UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

FRANCO ALMADA VALDEZ

OBJECTIVE VIDEO QUALITY ASSESSMENT CONSIDERING FRAME AND DISPLAY TIME VARIATION

Diploma Thesis

Prof. Dr. Sergio Bampi
Advisor

Prof. Dr. Raphael Guerra
Co-Advisor

Porto Alegre, July 2012

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
Rector: Carlos Alexandre Netto
Vice-rector: Rui Vicente Oppermann
Graduation pro-rector: Valquíria Linck Bassani
Informatics Institute Director: Luís C. Lamb
Comgrad Coordinator: Raul F. Weber
Informatics Institute Librarian-Chief: Beatriz Regina Bastos Haro

ACKNOWLEDGEMENTS

To my parents, Cristina Almada and Gerson Valdez, which had always, give me everything they could and more, for me to have conditions to be here today. Parents that transmitted me values that I will always carry for my whole life. To my uncle, Ricardo Valdez, and my grandmother, Gládis Lindmeyer, that helped me in my education when I was young.

I would like to thank all my friends that uphold on me for all these years. To my colleagues in Brazil that helped during my graduation, and to the colleagues I have met in Germany that always provided good advice for my work. To my professors, Sergio Bampi, Gerhard Fohler and Raphael Guerra, that gave me the opportunity to work and study in this really interesting research area.

And I cannot forget my woman, my friend, Greice Nunes, the one that stayed with me every day in this last year, when I was ten thousand kilometers away, listening to my problems and my doubts, cheering me up when I needed most, loving me unconditionally.

Finally, I would like to express my gratitude for everyone that helped me during this last year, the ones that stayed at my side even when I was far away, the ones that I really tend to follow for the rest of my life.

Thank you all!

CONTENTS

LIST OF ABBREVIATIONS

ATSC	*Advanced Television System Committee*
AVC	*Advanced Video Coding*
CODEC	*COder/DECoder*
CDVL	*Consumer Digital Video Library*
DTS	*Decoding Time Stamp*
DVB	*Digital Video Broadcasting*
DVD	*Digital Versatile Disc*
EDTV	*Extended Definition Television*
FPS	*Frames-Per-Second*
GB	*Giga Bytes*
GOP	*Group of Pictures*
HDTV	*High Definition Television*
IPTV	*Internet Protocol Television*
ISDB	*Integrated Services Digital Broadcasting*
ISO	*International Organization for Standardization*
ITU-T	*International Telecommunication Union – Telecommunication*
JVT	*Joint Video Team*
MB	*Mega Bytes*
MOVIE	*MOtion-based Video Integrity Evaluation*
MPEG	*Motion Pictures Experts Group*
MSE	*Mean Square Error*

NTIA	*National Telecommunications and Information Administration*
NTSC	*National Television System Committee*
PAL	*Phase Alternating Line*
PES	*Packetized Elementary Stream*
PEVQ	*Perceptual Evaluation of Video Quality*
PS	*Program Stream*
PSNR	*Peak Signal-to-Noise Ratio*
PTS	*Presentation Time Stamp*
RGB	*Red Green Blue*
SDTV	*Standard Definition Television*
SSIM	*Structural Similarity Index Metric*
TS	*Transport Stream*
UFRGS	*Universidade Federal do Rio Grande do Sul*
VCEG	*Video Coding Experts Group*
VFD	*Variable Frame Delay*
VQEG	*Video Quality Experts Group*
VQM	*Video Quality Metric*
YCbCr	*Luminance, Chrominance Blue, Chrominance Red*

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Video quality assessment is becoming more important in current digital video application. Most of the video encoders and decoders focus in providing more visual quality in less amount of digital information, being the compression one of the most important solutions for making possible the transmission of high quality videos in the current digital transmission media. Sometimes, however, not only video compression is good enough for video quality transmission, maybe the available network bandwidth or the ability to process the video by a low-end processor is not sufficient for continuous playback. Many techniques that consider time variabilities of different sorts need to be developed to keep a digital video watchable.

This work focus on a new timeline approach to compare the quality of videos for which a presentation time stamp variability is allowed. Based on video quality metrics already established the new approach can be used to help the decision about the most suitable method for real-time video transmission. Among the many possible techniques able to cope with time stamp variability, one can mention the variable delay time in video presentation, or the use of frame skipping technique, or even more compression applied to reduce the bandwidth and/or processing - improving the perceived video quality under scarce resource availability.

A new timeline approach to calculate the current video quality metrics used in most of the encoders and decoders available was developed. This timeline approach takes into account not only the frames themselves, as a frame-by-frame evaluation, but both the frame and presentation time stamp are considered in the proposed approach to the problem. The quality comparison is based on a new discretization, which hopefully serves better the quality of digital video without losing perceptual quality in the process. All the work is tested in an offline solution program that is able to give numeric quality index to the tested sample videos. Those samples are selected from the current distribution of the German digital video television system, encoded in MPEG-2.

Keywords: video quality metrics, video processing, digital television.

1. INTRODUCTION

During the last decade an incredible expansion in transmission of digital video content was presented, with digital video television being one of the most important ways of distributing this content. Currently in Germany almost half of the homes have access to digital television **(DVB, 2008)**, transmitted via digital terrestrial (DVB-T), digital cable (DVB-C), digital satellite (DVB-S/S2) or IPTV.

Since this work focus in videos distributed by the German digital television with its current standards and technologies, basically the MPEG-2 Transport Stream video was used as main source of coding for this work. This standard and others details of the system are explained in later chapters of this work.

In this new era of information with vast digital applications, with video transmission being one of the most common uses of media communication systems, some problems may occur in the reception of digital video content. Maybe the video transmission has been corrupted (distorted) or maybe the receptor is not able to decode the video in real-time, etc. These problems normally have solution, and these solutions are called real-time video adaptation, which represents the use of techniques that permits the video to be played in the best possible way according to restrictions of time, processing and/or network requirements.

The main objective of this work is to give a video quality index taking in account the presentation time stamps variation. Already proven video quality metrics that can predict the overall quality of the playback video are used. Those metrics tested are computed in a new timeline approach for video quality assessment. In this work the use of these metrics was done to raise the perceptual quality assessment of videos with variable delays. This evaluation tried to fit better the human visual perception; exploring the video quality techniques without losing perceptual visual quality in the process.

All selected metrics were classified and detailed. The proposed timeline approach for using these metrics is compared to the classic approach used for testing video quality today. All the source code for the metrics is available for download and when necessary for better understanding are given. Much of already developed material is used in this work and cited when referenced. Offline development is completely available for access.

This text was built in a structured way that follows as this. Chapter 2 deals with the description of the European digital television system used in the source video, some explanation about how works a video and how it is composed in the current systems. Used standards and information regarding digital video quality are showed. In chapter 3 the description and classification of the used video quality metrics are explained. Chapter 4 brings the way these metrics in today measure systems are used. The classical approach for using them is compared with the proposed new timeline approach for using video quality metrics with variable presentation time stamps in videos. In chapter 5 the implementation of the algorithm and the offline solution is proposed, in this chapter also the test methodology and parameters are detailed. Chapter 6 gives the results, with graphs comparisons and quality analysis of the proposed timeline solution. Chapter 7 brings the conclusions of this work.

2. VIDEO CONCEPTS AND STANDARDS

For better understanding of video related standards some concepts and important information about digital video processing are presented. These concepts are related to modern digital television systems and were considered in this work. This brief introduction has purposes in only explaining basically what is useful for comprehension of this thesis.

2.1 BASIC CONCEPTS

The definition of video is a sequence of images. In the current European format exactly 25 images (frames) are used to construct 1 second of video. However this can vary depending on the format adopted by the video system. In this thesis 25 FPS (frames per second) format was used. Below an image shows the video concept.

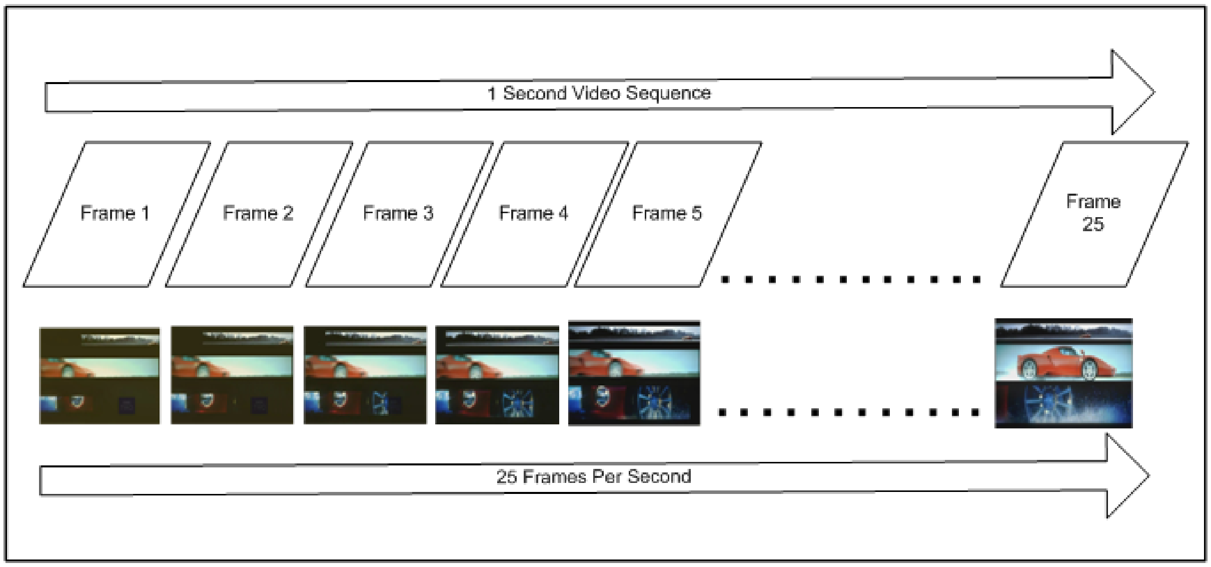


Figure 2.1: Video sequence of one second adopting 25 FPS

Each of these images presented has a fixed resolution; these resolutions represent the size of the picture in terms of lines and columns, or as typically seen width vs. height. Video resolution adopts the size of an image in pixels. Where pixel is the smallest point that can be represent in a digital image. In a color image each of these pixels has a representation of three colors; normally RGB (red-green-blue) and each of these pixels can have different color tones represented by different bits. Other pixel is "YCbCr" (luminance, chrominance blue and red).

Below an image shows different resolutions typically seen in current digital video systems:

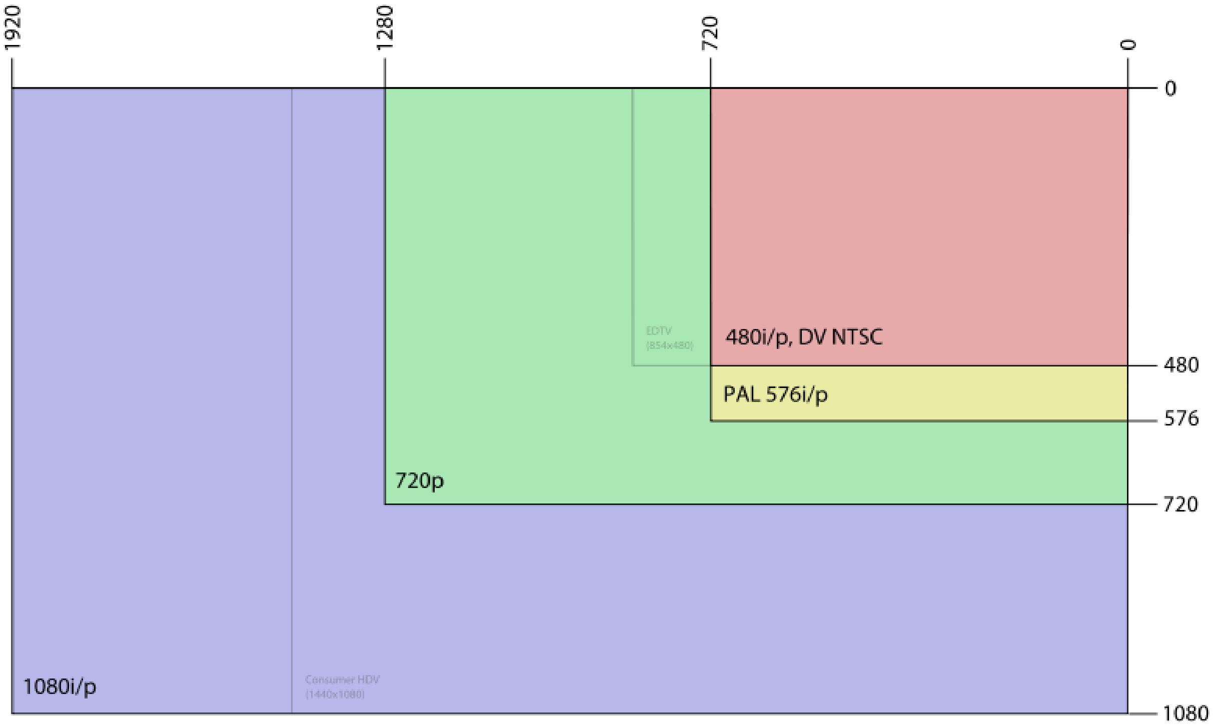


Figure 2.2: Typical Resolutions ***(NWE, 2011)***

The amount of data required for a video is exactly the resolution of the images that build it multiplied by the number of bits used for each pixel. The following equation expresses the size of an image in terms of digital space in bits.

$$Image\ Size\ =\ width\ x\ height\ x\ color\ bits$$

Considering that one second of video has a defined number of frames it is possible to calculate the video size like this:

$$Video\ Size\ =\ width\ x\ height\ x\ color\ bits\ x\ frames\_per\_second$$

Below one table listing the most common image resolutions used in digital television and the respective size in pixels and bits:

	WIDTH	HEIGHT	PIXELS	16-bit COLOR	24-bit COLOR
NTSC	720	480	345600	5529600	8294400
PAL	720	576	414720	6635520	9953280
HD READY	1280	720	921600	14745600	22118400
FULL HD	1920	1080	2073600	33177600	49766400

Table 2.1: Common Image Resolutions ***(*WANG Z. AND BOVIK A. C.**, ***2009)***

An image with 720x576 resolution (PAL) and 16 color bits has 6635520 bits or 0.79MB, and a video with the same resolution should have 25 times that (if 25 FPS), or exactly 19.77MB in one second, simplifying 19.77MB/s. That is a really impressive amount of data and it is really difficult to transmit and work with such data using current computers and networks.

When using a high definition video with 1920x1080 resolution and 24 color bits having 30 frames-per-second the bandwidth is 177.97MB/s. That is really impractical data to work with, no computer or network with narrow band (6 MHz) have the structure to handle this kind of data.

With huge amount of data that videos produce it is needed a way to compress these videos to a size more suitable for processing and transmitting. In Germany the Digital Video Broadcasting ***(DVB, 2008)*** is a suite of international open standards that determine the way like digital video television must be processed and used, from the compression to the transmission. Both topics are explained in the next subsection.

2.2 VIDEO COMPRESSION AND TRANSMISSION

Compression area is basically divided in two types of compression ***(BOSI, 2002)***: lossless and lossy. In first type the compression is without quality losses and the retrieve of the original image/video is possible. A small margin of compress ratio is reached in this case, at best 3:1 ratio ***(BOSI, 2002)***. In lossless compression Huffman coding is a most common example.

In the second type there is accepted perceptual loss of quality in video and there is no way to recover the original video. In this method a bigger margin of compress ratio needed for current video demand is possible, 100:1 ratio or higher ***(BOSI, 2002)***. In lossy compression problems are faced, like the tradeoff "compression vs. quality", one of the important problems in video compression today. These lossy methods introduce some artifacts.

Many different lossy compression standards have been published in the last years. In this work however only the one used in the German digital television system, the Motion Pictures Expert Group part 2 is presented ***(ITU-T, 2008)***.

2.2.1 Motion Pictures Expert Group Part 2 Standard

MPEG-2 is an international standard used as the format of digital television signals that are broadcasted by terrestrial, cable and some satellite TV systems. It also specifies the format of movies and other programs that are distributed on DVD. Normally, TV stations, TV receivers, DVD players and other equipment are usually designed to this standard ***(DVB, 2008)***.

The standard current defines two distinct container formats. One is the MPEG transport stream (TS), designed to carry digital video and audio over possibly lossy media (broadcasting) like ATSC, DVB, ISDB and HDV. The other MPEG-2 container defines the MPEG program stream (PS), a format designed for supporting file-based media such as hard disk drives, optical discs and flash memory.

Other related features supported by MPEG-2 are: interlaced video, where every frame is composed by two fields; and progressive video, where every picture is a complete frame. The standard supports subsampling of chrominance values too, due the fact that the eye is more suitable for the luminance, considering part of the chrominance to reduce data rate. It currently supports 4:2:2 (half chrominance can be removed), 4:2:0 (three quarters of chrominance removed) and 4:4:4 (no chrominance removed.

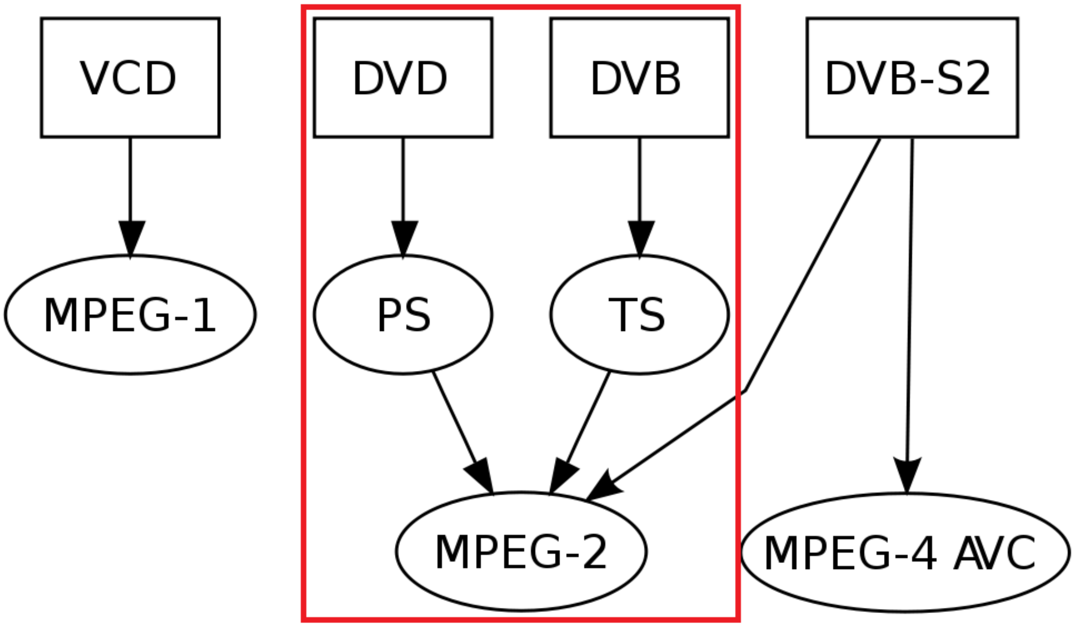


Figure 2.3: Typical MPEG-2 containers ***(HORKY, 2010)***

2.2.2 MPEG-2 DESCRIPTION

The MPEG-2 specifies that raw frames can be compressed into three kinds of frames: intra-coded frames (I-frames), predictive-coded frames (P-frames), and bidirectional-predictive-coded frames (B-frames).

I-frame is a compressed version of a single frame, i.e I-frames do not depend on data from the preceding or the following frames. Briefly, each raw frame is divided into “8 pixels by 8 pixel blocks”. The data in each block is transformed by a discrete cosine transform (DCT). The result is an “8 by 8 matrix of coefficients”. The transformation converts spatial information into frequency variations, but it doesn’t change the information in the block; the original block can be reconstructed exactly by applying the inverse cosine transform. The reason of doing this process is that the image can be simplified by quantizing the coefficients; however it loses some subtle details in brightness and color. In next step, the quantized coefficient matrix is compressed. Normally, every 15th frame ***(DVB, 2008)*** there is an introduction of an I-frame. P-frames and B-frames might follow an I-frame as, “IBBPBBPBBPBB(I)”, to form what is called Group Of Pictures (GOP).

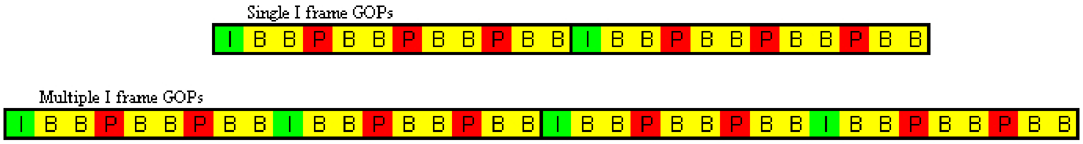

Single I frame GOPs
I B B P B B P B B P B B I B B P B B P B B P B B
Multiple I frame GOPs
I B B P B B P B B I B B P B B P B B I B B P B B P B B I B B P B B P B B

Figure 2.4: GOP Structure ***(ITU-T, 2008)***

Predictive-frames provide more compression than I-frames because they take advantage of previous I-frame or P-frame data, called as reference frame. To generate a P-frame, the previous reference frame must be first reconstructed, just as it would be in a receiver. The frame being compressed is divided into “16 pixel by 16 pixel macro-blocks” and for each macro-block, the reconstructed reference frame is searched to find the “16 by 16 macro-block” that best matches with the macro-block being compressed. The offset of both positions is encoded as a "motion vector". However the match between two macro-blocks normally is not equal and to adjust this, the encoder calculates the difference of all corresponding pixels of both macro-blocks. This residual data is appended to the motion vector and the result sent to the receiver or stored on the video for each macro-block being compressed. Sometimes no suitable match is found ***(ITU-T, 2008)*** and the macro-block is treated like an I-frame macro-block.

The processing of B-frames is similar to P-frames except that B-frames can consider subsequent reference frame as well as the picture in a preceding reference frame.

The MPEG-2 video supports different applications from mobile to high quality video, and for many of these applications it is too expensive or unpractical to support the hole standard, so, to allow such applications to support only subsets of it, the standard defines profiles and level.

The MPEG-2 profile defines a subset of features such as compression algorithm, chroma format, and more. The level defines the subset of quantitative capabilities such as maximum bit rate, maximum frame size, and more.

The table below shows the most used profiles:

	Name	Picture Coding	Chroma Format	Aspect Ratios	Scalable modes	Intra DC Precision
SP	Simple profile	I, P	4:2:0	square pixels, 4:3, or 16:9	none	8, 9, 10
MP	Main profile	I, P, B	4:2:0	square pixels, 4:3, or 16:9	none	8, 9, 10
SNR	SNR Scalable profile	I, P, B	4:2:0	square pixels, 4:3, or 16:9	SNR (signal-to-noise ratio) scalable	8, 9, 10
Spatial	Spatially Scalable profile	I, P, B	4:2:0	square pixels, 4:3, or 16:9	SNR- or spatial-scalable	8, 9, 10
HP	High profile	I, P, B	4:2:2 or 4:2:0	square pixels, 4:3, or 16:9	SNR- or spatial-scalable	8, 9, 10, 11
422	4:2:2 profile	I, P, B	4:2:2 or 4:2:0	square pixels, 4:3, or 16:9	none	8, 9, 10, 11
MVP	Multi-view profile	I, P, B	4:2:0	square pixels, 4:3, or 16:9	Temporal	8, 9, 10

Table 2.2: MPEG-2 Profiles ***(WANG Z., 2003)***

2.2.3 MPEG-2 Transport Stream

Transport Stream is a standard format for transmission and storage of audio, video, and Program and System Information Protocol (PSIP) data.

Transport stream specifies a container format encapsulating packetized elementary streams, with error correction and stream synchronization features for maintaining transmission integrity when the signal is degraded (packet loss).

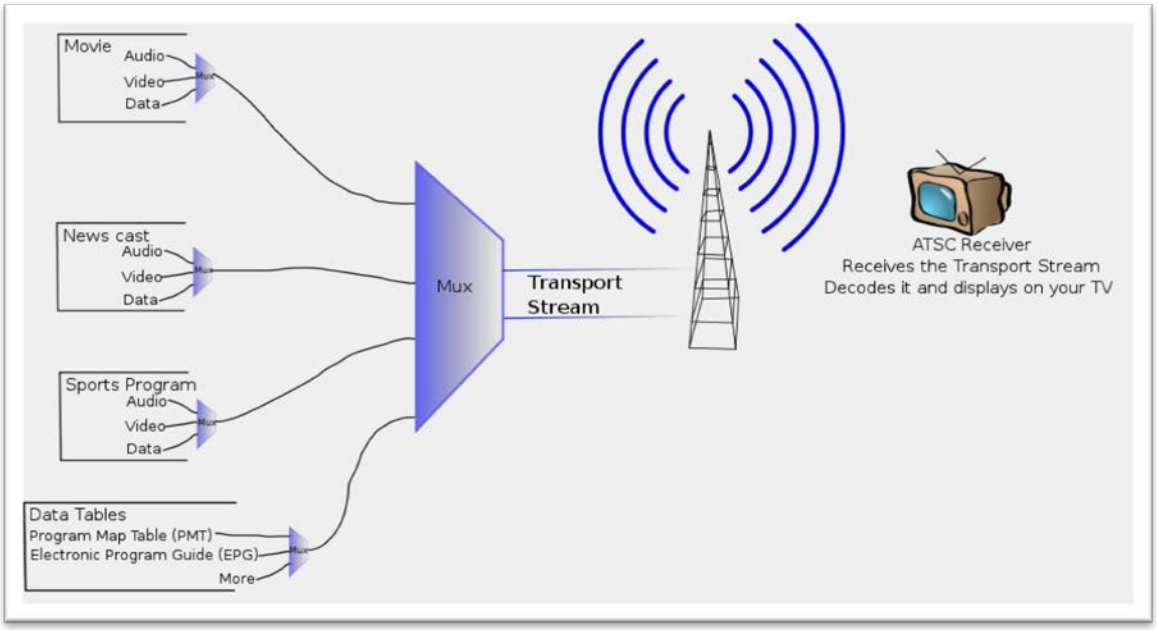


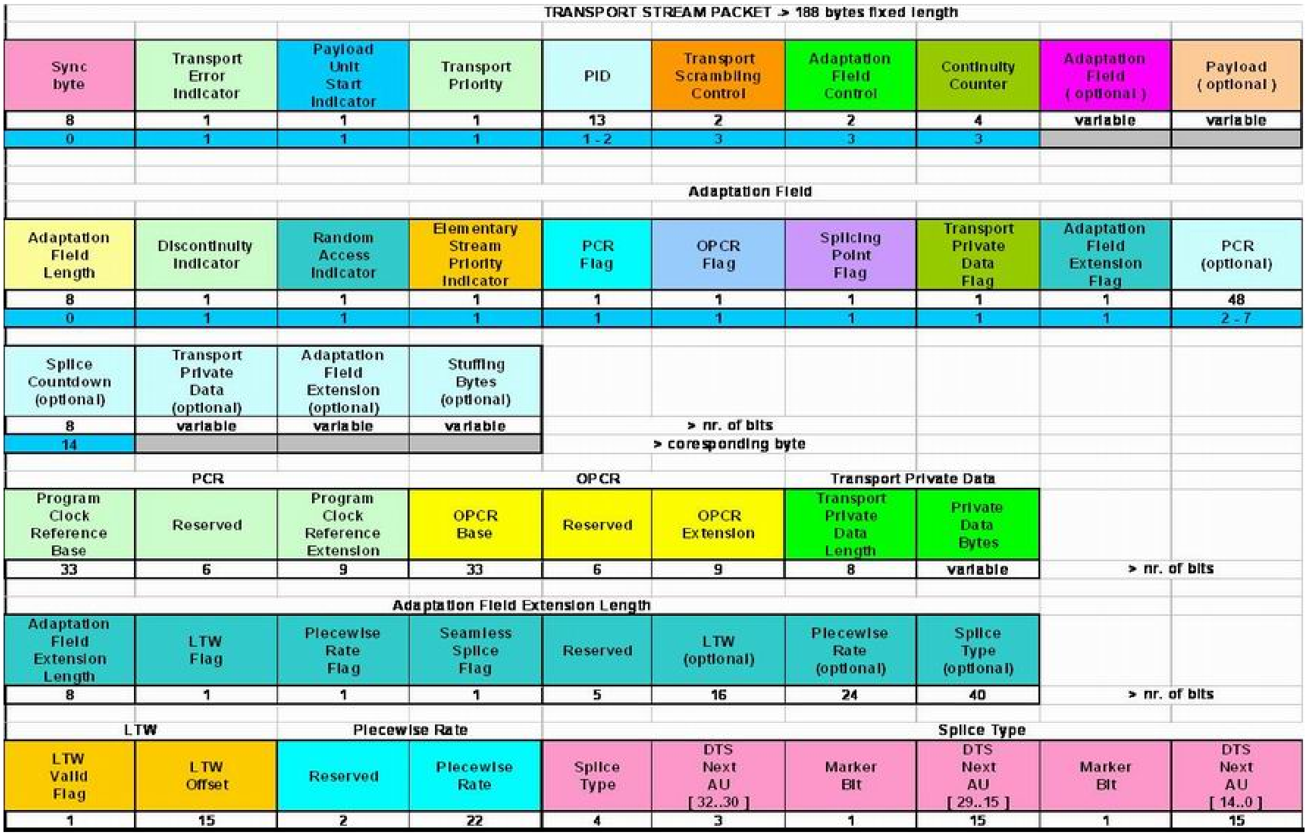
Figure 2.5: Transport Stream Transmission ***(CARDENAS, 2006)***

Transport stream is processed by the receiver in layers. An example stream containing video may be processed as follows:

- Composition of the various programs;
- Packetized elementary stream (PES);
- Elementary stream (ES) — audio or video;
- Group of pictures (GOP) — providing random access points;
- Slice — preventing an error from being propagated through intra prediction;
- Macro-block—consisting of 6 to 12 DCT blocks;
- Encoding block or just block—a DCT encoding block, 8x8 pixels.

A Transport Stream consists of fixed length packets. There can be more sizes but the standard one is 188 bytes ***(ATSC, 2003)***. A packet can contain info from a Program Stream; can be a null packet, a Program Association Table or a Conditional Access Table. The PID is the field deciding the packet that is used for audio and/or video.

Below is the composition of the transport stream:



TRANSPORT STREAM PACKET -> 188 bytes fixed length									
Sync byte	Transport Error Indicator	Payload Unit Start Indicator	Transport Priority	PID	Transport Scrambling Control	Adaptation Field Control	Continuity Counter	Adaptation Field (optional)	Payload (optional)
8	1	1	1	13	2	2	4	variable	variable
0	1	1	1	1 - 2	3	3	3		
Adaptation Field									
Adaptation Field Length	Discontinuity Indicator	Random Access Indicator	Elementary Stream Priority Indicator	PCR Flag	OPCR Flag	Splicing Point Flag	Transport Private Data Flag	Adaptation Field Extension Flag	PCR (optional)
8	1	1	1	1	1	1	1	1	48
0	1	1	1	1	1	1	1	1	2 - 7
Splice Countdown (optional)	Transport Private Data (optional)	Adaptation Field Extension (optional)	Stuffing Bytes (optional)						
8	variable	variable	variable	> nr. of bits					
14				> coresponding byte					
PCR			OPCR			Transport Private Data			
Program Clock Reference Base	Reserved	Program Clock Reference Extension	OPCR Base	Reserved	OPCR Extension	Transport Private Data Length	Private Data Bytes		
33	6	9	33	6	9	8	variable	> nr. of bits	
Adaptation Field Extension Length									
Adaptation Field Extension Length	LTW Flag	Piecewise Rate Flag	Seamless Splice Flag	Reserved	LTW (optional)	Piecewise Rate (optional)	Splice Type (optional)		
8	1	1	1	5	16	24	40	> nr. of bits	
LTW		Piecewise Rate		Splice Type					
LTW Valid Flag	LTW Offset	Reserved	Piecewise Rate	Splice Type	DTS Next AU [32..30]	Marker Bit	DTS Next AU [29..15]	Marker Bit	DTS Next AU [14..0]
1	15	2	22	4	3	1	15	1	15

Figure 2.6: Transport Stream Fields ***(BOGDAN, 2004)***

The most important fields that are going to be used for this work are the PTS (Presentation Time Stamp) and DTS (Decoding Time Stamp) both represent the time which an access unit should be instantaneously removed from buffer and decoded or presented. DTS differs from PTS only in some special B-frames ***(ATSC, 2003)***. According to ATSC both times are entered in the bit stream at intervals not exceeding 700ms.

2.3 DIGITAL VIDEO ARTIFACTS

A video artifact is an undesirable feature in a video, can be achieved in an original video or in a distorted video. Video artifacts can be introduced during capture, transmission, storage and display, can be produced by any lossy image/video processing algorithm that is applied along the way ***(FARIAS, 2010)***.

Normally artifacts in their physical understanding are very complex and can be very difficult to describe. Most of them have more than one perceptual visual inconsistency; however it is possible to find artifacts that are relatively pure in effect. Below there is a list with the most common artifacts ***(FARIAS, 2010)***:

Blurring – it is a loss of spatial detail and reduction of edge sharpness. During the compression the blur is caused by the suppression of high-frequency coefficients in quantization process.

Blocking - a type of artifact characterized by a block pattern visible in the video. It is result of the independent quantization of individual blocks in DCT coding schemes, these leads to discontinuities at the boundaries of near blocks. The blocking effect is the most annoying artifact found in a compressed video, depending on the periodicity and extent. Some recent video codecs use a de-blocking filter to reduce this artifact.

Staircase Effect – appear of slanted lines due to the loss of higher frequencies of DCT coefficients that are not suitable for lines oriented differently from horizontal and vertical.

Bleeding – knows as a smearing of colors in areas with different chrominance in contrast. It is the result of high-frequency coefficients of Chroma components.

Jitter – result of skipping or delaying regularly video frames to reduce the amount of video data that is being used by the system, the effect is a slow motion of the video, instead of being smooth and continuous.

Packet Loss – occurs during transmission where several frames of the video are lost, this introduce a break in the video that reduces the perceived quality.

Ringing – it is perceived in high contrast edges and is a result of quantization irregularities during the reconstruction of the video, can be seen in luminance and chrominance.

Flickering – in high texture scenes, with varying quantization factors over time during compression of texture blocks that result is the flickering effect.

Mosquito Noise – temporal artifact during encoding that is seen in smooth textured regions as luminance or chrominance fluctuations around high contrast edges or moving objects, it is a result of coding differences for the same area of a scene in consecutive frames of a video.

Below the pictures shows some of the artifacts presented:

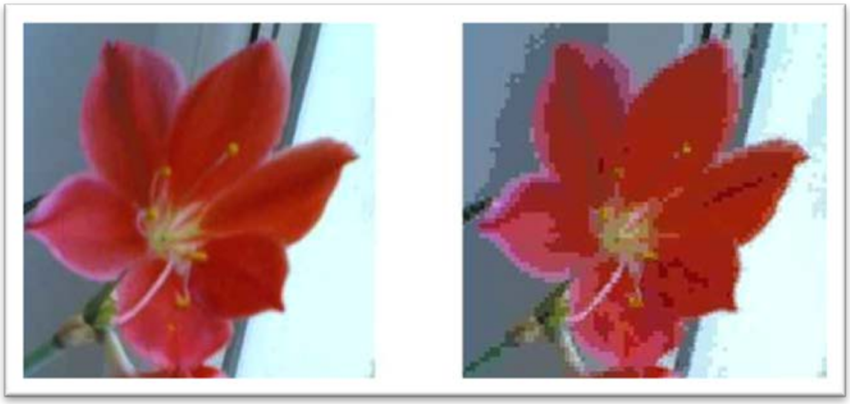


Figure 2.7: Blocking effect ***(GONZALES AND WOODS, 2002)***

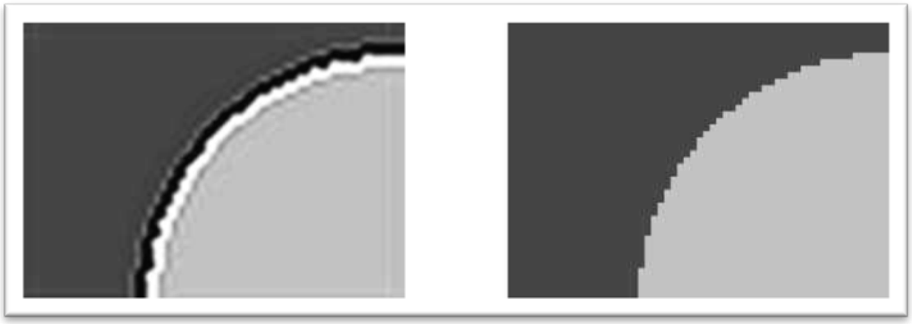


Figure 2.8: Ringing effect ***(GONZALES AND WOODS, 2002)***

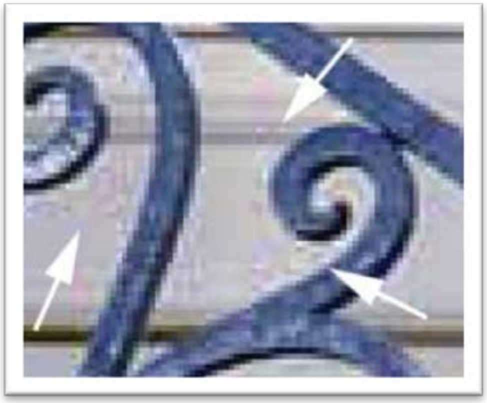


Figure 2.9: Mosquito noise ***(GONZALES AND WOODS, 2002)***

3. VIDEO AND IMAGE QUALITY METRICS

One important subject of video quality metrics is “how” the metrics works.

Video and image quality metrics are currently classified in different manners, first comes the classification of the metrics in subjective and objective. The main goal is to work with objective video quality metrics and this will be detailed in the second section.

3.1 SUBJECTIVE VIDEO QUALITY ASSESSMENT

Subjective video quality assessment ***(ITU-T, 2008)*** is not an automated video metric, it is a collection of psychophysical experiments that try to represent and measure the quality of a video. With a defined number of people used to watch a video and give a score to it. Gathered with this information the average is calculated and there is a number called Mean Observer Score (MOS) or Mean Opinion Score.

Normally subjective evaluations are expensive and time-demanding; all the experiments should be designed and executed by hand, later the data analysis consumes more time to be classified and evaluated. It is dependent to the number of the observers, equipment, calibration, physical space, etc. As these general problems for creating a subject analysis occur and can’t be prevented, it is needed some methodology to get the most out from the resources available, and these methodologies are recommendations from the ITU (International Telecommunication Union).

The most popular and important assessment procedures are listed ***(ITU-T, 2008):***

- Double Stimulus Continuous Quality Scale (DSCQS);
- Double Stimulus Impairment Scale (DSIS);
- Single Stimulus Continuous Quality Evaluation (SSCQE);
- Absolute Category Rating (ACR) or Single Stimulus Method (SSM);
- Degradation Category Rating (DCR);
- Pair Comparison (PC);

These subjective analysis methods are considered the most precise way to evaluate the quality of a video, because gives a real perception of the human vision, and human vision is the most accurate evaluator of video quality. It is the biggest advantage of subjective video analysis; the quality is always conforming to the human vision opinion.

It is impossible however to use subjective video analysis for every test in daily video quality judgment; which is why objective video quality metrics have been developed and used. They were meant to evaluate the quality of a video without the need of a human subject test, but always considering the human visual stimulus, or trying to do that. All the objective metrics to be considered good in quality are compared to subjective results and later used in video target applications.

An example that shows how the subjective video quality and objective video quality are related, with the comparison of the average opinion score and the results from an objective video quality metric PSNR (peak-to-signal noise ratio), which is explained in next subchapters, can be seen below:

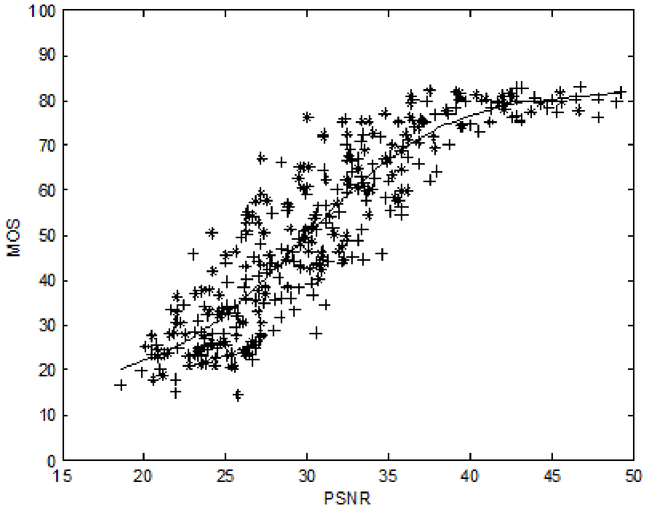


Figure 3.1: MOS vs PSNR ***(WANG Z., 2004)***

Objective video quality metrics try to automatically estimate average viewer opinion on a video processed by the system as subjective video assessment; however they have some problems to correlate with the human vision parameters and non-linear behavior.

VQEG (Video Quality Experts Group) created a subjective analysis database to be used for anyone who wants to submit a new video quality assessment method; it is free for non-commercial use and available online for comparisons, every year they update the database. Submissions and reports containing the news about objective video quality metrics are published ***(VQEG, 2003)***.

3.2 OBJECTIVE VIDEO QUALITY ASSESSMENT

Objective video quality assessment is intended to give a quality number from a video that is suitable and comparable to the human vision in most of the cases. Instead of being expensive and time-demanding however, objective video quality metrics are an automated way of calculating and extracting the quality of a video. They're used in different situations ***(FARIAS, 2010)***:

- Monitor video quality;
- Compare performance of different video systems;
- Optimize algorithms and parameter settings for video processing system.

Objective video quality metrics are normally classified in three categories considering the availability of the original video signal ***(FARIAS, 2010)***:

- **Full Reference (FR)** – original and distorted video available;
- **Reduced Reference (RR)** – original video unavailable (only description), distorted video and some parameters available;
- **No-Reference (NR)** – only distorted video available.

Below is a description of each category in block diagrams:

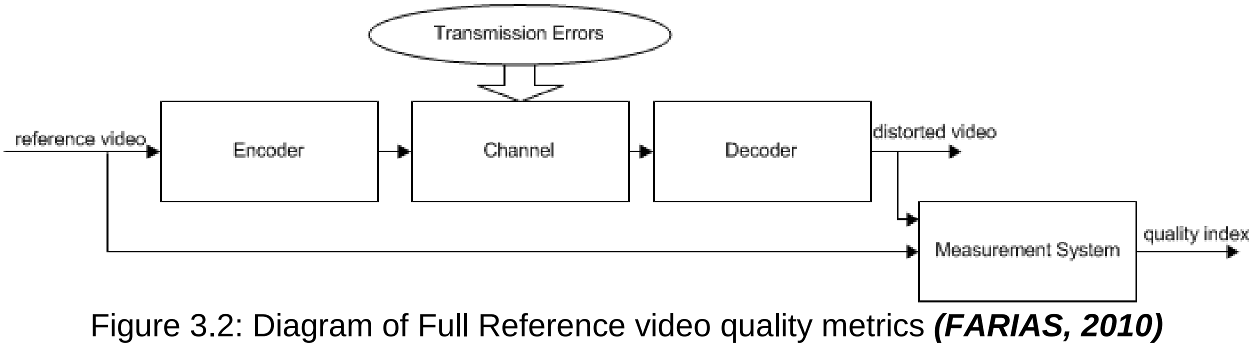


Figure 3.2: Diagram of Full Reference video quality metrics ***(FARIAS, 2010)***

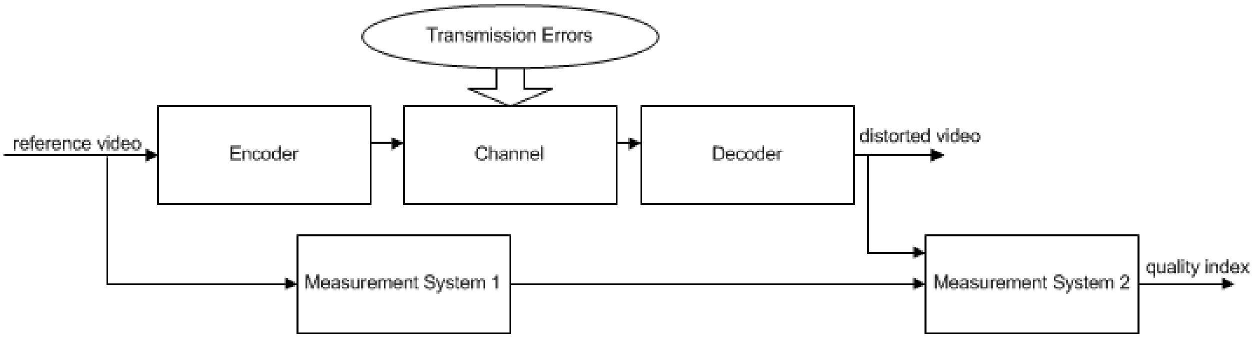


Figure 3.3: Diagram of Reduced Reference video quality metrics ***(FARIAS, 2010)***

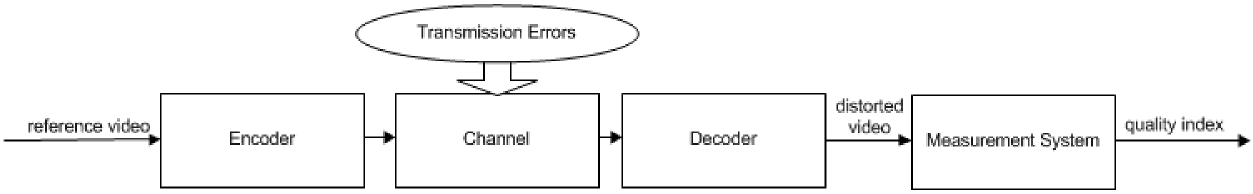


Figure 3.4: Diagram of No-Reference video quality metrics ***(FARIAS, 2010)***

Usually these classes of objective video quality metrics are used in different situations. FR metrics are more used in offline solutions, where there is a detailed and accurate measure of the video quality. RR and NR metrics are more targeted where the computational requirements lack of an immediate source reference.

Objective video quality metrics can still be classified according to the way they approach for estimating the video impairments ***(FARIAS, 2010)***.

- **Error Sensitivity (ES)** – analyze visible differences between test and reference videos, used in FR metrics, capable of doing pixel-by-pixel difference;
- **Feature Extraction (FE)** – uses higher-level features that doesn't exist in the reference video, used in RR and NR metrics, uses information previously known from reference video.

Also, objective video quality metrics can be classified also by the type of information they take in account when processing the video ***(FARIAS, 2010)***:

- **Picture Metrics or Perceptual Metrics (PM)** – uses the Human Visual System;
- **Data Metrics (DM)** – measure only the signal fidelity without considering the content or correlation to the human visual system.

3.2.1 Mean Square Error and Peak signal-to-noise Ratio

These two metrics measure the physical differences between two signals regarding the content, so they're not exclusive from video quality metrics; they're also used in signal analysis for electronics and other engineering areas. These are the most used fidelity metrics and are defined as:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(X_i - Y_i)^2$$

$$PSNR = 10\ .\log_{10}\frac{L^2}{MSE}$$

Where N is the total number of pixels in the video, L is the dynamic range of allowable image pixel intensity. For example, images with 8 bits per pixel of gray-scale, $L = 2^8 - 1 = 255$. And X_i and Y_i are the i-th pixels in the original and distorted video, respectively.

PSNR is the most widely used objective video quality metric and is useful if images having different dynamic ranges are being compared, but otherwise contains no new information relative to the MSE. These two metrics are very popular in the image processing community because of their physical significance and simplicity but can only predict subjective rating with reasonable accuracy if the comparisons are made for the same content, same technique or same type of artifact.

That is why over the years PSNR and MSE are criticized ***(WINKLER, 1999)***; because there values do not perfectly correlate with a perceived visual quality due to the non-linear behavior of the human visual system. These simple metrics don't consider the relationships among pixels in an image or frames, and also don't consider how spatial and frequency content of the impairments are perceived by human viewers. And for these reasons other metrics have been submitted over the past years, always trying to reach the best human visual experience index with higher correlations to the subjective quality assessment.

3.2.2 Structural Similarity Index Metric (SSIM)

The Structural Similarity Index Metric ***(WANG Z. AND BOVIK, 2004)*** is based on the idea that natural images are highly structured, and these images have strong relationships among themselves, which carry information about the structures of the objects in the scene. The main idea behind this is that human visual system is highly specialized in extracting structural information from the viewing field and it is not specialized in extracting the errors. So, the measurement on structural distortion should give a better correlation to the subjective impression.

The similarity between a reference image and the distorted one is estimated. SSIM algorithm measures the luminance $l(x,y)$, contrast $c(x,y)$, and structure $s(x,y)$, of the distorted image y and the corresponding reference image x, using the following equations:

$$l(x,y) = \frac{2\mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1'}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2'}$$

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3'}$$

Where C_1, C_2 , C_3 are small constants given by $C_1 = (K_1.L)^2$, $C_2 = (K_2.L)^2$, and $C_3 = C_2/2$. L is the dynamic range of the pixels (for example, images with 8 bits per pixel of gray-scale, $L = 2^8 - 1 = 255$), $K_1 \ll 1$, and $K_2 \ll 1$.

The general formula of the SSIM is given by:

$$SSIM(x,y) = [l(x,y)]^{\alpha}.[c(x,y)]^{\beta}.[s(x,y)]^{\gamma}$$

Where α, β and γ are parameters that define the relative importance of the luminance, contrast, and structure components. However, if α=β=γ=1, the above equation is reduced to:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

The SSIM has a range of values varying from “0” and “1”, with “0” being the worst possible value and “1” meaning the perfect score (exactly the same quality). A block diagram of the SSIM algorithm is depicted above:

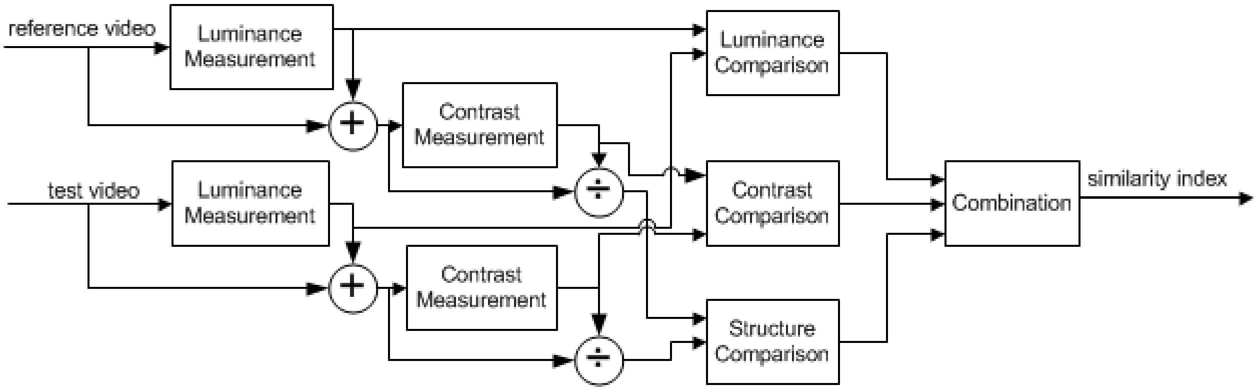


Figure 3.5: SSIM Diagram **(FARIAS, 2010)**

The diagram above is a representation of the formulas from SSIM, and is adding the luminance measurement to contrast measurement, dividing the result of both measurements for the structure comparison. The other comparisons are directly calculated from both videos. The final result is the combination of all comparisons.

3.3 OBJECTIVE VIDEO QUALITY METRIC SELECTION

With the previous knowledge of quality metrics above it is possible to continue with the selection of the metrics and show the reasons to choose them. It is possible to check all the details of the metrics online, because none of the metrics selected in this thesis were developed in this thesis; they are only used in the timeline approach that takes the frame and presentation time variation in the calculation process of video assessment. The metrics and its original implementation and meaning were untouched.

First, only FR metrics, because they can give the best results possible in terms of visual quality and because there is the access to the original and modified video available for this work were selected. In this work only ES metrics both PM and DM are used.

The table below was made in this work, research was done on each respective metric website, shows some of the criteria to select metrics for this monography:

	Physical Meaning?	Suitable Human Vision?	Code Available?	Suitable for Real-Time Application?	Standard?	Frame-by-Frame Calculation?	Complexity
MSE/PSNR	Yes	No	Yes	Yes	Yes	Yes	Low
SSIM	Yes	Yes	Yes	Yes	No	Yes	Medium
VQM	Yes	Yes	Yes	No	Yes	Yes	Very High
MOVIE	Yes	Yes	Yes	No	No	Yes	Very High
PEVQ	No	Yes	No	No	Yes	No	Very High
CZD	Yes	No	No	Yes	No	Yes	Low

Table 3: Video Metric Selection

The selection on MSE/PSNR and SSIM was made. The first one was selected because it is current the most used metric for video quality assessment **(VQEG, 2003)**, is a proven standard and is suitable for test purposes. The second was selected because it is a really promising metric with good visual correlation according to VQEG, is easy to compute for the needs of this work and is been adopted by most of the video encoders and decoders available.

The other metrics: MOVIE, PEVQ were discarded because they have problems with high complexity and are not yet evaluated in latest VQEG report. VQM is considered by VQEG report as one of the best metrics available, however this metric was created and implemented by NTIA and is patented by the US government. CZD has poor correlation with visual quality and is based in PSNR. Therefore there is no need of using it, as it is similar to other already used metric.

4. APPROACHES FOR VIDEO QUALITY METRICS

In this chapter the classical approach for computing the quality of a video using the metrics presented before is presented. The introduction of the new hypothetical timeline approach for computing the video quality using the same metrics is presented too.

Both approaches use the video quality metrics already cited in this work. The way these metrics are used, however, can vary in a video evaluation. The classical approach is an average of frame quality for the video, without considering the display time of the frame. The new timeline approach focus in using this frame time to calculate the differences from different frames, as a result it is introduced a new concept to the video quality evaluation, that is “quality with variable display time”.

The proposed timeline approach is not a new video quality metric. It is only a new way of computing the quality of a video using the already developed metrics that have proven relation to video quality assessment or have been used in the field of video quality.

4.1 CLASSICAL APPROACH FOR COMPUTING VIDEO QUALITY

The classic way of doing any quality measurement is computing frame by frame the quality of the images and then using some method (average, weight pooling, etc.) to obtain a single number (index) that will give the quality of the video sequence.

This strategy is quite reasonable because a video is a sequence of frames then it is expected that the average quality of the video is the average of the frames quality. This strategy however does not take in account the time the frames are presented.

Classical approach for each metric selected in the last chapter is presented in the next charts. The “test scenario” has only 8 frames. The analysis occurs with two videos without any kind of frame time variation (delay); identical videos (compressed) having exactly the same video quality score for each metric. It is like comparing the video with itself and not having any difference between them.

Traditional Approach – Frame-by-Frame Evaluation

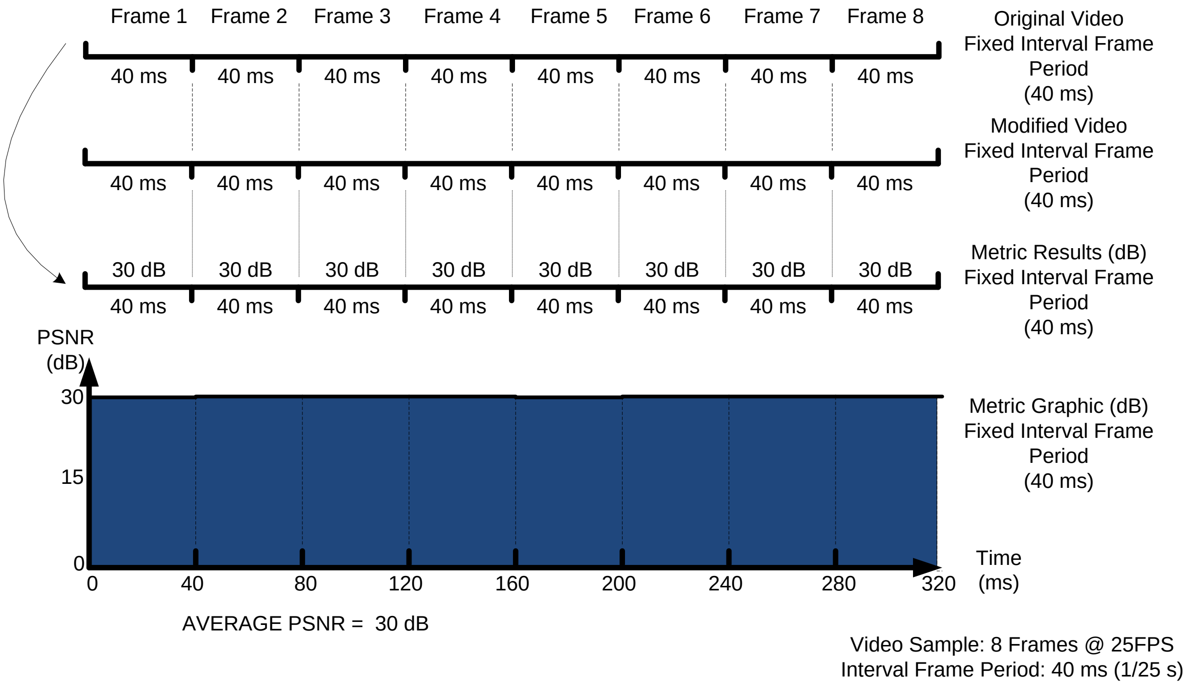


Figure 4.1: PSNR Frame-by-Frame Fixed Time

Following, only the evaluation metric is changed (from PSNR to SSIM) and there is the perfect score for each one:

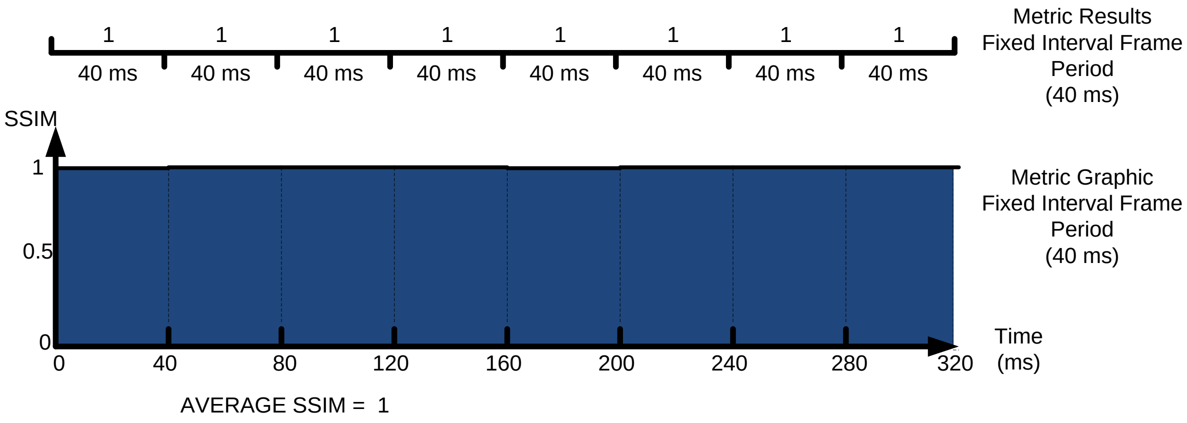


Figure 4.2: SSIM Frame-by-Frame Fixed Time

The results are pretty correct, there are two identical videos and it is obviously that the video quality will be exactly the same.

However if two videos differing one from another only from the time that the frames are displayed are tested; this is how this classical frame by frame approach works without taking display time in consideration. The results are described below:

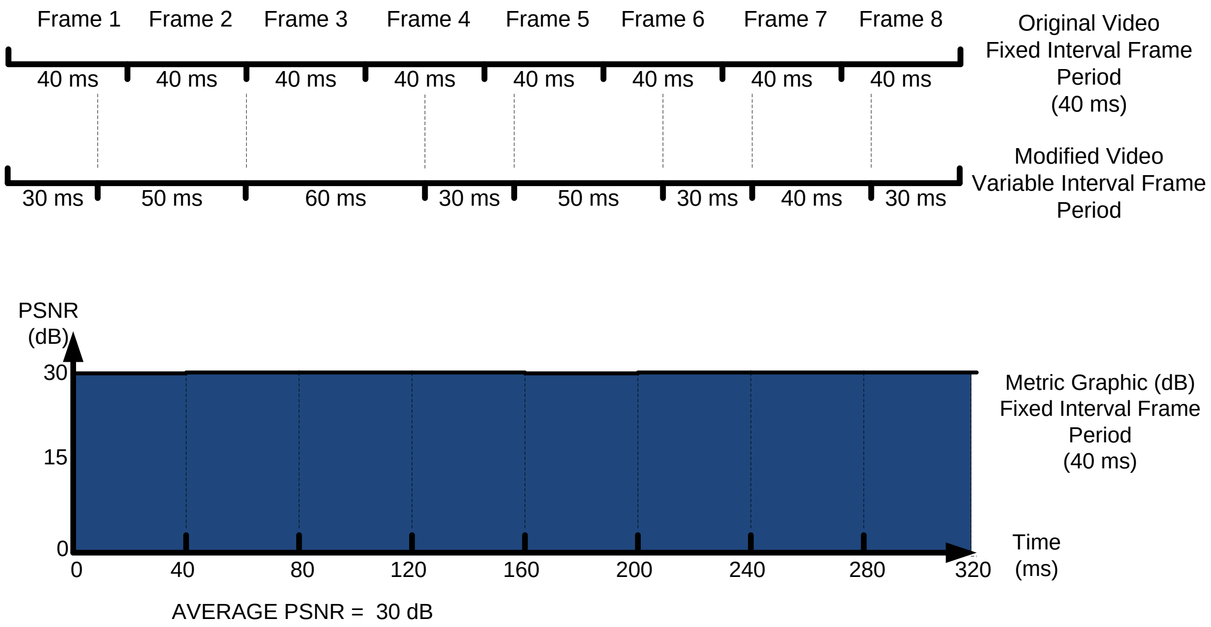


Figure 4.3: PSNR Frame-by-Frame Variable Time

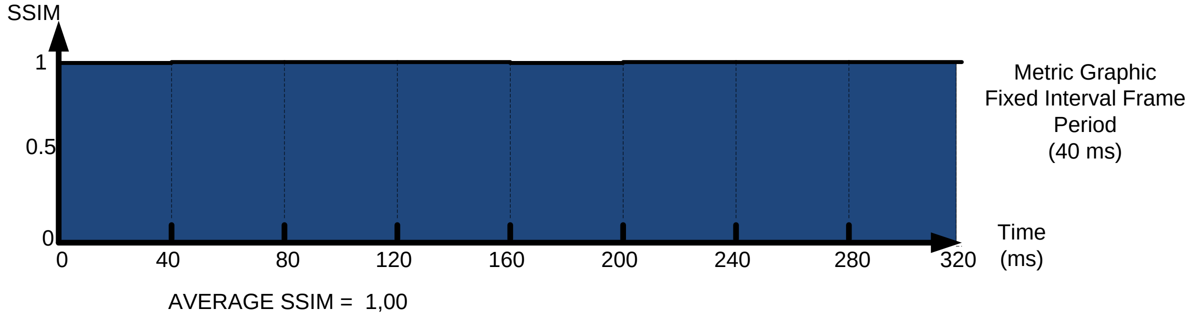


Figure 4.4: SSIM Frame-by-Frame Variable Time

There is a problem identified; the video quality score is still considered perfect by the video quality metrics. This is certainly not right, because watching the video and judging by the eye (subjectively) shows clearly that the quality of the videos is different **(GUERRA, 2011)**.

This is not faulting of the video quality metric itself; it is a problem in how the video quality metrics are calculated. The approach used in today video quality metrics calculation is not suitable for variable time delay in videos. The new timeline approach proposed next tries to solve this problem, giving a more suitable video quality measure of a video using the same metrics already used; considering time variation now.

4.2 TIMELINE APPROACH FOR COMPUTING VIDEO QUALITY

Previous research **(HUYNH-THU Q. and GHANBARI M., 2006)**, regarding the display time delay and subjective analysis, has demonstrated that the video quality has direct correlation with the time that a frame is presented in a video. That is, the quality of the video is not only related to the images themselves, but also to when they are presented in the display.

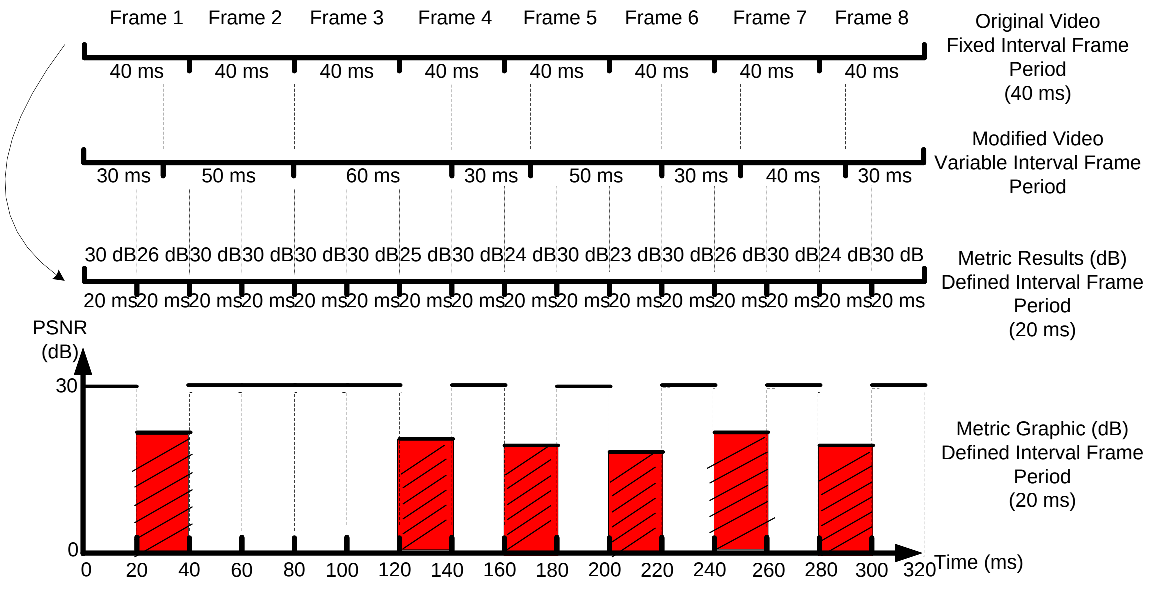
The main purpose of this work is to develop a new approach for calculating the video quality that takes in account not only a frame by frame result but a timeline search that will give a more precise result in average, because in the new approach it will take the presentation time stamp in consideration. A new quality metric is not being proposed. The existing metrics are used to calculate the quality considering the display time.

4.2.1 FIXED TIMELINE EVALUATION APPROACH

For the first attempt to create a new timeline approach, a defined fixed interval period for calculating the quality of the video is used. As this interval time is fixed and also small, there is the need to calculate quality more times than the expected frame delay. This timeline with fixed calculation will be discarded later. This approach demands redundant calculation, because it is calculating many frames that do not need to be compared more than one time, this can be seen below:

New Approach – Timeline Evaluation

Fixed Time Evaluation

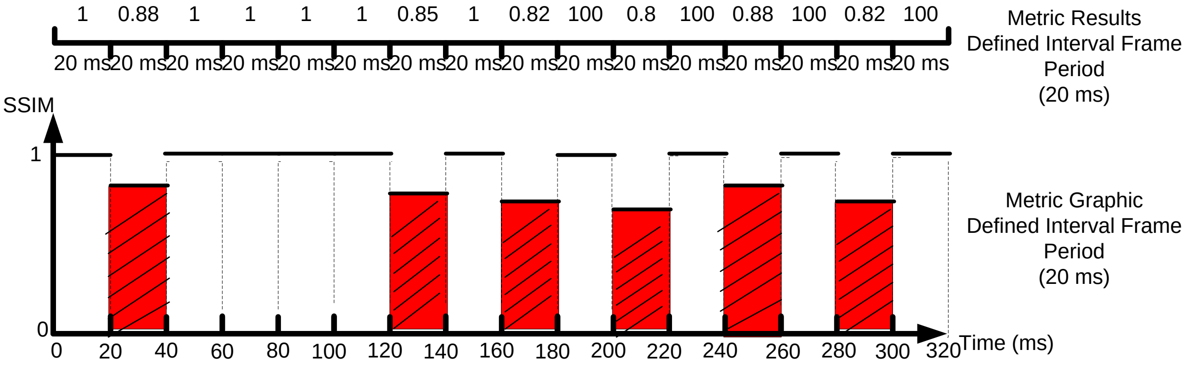


AVERAGE PSNR ERROR FROM FRAME TIME VARIATION = 24.66 dB

Video Sample: 8 Frames @ 25FPS
Interval Frame Period: 40 ms (1/25 s)

Figure 4.5: PSNR Timeline Fixed

In this fixed time evaluation the average of the differing frames was computed as the average distortion error from time variation. This average calculation adopts SSIM metric too:



AVERAGE SSIM ERROR FROM FRAME TIME VARIATION = 0.841

Figure 4.6: SSIM Timeline Fixed

Considering the display time of a video, using the same metrics that were used in the classical approach; this method is better than the classical one. The new approach recognizes the variable frame delay from a certain video, giving a quality index to the error. However, this approach has some problems with unneeded computation.

4.2.2 VARIABLE TIMELINE EVALUATION APPROACH

The proposed timeline approach was refined and then a variable timeline solution version was reached. This approach is more suitable for considering variable delays. This can be seen below:

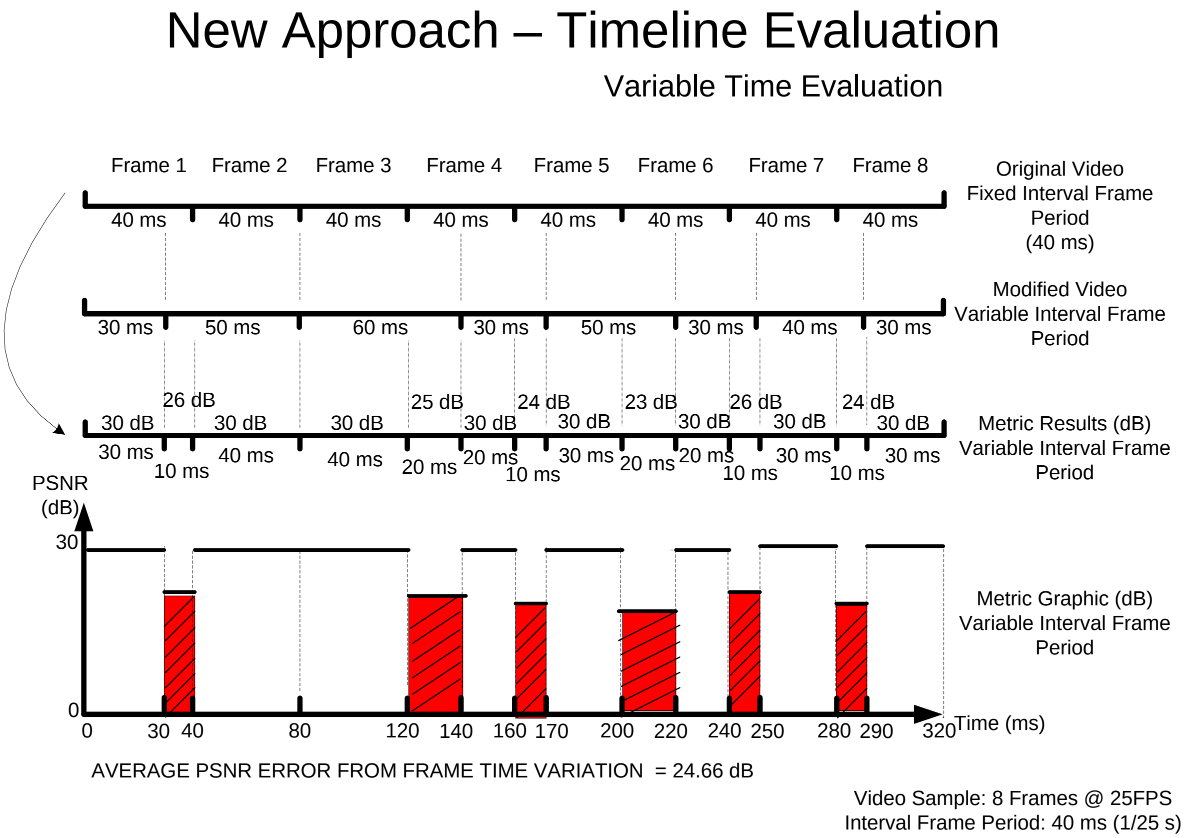


Figure 4.7: PSNR Timeline Variable

In this variable time evaluation the average of the differing frames was computed as the average distortion error from time variation. The average calculation was made for SSIM metric too, as can be seen below:

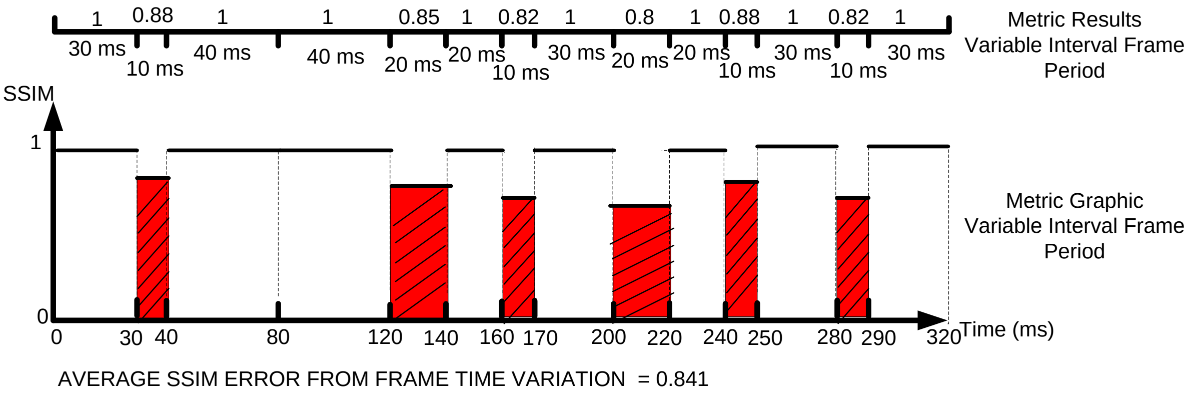


Figure 4.8: SSIM Timeline Variable

The approach with variable time evaluation is capable of recognizing the error caused by variable display time too. The results showed above demonstrate that this approach has the same quality index error from time variation when compared to the fixed time evaluation. The computation however is much less intense, because the quality calculation is only done when needed, according only to the variable time, not every fixed time. This variable timeline solution is more suitable for applications that demand less computation work and thus being the subject for development of the algorithm.

During the evaluation of the average error from frame time variation it is needed to confirm that the quality of a modified video is worst only regarding the time display factor. The quality of the modified video however is not only the average error from time variation, but the average error from the video sequence considering the new variable delay.

The quality of the test video below is described for each metric:

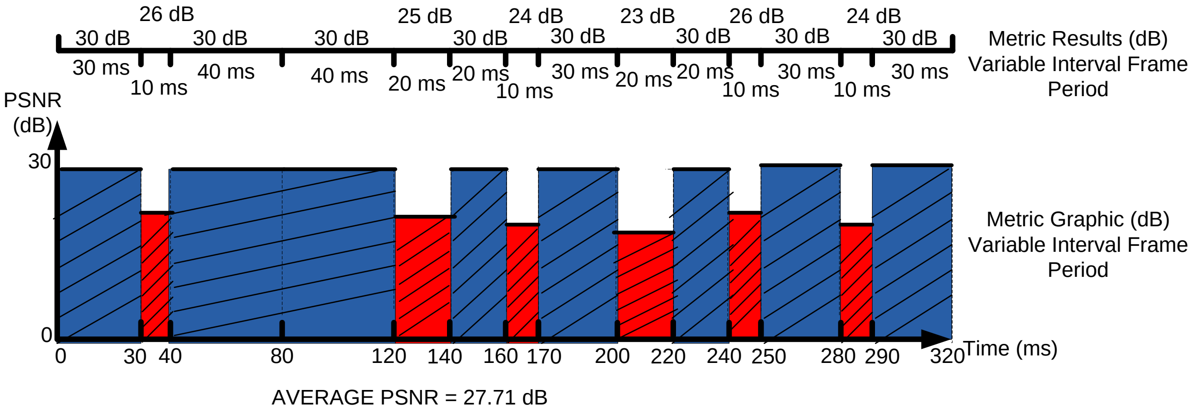


Figure 4.9: Average PSNR Timeline Variable

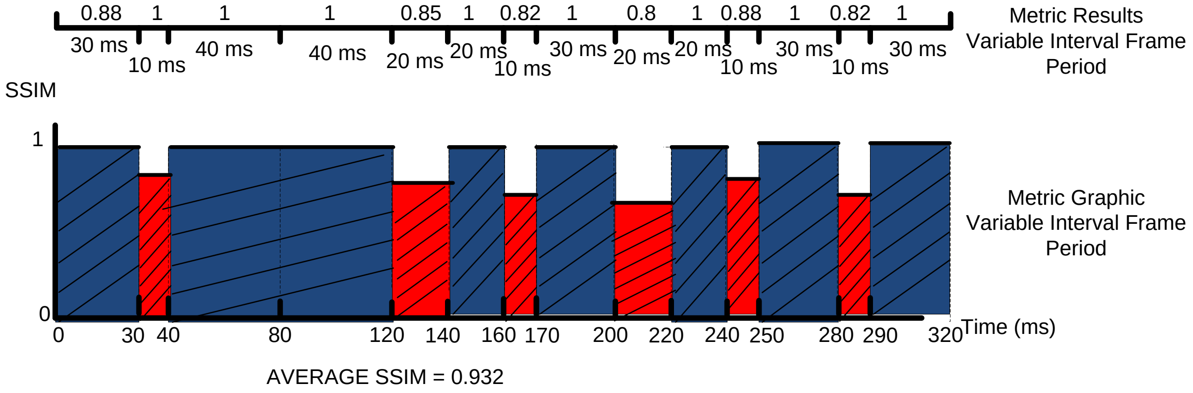


Figure 4.10: Average SSIM Timeline Variable

The variable timeline approach is really considering the time variation from the frames, because there is not the perfect score when comparing two identical videos that differs only the display time, there is a quality index that tries to match properly the quality error introduced by the time variation.

In this chapter only a hypothetical timeline approach is brought; the results presented during the timeline are not precise or concise to a real implementation of the timeline approach.

Below a table comparing the classical approach with the perfect score and the variable timeline approach considering the time display frame as a new variable to the problem is presented:

	Classical Approach	Variable Time Approach
PSNR(dB)	30*	27.71
SSIM	1	0.932

Table 4: Approaches Comparison

*In this hypothetical analysis the superior limit from PSNR was considered 30dB, however this value can be higher, depends on the previous established maximum quality for PSNR. In this case it was assumed 30dB as maximum without any further consideration, because the hypothetical analysis is not a real implementation.

5 ALGORITHM AND OFFLINE SOLUTION

This chapter presents an algorithm for video quality assessment that was developed using the quality metrics already available and the new timeline approach, specifically the variable timeline approach explained before. First, an algorithm is modeled without considering its implementation aspects, only the logic needed to solve the problem.

In the second part all the aspects of the offline solution will be explained, including the hardware used, platform, programming language, video samples standard, problems, etc. The test methodology and parameters that are used in the next evaluation chapter are introduced in the third part of this chapter.

5.1 STRUCTURED ALGORITHM

The structured algorithm shown in figure 5.1 does not consider aspects of implementation (video standards, compression rate, data structures, etc.). The algorithm in the figure considers the comparison of two videos, with every frame from the video carrying the time of beginning and ending. With that information it is possible to present a solution for the problem of computing the quality differences between and original video with fixed presentation time stamp, and the one which has variable presentation time stamp.

Below, in the next page, a flowchart representing the developed structured algorithm is presented.

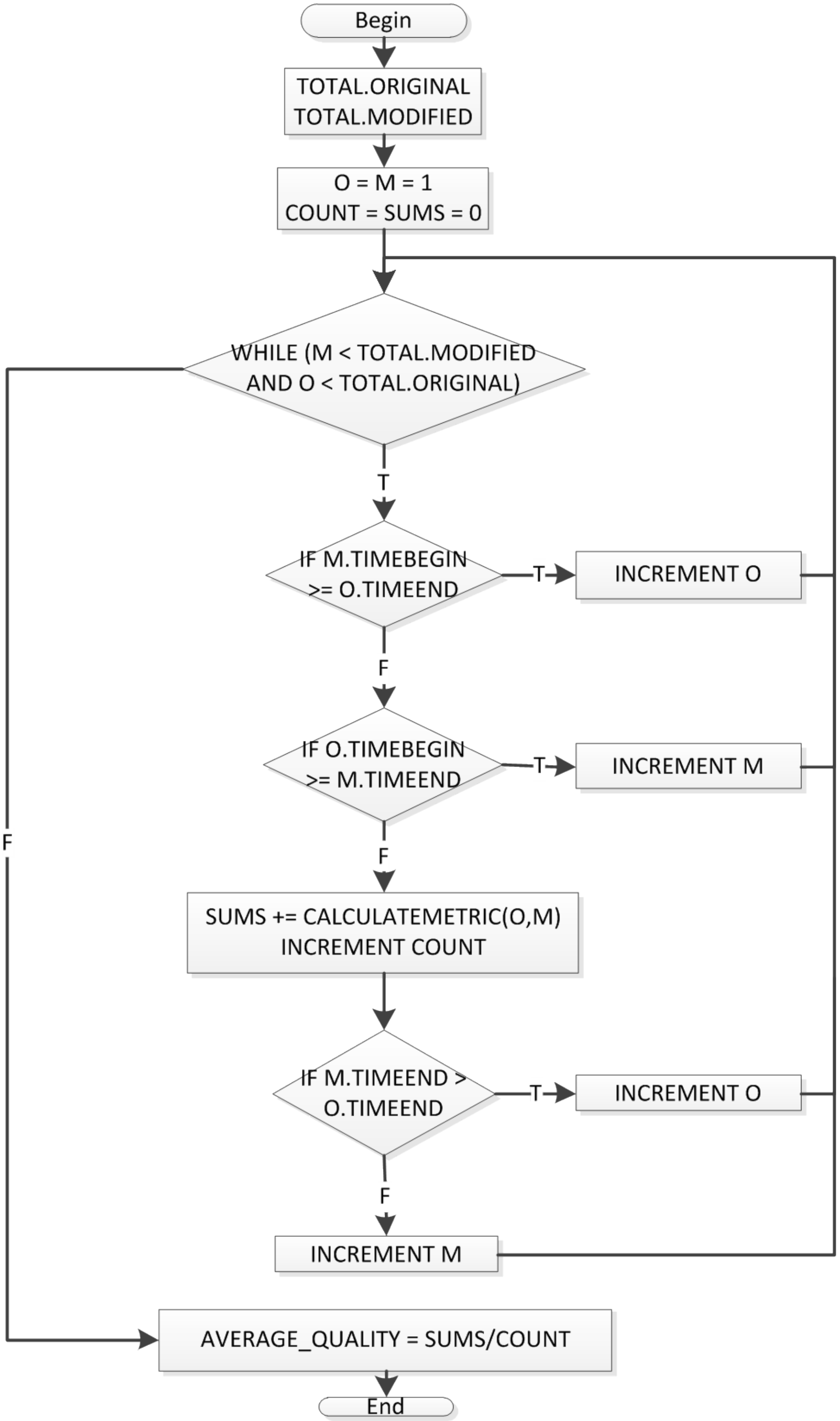


Figure 5.1 Algorithm for video quality assessment considering frame and time variation

5.2 OFFLINE SOLUTION

The computer system below was used to implement the objective video quality metrics with the proposed timeline approach:

Processor: Intel Quad Core Q9550 3.4GHz

Motherboard: Asus ROG X38 LGA 755

Memory: 2 x 4GB DDR2 1066MHz Mushkin

Hard Disk: 2 x 1 TB Seagate 7200 64mb (RAID 0)

Graphics Card: Nvidia GTX480 1536MB

Operational System: Windows 7 Professional

Figure 5.2 Computer system configuration used

The programming platform and language selected was MATLAB 2012a for the reasons explained below.

The choice of MATLAB as a programming language and platform was based on the simplicity and capability to process video and images as direct objects; everything can be directly processed with the built-in function of MATLAB thus reducing the amount of necessary programming effort to work with this kind of data.

MATLAB 2012a has already implemented the Video Render class that is capable of reading all the proposed video for this work. It was not necessary to create a parser for MPEG-2 Transport Stream. MATLAB also has a lot of predefined manipulation functions and filters that turn quite simple to handle the metrics that are used.

The metrics PSNR and SSIM have been both widely implemented in MATLAB, and diffused over the internet, and their codes are available for download and use.

Below is the code from MATLAB for PSNR (a direct implementation from the metric presented in chapter 3, developed for this work) and SSIM index with the reference from where it was taken, directly from the creator’s website **(WANG Z., 2003)**, the code was modified for the purposes of this work but the core solution was left intact. The main function of the proposed work is a representation from the structured algorithm showed in the last chapter and is not explained again here.

```
PSNR (original,modified)
R=original-modified
// MSE
MSE=sum(sum(R.^2))/(size(original,1)*size(original,2)); % MSE
// PSNR
if MSE>0
        PSNR=10*log10(255^2/MSE);  else
        PSNR=Inf;
```

Figure 5.3 MSE/PSNR Implementation **(WANG Z., 2003)**

```
SSIM (original, modified)
[M N] = size(original);
window = fspecial('gaussian', 11, 1.5);
        K(1) = 0.01;            // default settings
        K(2) = 0.03;            //
        L = 255;                //
C1 = (K(1)*L)^2;
C2 = (K(2)*L)^2;
window = window/sum(sum(window));
original = double(original);
modified = double(modified);
ssim_map = filter2(window, original, 'valid');       // gx
w1 = filter2(window, modified, 'valid');             // gy
w2 = ssim_map.*w1;                                   // gx*gy
```

```
w2 = 2*w2+C1;                                   // 2*(gx*gy)+C1 = num1
w1 = (w1-ssim_map).^2+w2;                       //(gy-gx)^2+num1 = den1
ssim_map = filter2(window, original.*modified, 'valid');         // g(x*y)
ssim_map = (2*ssim_map+(C1+C2))-w2;         //2*g(x*y)+(C1+C2)-num1 = num2
ssim_map = ssim_map.*w2;                           // num
original = original.^2;                          // x^2
modified = modified.^2;                            // y^2
original = original+modified;                          //x^2+y^2
if (C1 > 0 && C2 > 0)
        w2 = filter2(window, original, 'valid');          // g(x^2+y^2)
        w2 = w2-w1+(C1+C2);                             // den2
        w2 = w2.*w1;                                // den
        ssim_map = ssim_map./w2;                      // num/den = ssim
else
        w3 = filter2(window, original, 'valid');          // g(x^2+y^2)
        w3 = w3-w1+(C1+C2);                             //den2
        w4 = ones(size(w1));
        index = (w1.*w3 > 0);
        w4(index) = (ssim_map(index))./(w1(index).*w3(index));
        index = (w1 ~= 0) & (w3 == 0);
        w4(index) = w2(index)./w1(index);
        ssim_map = w4;
        end
        SSIM = mean2(ssim_map);
```

Figure 5.4 SSIM Implementation **(WANG Z., 2003)**

The main problem of the implementation part was to access the presentation time stamp from a Transport Stream video (the PTS contains the exactly time the frame will be displayed), because this information is not accessible by any function of MATLAB. A simple solution was to create the timeline in an array for the video, based on the variable delay; this simplifies the content access because there is already an imposition of the original video PTS (40ms) every frame.

This means that an array with the original video PTS like (0, 40, 80, 120, 160, 200, 240 …) was constructed for each frame of the video, meaning that each position corresponds to the frame and the current PTS.

For the alternative video presentation, the modified PTS based on the original PTS was calculated, so that only the difference in time for each modified frame was computed according to the delay text file. A new array for each modified video PTS like (0, 30, 60, 80, 150, 200 …) was created and this array is the sum from the original (0, 40, 80, 120, 160, 200, 240 …) and the delay.

Other important detail is the superior limit from PSNR in the implementation, according to **(VQEG, 2003)** a PSNR value around 35dB is considered good in video quality evaluation (human eye cannot detect important differences beyond this point), that is why in this work was set 35dB as the top quality for the video comparison using PSNR, this limit during tests was never reached corroborating to VQEG.

5.3 TEST METHODOLOGY

To test the quality of modified videos with variable delays it was used some sample videos available at the Consumer Digital Video Library (CDVL), all the videos have the European Standard video resolution **(DVB, 2008)** used current by the German Digital Television (720x576 @25FPS).

All the test videos were compressed with the German Digital Television Standard for Digital Video Broadcasting over Aerial Transmission (public digital television), this means that all the videos were compressed with MPEG-2 and encapsulated in TS (transport streams). This encoding part was not done in this work; all the encoded samples were retrieved from the Real-Time Systems Lab during the time this work was made in Germany. After that, for each video were imposed variable delays according to Raphael Guerra’s work **(GUERRA, 2011)**.

For the test analysis Guerra made the delay adjustments, resulting in three kinds of delay. The author provided three text files containing the number of delays (quantity) and delay times (microseconds) that must be applied to the modified video presentation time stamps, resulting in three modified versions of each video.

According to his work, Guerra generated three different kinds of delay aiming for each one a different delay quality.

The average delay for each kind of delay was calculated using this formula:

$$Average\ Delay = \frac{1}{N}\sum_{i=1}^{N}|PTS_{oi} - PTS_{mi}|$$

Where “N” is the total number of frames that the delay is being imposed.

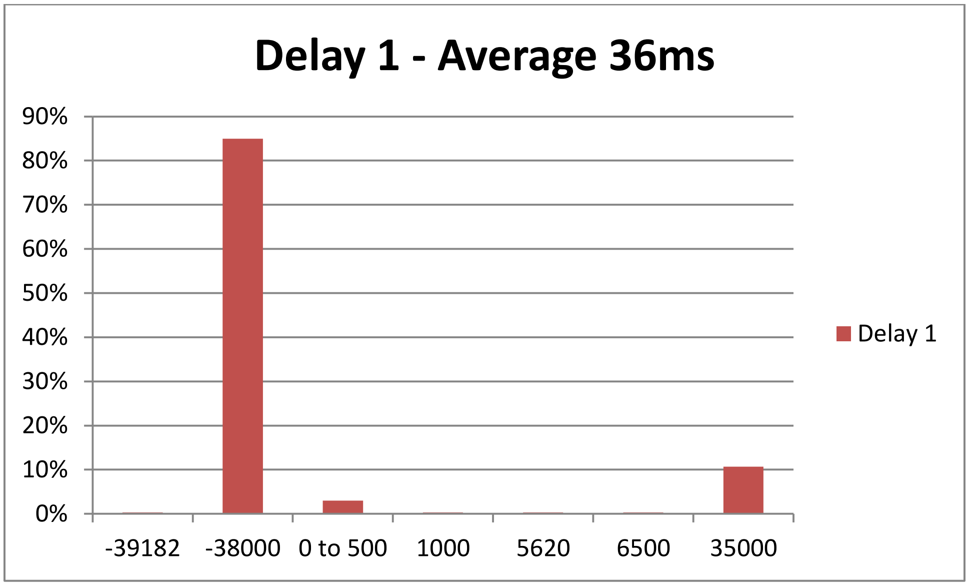


Figure 5.5 Delay 1 Distribution. Average delay imposed is 36ms.

Delay times in microseconds (µs).

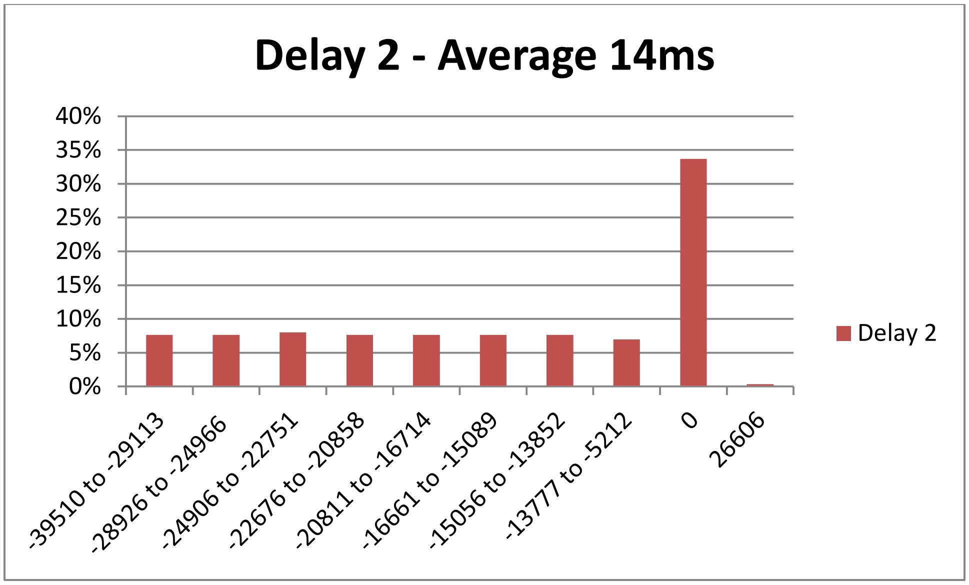


Figure 5.6 Delay 2 Distribution. Average delay imposed is 14ms.

Delay times in microseconds (µs).

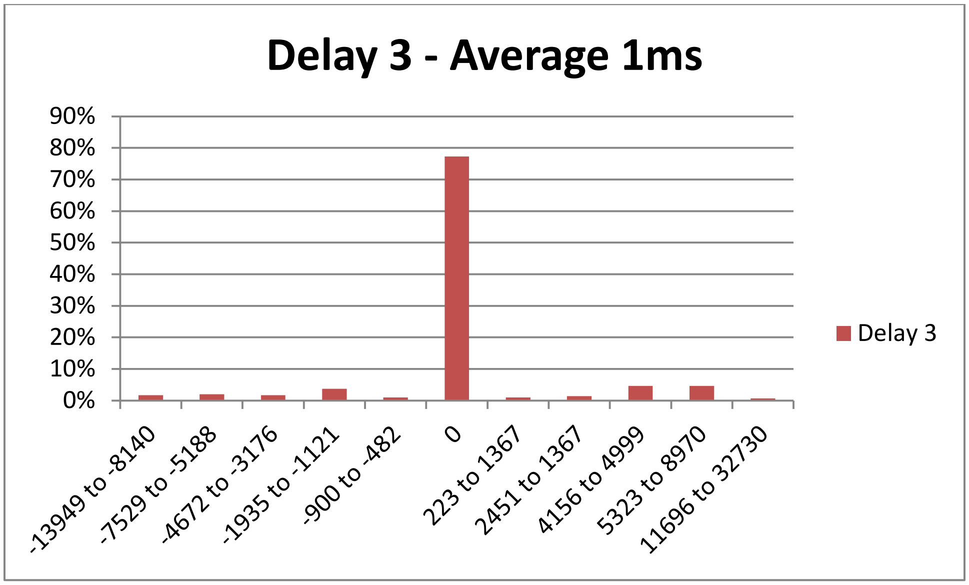


Figure 5.7 Delay 3 Distribution. Average delay imposed is 1ms.

Delay times in microseconds (µs).

According to **(GUERRA, 2011)** the Delay 1 group of 5 videos should represent the worst video quality judged by subjective analysis made in his work. The Delay 2 group should represent an intermediate quality and Delay 3 group should have the best quality, all when compared to the video with no delay (perfect video).

The next chapter will show the results for each video and the types of delay, the resulting graphics for each delay and metric. Tables comparing the objective video quality achieved from each delay will be presented later.

Each test was performed according to the model as following sequence.

A total of 5 videos, 4 with 300 frames per video, were used.

- **Graphics DELAY 1 (PSNR and SSIM): Video (1,2,3,4,5)**
- **Graphics DELAY 2 (PSNR and SSIM): Video (1,2,3,4,5)**
- **Graphics DELAY 3 (PSNR and SSIM): Video (1,2,3,4,5)**

Video: NAME (Number of Frames, Time)

6 RESULTS AND COMPARISONS

This chapter presents the results and comparisons for all the video quality tests, according to the test methodology proposed in the previous chapter.

The video samples used in this work were compressed in MPEG-2 compliance, and were broadcasted in the German digital television, at the resolution of 720 by 576 pixels per frame, in progressive mode. They are presented below:

HORSECAB.TS (300 FRAMES, 12S)

RALLY.TS (300 FRAMES, 12S)

SPLASH.TS (300 FRAMES, 12S)

WALK.TS (250 FRAMES, 10S)

WATERSKIING.TS (300 FRAMES, 12S)

Figure 6.1: Test Videos Sequences Image

The results are divided for each Delay and Metric, and for each video the PSNR and SSIM metrics are shown. Because the resulting tests are numerous, resulting in 30 graphs, in this section only one video test for each delay and associated metric will be explained. The remaining graphs will be added to the appendix A at the end of this monography.

DELAY 1 – PSNR and SSIM

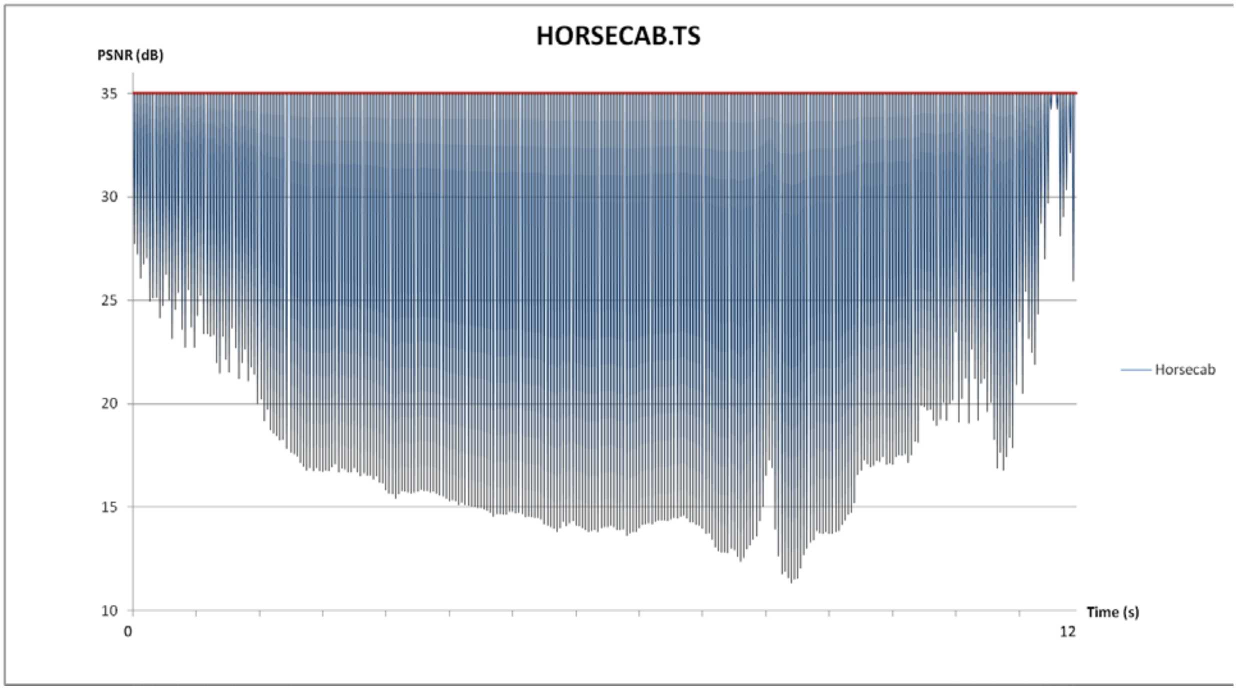


Figure 6.2: Delay 1 (PSNR) - Horsecab Video

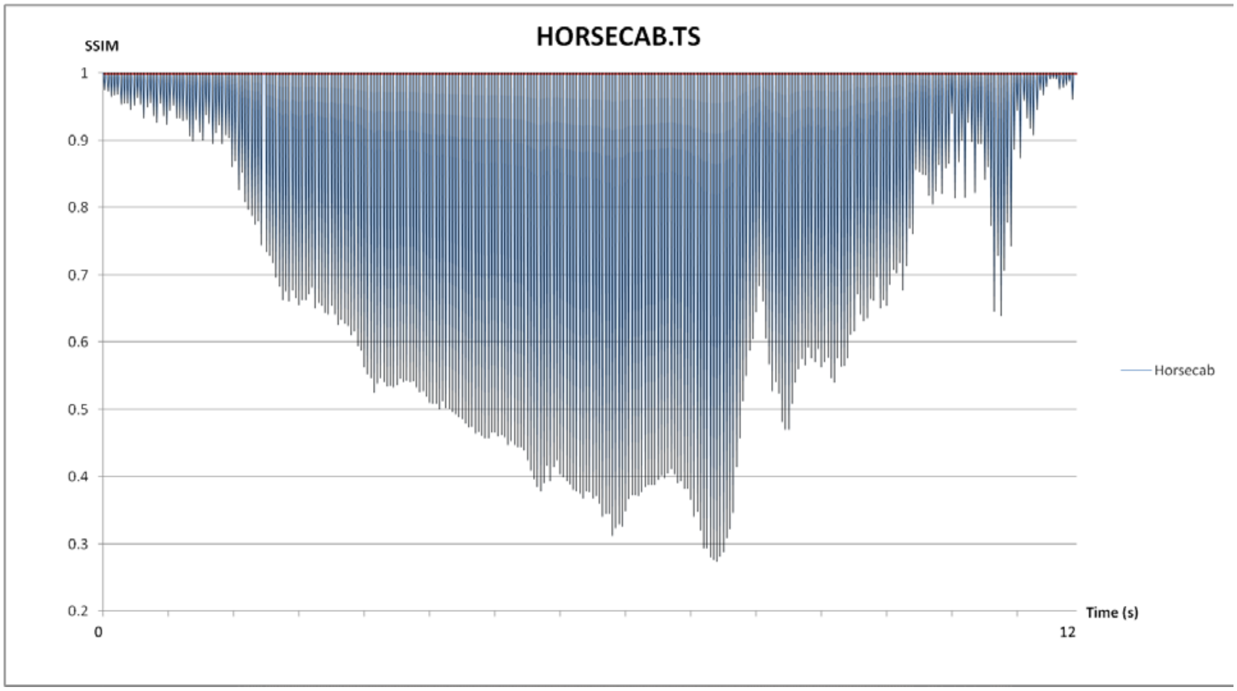


Figure 6.3: Delay 1 (SSIM) - Horsecab Video

In these two graphs no matter which metric was used, PSNR or SSIM, it is possible to see that the Delay 1 (with an absolute average of 36ms delay) is introducing a lot of computation to the metrics; in fact using the Delay 1 is introducing along the timeline 599 comparisons between frames, the number of comparisons is directly related to the quantity of delay imposed.

With more comparisons imposed by the Delay, the quality of the video is reduced proportionally to the delay, having an average result of 26.30dB for the PSNR metric and 0.82 for the SSIM metric.

For better understanding of the quality detection imposed by the Delay 1 it is recommended to view the result of all graphs available at the end of this work, attached to the Appendix A.

DELAY 2 – PSNR and SSIM

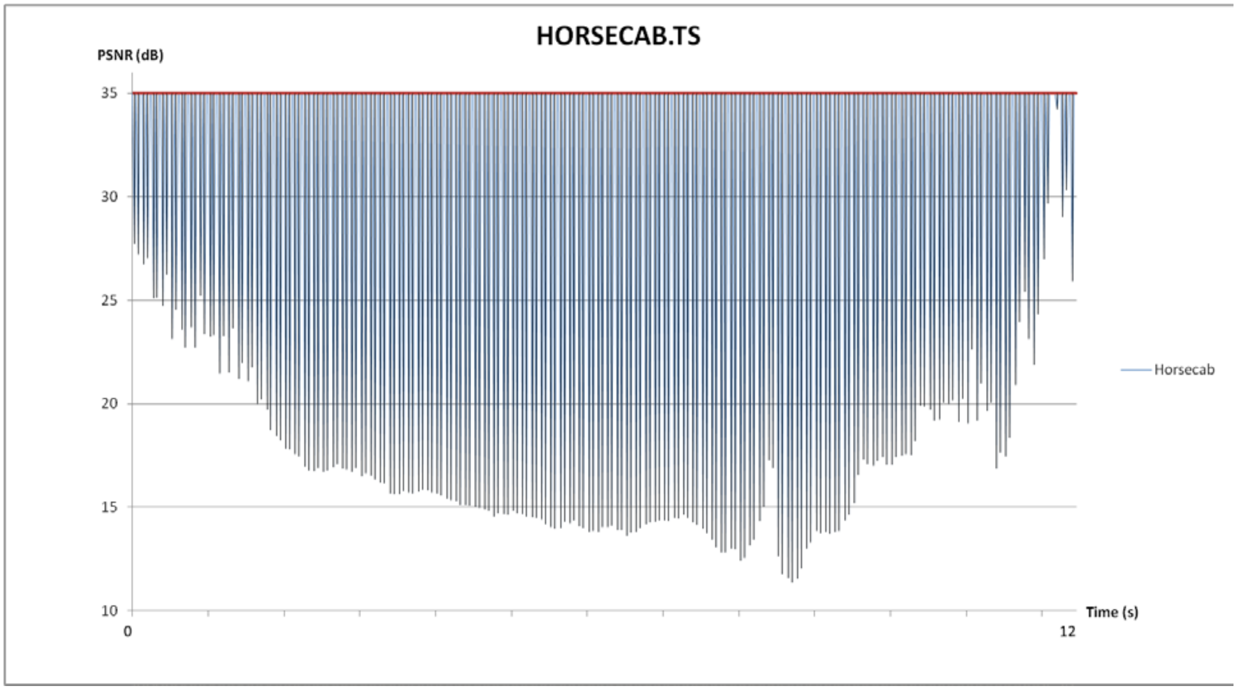


Figure 6.4: Delay 2 (PSNR) – Horsecab Video

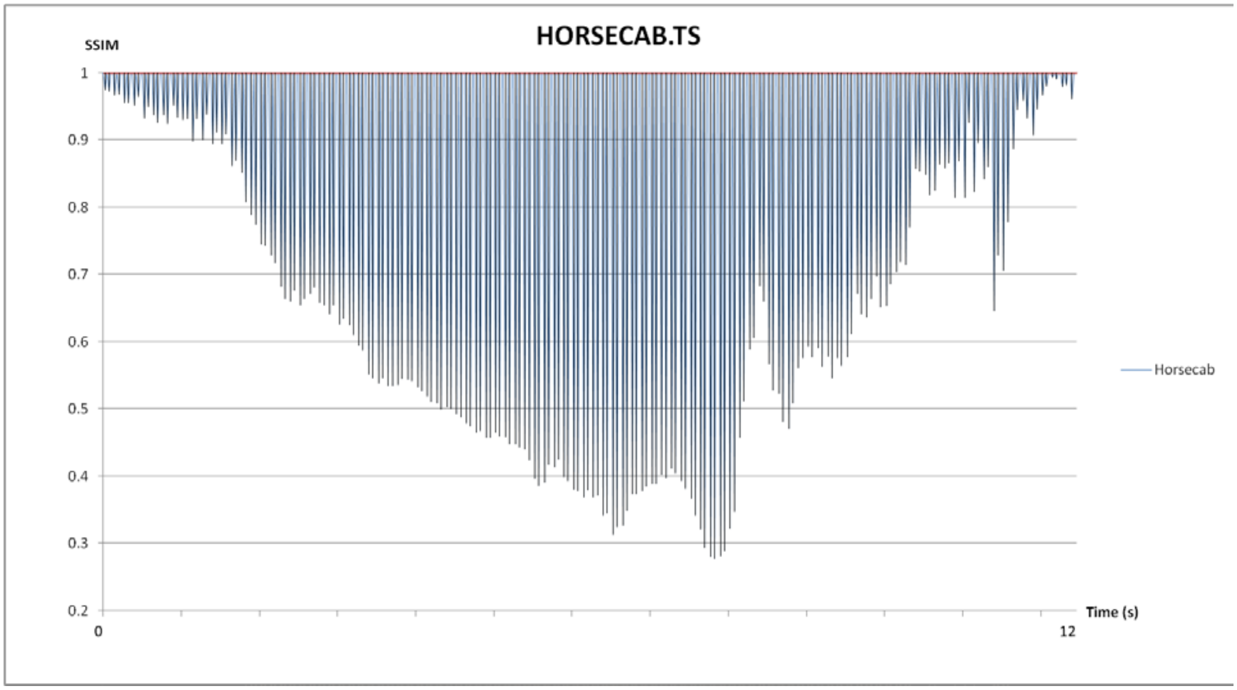


Figure 6.5: Delay 2 (SSIM) – Horsecab Video

In the graphs above regarding which metric was used, PSNR or SSIM, it is possible to see that the Delay 2 (with an absolute average of 14ms delay) is less computationally demanding to the metrics when compared to the previous delay; in fact using the Delay 2 is introducing along the timeline 499 comparisons between frames, the number of comparisons is directly related to the quantity of delay imposed.

With less comparisons imposed by the Delay 2, the quality of the video is enhanced, having an average result of 28.03dB for the PSNR metric and 0.85 for the SSIM metric.

For better understanding of the quality detection imposed by the Delay 2 it is recommended to view the result of all graphs available at the end of this work, attached to the Appendix A.

DELAY 3 – PSNR and SSIM

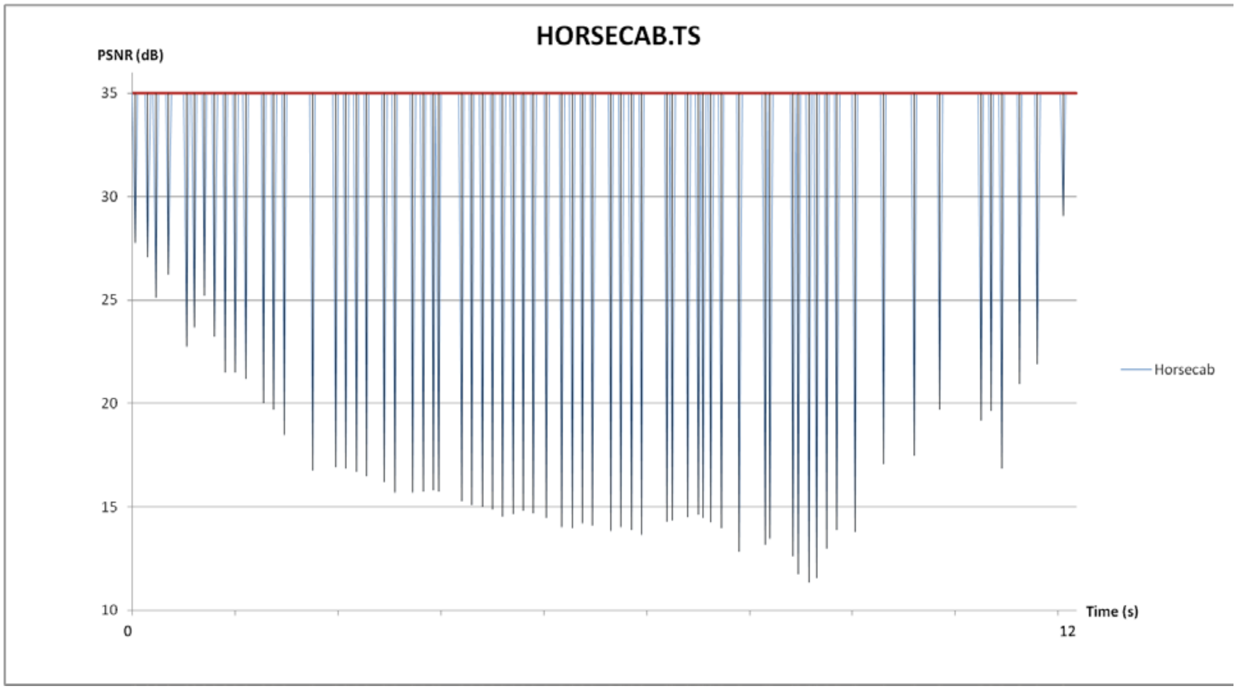


Figure 6.6: Delay 3 (PSNR) – Horsecab Video

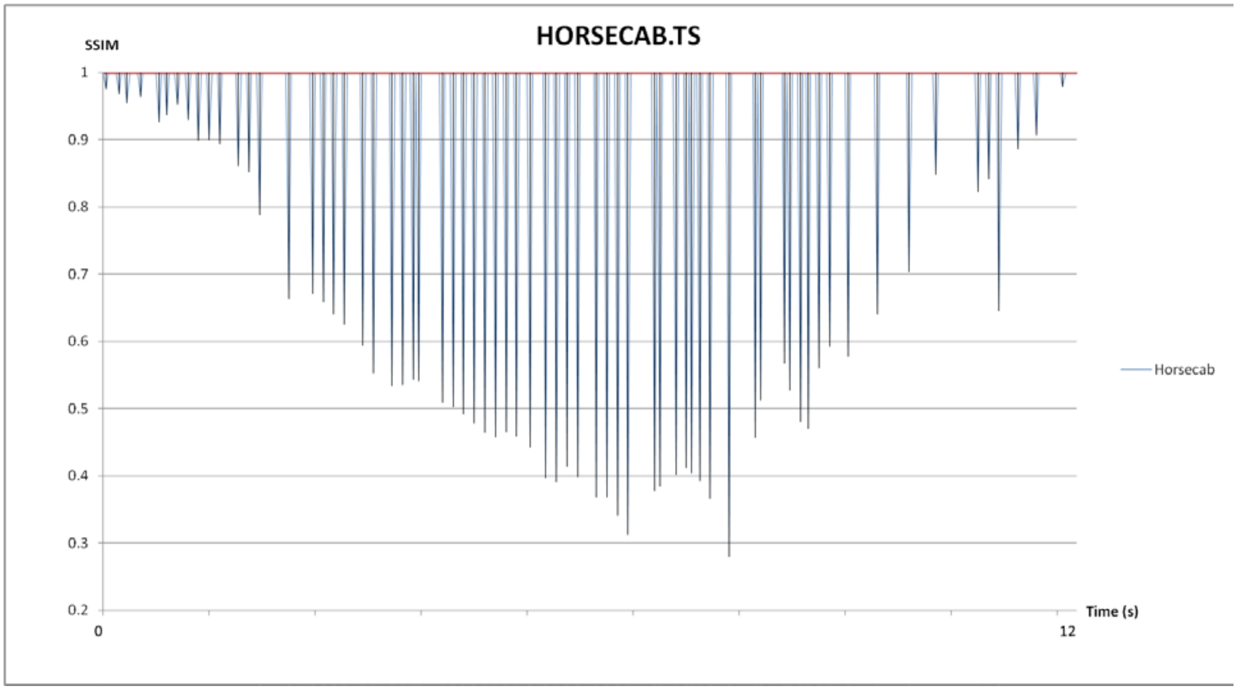
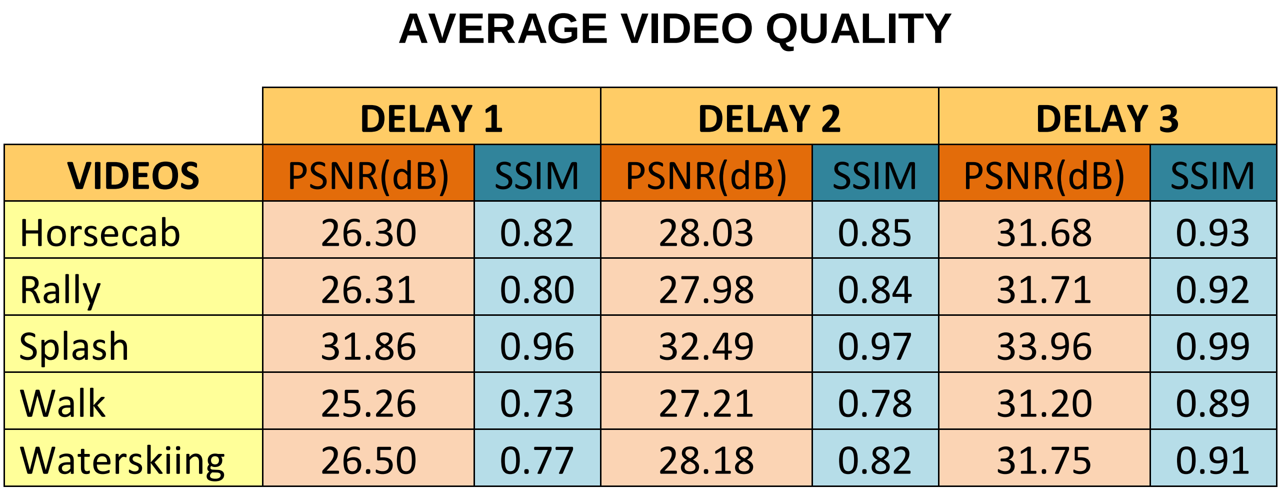


Figure 6.7: Delay 3 (SSIM) – Horsecab Video

In the last two graphs above regarding which metric was used, PSNR or SSIM, it is possible to see that the Delay 3 (with an absolute average of 1ms delay) is demanding much less computation to the metrics when compared to the previous two delays; in fact using the Delay 3 is introducing along the timeline only 368 comparisons between frames, the number of comparisons is directly related to the quantity of delay imposed.

With less comparisons imposed by the Delay 3, the quality of the video is enhanced again, having an average result of 31.68dB for the PSNR metric and 0.93 for the SSIM metric.

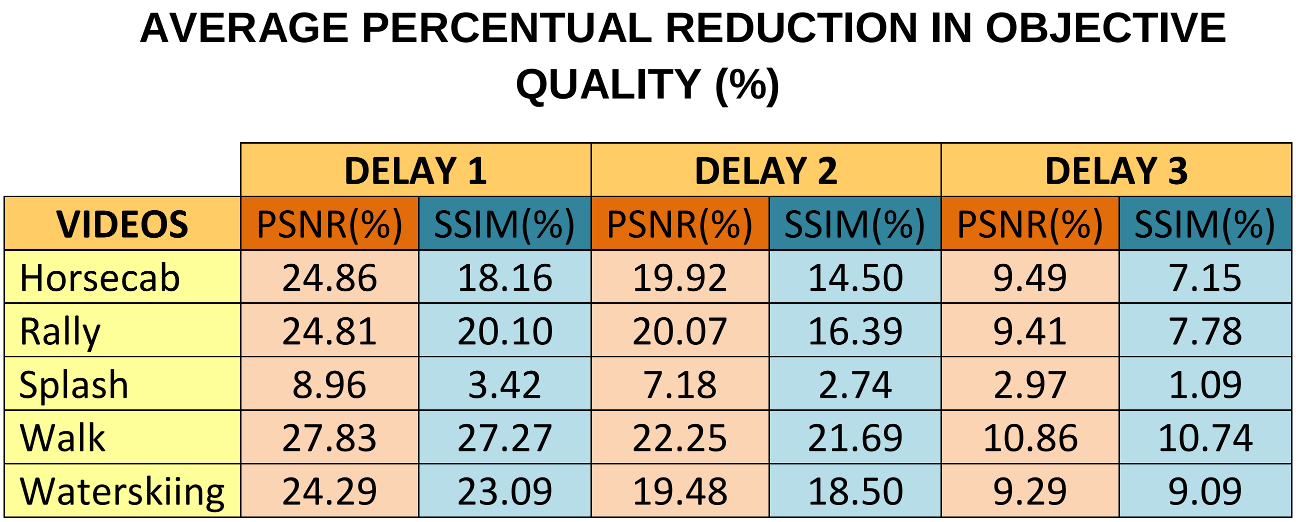
For better understanding of the quality detection imposed by the Delay 3, it is recommended to view the result of all graphs available at the end of this work, attached to the Appendix A.



AVERAGE VIDEO QUALITY

	DELAY 1		DELAY 2		DELAY 3	
VIDEOS	PSNR(dB)	SSIM	PSNR(dB)	SSIM	PSNR(dB)	SSIM
Horsecab	26.30	0.82	28.03	0.85	31.68	0.93
Rally	26.31	0.80	27.98	0.84	31.71	0.92
Splash	31.86	0.96	32.49	0.97	33.96	0.99
Walk	25.26	0.73	27.21	0.78	31.20	0.89
Waterskiing	26.50	0.77	28.18	0.82	31.75	0.91

Table 5.1: Average Video Quality – All Videos, Delays, Metrics



AVERAGE PERCENTUAL REDUCTION IN OBJECTIVE QUALITY (%)

	DELAY 1		DELAY 2		DELAY 3	
VIDEOS	PSNR(%)	SSIM(%)	PSNR(%)	SSIM(%)	PSNR(%)	SSIM(%)
Horsecab	24.86	18.16	19.92	14.50	9.49	7.15
Rally	24.81	20.10	20.07	16.39	9.41	7.78
Splash	8.96	3.42	7.18	2.74	2.97	1.09
Walk	27.83	27.27	22.25	21.69	10.86	10.74
Waterskiing	24.29	23.09	19.48	18.50	9.29	9.09

Table 5.2: Average Percentual Reduction – All Videos, Delays, Metrics

After all the graph results it is possible to infer that the variable timeline approach is recognizing the errors caused by the variable delay introduced in the video, considering the presentation time stamps variation and frames compared to the original video with a reference quality, that is, with no delays in the presentation time stamps.

Table 5.1 presents an average video quality for all five videos tested. Table 5.2 presents the average percentual reduction for the same videos, considering the percentual reduction on the objective quality, using 35dB as the reference quality for the PSNR, and SSIM equal to 1 as the reference quality for this metric.

Observing the average quality from the videos regarding the metrics used it is possible to see that the same kind of delay introduces in average the same loss of quality in the videos, except for “Splash”. The explanation for this is that the algorithm recognizes not only the presentation time stamps but the frames from the video, meaning that different kinds of video can represent great variation in quality depending on the introduced delay.

Videos with a lot of movement tend to have more losses during the introduction of delays, while videos with steady camera with little variation or movement tend to have better quality when submitted to delayed PTS. This can be seen in the videos because all those except “Splash” have increased movement scenes, and the quality from all those videos in average is the same.

This assessment has confirmed that the more delay is imposed to a video the worst the quality is, and the percentual reduction introduced by this delay depends on the video characteristics also.

7 CONCLUSIONS

This work presented a new approach, called variable timeline approach for calculating Video Quality Metrics, using already proven quality metrics tested by the VQEG (PSNR, SSIM).

Some principles of video standards, video artifacts, video quality, current calculation methods for video quality assessment and detailed development of the variable timeline approach were described. Results and comparisons that can relate to frame delay time with video quality were generated.

Since the beginning of this work there was no method to consider the presentation time stamp from the frames to measure the quality of a video. During the time this work was developed, the NTIA - creators of the VQM metric - published a new Video Quality Model with variable frame delay in September, 2011, which uses this idea of considering the variable display time of a video. The results from this new model are capable of achieving better visual quality results than the classical ones, but this model has not been validated or standardized yet.

This work tested the variable timeline approach using five different videos, with three different kinds of delay, and using two different video quality metrics. The new method was used to evaluate every imposed delay and after all the results it is possible to conclude that the proposed approach of this work is capable of detecting the differences caused by variable presentation time stamps in videos when using SSIM and PSNR metric, giving a result that is more suitable to the perceptual visual quality according to the video quality metric used.

One problem that was not detected by the tests was the intensity of each delay, because analyzing the structure of the algorithm it is possible to predict that the same kind of delay in quantity, only varying it is intensity (between 0ms and 40ms), will give the same result, and this is inadequate for solving the problem of video quality considering frame and time. So, the way the variable presentation time was taken in account by the current metrics is not the perfect solution for calculating the video quality with variable presentation time in videos.

The reason for this lack of full quality that the new approach is capable of detecting only the types of variable delay in quantity, number of delays imposed to the video; using a good approximation, calculating the differences from frames it differs along the display time of the modified video, but is not considering the time duration itself in the calculation of the problem. So, two different delays in time can have the same quality result depending in how long is the delay, because the metric is only recognizing the type of the delay. Other problem detected is that the program doesn’t allow frame reordering or frame skipping, because that would interfere with the timeline calculation used. Remember, it is not created a new video quality metric that adds the presentation time as a new video quality parameter.

Probably doing this, calculating the metrics along time, considering each frame time duration and adding this time duration to the metrics would achieve much better video quality results; however this would change the metric itself and thus was not considered for this work.

Possible future works can refine this work and achieve a new video quality metric, as described above, and can adapt the variable timeline approach to be used with different types of video and standards.

REFERENCES

CDVL – Available: <www.cdvl.org> Access: October 2011

MATLAB – The Language of Technical Computing – Available: <www.mathworks.com>

MSU VIDEO QUALITY MEASUREMENT TOOL – Available: <www.compression.ru>

"ATSC Standard: Transport Stream File System Standard" Advanced Television Systems Committee, 1750 K Street, N.W. Suite 1200Washington, D.C. 20006 www.atsc.org - Doc. A/95, 25 February 2003

BERTS JOHAN AND PERSSON ANDERS. **"Objective and subjective quality assessment of compressed digital video sequences"** Master Thesis performed at Ericsson Microwave Systems AB, Department of Signals and Systems, Chalmers University of Technology**,** Göteborg, Sweden, 1998

BOGDAN – Available:<http://www.codeproject.com/Articles/6981/Transport-Stream-Analyzer-for-HDTV-standard> Access May 2011

BOSI, M. and GOLDBERG, R. E. Introduction to Digital Audio Coding and Standards. Kluwer Academic Publishers, 2002.

BRANDEN LAMBRECHT C. J. AND VERSCHEURE O.. **"Perceptual Quality Measure using a Spatio-Temporal Model of the Human Visual System"** Proc. SPIE Vol. 2668, p. 450-461, March, 1996.

CARDENAS – Available:<en.wikipedia.org/wiki/File:MPEG_Transport_Stream_HL.svg> Acess: April 2011

DVB – Available:< http://www.dvb.org/> Acess: May 2011.

FARIAS, MYLÈNE C. Q.. **"Video Quality Metrics"**, **ISBN 978-953-7619-70-1**

GONZALEZ, R. C. AND WOODS, R. E. "**Digital Image Processing, 2nd ed**." Prentice Hall, Upper Saddle River, NJ.2002

GUERRA, RAPHAEL, **"A gravitational task model for target sensitive real-time applications"**, Technische Universitaet Kaiserslautern, Ph.D. Thesis, June 2011.

HORKY – Available: <en.wikipedia.org/wiki/File:MPEG.svg> Access: April 2011

HUYNH-THU Q. AND GHANBARI M., **“Impact of jitter and jerkiness on perceived video quality”** in Proc. of International Workshop on Video Processing Quality Metrics, Scottsdale, USA, Jan. 2006

ITU-T REC. J.247, **“Objective Perceptual Multimedia Video Quality Measurement in the Presence of a Full Reference”** International Telecommunication Union, Geneva, 2008.

NIRANJAN D. ET AL. **“Image Quality Assessment Based on a Degradation Model”** IEEE Transaction on Image Processing, VOL. 9, NO. 4, April 2000.

NWE – Available: <www.newworldencyclopedia.org/entry/HDTV> Access: April 2011

PINSON M. AND WOLF S., **“A New Standardized Method for Objectively Measuring Video Quality”** IEEE Transactions on Broadcasting, VOL. 50, NO.3, pp. 312-322, September, 2004.

VQEG.“**Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment.”** August 2003. Available: <www.vqeg.org> Access: July 2001

WANG Y., **“Survey of Objective Video Quality Measurements”** EMC Corporation Hopkinton, MA 01748, USA, 2006.

WANG Y., **“A Novel Quality Metric for Compressed Video Considering both Frame Rate and Quantization Artifacts”** in Proc. of VPQM’09, Scottsdale, AZ, USA, January 2009.

WANG Z., SHEIKH H. R. AND BOVIK A. C., **“Objective video quality assessment,”** in The Handbook of Video Databases: Design and Applications (B. Furht and O. Marqure, eds.), CRC Press, pp. 1041-1078, Sept. 2003.

WANG Z., BOVIK A. C. AND SIMONCELLI E. P., **“Structural Approaches to image quality assessment,”** in Handbook of Image and Video Processing (Al Bovik, ed.), 2nd edition, Academic Press, June 2005.

WANG Z., SIMONCELLI E. P. AND BOVIK A. C., **“Multi-scale structural similarity for image quality assessment,”** Invited Paper, IEEE Asilomar Conference on Signals, Systems and Computers, Nov. 2003. [Matlab Code] [Java Code 1, Java Code 2]

WANG Z., LU L., AND BOVIK A. C.**, “Video quality assessment based on structural distortion measurement,”** Signal Processing: Image Communication, special issue on “Objective video quality metrics”, vol. 19, no. 2, pp. 121-132, Feb. 2004.

WANG Z. AND LI Q.**, "Video quality assessment using a statistical model of human visual speed perception,"** Journal of the Optical Society of America A, vol. 24, no. 12, pp. B61-B69, Dec. 2007.

WANG Z., BOVIK A. C., SHEIKH H. R. AND SIMONCELLI E. P. , **"Image quality assessment: From error visibility to structural similarity,"** IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, Apr. 2004

WANG Z. AND BOVIK A. C. , **"Mean squared error: love it or leave it? - A new look at signal fidelity measures,"** IEEE Signal Processing Magazine, vol. 26, no. 1, pp. 98-117, Jan. 2009.

WANG Z., BOVIK A. C. AND LU L., **"Why is image quality assessment so difficult?"** IEEE International Conference on Acoustics, Speech, & Signal Processing, May 2002.

WANG Z., LU L., BOVIC A. C. **"Video quality assessment using structural distortion measurement"** Signal Processing: Image Communication, special issue on "Objective video quality metrics", vol. 19, no. 2, pp. 121-132, February 2004.

WANG Z. AND BOVIK ALAN C. "**Modern Image Quality Assessment**." **ISBN: 1598290223**

WANG Z., **"Objective image/video quality measurement – a literature survey"** EE 381K: Multidimensional Digital Signal Processing.

WINKLER S., **"Digital Video Quality - Vision Models and Metrics."** John Wiley & Sons, Ltd, 2005.

WINKLER S., **"Video quality and beyond"** in Proc. European Signal Processing Conference, Poznan, Poland, September 3-7 2007.

APPENDIX A - OBJECTIVE MEASURAMENT RESULTS

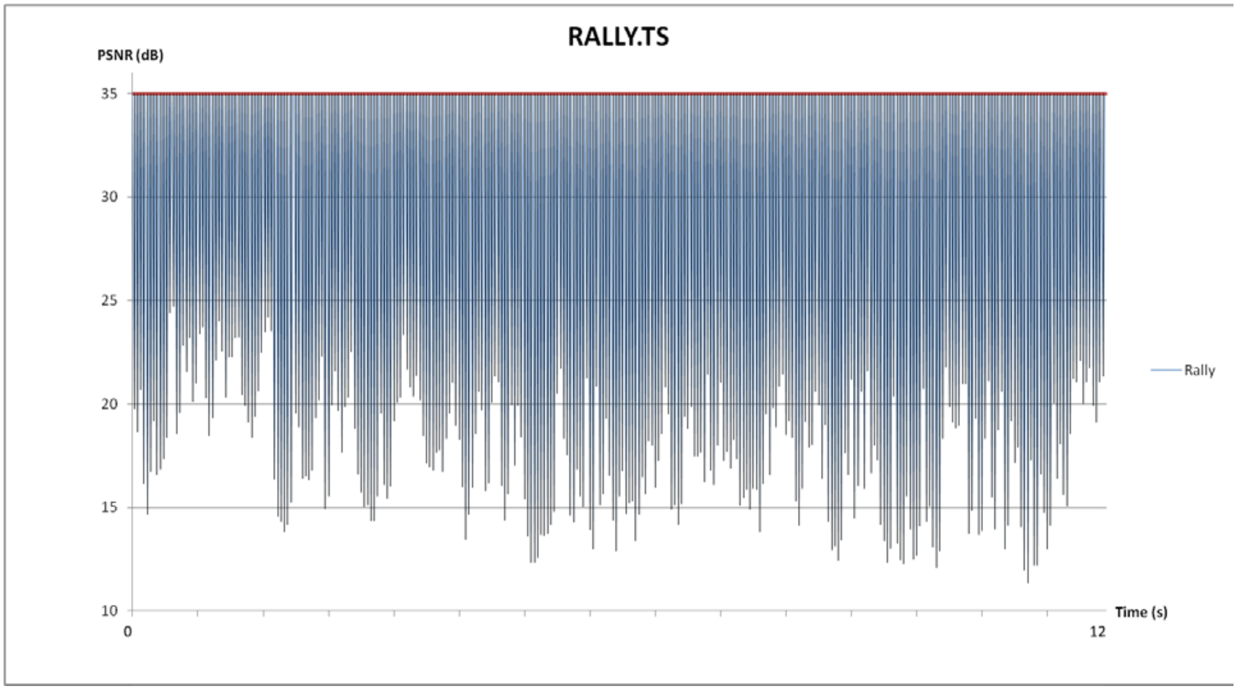


Figure 7.1: Delay 1 (PSNR) - Rally Video

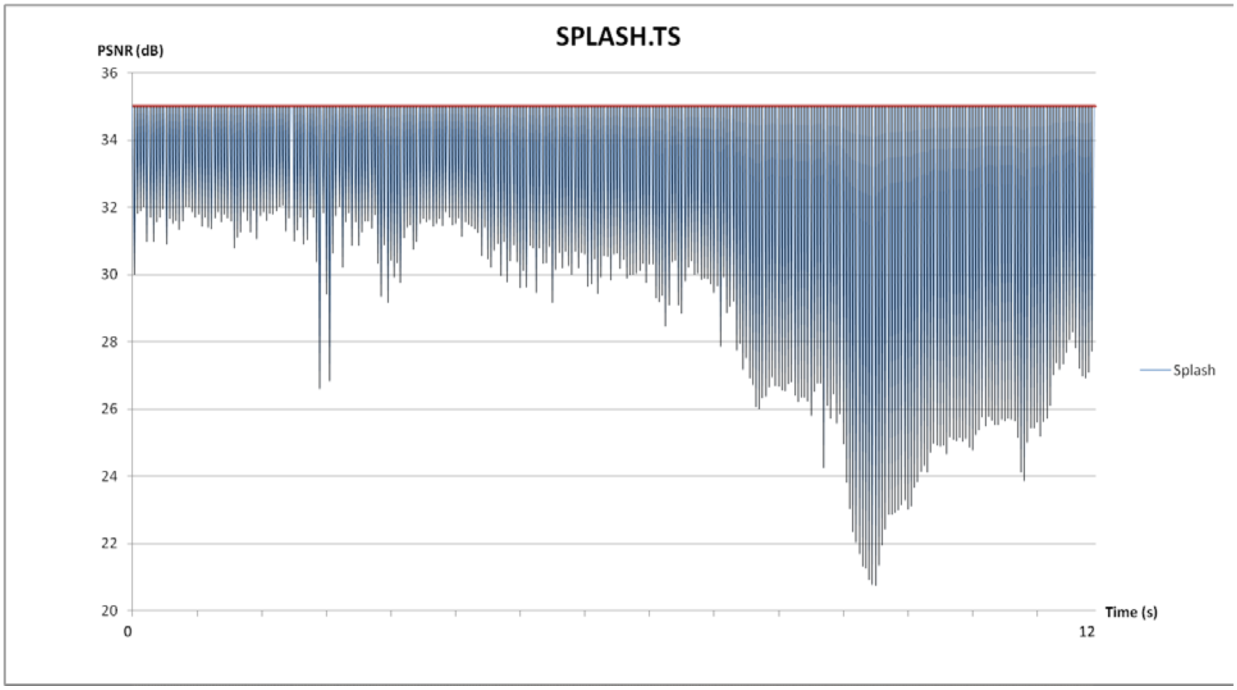


Figure 7.1.1: Delay 1 (PSNR) - Splash Video

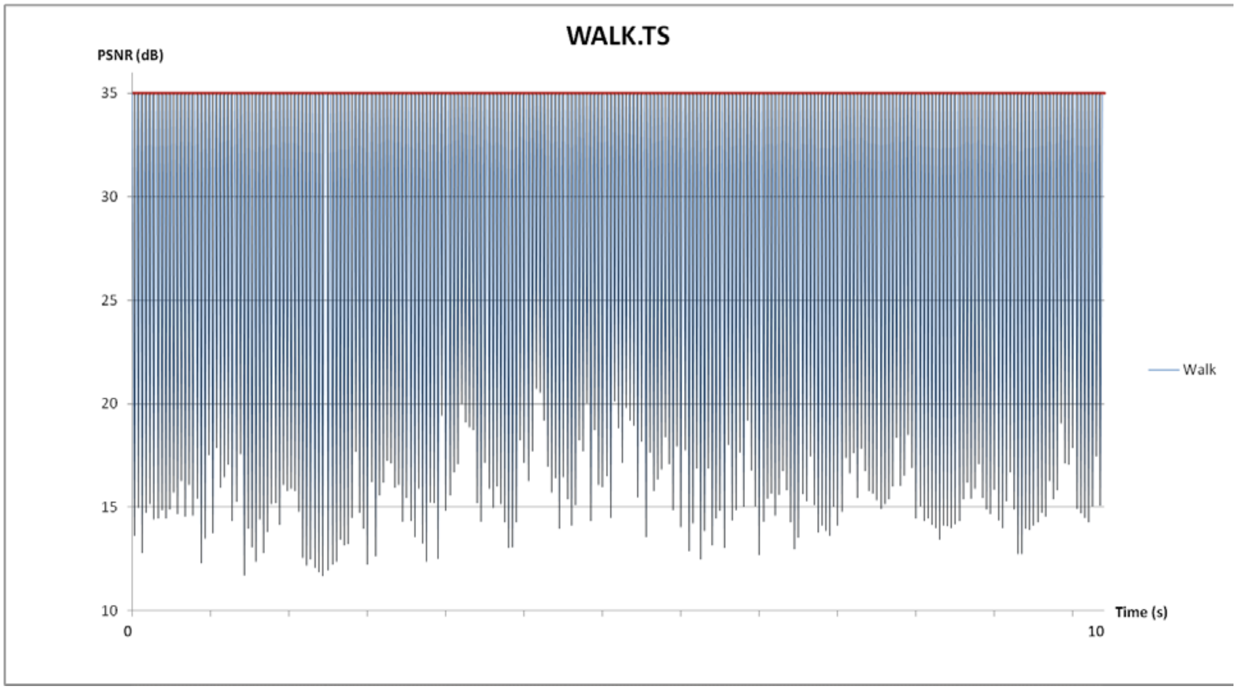


Figure 7.1.2: Delay 1 (PSNR) - Walk Video

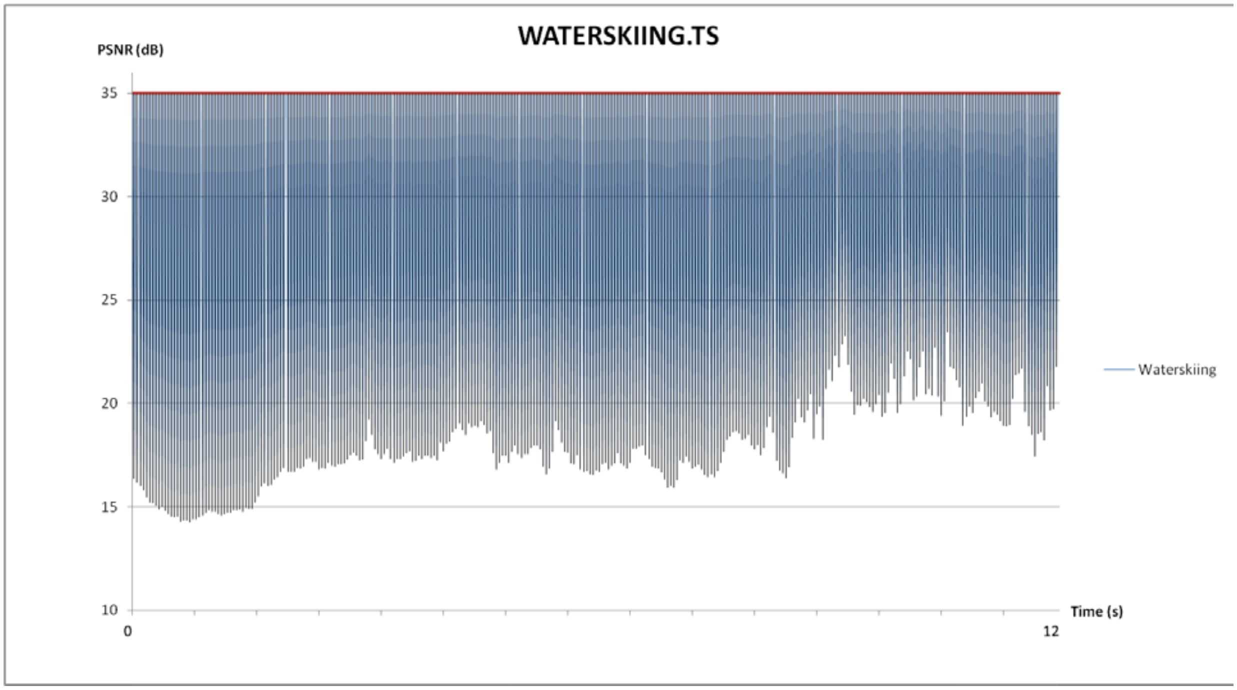


Figure 7.1.3: Delay 1 (PSNR) - Waterskiing Video

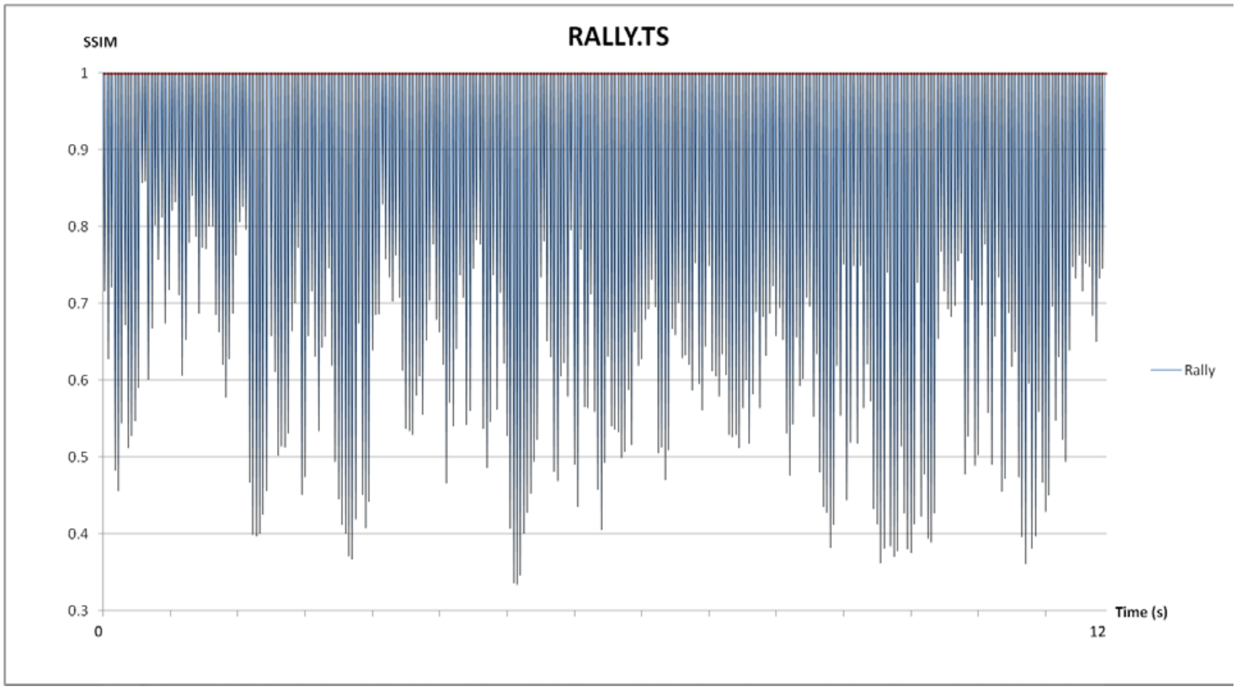


Figure 7.2: Delay 1 (SSIM) – Rally Video

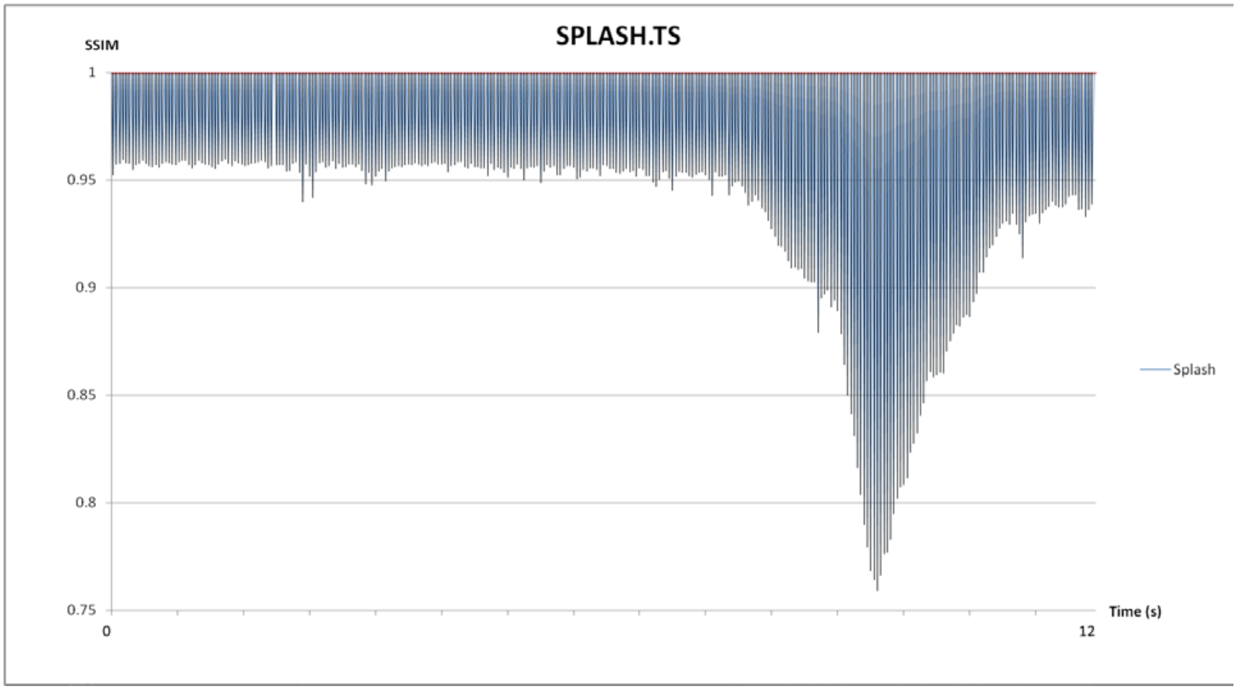


Figure 7.2.1: Delay 1 (SSIM) - Splash Video

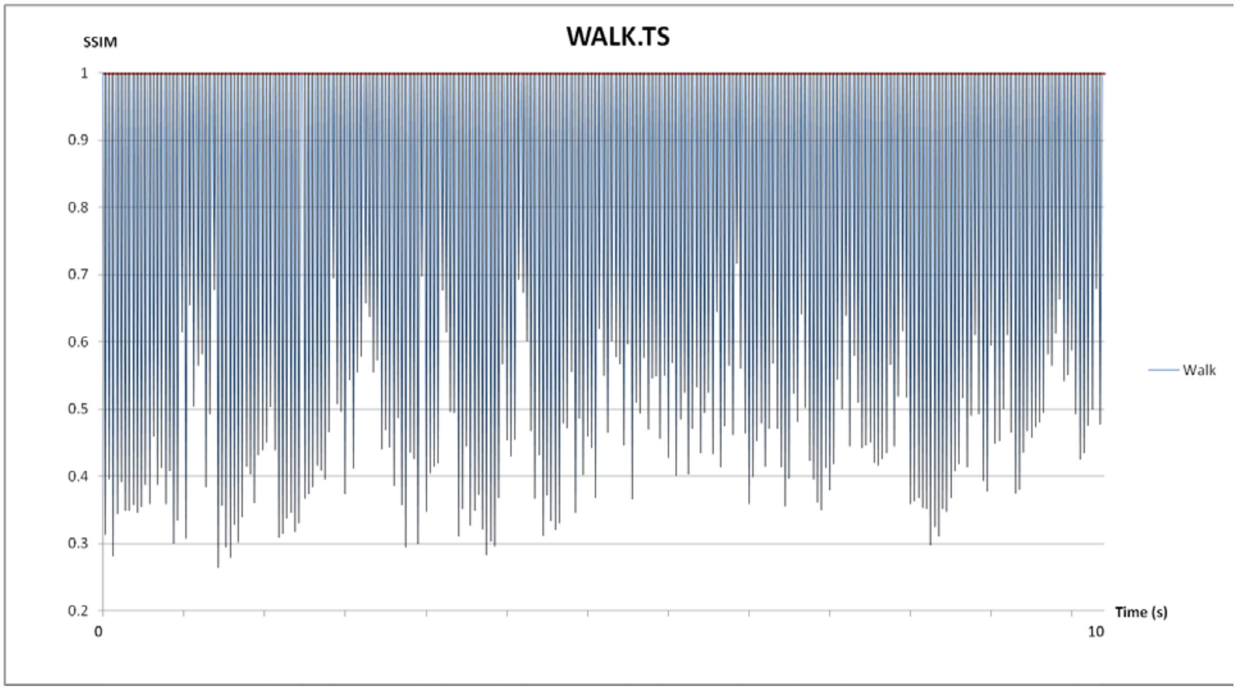


Figure 7.2.3: Delay 1 (SSIM) - Walk Video

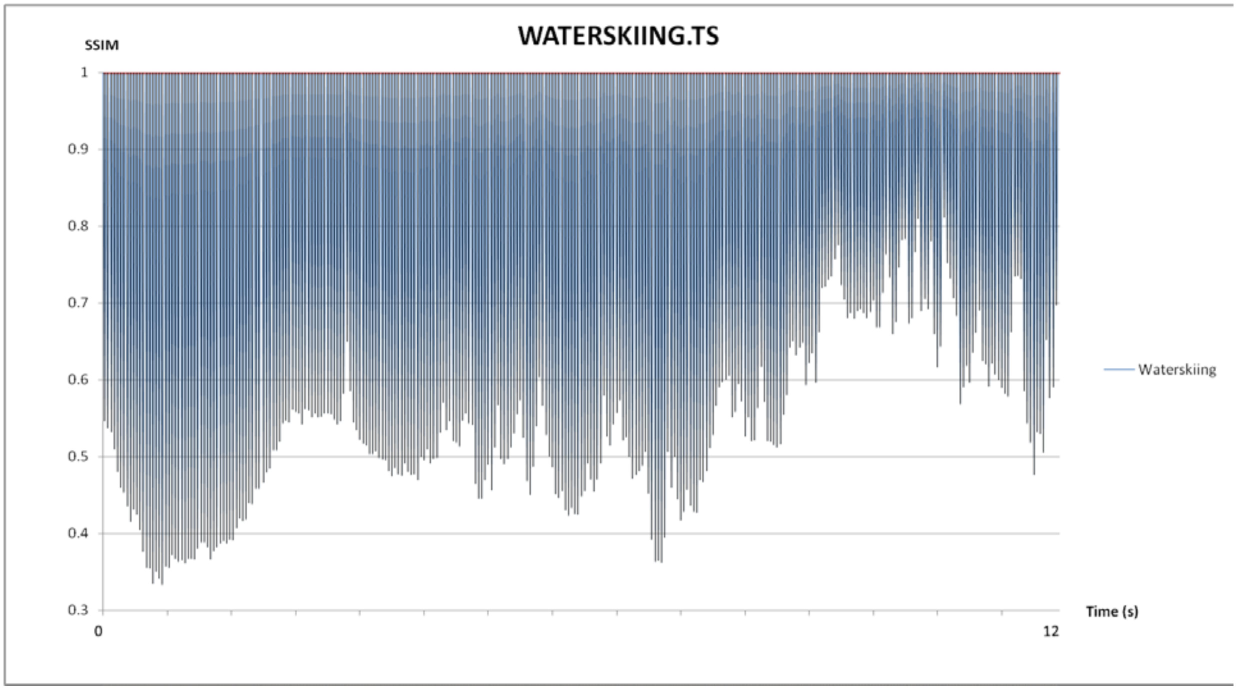


Figure 7.2.4: Delay 1 (SSIM) – Waterskiing Video

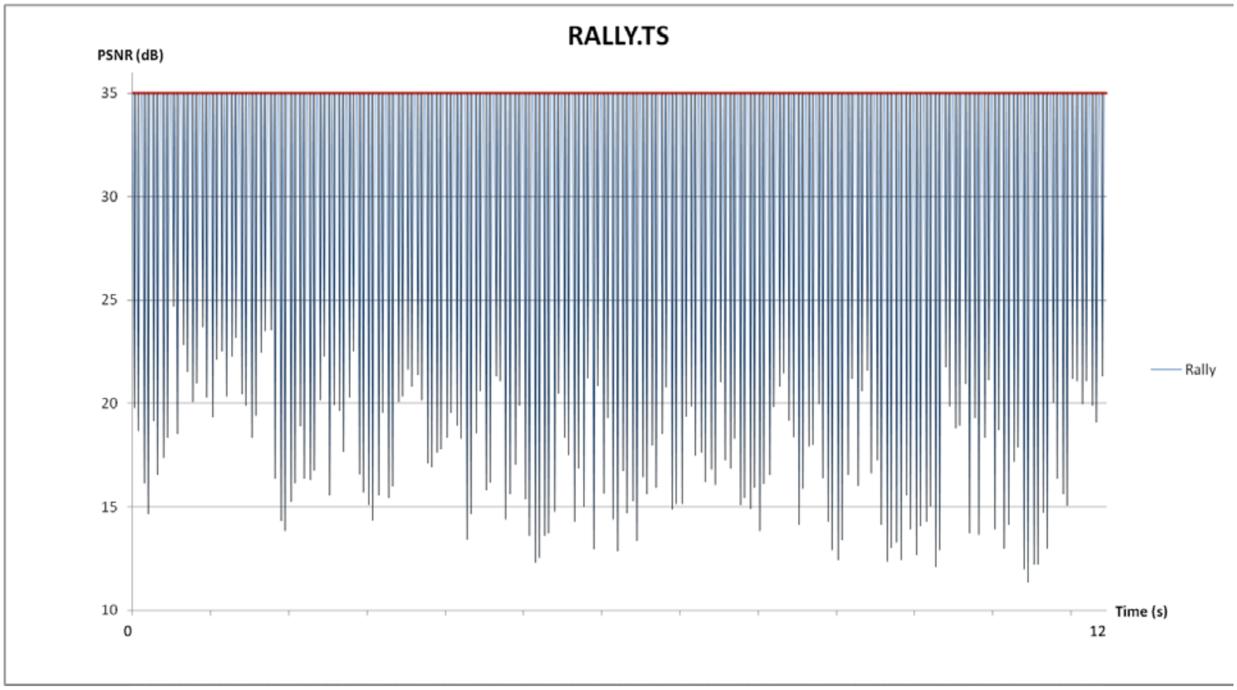


Figure 7.3: Delay 2 (PSNR) – Rally Video

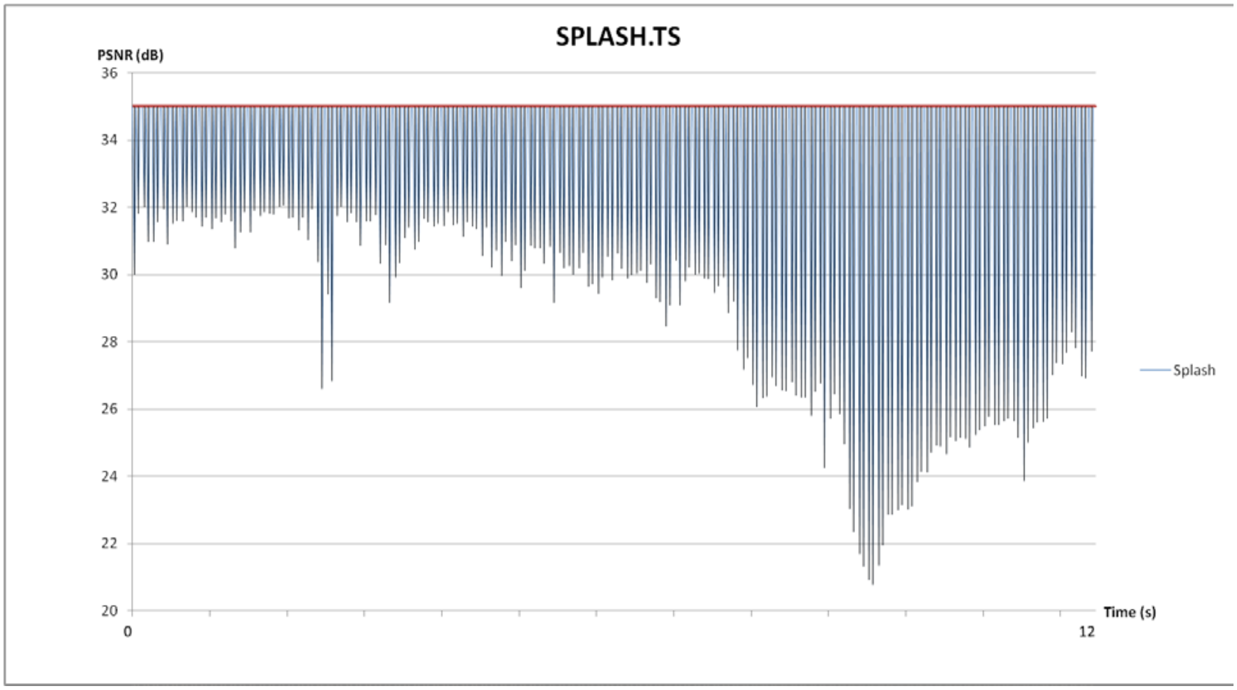


Figure 7.3.1: Delay 2 (PSNR) – Splash Video

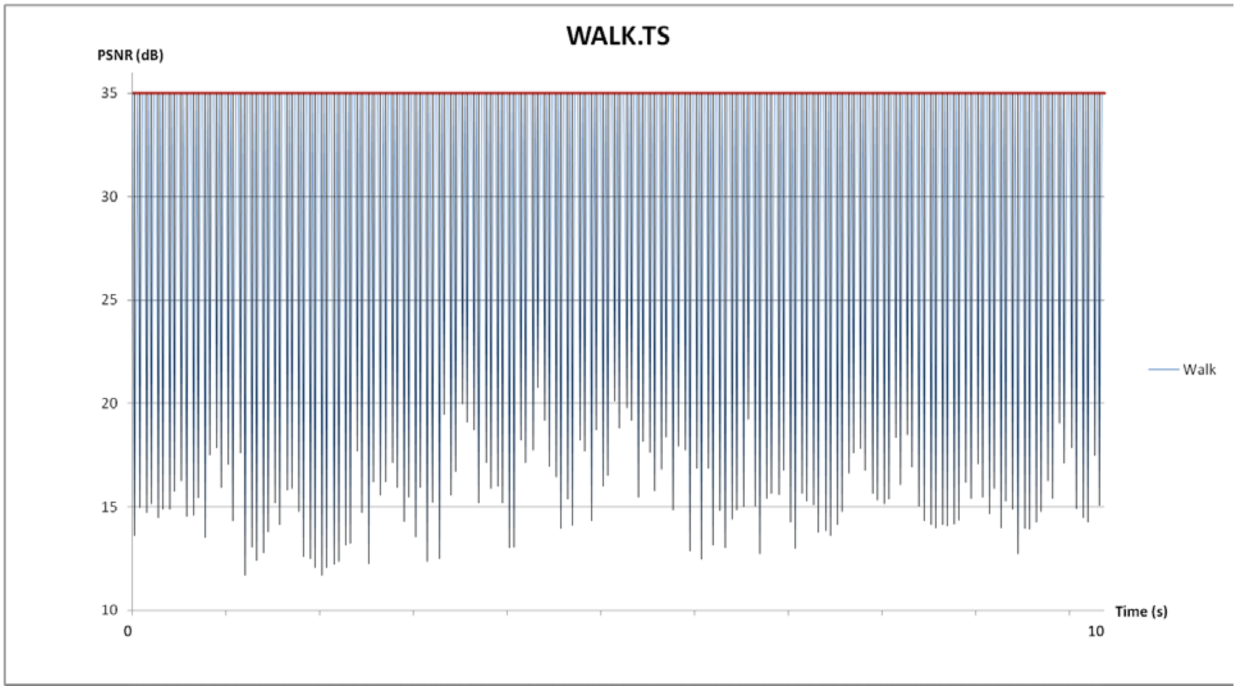


Figure 7.3.2: Delay 2 (PSNR) – Splash Video

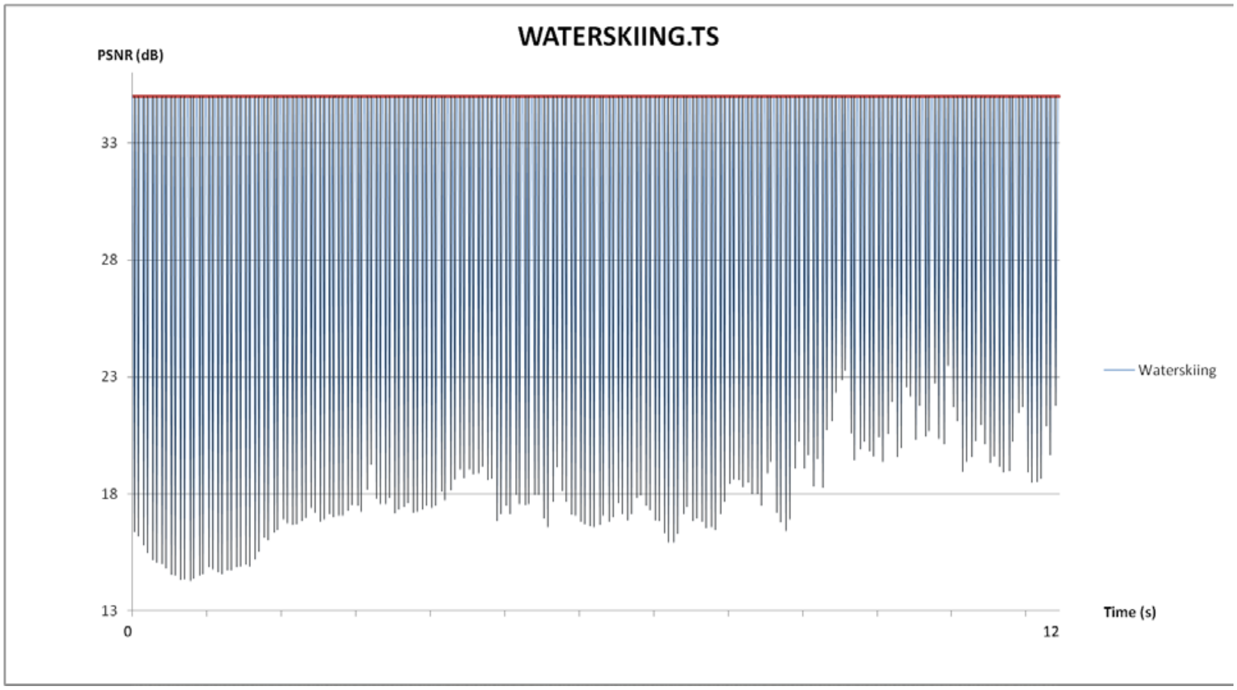


Figure 7.3.3: Delay 2 (PSNR) – Waterskiing Video

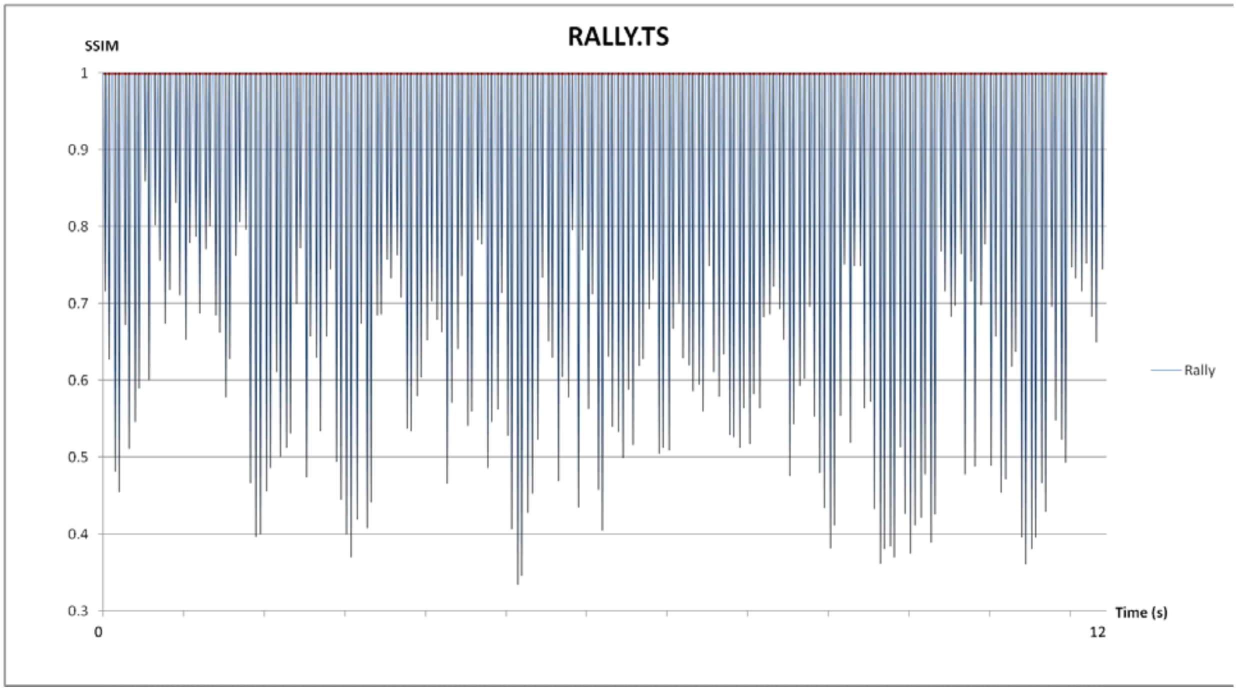


Figure 7.4: Delay 2 (SSIM) – Rally Video

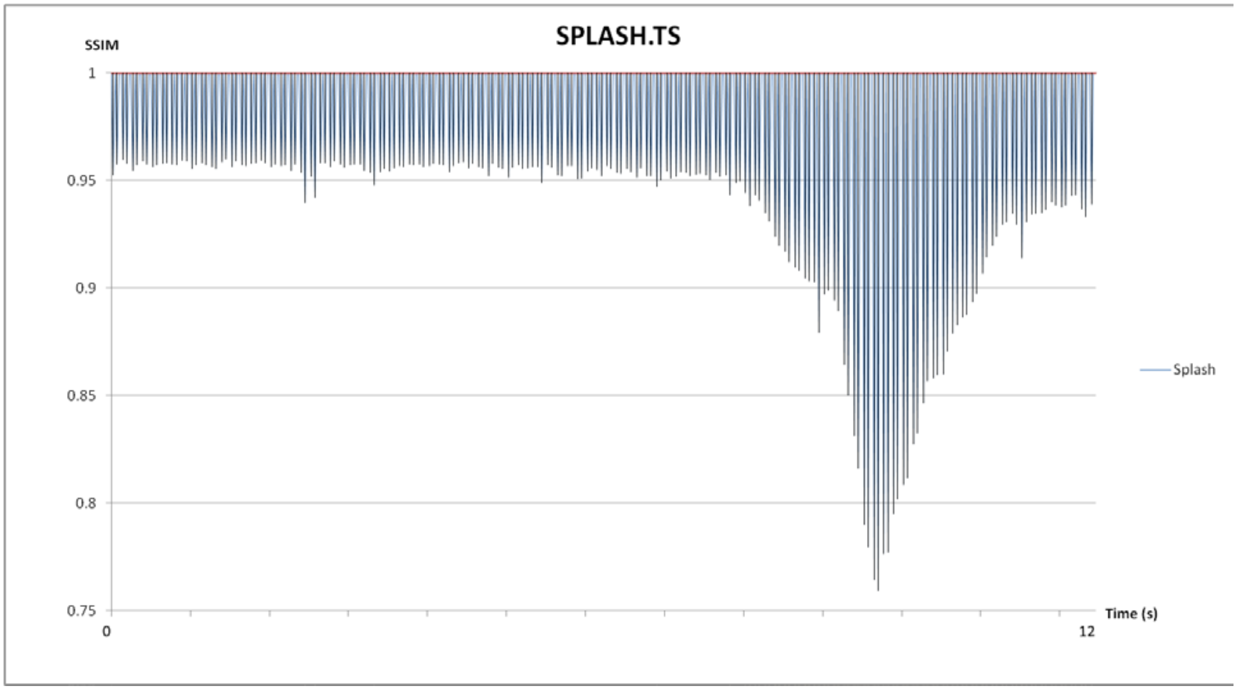


Figure 7.4.1: Delay 2 (SSIM) – Splash Video

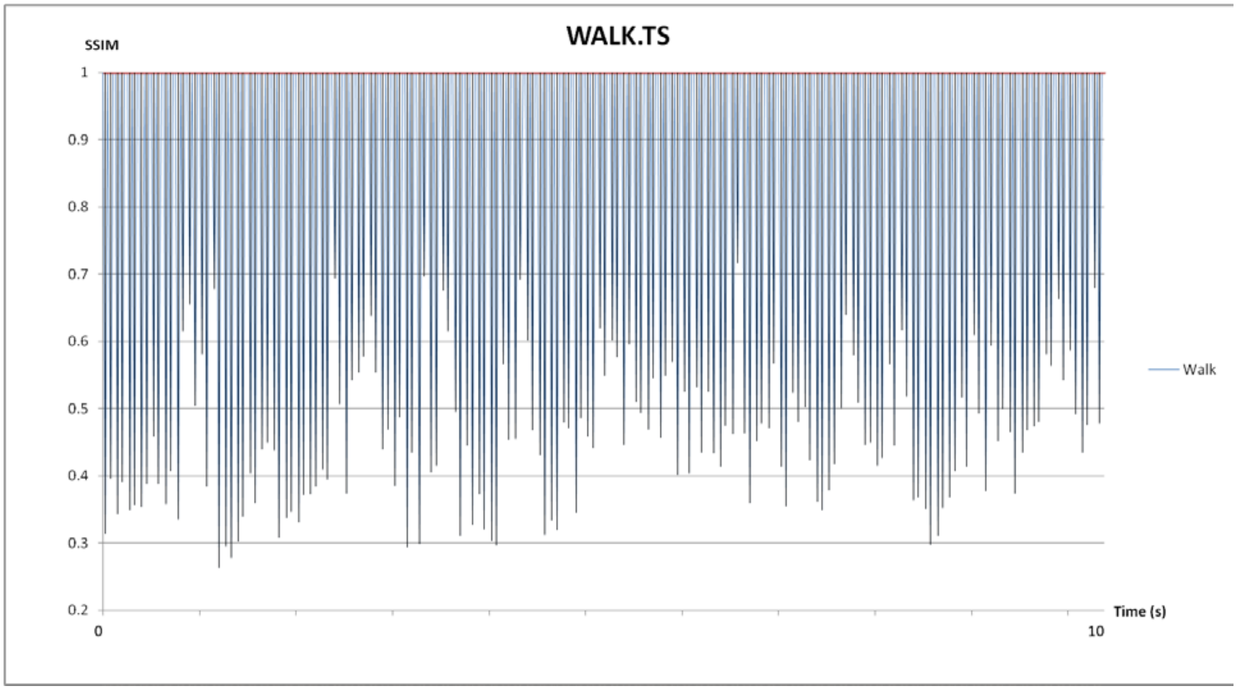


Figure 7.4.2: Delay 2 (SSIM) – Walk Video

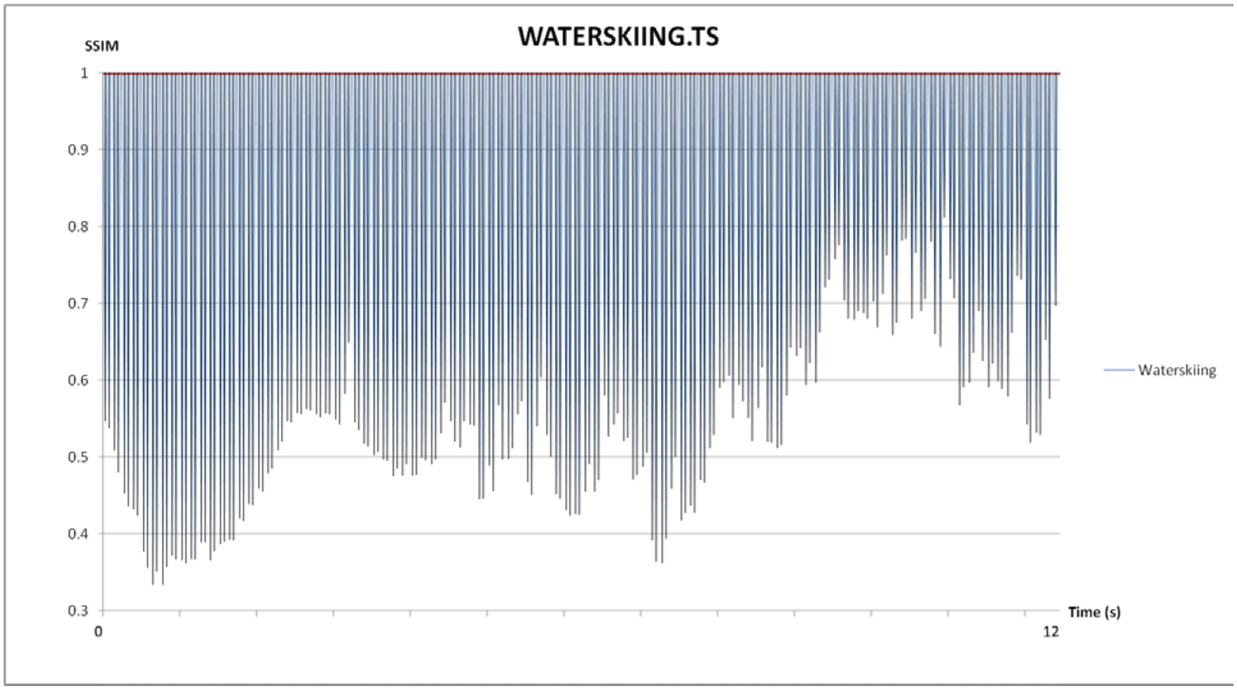


Figure 7.4.3: Delay 2 (SSIM) – Waterskiing Video

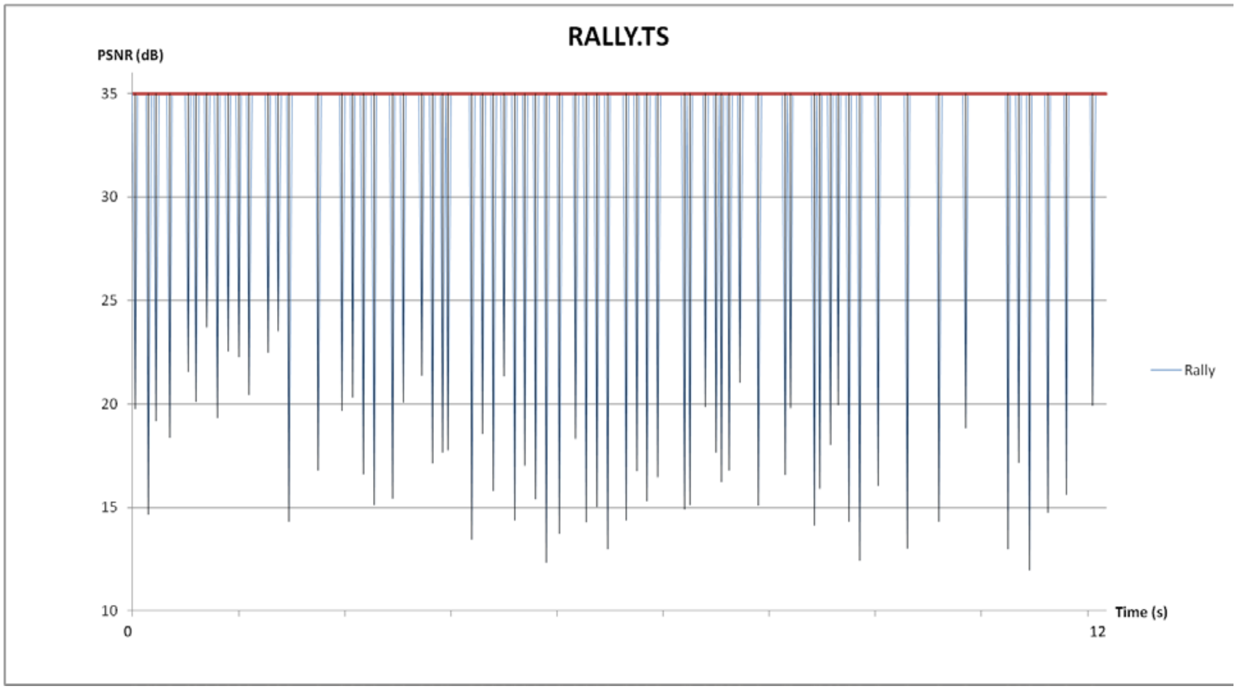


Figure 7.5: Delay 3 (PSNR) – Rally Video

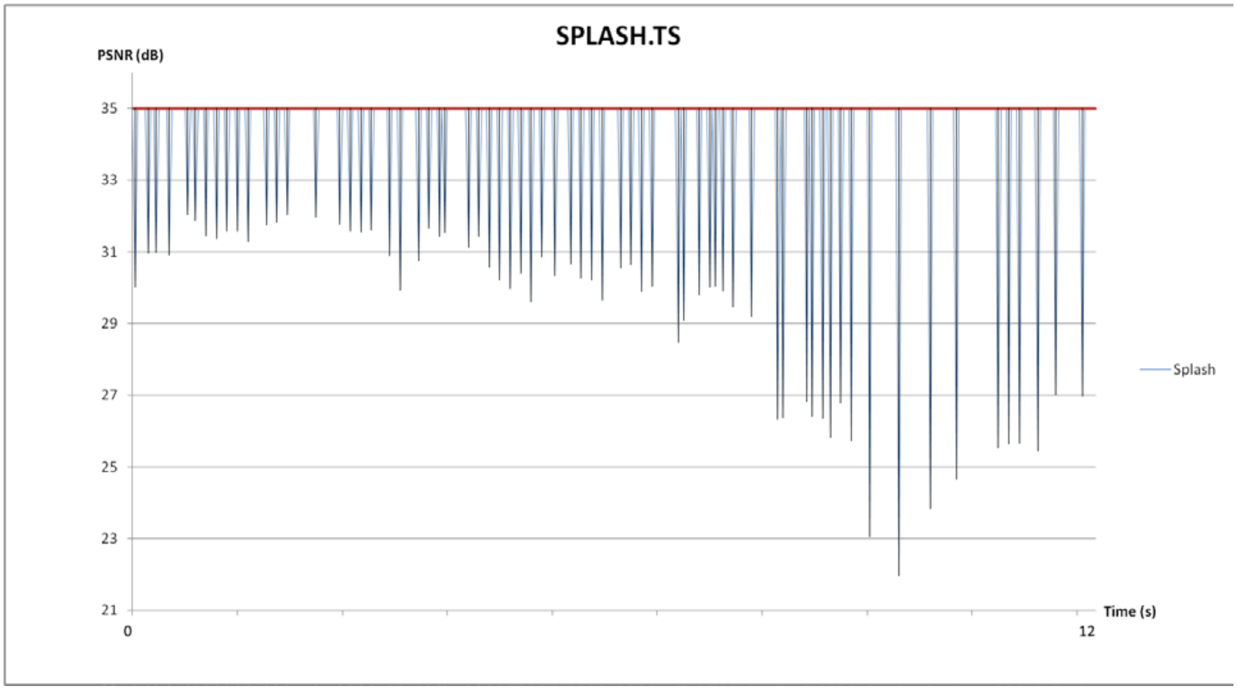


Figure 7.5.1: Delay 3 (PSNR) – Splash Video

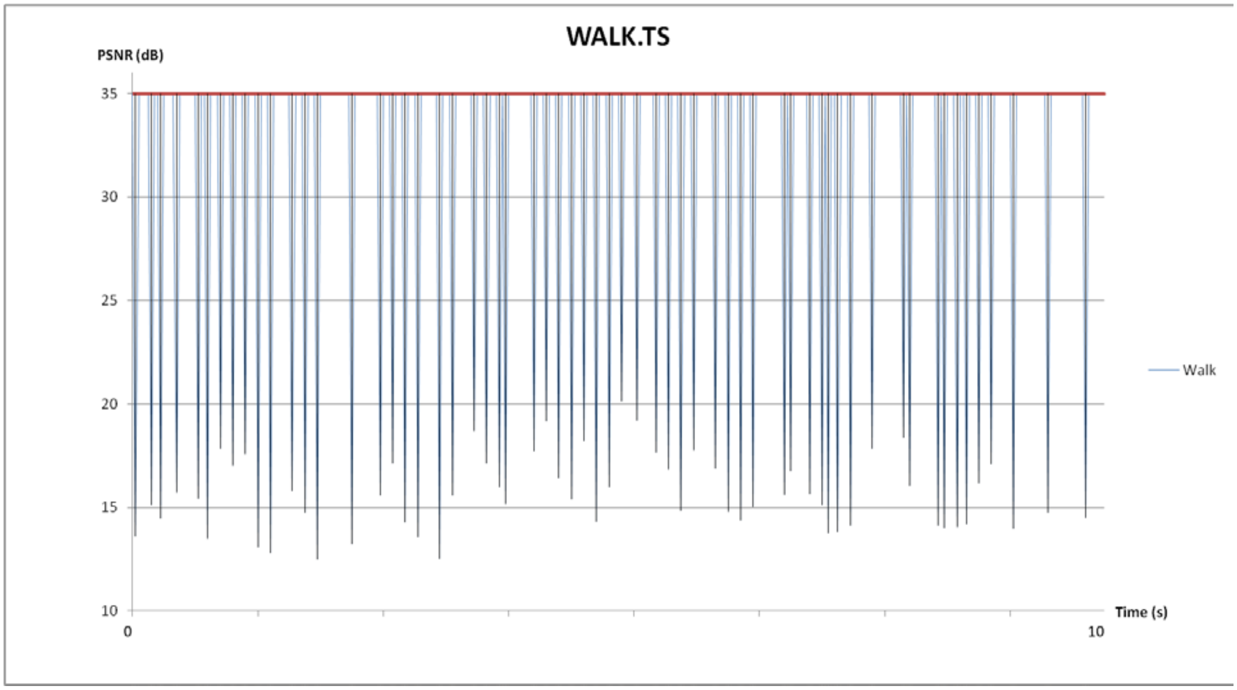


Figure 7.5.2: Delay 3 (PSNR) – Walk Video

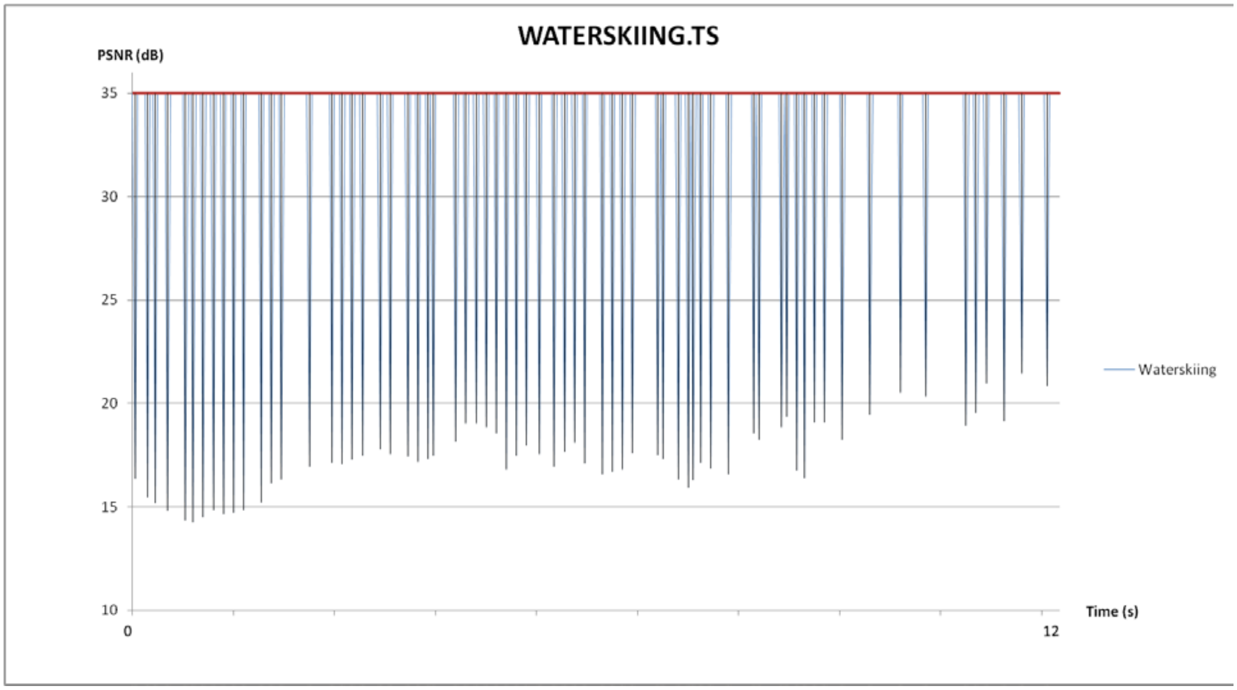


Figure 7.5.3: Delay 3 (PSNR) – Waterskiing Video

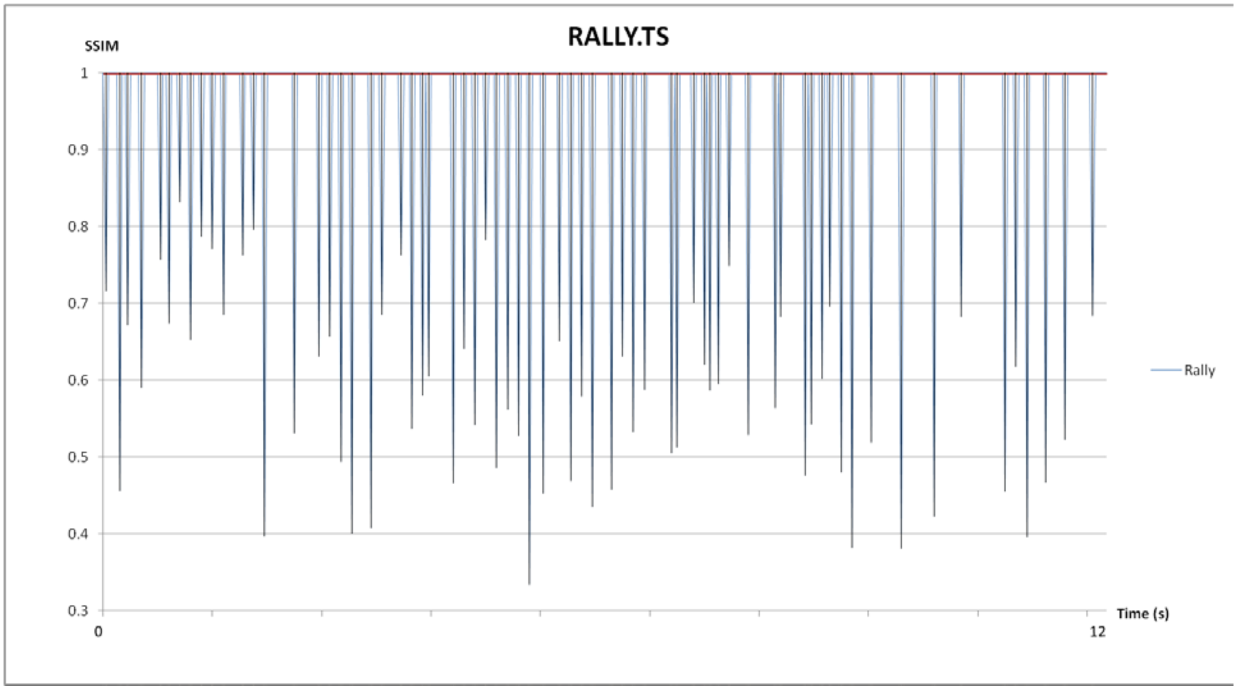


Figure 7.6: Delay 3 (SSIM) – Rally Video

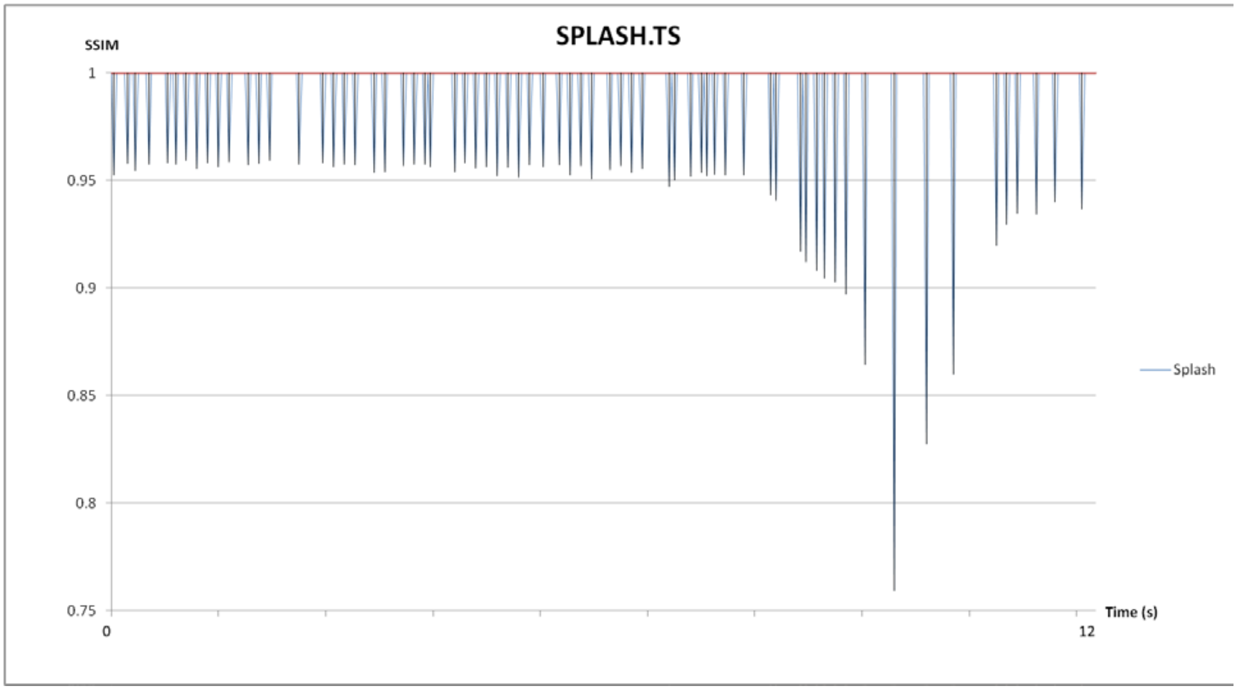


Figure 7.6.1: Delay 3 (SSIM) – Splash Video

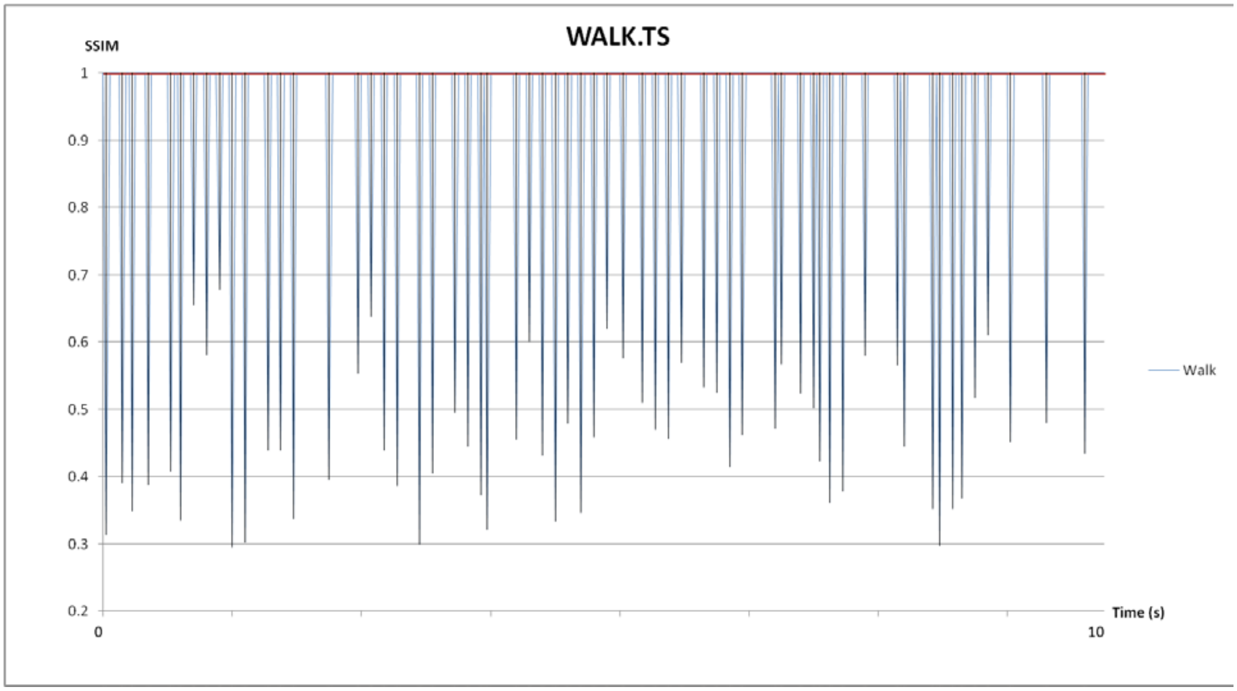


Figure 7.6.2: Delay 3 (SSIM) – Walk Video

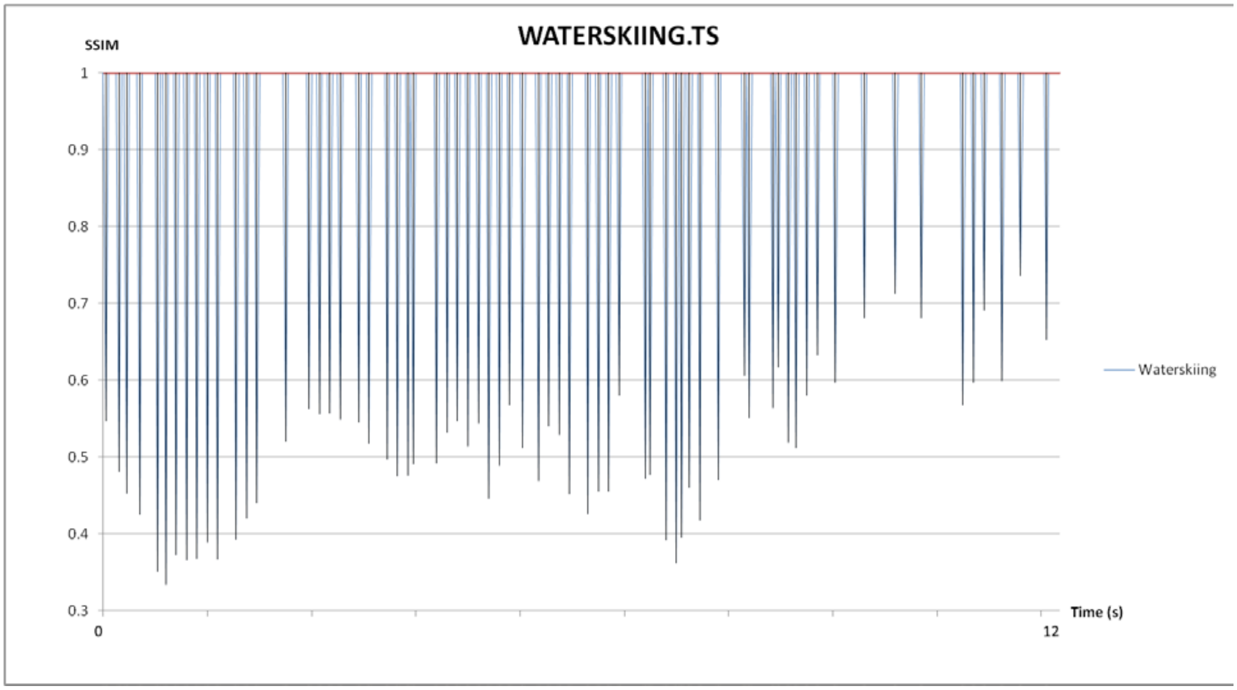


Figure 7.6.3: Delay 3 (SSIM) – Waterskiing Video