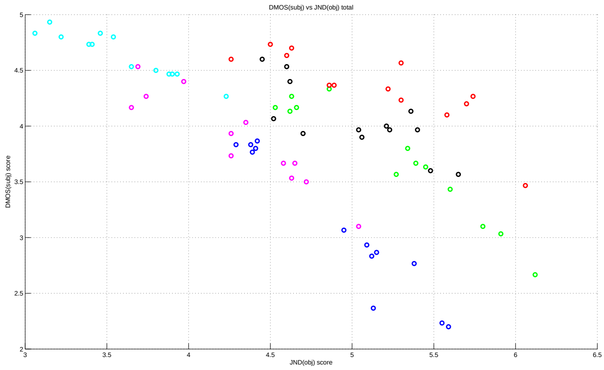


Figure 3.12: Relationship between the results for JND and subjective DMOS. Original found in figure C.10(c)

The JND method seems a little more complicated when trying to find a dependency, compared to DMOS and PSNR. While there were significant grouping of the contents for the previous methods, the contents are for JND collected in a large group with large distances between the different contents. All the results are mixed together, but still not gathered close enough to achieve a good correlation with any of the fitting functions.

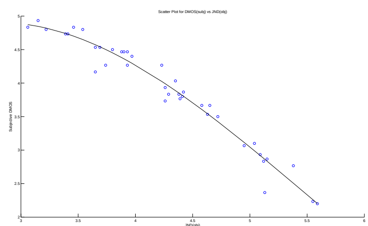
JND shows a tendency to be dependent on both the spatial and temporal information. The grouping of contents performed are done by using the mean values in figure 3.1, and separate the contents into each group of low/high spatial/temporal information. Each square in the figure classifies each group to evaluate. The classification is determined after analysis of figure 3.12 with respect to figure 3.1.

The results for the classes with low spatial information are very good, with values approximately 0.98. There is a good relation between the subjective and objective results. For the Low-Low class the SROCC is above 0.9, which indicate a good clustering. For the three groups, with only one content, the SROCC is beneath 0.8, which is not so good. The correlation for the high spatial information content is not that good, as with both PSNR and DMOS. There is a possibility that the low number of samples, which is also indicated by the low SROCC values, affects the correlation.

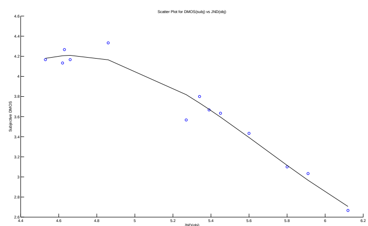
The fitting functions in figure 3.13 show that they can not be used as mapping functions for three of the classes. As with the earlier dependent fitting functions, the number of samples is too low to find a suitable fitting function.

Table 3.9: Correlation results for the four groups of contents with JND measurement. The groups are assigned with Spatial-Temporal classification divided by the mean value. Low is the classification from mean and decreasing, while high is from mean and increasing.

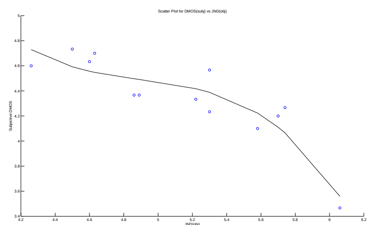
		Without fitting		Exp fitting		Pol fitting	
Group	Model	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC
Low-Low	JND	0.96926	0.92485	0.98108	0.92485	0.98098	0.92485
Low-High	JND	0.95749	0.78571	0.9793	0.78571	0.98237	0.78571
High-Low	JND	0.83408	0.78434	0.90491	0.78434	0.90921	0.78434
High-High	JND	0.77925	0.78571	0.77916	0.78571	0.82756	0.78571



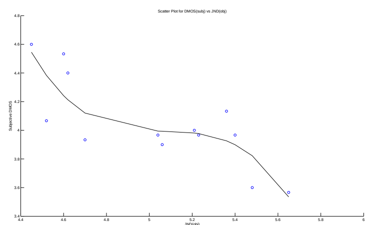
(a) Low spatial/Low temporal information



(b) Low spatial/High temporal information



(c) High spatial/Low temporal information



(d) High spatial/High temporal information

Figure 3.13: Fitting functions for the individual groups for objective JND. Original is found in figure C.14.

It is still a possibility that the functions would be more fitted if the numbers of samples were increased.

Both JND and DMOS show a promising relationship between subjective and objective result. The amount of samples needs to be increased to cover all the possible spatial and temporal information to make a correct classification for the dependent mapping functions. PSNR was not expected to be able to result in a good enough relationship, based on earlier research. PSNR shows a good relationship for contents with low spatial information, but was inaccurate for the higher spatial information. The pseudo-subjective methods prove to relate more to subjective results. They are still not good enough to assess the subjective results for all kinds of contents, but with dependent mapping functions they can be useful to assess how people would perceive the quality of a sequence. The relationship for objective/subjective results seems to decrease for an increase in the spatial information. DMOS looks to be dependent only on the spatial information, while JND looks dependent on both spatial and temporal information.

CHAPTER 4

Conclusion

The objective quality assessment methods are not sufficient to provide results that correlate well with subjective perceived quality assessment. This applies when all contents are used in the analysis. However, the thesis proves that objective quality assessment methods can be used to assess perceived video quality with mapping functions that are dependent on the content of evaluation.

The results are only investigated for H.264/AVC compressed material with a resolution of 1080i. Possible effect of other compression methods or resolution is not analysed in this thesis.

In this thesis the subjective DMOS proved the best basis for analysis with the single stimulus method. Subjective results are dependent on what kind of content the subjects are watching. MOS does not take the reference into account and proved to be a poor candidate.

The difference between the distorted videos and a hidden reference gives a better result for use in analysis with the full reference methods; hence the DMOS is the best candidate in this thesis.

PSNR showed a low correlation for all content aggregated in the analysis. The fitting functions tested were not suitable for use, because it did not represent a good correlation between subjective and objective results. The spatial dependent analysis showed an improvement, but the results are not accurate enough to be used. PSNR can be used for contents with low spatial information, but it fails when the spatial information is high. PSNR is found independent on the temporal information, but, as previous research concludes, PSNR are not well correlated to subjective results.

DMOS gave the highest correlation result, 0.75, in the analysis for all content aggregated. This result is not good enough to prove that DMOS can be used with a mapping function to assess perceived quality, without dependencies for the mapping function. Analysis of the use of DMOS with a spatial dependency shows a possible objective/subjective relationship. DMOS showed no dependency on the temporal information. DMOS is well correlated to the subjective results for groups of content dependent on spatial information. Unfortunately, the amount of different contents in this analysis is to small to conclude with useable dependent mapping functions.

JND achieved a lower correlation than DMOS, with 0.66 as its best result for all content aggregated. JND showed a good dependency of both spatial and temporal information. The analysis of dependent mapping function resulted in a 0.98 correlation to the content in the low spatial and low temporal information group. The results was the best achieved correlation result and the fitting function from figure 3.13(a) show a function with a reasonable progression and can be used as a mapping function.

On the whole it was not proven that the methods could be used in assessment of perceived quality, independent of the content analysed. The pseudo subjective methods, JND and DMOS, gave results that are promising regarding the findings of a relation between subjective and objective results. However, this applies only when the mapping functions are made dependent on the spatial and temporal information of the content. All methods tend to have a lower correlation for contents with high spatial information, compared contents with low spatial information. This might indicate that high spatial information sequences are difficult for non-experienced viewers to assess.

Observations

The overall results indicate that bitrate is the compression setting that has the highest influence on perceived quality.

With limited available bitrate, a reduction of horizontal resolution might improve the perceived picture quality.

4.1 Future Work

Recommended work for the future is to investigate the dependencies on the spatial/temporal information to create dependent mapping functions for assessment of perceived video quality. A larger sample of contents is needed to be able to find mapping functions and suitable grouping of contents. This thesis results show a good indication that the pseudo subjective DMOS and JND methods might be used in assessment of perceived video quality with a video quality analyser. Future work should also include other compression algorithms to look for the same kind of results as in this thesis.

Other future work can investigate the relationship between compression vs. resolution and their impact on perceived quality.

Bibliography

[1] L. Haglund, "The svt high definition multi format test set," tech. rep., SVT, February 2006.

[2] ITU, "Itu-r bt 500-11: Methodology for the subjective assessment of the quality of television pictures," *Question ITU*, vol. 211, p. 11, 2002.

[3] VideoClarityInc., "Clearview users guide." webpage, http://www.videoclarity.com/PDF/ClearViewUserGuide.pdf, April 2010.

[4] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *Signals, Systems and Computers, 2003. Conference Record of the Thirty-Seventh Asilomar Conference on*, vol. 2, 2003.

[5] J. Lubin and D. Fibush, "Contribution to the ieee standards subcommittee: Sarnoff jnd vision model," 1997.

[6] A. Webster and F. Speranza, "Final report from the video quality experts group on the validation of objective models of multimedia quality assessment, phase i," tech. rep., VQEG, September 2008.

[7] P. Corriveau and A. Webster, "Final report from the video quality experts group on the validation of objective models of video quality assessment, phase ii," tech. rep., VQEG, August 2003.

[8] I. T. UNION, "Itu-t recommendation p.910 subjective video quality assessment methods for multimedia applications." pdf, September 2009.

[9] G. Cermak, L. Thorpe, and M. Pinson, "Test plan for evaluation of video quality models for use with high definition tv content," *Video Quality Experts Group (VQEG)*, 2009.

[10] F. De Simonea, M. Tagliasacchib, M. Naccarib, S. Tubarob, and T. Ebrahimia, "A h. 264/avc video database for the evaluation of quality metrics," 2010.

[11] E. Gelasca, "Full-reference objective quality metrics for video watermarking, video segmentation and 3d model watermarking," tech. rep., Tech. Rep., EPFL, September 2005, Ph. D. Thesis, 2005.

[12] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of psnr in image/video quality assessment," *Electronics Letters*, vol. 44, p. 800, 2008.

[13] P. Corriveau, J. Lubin, J. Pearson, A. Webster, *et al.*, "Video quality experts group: Current results and future directions," in *Proc. SPIE Visual Comm. and Image Processing*, vol. 4067, pp. 2–15, 2000.

[14] S. Winkler, *Digital video quality: vision models and metrics.* Wiley, 2005.

APPENDIX A

Encoder configuration

The Tandberg EN5990 encoder are made for HD (H.264/AVC). The parameters used for configuration are explained below.
Profile: Which MPEG4-profile that is in used.
Mode: Is a kind of how fast the encoding should be done. In this thesis the standard option are used. This tells how much delay can be expected when codign the video.
VBR: Variable bit rate. This option is not used to achieve a constant bitrate on the video. VBR are used in broadcasting, but there in combination with statical multiplex.
Aspect Ratio: The aspect ratio of the video compressed. In broacasting usual 16:9 or 4:3. For the HD material used in this thesis the 16:9 is used, also known as "widescreen".
Bitrate: The bitrate of the video coded given i Mbps. The range of possible settings are 1 - 18.28 Mbps.
GOP length: The GOP length can range from 12 to ... 12 was chosen since this settings is common used in broadcasting.
GOP Structure: Is the structure of the GOP. Options are IP, IBP, IBBP and IBBBP. IP and IBBBP are chosen to contain the *"best"* and *"worst"* setting.
Resolution: A compression done on the horizontal resolution of the video. The choice is between 960 to 1920 pixels in the compressed video.
Bandwidth: Is an option on filtering the image. Sharp contains no filtering.

APPENDIX B

Equations

B.1 Screening equations

Counters P_i and Q_i are made for each subject. The equations presented are copied from [2]. If the equation

$$\frac{P_i + Q_i}{J \times K \times R} > 0.05 \text{ and } \left| \frac{P_i - Q_i}{P_i + Q_i} \right| < 0.3 \tag{B.1}$$

is fulfilled, then the subject i will be rejected.

To run the test for each subject, a mean value, standard deviation and a kurtosis coefficient(equationB.2) must be calculated for each video sequence evaluated by the subjects.

$$\beta_{2jkr} = \frac{m_4}{(m_2)^2} \text{ with } m_x = \sum_{i=1}^{N} \frac{(u_{ijkr} - \bar{u}_j kr)^x}{N} \tag{B.2}$$

For each subject, i, P_i and Q_i are calculated for j,k,r = 1,1,1 to J,K,R:

$$
\begin{aligned}
&\text{if } 2 \leq \beta_{2jkr} \leq 4 \\
&\quad \text{if } u_{ijkr} \geq \bar{u}_j kr + 2 \times S_{jkr} \text{ then } P_i = P_i + 1 \\
&\quad \text{if } u_{ijkr} \leq \bar{u}_j kr - 2 \times S_{jkr} \text{ then } Q_i = Q_i + 1 \\
&\text{else} \\
&\quad \text{if } u_{ijkr} \geq \bar{u}_j kr + \sqrt{20} \times S_{jkr} \text{ then } P_i = P_i + 1 \\
&\quad \text{if } u_{ijkr} \leq \bar{u}_j kr - \sqrt{20} \times S_{jkr} \text{ then } Q_i = Q_i + 1
\end{aligned}
$$

These equations are run for each subject for the score for every video sequence evaluated. The counters, P_i and Q_i, are afterwards used in equation B.1 to control if one observers scores should be rejected or kept.

The calculations make use of the following information from the subjective test:

N : Number of observers (30)
J : Number of test conditions including reference (14)
K : Number of test pictures or sequences (6)
R : Number of repetitions (1)

B.2 PSNR

$$\text{PSNR} = 10 \log \frac{m^2}{\text{MSE}} \tag{B.4}$$

$$\text{MSE} = \frac{1}{XYZ} \sum_t \sum_x \sum_y [I(t,x,y) - \tilde{I}(t,x,y)]^2 \tag{B.5}$$

Where m is the maximum value that a pixel can take[1] and for pictures of size $X \times Y$ with T frames in the sequence [14].

This equation calculates differences between the values of pixels in two pictures or sequences I (reference) and $\tilde{I}$ (processed).

[1] The maximum value that a pixel can take is equal to $2^{\text{number of bits for a pixel in the image}}$

B.3 Correlation

Pearson Linear Correlation Coefficient is a measure of the correlation (functional dependence) between two variables X and Y, giving a value between +1 and -1. In this thesis only positive values are used. 0 represent no correlation and 1 represent a perfect correlation.

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y} \tag{B.6}$$

Spearman Rank Order Correlation Coefficient is a non-parametric measure of statistical dependence between two variables, giving a value between +1 and -1. In this thesis only positive values are used. 0 represent no correlation and 1 represent a perfect correlation.

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \tag{B.7}$$

APPENDIX C

Graphs

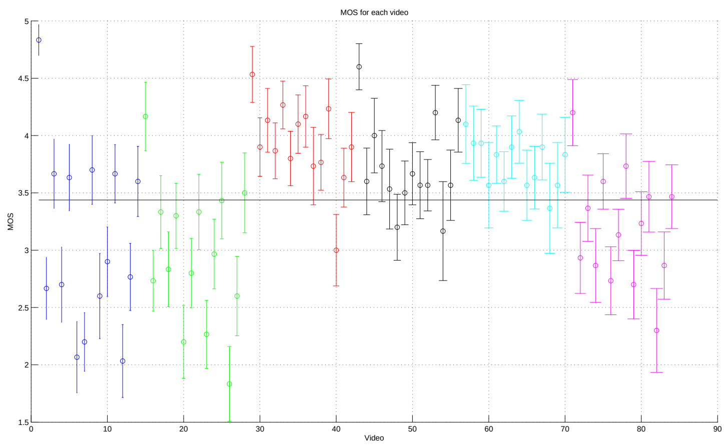


Figure C.1: MOS score for each video. Sequences for each content are in order A-M, with the reference as the first sequence. Colouring for each sequence can bee seen in section D. 95% Ci is included using the *errorbar* matlab function. Each content is presented in figure C.3.

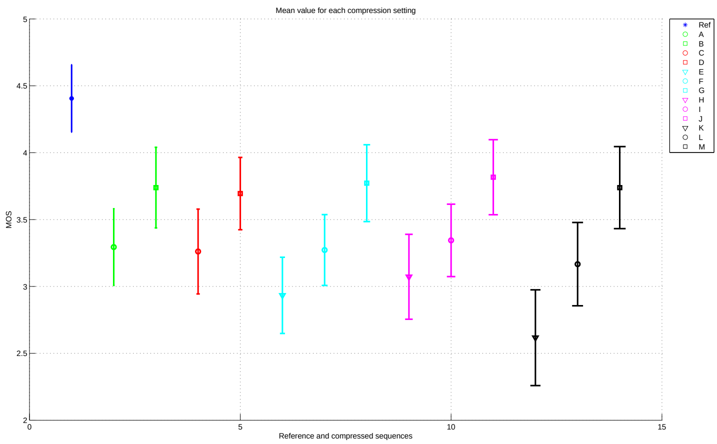
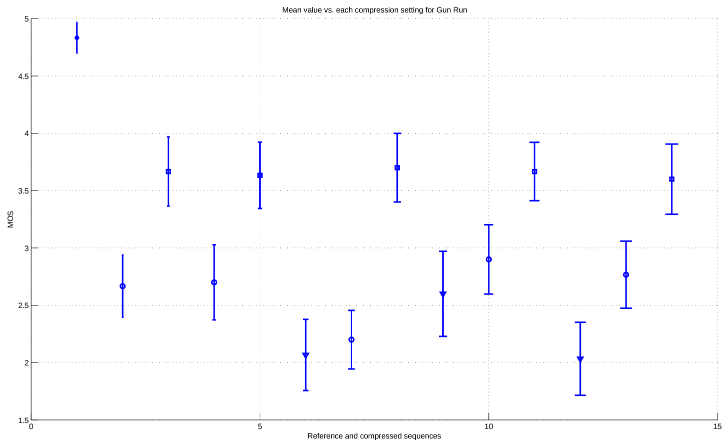
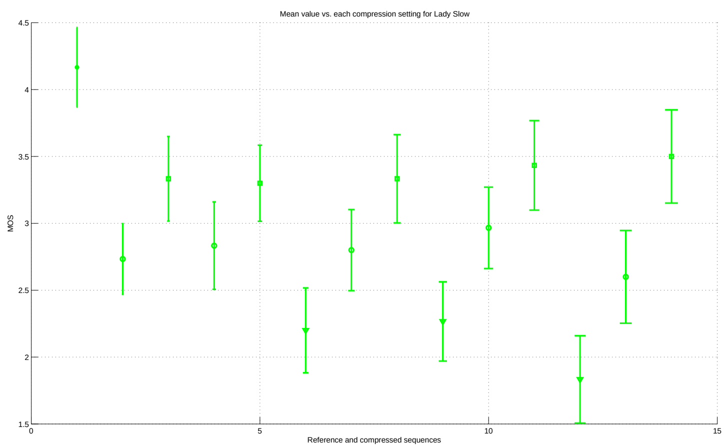


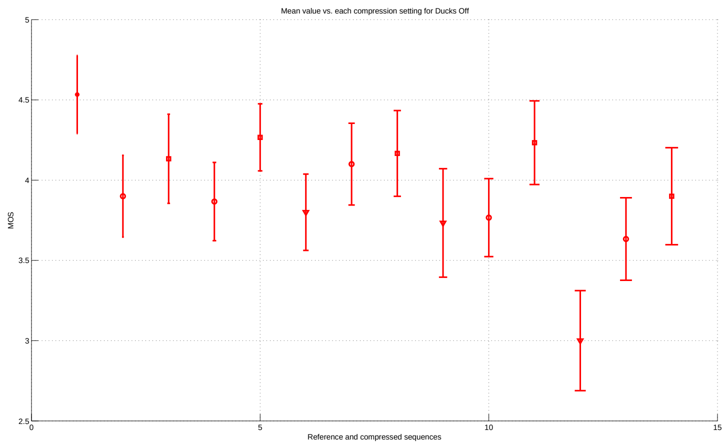
Figure C.2: MOS for each compression setting. The results presented are for each compressed sequences according to table 2.2. The colouring represent the setting changes that are not bitrate. The marks for the MOS value represents the different bitrates. 95% CI is included. Triangle = 8 Mbps, circle = 10 Mbps and square = 15 Mbps.



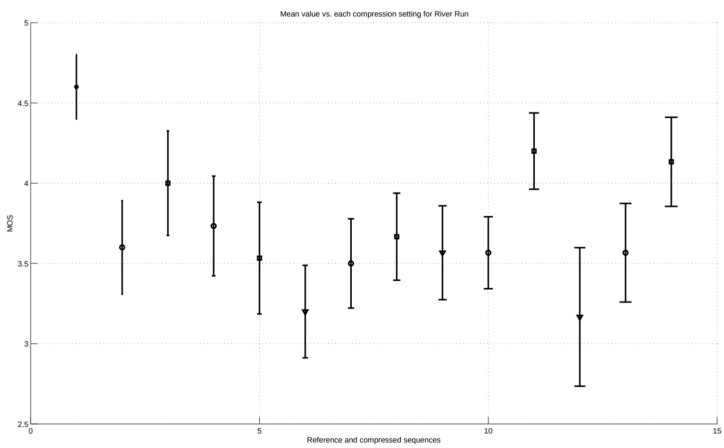
(a) MOS for Gun Run for each compression setting. The marks for the MOS value represents the different bitrates. 95% CI is included. Triangle = 8 Mbps, circle = 10 Mbps and square = 15 Mbps.



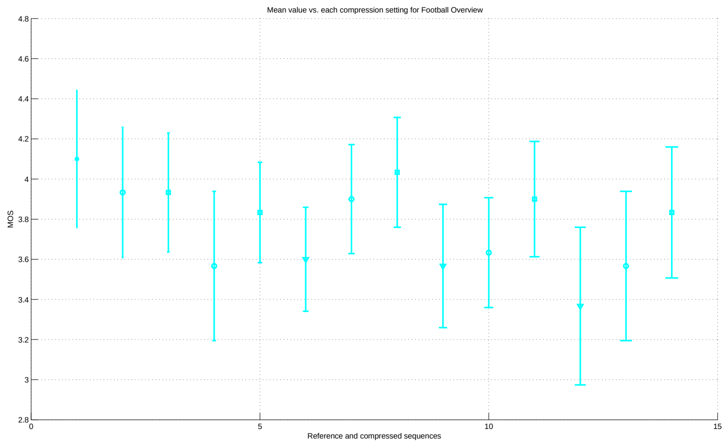
(b) MOS for Lady Slow for each compression setting. The marks for the MOS value represents the different bitrates. 95% CI is included. Triangle = 8 Mbps, circle = 10 Mbps and square = 15 Mbps.



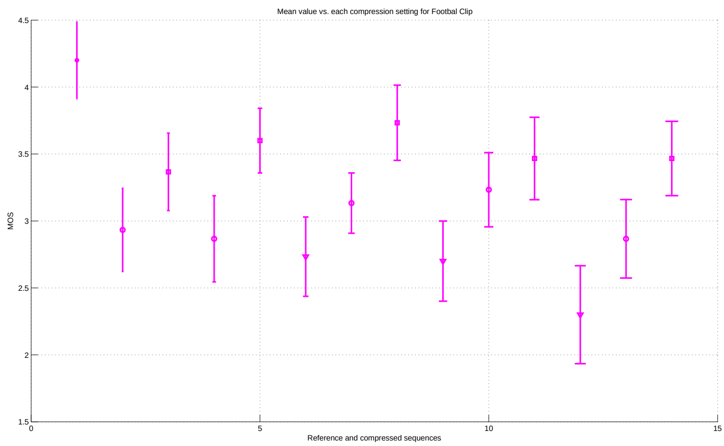
(c) MOS for Ducks Off for each compression setting. The marks for the MOS value represents the different bitrates. 95% CI is included. Triangle = 8 Mbps, circle = 10 Mbps and square = 15 Mbps.



(d) MOS for River Run for each compression setting. The marks for the MOS value represents the different bitrates. 95% CI is included. Triangle = 8 Mbps, circle = 10 Mbps and square = 15 Mbps.

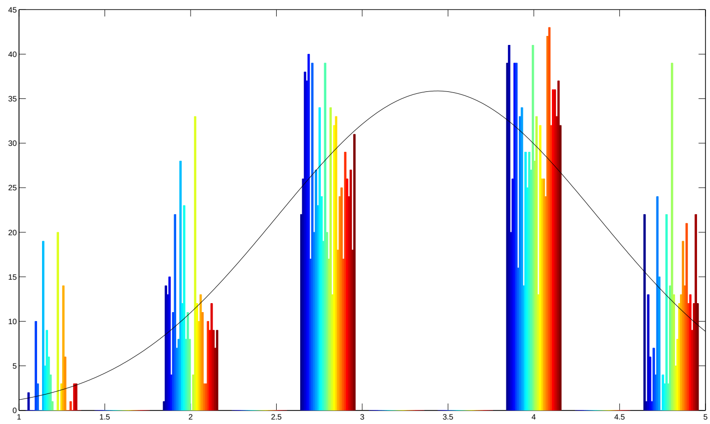


(e) MOS for Football Overview for each compression setting. The marks for the MOS value represents the different bitrates. 95% CI is included. Triangle = 8 Mbps, circle = 10 Mbps and square = 15 Mbps.

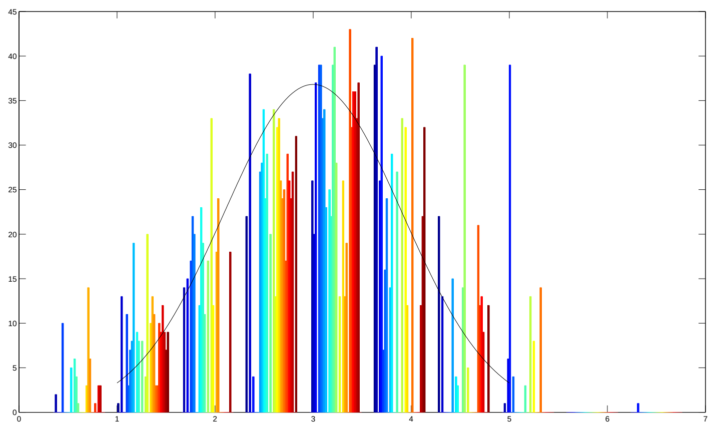


(f) MOS for Football Clip for each compression setting. The marks for the MOS value represents the different bitrates. 95% CI is included. Triangle = 8 Mbps, circle = 10 Mbps and square = 15 Mbps.

Figure C.3: All contents MOS plotted with respect to the compression settings. The compressing code, showed in table 2.2 is in increasing order, with the reference first in the graphical presentation. 95% CI is included using the *errorbar* matlab function.

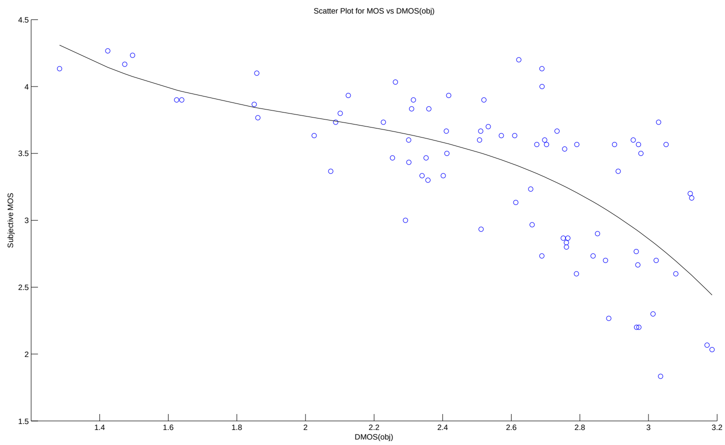


(a) Original scores. $\mu = 3.4377, \sigma = 0.9343$

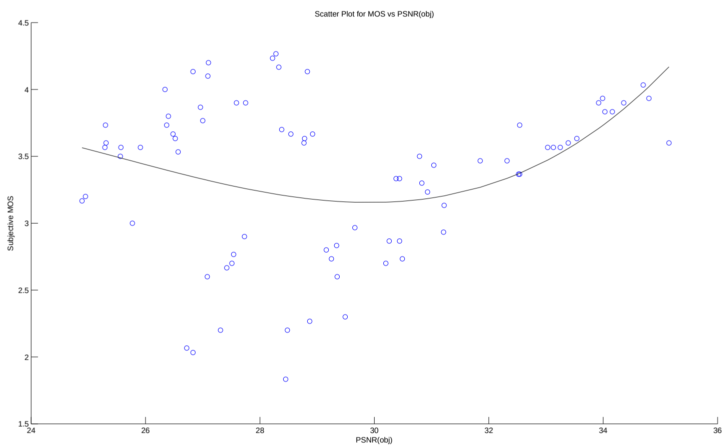


(b) Normalised scores. $\mu = 3, \sigma = 0.9103$

Figure C.4: Histogram of the votes for every sequence used. The black line is a normal distribution made by the mean and standard deviation for all contents.

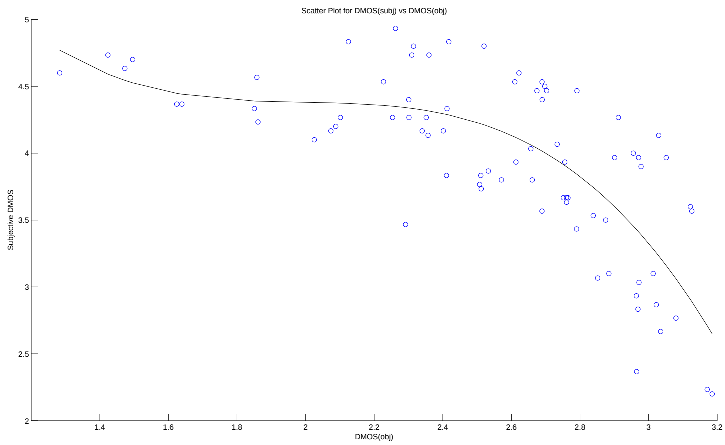


(a) MOS vs. DMOS(objective)

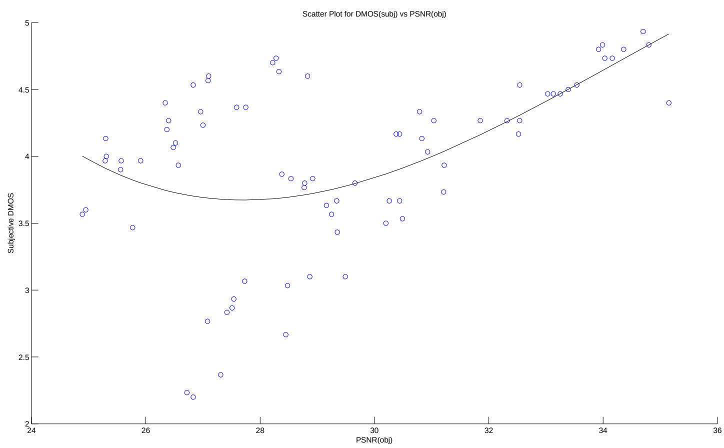


(b) MOS vs. PSNR(objective)

Figure C.5: Plot of the subjective relationship with DMOS and PSNR, with their fitting function included. The resulting correlation are found in table 3.3. Both fitting functions are made by the polynomial function.



(a) DMOS(subjective) vs. DMOS(objective)



(b) DMOS(subjective) vs. PSNR(objective)

Figure C.6: Plot of the subjective DMOS relationship with DMOS and PSNR, with their fitting function included. The resulting correlation are found in table 3.4. Both fitting functions are made by the polynomial function.

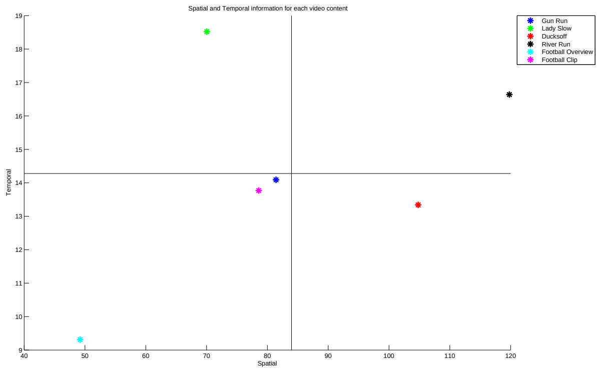


Figure C.7: The spatial and temporal information relationship between the different reference contents used in testing. The lines indicate the mean value across test material. Only reference videos are measured.

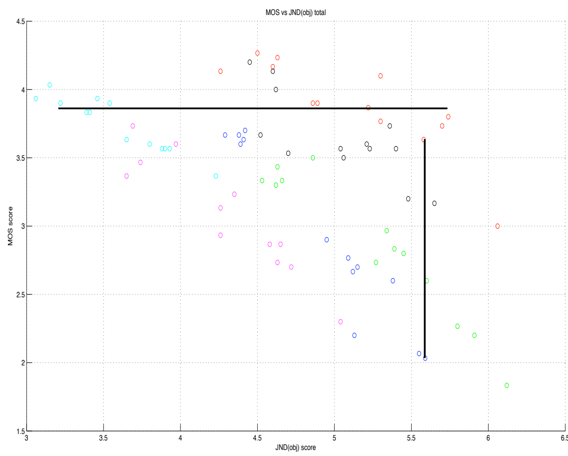
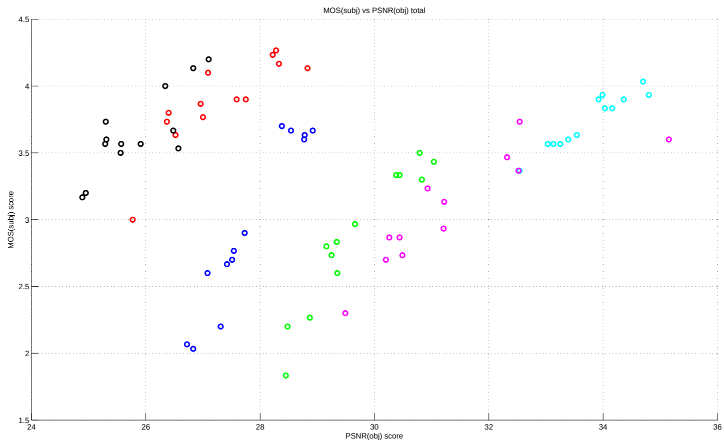
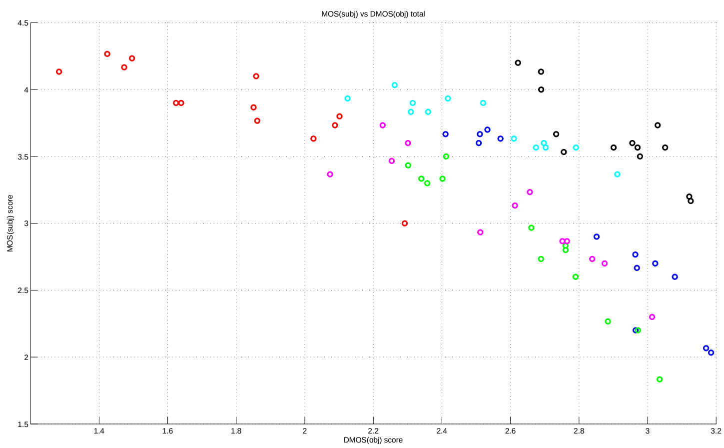


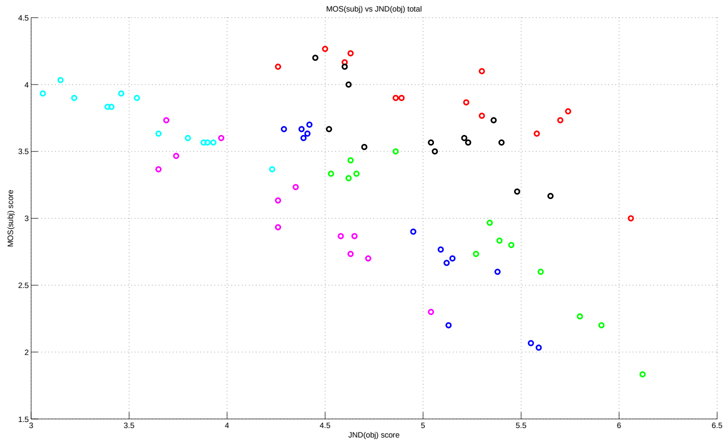
Figure C.8: Example of the variation in objective results, for a specific subjective value, and vice versa. The black lines indicate the differences. It is the relationship between MOS and JND that are represented.



(a) MOS vs. PSNR

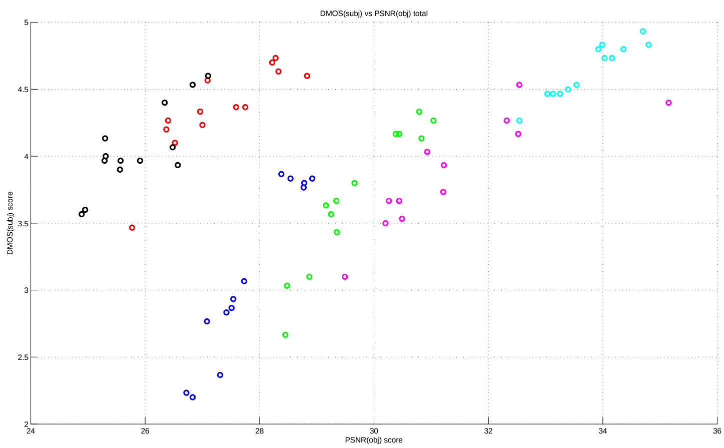


(b) MOS vs. DMOS

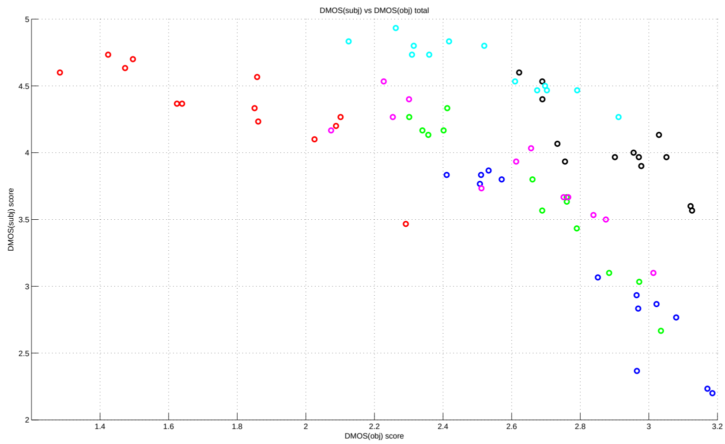


(c) MOS vs. JND

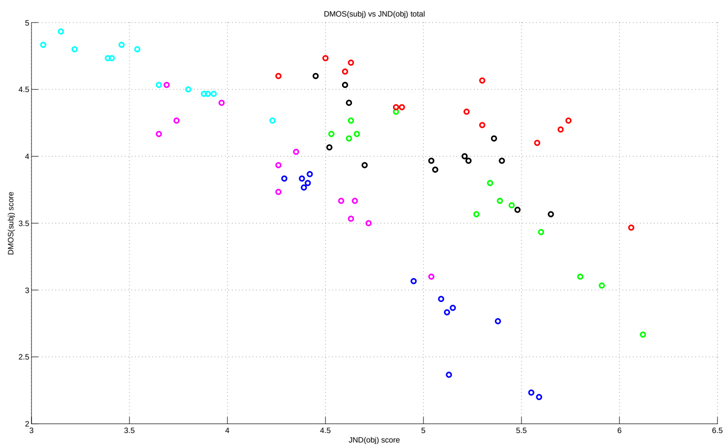
Figure C.9: MOS vs. objective results are plotted. Independent colour for each content described in Appendix D.



(a) DMOS(subjective) vs. PSNR

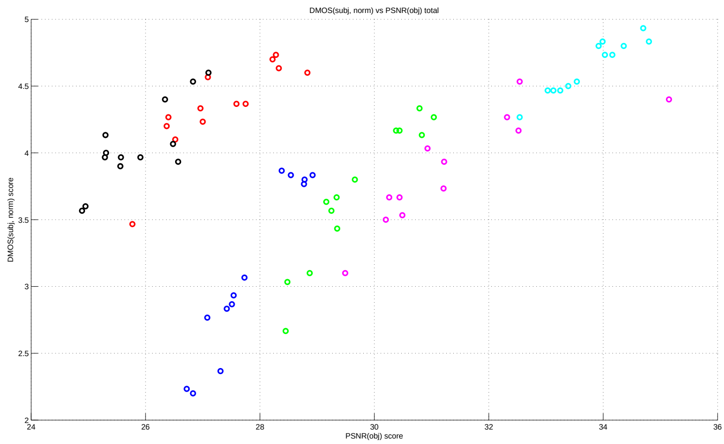


(b) DMOS(subjective) vs. DMOS

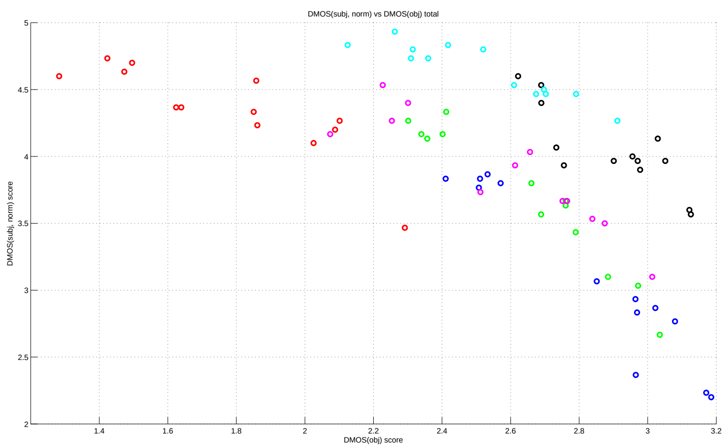


(c) DMOS(subjective) vs. JND

Figure C.10: Subjective DMOS vs. objective results are plotted. Independent colour for each content described in Appendix D.



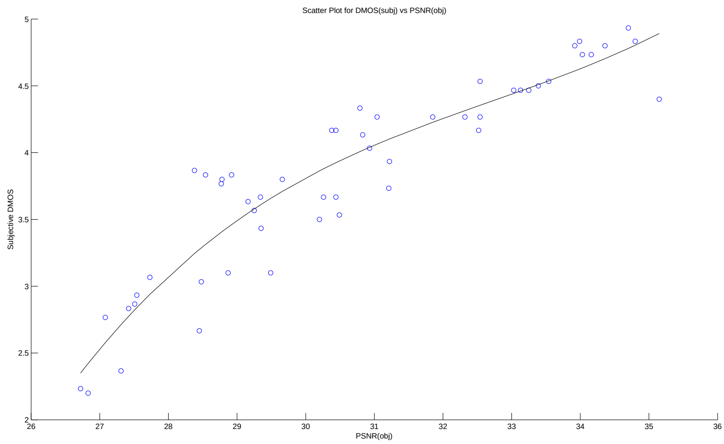
(a) DMOS(subjective, normalised) vs. PSNR



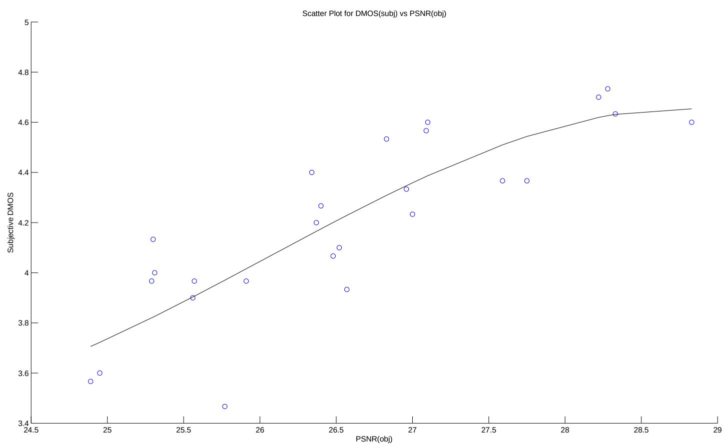
(b) DMOS(subjective, normalised) vs. DMOS

(c) DMOS(subjective, normalised) vs. JND

Figure C.11: The normalised subjective DMOS vs. objective results are plotted. Independent colour for each content described in Appendix D.

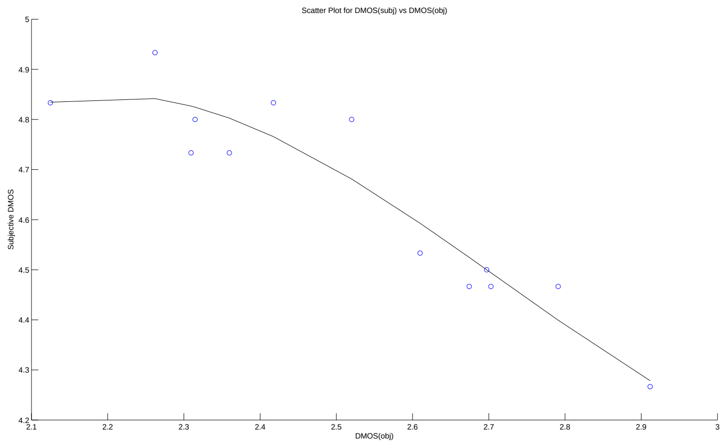


(a) Low spatial information

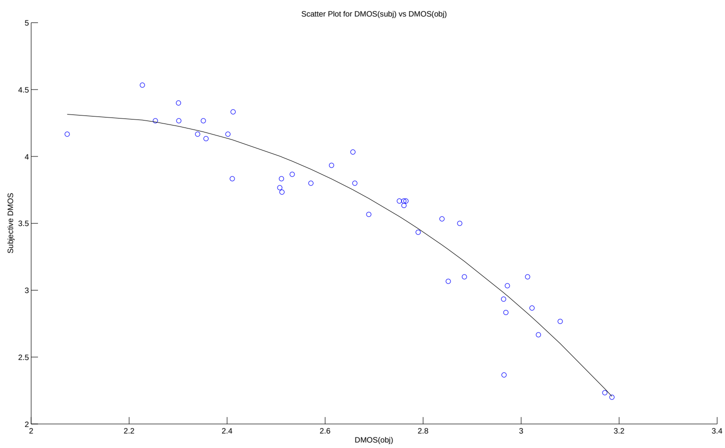


(b) High spatial information

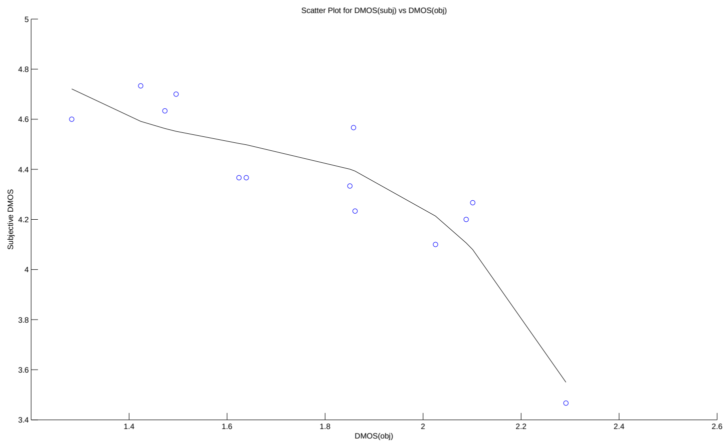
Figure C.12: Fitting functions for the individual groups of objective PSNR.



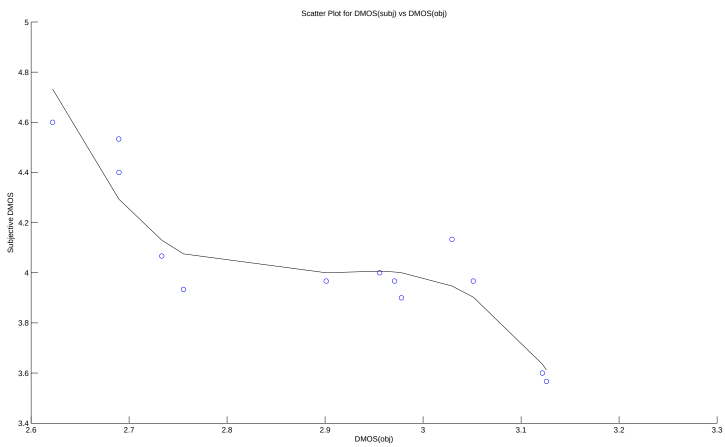
(a) <65 spatial information



(b) 65-85 spatial information

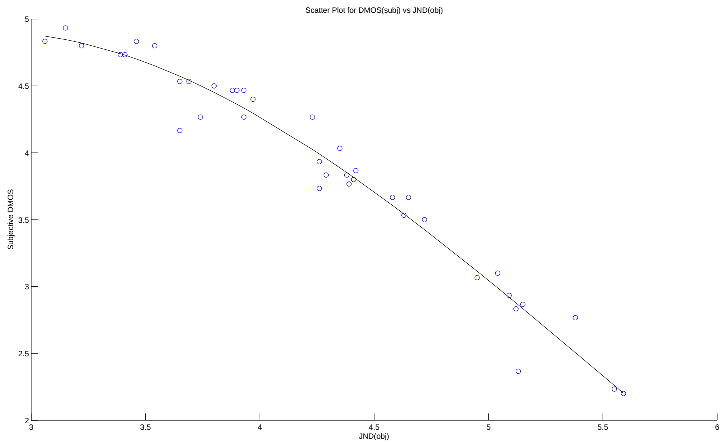


(c) 85-105 spatial information

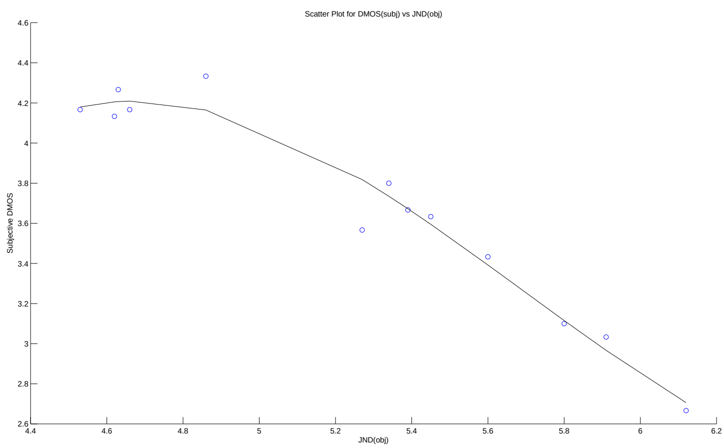


(d) >105 spatial information

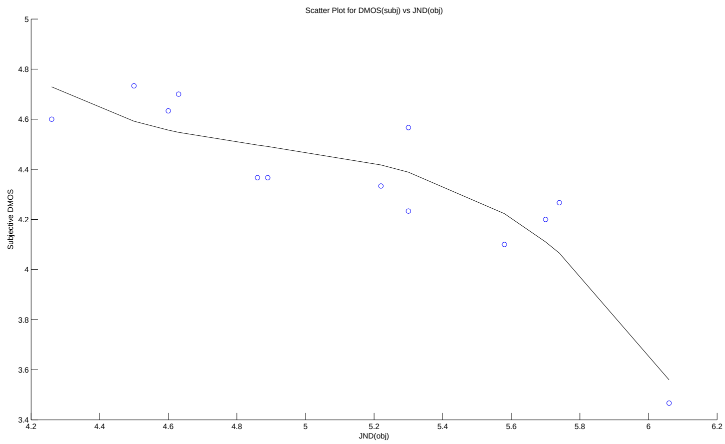
Figure C.13: Fitting functions for the individual groups of objective DMOS.



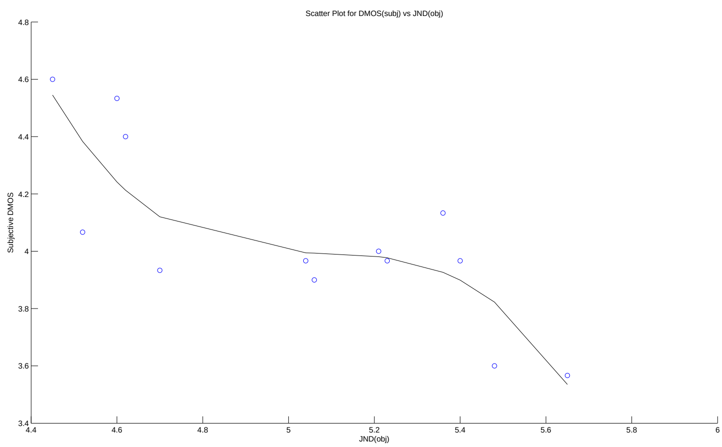
(a) Low spatial/Low temporal information



(b) Low spatial/High temporal information



(c) High spatial/Low temporal information



(d) High spatial/High temporal information

Figure C.14: Fitting functions for the individual groups of objective JND.

APPENDIX D

Sequences

A description of each of the sequences used in the training, stabilisation and test. All sequences are either taken from the SVT Fairytale or the finals between Molde and Aalesund fall 2009.

Training sequences:

- **Stockholm :** A video from a helicopter-borne camera flying over Stockholm. Sequence contains some movement and details. Picture seen in figure E.1(a).
- **People Gravel :** A sequence containing people moving around on gravel in a park. A variety of colour, movement and distance to objects. Picture seen in figure E.1(b).

Stabilisation sequences

- **Football 3 :** A sequence of a Molde offence. Molde move the ball against Aalesund's goal. Lot of movement, details in crowd and on the field. Picture seen in figure E.2(a).
- **Into Tree :** A sequence recorded from a helicopter flying towards a mansion, and zooming into a large tree beside the mansion. Some movement, but lots of details in the tree. Picture seen in figure E.2(b).

Test sequences:

- **River Run :** A group of people running beside a river. The camera is following the group from the other side of the river. Trees are passing by when the camera is moving. Movements in several different depth planes. Picture seen in figure E.3(f).
- **Football Clip :** The sequence starts with a close-up of an Aalesund player who dribbles a Molde player and make a pass into the middle of the field. The picture switches to an overview. Aalesund tries a kick against the goal, but Molde manages to defend. The close up camera contains a lot of details on the players in focus. The overview have a lot of movement and colourful crowd. Picture seen in figure E.3(b).
- **Gun Run :** The sequence starts with a close up picture on a gun firing. The gun moves out from the picture and the sequence shift to an overview of a crowd running on grass. Details on the gun in focus and lots of details and movement in the crowd. Picture seen in figure E.3(d).
- **Lady Slow :** The sequence starts with a woman in front of a tree cheering. The picture switches to a close-up of a running crowd. Many details in the grass. Picture seen in figure E.3(e).
- **Ducks Off :** A sequences of five ducks in water. The ducks fly away, and produces a lot of detailed movement in the water. Picture seen in figure E.3(a).
- **Football Overview :** A free kick from Aalesunds own half. The ball is kicked high into the other half of the pitch. The ball going backwards and forwards before moving down the pitch. The picture contains a lot of movements and many details in the pitch and crowd. Picture seen in figure E.3(c).

Content information:

Colour description of the different contents:
Blue – Gun Run, **Green** – Lady Slow, **Red** – Ducks off, **Black** – River Run, **Cyan** – Football Overview, **Magenta** – Football Clip

Table D.1: Results of a measurement of the spatial and temporal information. The results are measured according to [8] by the VQA system

	Gun Run	Lady Slow	Ducks Off	River Run	Football O	Football C
Spat	81.420	70.050	104.79	119.79	49.200	78.580
Temp	14.090	18.520	13.340	16.640	9.3100	13.770

APPENDIX E

Pictures

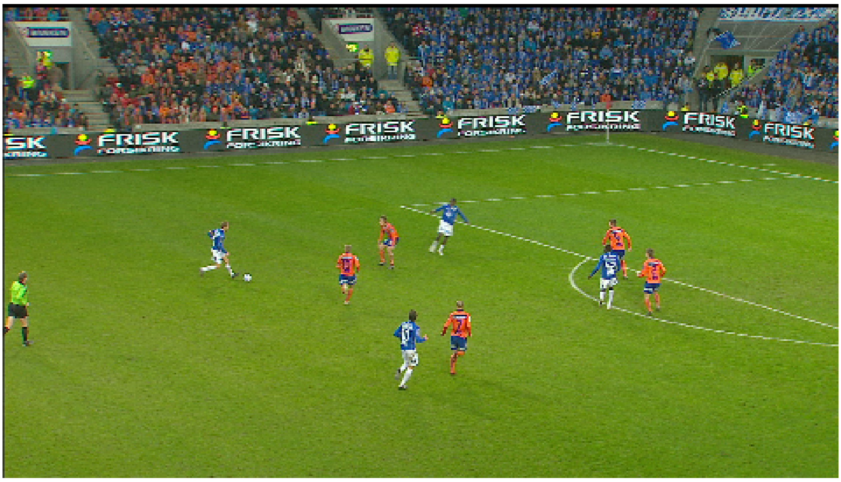
E.1 Training sequences

(a) Stockholm

(b) People Gravel

Figure E.1: Picture taken from the training sequences

E.2 Stabilisation sequences



(a) Football 3

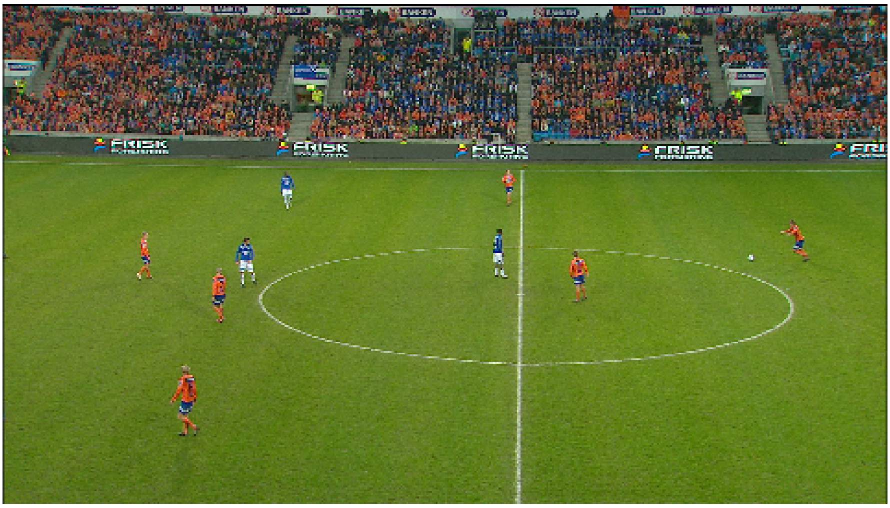
(b) Into Tree

Figure E.2: Picture taken from the stabilisation sequences

E.3 Test sequences

(a) Ducks Off

(b) Football Clip



(c) Football Overview

(d) Gun Run

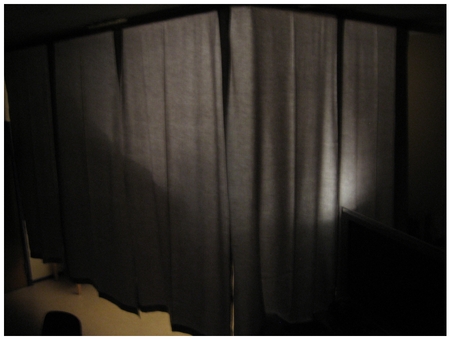
(e) Lady Slow

(f) River Run

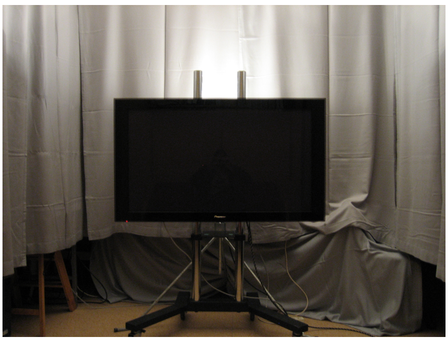
Figure E.3: Picture taken from the test sequences

E.4 Evaluation area pictures

(a) The seating of the subjects for the subjective testing



(b) Evaluation area in the dark, seen from outside. Outside the area, there were no light sources



(c) The screen shown from the seating position

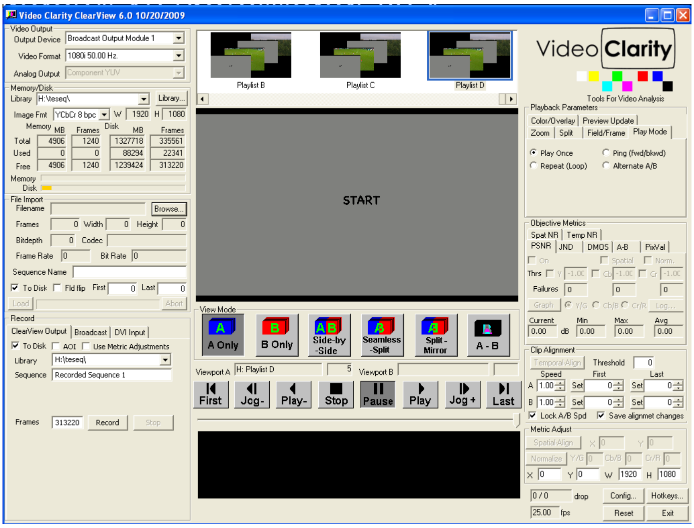
(d) The screen, floor and the gap between the floor and the curtains

(e) The placement of the lamp behind the screen

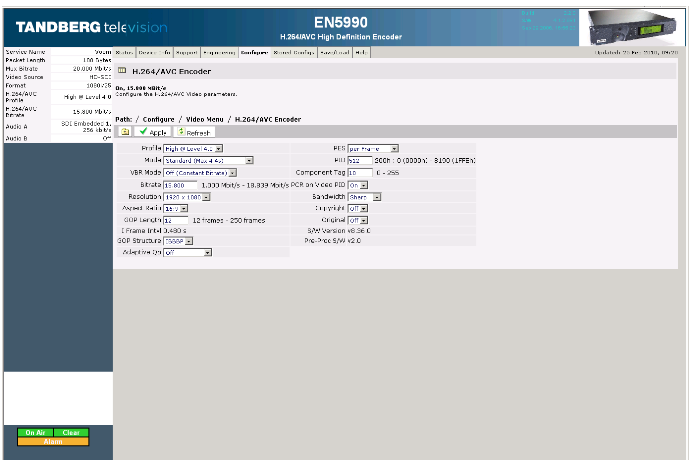
(f) The room seen from the subjects viewpoint

Figure E.4: The evaluation area

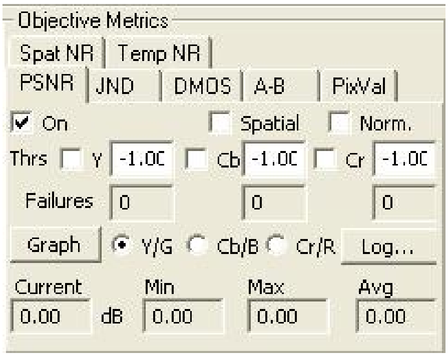
E.5 Screenshot of software



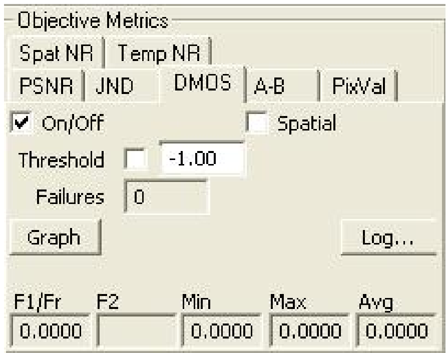
(a) Screenshot of the clearview software used for video editing and objective evaluation



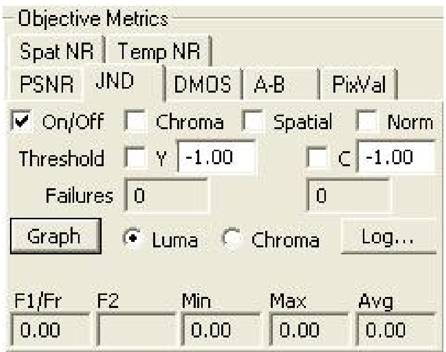
(b) Screenshot of the Tandberg EN5990



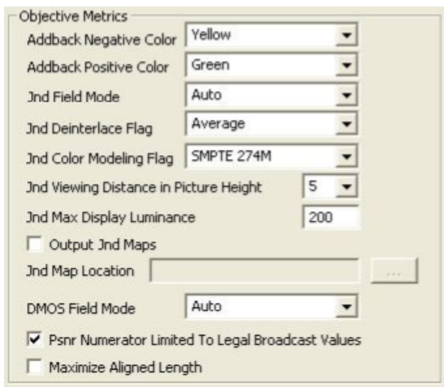
(c) PSNR setting for performing objective measurements



(d) DMOS setting for performing objective measurements



(e) JND setting for performing objective measurements

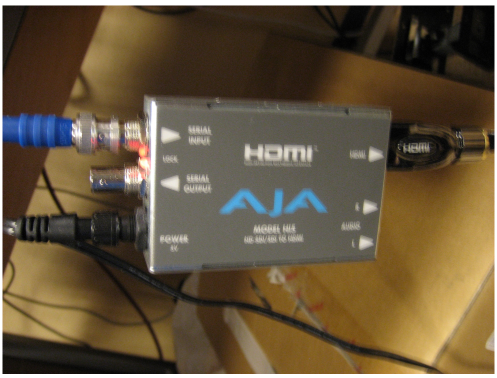


(f) Objective measurement configuration box

Figure E.5: Screenshot of the software used

E.6 Equipment

(a) The VQA system is located under the monitor. The decoder is hardware on top, on the left side of the picture. The encoder is below the decoder.



(b) The HD-SDI to HDMI converter used to play out from the VQA system on the monitor

Figure E.6: The VQA system and HD-SDI to HDMI converter

APPENDIX F

Subjective and objective testing sheets and results

On the following pages the instructions to the subjects, voting sheet and all results from the subjective testing will be presented.

Subjective intruction
The fist two pages contain the instructions given to the subjects before the subjective evaluation.

Voting sheet
The voting sheet that the subjects used for answering are presented on the next two pages.

Subjective result
On the next three pages all the subjects votes for the different sequences are presented, with the calculation of mean, standard deviation and confidence interval.

Objective result
The last page contain the objective results given by the VQA system.

Welcome.

Thank you for participating in our experiment on visual subjective quality evaluation. The test room is built to conform international recommendations regarding subjective evaluation of visual quality. This experiment is part of a master thesis regarding video quality. The aim is to improve broadcasting picture quality.

You are to evaluate a series of video sequences shown on the monitor in the room. Each sequence will be followed by a three (3) second grey screen. This is the voting time. Please grade each video sequence without making any comparison to the other sequences. The grey screen is for you to make your mark, so the quality should be determined during the video sequence. The grade is based on your experience of the whole ten seconds sequence.

The session will consist of the following:

- A test session of five sequences to get you familiar with the method.
- A short break where you can ask questions.
- A 19 minutes long “movie” which is the assessment session.
  Please keep focused through the session, and use your grading sheet.
  The test cannot be stopped after it has been started
    - All the video sequences last for 10 seconds.
    - You will only be able to see it once.
    - During these 10 seconds you have to decide the perceived quality of the test.
    - Followed by 3 seconds of grey screen.
      Make your grade on the sheet during the three seconds .
    - The grey screen shows the letter for the sequence, after each sequence, to help you grade the correct sequence.

The quality grading scale is represented by numbers from 1 to 5. The quality of each video sequence shown on the screen shall be judged by vote one of the numbers.

5 represent **Excellent** quality, which means there are no perceived distortion.
4 represent **Good** quality, which means that you see at least one distortion in the entire sequence but must pay attention to notice it.
3 represent **Fair** quality, which means that you see some distortion spread all over the sequence without effort.
2 represent **Poor** quality, which means that you see many noticeable, strong artifacts that cannot be overlooked.
1 represent **Bad** quality, which means you see a disaster, bothersome video sequence.

When giving a grade you should make an X inside the box next to the number you rank the quality. If you cross in wrong box, make a circle around it and make an X in the correct box.

Remember to switch off your phone and give it to the Test manager to make sure that the phone does not interrupt the test. Remain seated all the time during the test until it finishes. We need your judgement of the quality without comparison. Please remember to make your decision of the quality of a sequence based on that sequence and do not compare it to others in the session.

Thank you for participating!

Video Quality **NTNU** **Formal Subjective Test**

Trondheim 2010 March ……

	A	B	C	D	E	F	G	H	I	J
Excellent	5	5	5	5	5	5	5	5	5	5
Good	4	4	4	4	4	4	4	4	4	4
Fair	3	3	3	3	3	3	3	3	3	3
Poor	2	2	2	2	2	2	2	2	2	2
Bad	1	1	1	1	1	1	1	1	1	1

	K	L	M	N	O	P	Q	R	S	T
Excellent	5	5	5	5	5	5	5	5	5	5
Good	4	4	4	4	4	4	4	4	4	4
Fair	3	3	3	3	3	3	3	3	3	3
Poor	2	2	2	2	2	2	2	2	2	2
Bad	1	1	1	1	1	1	1	1	1	1

	U	V	W	X	Y	Z	AA	AB	AC	AD
Excellent	5	5	5	5	5	5	5	5	5	5
Good	4	4	4	4	4	4	4	4	4	4
Fair	3	3	3	3	3	3	3	3	3	3
Poor	2	2	2	2	2	2	2	2	2	2
Bad	1	1	1	1	1	1	1	1	1	1

	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN
Excellent	5	5	5	5	5	5	5	5	5	5
Good	4	4	4	4	4	4	4	4	4	4
Fair	3	3	3	3	3	3	3	3	3	3
Poor	2	2	2	2	2	2	2	2	2	2
Bad	1	1	1	1	1	1	1	1	1	1

	AO	AP	AQ	AR	AS	AT	AU	AV	AW	AX
Excellent	5	5	5	5	5	5	5	5	5	5
Good	4	4	4	4	4	4	4	4	4	4
Fair	3	3	3	3	3	3	3	3	3	3
Poor	2	2	2	2	2	2	2	2	2	2
Bad	1	1	1	1	1	1	1	1	1	1

	AY	AZ	BA	BB	BC	BD	BE	BF	BG	BH
Excellent	5	5	5	5	5	5	5	5	5	5
Good	4	4	4	4	4	4	4	4	4	4
Fair	3	3	3	3	3	3	3	3	3	3
Poor	2	2	2	2	2	2	2	2	2	2
Bad	1	1	1	1	1	1	1	1	1	1

	BI	BJ	BK	BL	BM	BN	BO	BP	BQ	BR
Excellent	5	5	5	5	5	5	5	5	5	5
Good	4	4	4	4	4	4	4	4	4	4
Fair	3	3	3	3	3	3	3	3	3	3
Poor	2	2	2	2	2	2	2	2	2	2
Bad	1	1	1	1	1	1	1	1	1	1

Name:………………………………... Gender: M ☐ F ☐ Vision Correction: Y ☐ N ☐

Age: Sequence:

Video Quality **NTNU** **Formal Subjective Test**

Trondheim 2010 March ……

		BS		BT		BU		BV		BW		BX		BY		BZ		CA		CB
Excellent	5		5		5		5		5		5		5		5		5		5	
Good	4		4		4		4		4		4		4		4		4		4	
Fair	3		3		3		3		3		3		3		3		3		3	
Poor	2		2		2		2		2		2		2		2		2		2	
Bad	1		1		1		1		1		1		1		1		1		1	

		CC		CD		CE		CF		CG		CH		CI		CJ		CK
Excellent	5		5		5		5		5		5		5		5		5	
Good	4		4		4		4		4		4		4		4		4	
Fair	3		3		3		3		3		3		3		3		3	
Poor	2		2		2		2		2		2		2		2		2	
Bad	1		1		1		1		1		1		1		1		1	

Name:……………………………….. Gender: M ☐ F ☐ Vision Correction: Y ☐ N ☐

Age: Sequence:

Seq number	Subjects 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	Mean	Std. Dev.	CI
Gun Run	5	5	5	4	4	5	5	5	5	4	5	5	5	5	5	5	5	4	5	5	5	5	4	5	5	5	5	5	5	5	4,833333	0,379049	0,1356385
Gun Run A	4	2	3	3	4	2	3	3	4	2	2	2	3	2	3	3	3	1	2	2	2	3	3	3	2	3	3	2	4	2	2,666667	0,758098	0,2712769
Gun Run B	3	4	4	5	3	4	4	4	4	2	3	3	4	2	5	4	4	2	4	3	4	5	4	4	3	4	3	3	5	4	3,666667	0,8441823	0,3020812
Gun Run C	4	2	2	4	3	4	3	4	3	2	1	2	3	3	3	4	3	1	2	2	1	2	2	4	2	3	3	3	3	3	2,7	0,9153857	0,3275606
Gun Run D	4	4	4	3	3	4	3	4	3	2	3	3	4	3	5	5	4	2	4	3	4	4	5	5	3	4	4	3	4	3	3,633333	0,8087169	0,2893903
Gun Run E	4	2	3	2	3	1	2	2	2	1	1	1	2	1	2	3	3	1	2	1	2	1	4	3	2	2	2	2	3	2	2,066667	0,868345	0,3107275
Gun Run F	3	2	2	2	2	1	1	3	2	1	3	2	3	2	3	3	3	1	2	2	1	2	2	3	3	3	2	2	2	3	2,2	0,7143842	0,2556344
Gun Run G	5	4	4	4	4	4	2	5	3	4	3	4	4	2	5	4	4	2	3	3	4	4	4	4	3	4	3	3	5	4	3,7	0,83666	0,2993894
Gun Run H	3	3	4	3	2	1	2	4	4	2	2	3	3	1	3	3	4	1	2	2	1	2	5	3	2	2	2	3	4	2	2,6	1,0372377	0,3711639
Gun Run I	3	2	3	5	4	2	3	4	4	2	3	2	3	2	4	4	3	1	3	3	2	3	3	3	3	2	2	3	3	3	2,9	0,8448628	0,3023247
Gun Run J	3	3	4	4	3	4	4	4	4	2	4	3	4	3	4	5	4	2	3	4	4	4	4	4	4	4	3	3	4	5	3,666667	0,711159	0,2544803
Gun Run K	3	3	3	2	2	1	2	2	4	1	1	2	3	2	1	3	2	1	2	1	1	2	3	3	1	3	1	1	3	2	2,033333	0,889918	0,3184472
Gun Run L	3	2	2	2	3	4	3	4	4	2	2	1	3	2	2	3	3	2	2	3	2	4	4	4	3	3	2	3	3	3	2,766667	0,8172002	0,2924259
Gun Run M	3	3	4	2	4	2	3	4	4	2	3	4	4	3	4	4	3	3	4	4	2	4	4	5	4	5	4	4	5	4	3,6	0,8550055	0,3059542
Lady Slow	4	3	3	4	4	4	5	5	4	2	4	5	5	3	5	5	4	4	4	4	3	5	4	5	5	4	5	3	5	5	4,166667	0,8339078	0,2984046
Lady Slow A	4	3	2	3	4	3	3	2	2	1	2	2	2	3	4	3	3	2	4	2	3	2	3	3	3	3	3	3	2	3	2,733333	0,73968	0,2646862
Lady Slow B	4	3	3	3	4	5	3	4	4	1	3	3	1	3	4	4	3	2	4	4	4	4	3	4	3	4	4	3	3	3	3,333333	0,8840866	0,3163605
Lady Slow C	4	3	3	2	4	3	2	3	2	1	2	3	3	2	4	3	2	1	3	3	1	3	4	3	3	3	4	4	4	3	2,833333	0,9128709	0,3266607
Lady Slow D	4	3	2	3	3	3	3	4	3	1	4	4	4	3	4	5	3	3	3	4	2	3	4	4	3	4	3	3	4	3	3,3	0,7943768	0,2842589
Lady Slow E	4	2	2	2	4	1	2	3	3	1	1	2	1	3	2	3	3	2	2	1	1	2	3	3	2	3	2	1	3	2	2,2	0,8866831	0,3172896
Lady Slow F	2	2	3	4	4	2	3	2	3	1	2	2	1	3	3	3	3	2	4	3	4	3	3	4	3	3	2	3	4	3	2,8	0,846901	0,3030541
Lady Slow G	4	3	3	2	4	3	4	5	3	1	2	4	4	3	4	4	3	2	3	3	4	2	4	5	3	3	4	4	4	3	3,333333	0,9222661	0,3300226
Lady Slow H	4	1	2	3	4	2	2	2	3	1	2	1	1	2	3	3	2	2	2	2	3	1	3	3	2	2	2	3	3	2	2,266667	0,827682	0,2961768
Lady Slow I	4	2	2	3	4	3	3	3	3	1	3	3	3	3	3	3	3	2	3	4	3	1	4	4	2	4	3	4	2	4	2,966667	0,8502873	0,3042658
Lady Slow J	4	3	3	2	3	4	4	4	3	1	3	2	4	3	4	4	3	2	3	4	4	3	5	4	4	3	3	5	5	4	3,433333	0,9352607	0,3346726
Lady Slow K	4	1	2	3	4	1	1	2	2	1	1	1	1	1	2	3	3	1	2	2	1	1	3	2	2	2	1	1	2	2	1,833333	0,9128709	0,3266607
Lady Slow L	4	2	3	2	3	4	2	3	3	1	2	1	1	3	4	3	2	2	2	4	1	2	3	3	3	2	4	2	4	3	2,6	0,9684684	0,3465556
Lady Slow M	5	2	2	2	4	4	3	5	4	2	3	4	4	4	5	3	3	2	4	4	4	2	4	4	3	4	4	5	3	3	3,5	0,9737946	0,3484615
Ducks Off	5	3	5	4	4	3	5	5	5	4	3	4	5	5	5	5	5	4	5	4	5	4	5	5	5	5	5	5	5	4	4,533333	0,6814454	0,2438476
Ducks Off A	5	4	4	3	4	4	3	5	4	3	3	4	4	4	3	5	5	3	3	4	3	3	4	5	4	4	4	5	4	4	3,9	0,7119667	0,2547693
Ducks Off B	5	4	5	3	3	4	4	5	5	2	4	4	5	4	4	4	5	5	3	3	5	5	4	4	5	4	4	4	4	4	4,133333	0,7760792	0,2777113
Ducks Off C	5	4	4	3	3	4	3	4	3	3	3	4	5	4	4	5	5	4	4	4	3	4	4	4	4	3	3	4	4	5	3,866667	0,6814454	0,2438476

Ducks Off D	4	4	5	4	4	4	3	5	5	3	4	4	5	4	5	5	4	4	4	4	4	5	4	5	5	4	4	4	4	5	4,266667	0,5832923	0,2087246
Ducks Off E	5	4	4	4	4	4	3	5	4	2	3	4	4	4	3	4	4	4	3	5	4	3	3	4	4	4	4	4	4	3	3,8	0,6643638	0,2377352
Ducks Off F	4	4	4	3	4	4	3	5	4	3	5	3	5	4	4	5	5	4	3	3	4	5	4	5	4	5	4	4	5	4	4,1	0,7119667	0,2547693
Ducks Off G	5	3	5	5	3	4	4	4	5	3	4	4	4	4	5	5	5	4	3	4	5	4	3	4	5	4	5	3	5	4	4,166667	0,74664	0,2671768
Ducks Off H	4	4	5	3	3	3	2	4	5	3	3	3	4	4	4	5	5	2	4	4	5	4	3	4	4	2	5	3	5	3	3,733333	0,9444332	0,3379549
Ducks Off I	5	4	4	3	4	5	3	4	3	4	3	3	5	4	4	4	5	3	3	3	4	3	3	4	4	3	4	4	4	4	3,766667	0,6789106	0,2429406
Ducks Off J	4	4	5	5	4	3	3	5	5	3	4	4	4	4	5	5	5	5	3	5	5	5	5	4	4	3	4	4	4	4	4,233333	0,727932	0,2604823
Ducks Off K	4	4	4	4	3	2	2	3	2	2	2	3	3	3	4	4	4	3	3	3	5	3	2	3	4	2	2	2	3	2	3	0,8709883	0,3116734
Ducks Off L	4	4	5	4	4	4	2	4	4	3	4	2	4	4	3	5	4	4	3	3	4	4	3	4	3	3	3	4	4	3	3,633333	0,7183954	0,2570698
Ducks Off M	5	4	4	3	3	4	3	4	4	3	3	2	5	3	5	5	5	3	4	3	4	5	4	5	4	4	3	4	4	5	3,9	0,8448628	0,3023247
River Run	4	4	5	4	4	4	5	5	5	3	5	4	4	5	5	5	4	5	4	5	5	5	4	5	5	5	5	5	5	5	4,6	0,5632418	0,2015498
River Run A	3	4	5	4	4	2	4	4	3	2	4	3	5	3	3	4	4	3	3	4	3	3	4	5	4	3	3	4	5	3	3,6	0,8136762	0,2911649
River Run B	3	4	5	4	4	5	4	5	3	2	4	3	5	4	4	5	5	2	4	5	3	4	3	4	3	4	4	5	5	5	4	0,9097177	0,3255323
River Run C	5	4	3	4	3	3	3	4	3	2	3	2	5	3	3	5	4	5	4	5	3	4	4	4	5	4	4	3	4	4	3,733333	0,868345	0,3107275
River Run D	3	4	4	4	4	4	2	3	4	2	4	3	2	2	3	5	5	4	5	2	2	3	4	5	4	4	4	4	4	3	3,533333	0,9732042	0,3482503
River Run E	3	3	4	3	3	3	3	4	3	1	4	3	3	3	3	4	4	2	5	3	2	3	4	5	3	3	3	3	3	3	3,2	0,8051558	0,288116
River Run F	5	4	3	4	3	5	3	4	3	3	4	3	5	3	3	4	4	3	4	3	3	4	4	3	3	3	2	2	4	4	3,5	0,7768193	0,2779761
River Run G	4	4	3	4	3	4	3	5	4	2	4	2	4	4	3	5	4	4	4	4	4	4	4	4	3	4	3	2	4	4	3,666667	0,758098	0,2712769
River Run H	4	4	4	5	3	3	2	3	3	2	4	2	4	3	3	4	4	3	5	3	5	3	4	4	3	4	4	4	4	4	3,566667	0,8172002	0,2924259
River Run I	3	4	4	3	3	4	3	3	4	2	4	3	3	4	4	5	4	4	4	3	4	4	3	4	3	3	4	4	4	3	3,566667	0,6260623	0,2240294
River Run J	3	4	4	3	4	5	4	4	5	3	4	4	5	4	4	5	4	5	5	5	4	4	5	4	4	5	4	3	5	4	4,2	0,6643638	0,2377352
River Run K	5	4	3	3	5	1	2	3	4	2	4	1	3	2	2	4	4	2	5	2	1	3	4	4	2	4	4	4	4	4	3,166667	1,2058288	0,4314923
River Run L	4	3	4	3	4	4	4	4	2	2	2	2	4	4	3	5	4	3	4	4	2	5	4	4	4	4	4	4	4	3	3,566667	0,8583598	0,3071545
River Run M	4	3	5	5	4	4	4	5	4	2	3	4	4	4	4	5	4	4	4	5	5	5	5	5	3	4	4	3	5	4	4,133333	0,7760792	0,2777113
Football Overview	4	4	2	4	3	4	2	5	5	4	4	5	5	3	4	5	3	2	4	5	4	5	4	5	5	4	5	5	5	4	4,1	0,9595257	0,3433556
Football Overview A	5	4	3	3	3	3	3	5	5	4	3	4	4	3	4	5	3	2	4	3	4	5	5	5	4	5	3	5	4	5	3,933333	0,9071871	0,3246268
Football Overview B	3	4	4	4	3	4	4	4	4	3	4	4	4	4	4	5	3	1	4	4	5	5	4	4	5	5	4	4	5	3	3,933333	0,827682	0,2961768
Football Overview C	3	3	2	2	3	3	3	5	4	3	4	2	4	3	5	4	3	1	3	4	4	3	4	5	5	4	4	5	4	5	3,566667	1,0400044	0,372154
Football Overview D	4	4	3	3	4	4	3	5	4	4	3	4	5	3	4	4	3	2	4	4	4	4	3	5	4	4	4	5	4	4	3,833333	0,6989319	0,250105
Football Overview E	3	3	3	3	3	4	3	3	4	3	4	3	5	3	4	5	3	2	4	4	3	4	4	4	3	4	4	4	4	5	3,6	0,7239737	0,2590659

Football Overview F	3	4	5	4	4	4	4	4	4	4	4	4	5	3	3	5	4	2	3	5	5	4	5	4	3	3	4	4	3	4	3,9	0,7588558	0,2715481
Football Overview G	4	4	3	2	3	4	4	5	4	3	4	4	5	4	4	5	4	3	4	5	3	4	5	4	4	5	4	4	5	5	4,033333	0,7648905	0,2737075
Football Overview H	3	4	3	3	4	4	3	3	4	3	4	3	4	3	4	5	3	1	3	3	4	5	3	5	3	5	4	4	3	4	3,566667	0,8583598	0,3071545
Football Overview I	4	4	3	4	3	4	3	4	4	4	3	3	5	4	4	5	3	1	3	4	4	4	3	4	4	4	3	4	4	3	3,633333	0,7648905	0,2737075
Football Overview J	5	4	3	3	3	4	4	5	5	2	4	4	4	4	4	5	4	2	4	3	3	5	4	4	4	5	4	4	4	4	3,9	0,8030116	0,2873487
Football Overview K	5	4	3	3	3	4	2	4	4	2	3	3	4	4	2	5	3	1	3	5	1	4	4	3	4	5	3	4	2	4	3,366667	1,0980652	0,3929304
Football Overview L	5	3	3	2	3	2	4	4	5	4	3	4	5	3	4	4	4	1	4	3	1	4	5	4	4	4	4	4	4	3	3,566667	1,0400044	0,372154
Football Overview M	5	4	3	4	3	4	2	5	5	3	4	3	3	4	4	5	4	2	3	4	2	5	4	4	5	4	5	4	4	4	3,833333	0,9128709	0,3266607
Football Clip	5	4	3	3	4	4	3	5	5	3	5	4	5	4	5	4	4	3	4	3	5	5	5	5	4	4	5	5	5	3	4,2	0,8051558	0,288116
Football Clip A	4	3	3	3	3	1	2	3	3	1	3	2	2	3	4	4	4	1	3	3	3	3	3	4	4	3	3	3	3	4	2,933333	0,868345	0,3107275
Football Clip B	5	3	3	4	3	4	3	4	3	1	3	2	3	3	4	4	4	2	3	3	4	3	4	4	3	3	4	4	4	4	3,366667	0,8087169	0,2893903
Football Clip C	4	3	3	3	3	2	3	3	3	3	2	2	2	3	2	4	3	2	3	3	1	1	5	2	4	4	3	3	4	3	2,866667	0,8995529	0,321895
Football Clip D	4	4	3	3	3	3	3	4	4	3	3	2	4	4	4	5	4	3	4	4	4	3	3	4	4	4	3	4	5	3	3,6	0,6746647	0,2414212
Football Clip E	3	2	3	2	3	3	2	3	2	1	3	1	3	3	4	3	3	2	3	2	3	1	4	4	3	3	3	3	4	3	2,733333	0,827682	0,2961768
Football Clip F	4	3	3	3	4	2	3	4	4	4	3	2	3	3	4	3	3	2	3	3	2	3	4	3	3	3	4	3	3	3	3,133333	0,6288102	0,2250127
Football Clip G	4	4	3	3	4	5	2	4	4	4	4	3	3	4	4	5	3	2	5	3	3	3	4	4	4	4	4	4	5	4	3,733333	0,7849153	0,2808732
Football Clip H	4	2	3	3	3	1	2	2	3	3	3	2	2	3	3	4	3	1	3	4	1	2	3	3	3	4	3	3	2	3	2,7	0,83666	0,2993894
Football Clip I	5	3	3	4	3	3	3	3	3	2	3	2	3	3	4	5	4	2	4	3	3	3	3	4	4	3	2	3	3	4	3,233333	0,7738544	0,2769151
Football Clip J	4	3	3	3	4	3	3	3	3	4	3	4	2	3	4	4	3	1	3	3	3	5	5	4	4	4	4	3	5	4	3,466667	0,8603661	0,3078724
Football Clip K	3	2	2	4	2	1	1	3	3	3	3	1	4	1	2	4	4	1	2	3	1	2	4	2	2	2	1	2	2	2	2,3	1,0221681	0,3657714
Football Clip L	3	3	3	3	4	2	2	3	3	4	4	1	2	3	4	4	3	1	3	3	2	3	3	3	4	3	2	2	3	3	2,866667	0,8193072	0,2931799
Football Clip M	4	4	3	3	3	4	3	4	4	2	3	3	4	3	4	5	3	2	4	3	2	3	4	4	4	5	3	4	3	4	3,466667	0,7760792	0,2777113

Seq number	PSNR	DMOS	JND	Seq number	PSNR	DMOS	JND
Gun Run	100,0000	0,0000	0,00	River Run	100,0000	0,0000	0,00
Gun Run A	27,4200	2,9690	5,1200	River Run A	25,3100	2,9556	5,2100
Gun Run B	28,5400	2,5109	4,3800	River Run B	26,3400	2,6897	4,6200
Gun Run C	27,5100	3,0223	5,1500	River Run C	25,3000	3,0295	5,3600
Gun Run D	28,7800	2,5710	4,4100	River Run D	26,5700	2,7555	4,7000
Gun Run E	26,7200	3,1708	5,5500	River Run E	24,9500	3,1217	5,4800
Gun Run F	27,3100	2,9652	5,1300	River Run F	25,5600	2,9779	5,0600
Gun Run G	28,3800	2,5328	4,4200	River Run G	26,4800	2,7333	4,5200
Gun Run H	27,0800	3,0797	5,3800	River Run H	25,2900	3,0513	5,4000
Gun Run I	27,7300	2,8513	4,9500	River Run I	25,9100	2,9011	5,0400
Gun Run J	28,9200	2,4106	4,2900	River Run J	27,1000	2,6219	4,4500
Gun Run K	26,8300	3,1854	5,5900	River Run K	24,8900	3,1259	5,6500
Gun Run L	27,5400	2,9643	5,0900	River Run L	25,5700	2,9709	5,2300
Gun Run M	28,7700	2,5075	4,3900	River Run M	26,8300	2,6894	4,6000
Lady Slow	100,0000	0,0000	0,00	Football O	100,0000	0,0000	0,00
Lady Slow A	29,2500	2,6891	5,2700	Football O A	33,9900	2,4174	3,4600
Lady Slow B	30,4400	2,3399	4,5300	Football O B	34,8000	2,1248	3,0600
Lady Slow C	29,3400	2,7606	5,3900	Football O C	33,1300	2,6743	3,9000
Lady Slow D	30,8300	2,3570	4,6200	Football O D	34,0300	2,3094	3,3900
Lady Slow E	28,4800	2,9719	5,9100	Football O E	33,3900	2,6973	3,8000
Lady Slow F	29,1600	2,7609	5,4500	Football O F	33,9200	2,5201	3,5400
Lady Slow G	30,3800	2,4016	4,6600	Football O G	34,7000	2,2621	3,1500
Lady Slow H	28,8700	2,8842	5,8000	Football O H	33,0300	2,7910	3,9300
Lady Slow I	29,6600	2,6607	5,3400	Football O I	33,5400	2,6099	3,6500
Lady Slow J	31,0400	2,3014	4,6300	Football O J	34,3600	2,3147	3,2200
Lady Slow K	28,4500	3,0353	6,1200	Football O K	32,5400	2,9117	4,2300
Lady Slow L	29,3500	2,7899	5,6000	Football O L	33,2500	2,7028	3,8800
Lady Slow M	30,7900	2,4123	4,8600	Football O M	34,1600	2,3596	3,4100
Ducks Off	100,0000	0,0000	0,00	Football Clip	100,0000	0,0000	0,00
Ducks Off A	27,5900	1,6391	4,8900	Football C A	31,2100	2,5120	4,2600
Ducks Off B	28,8300	1,2829	4,2600	Football C B	32,5200	2,0736	3,6500
Ducks Off C	26,9600	1,8506	5,2200	Football C C	30,2600	2,7515	4,6500
Ducks Off D	28,2800	1,4236	4,5000	Football C D	35,1500	2,3008	3,9700
Ducks Off E	26,4000	2,1013	5,7400	Football C E	30,4900	2,8385	4,6300
Ducks Off F	27,0900	1,8580	5,3000	Football C F	31,2200	2,6131	4,2600
Ducks Off G	28,3300	1,4731	4,6000	Football C G	32,5400	2,2270	3,6900
Ducks Off H	26,3700	2,0879	5,7000	Football C H	30,2000	2,8747	4,7200
Ducks Off I	27,0000	1,8612	5,3000	Football C I	30,9300	2,6567	4,3500
Ducks Off J	28,2200	1,4959	4,6300	Football C J	32,3200	2,2536	3,7400
Ducks Off K	25,7700	2,2915	6,0600	Football C K	29,4900	3,0132	5,0400
Ducks Off L	26,5200	2,0253	5,5800	Football C L	30,4400	2,7649	4,5800
Ducks Off M	27,7500	1,6242	4,8600	Football C M	31,8500	2,3515	3,9300