Master Thesis

Adaptation of a Perceptual Video Quality Measure to Low Bitrate Multimedia Applications

Srinath Loganathan

Chair of Multimedia Communications and Signal Processing
Prof. Dr.-Ing. A. Kaup

University of Erlangen–Nuremberg

Guide:
Dipl.-Ing. M. Barkowsky

April 2005

Acknowledgments

My sincere gratitude to Prof. Dr.-Ing. André Kaup, Head of the Department, Multimedia Communications and Signal Processing, Friedrich Alexander University-Erlangen, who gave me an opportunity to do a Master thesis under his supervision. I remarkably thank Dipl.-Ing. Marcus Barkowsky, who supervised me with great patience. I always remember his suggestions, which lead to the success of my thesis, as building blocks for my career. I convey my warm hearted thankfulness to OPTICOM, Erlangen, who made my stay in Germany a very pleasant one. I am greatly obliged for their concern towards my stay & financial aspects.

My special thanks for the course coordinator Dr. Nedelko Grbic, Blekinge Institute of technology (BTH), Sweden, who encouraged me with valuable guidelines to select a thesis work. I am delighted to express my thanks to Dr.Benny Lvstrom, Institute of Telecommunications and Signal Processing, BTH, with whom I discussed my academics in a friendly manner before I started my thesis. During the due course of time, I came to know that he is my internal supervisor for my master thesis. It's an immense pleasure to work with him.

Contents

List of Figures

Chapter 1

Introduction

In every engineering approach, technology tries either to match human perfection or to overcome human limitations. Based on this aspect, the present work tries to estimate the behavior of human visual system in the name of video quality. Appraisal of image quality in video and image processing systems plays a vital role in deciding the quality of service, network maintenance and even to compare different service providers. These systems have a wide range of applications from security services till entertainment which includes digital television, internet video, multimedia messaging services e.t.c,.

In 1997, Video quality experts came up with a video quality research project in order to develop and standardize the required technology for assessing the performance of digital video processing systems as Video Quality Experts Group (VQEG). Video quality assessment can be done in two different ways namely, objective and subjective. Subjective video quality assessment is the best one which is done by a group of viewers, who evaluates the quality based on source and degraded video sequences. Whereas, objective video quality measurement could be done without the help of end user. Objective video quality measurement is done by a software which processes the video signals to produce the quality of video. This mode to estimate the video quality is more advantageous as it could provide real time quality monitoring for video applications. One such objective quality measurement software, Perceptual Video Quality Measure (PVQM) is taken as a base in this work and adapted to achieve better video quality assessment for low bit rate video sequences.

Several research groups worked to develop an objective model to estimate processed video. In the development phase, different models for objective video quality estimation, were submitted to VQEG. One among those models is Perceptual Video Quality Measure (PVQM) which had the highest variance weighted regression correlation (see Chapter 3) when compared to other competing models. PVQM is an objective assessment model which generates the quality measure based on the perceptual feature which could affect the viewers quality estimation. The promising estimate of the video quality could be achieved by handling the perceptual features rather than the traditional methods like Peak Signal to Noise Ratio (PSNR), as the modern generation codecs are time varying, signal adaptive and hence they reflect the perceptual features of viewers.

This thesis is explained in 6 chapters. Chapter 2 analyses the parameters that are generated within the algorithm in the process of quality measure calculation. Chapter 3 explains, how the output of an objective model is pre-processed and compared with the outputs of other models. Chapter 4 deals with various issues and a method to generate temporal measure. This chapter explains the basic features of an algorithm and the preceeding adaptations done in it. Chapter 5 focuses on the impact of temporal measure generated based on the local subjective test scores. Final chapter discusses the future model that could be done to generate a perceptual temporal measure.

Chapter 2

Perceptual Video Quality Measure (PVQM)

2.1 Introduction

This thesis work began with the analysis of the given software PVQM. This chapter describes the software. A detailed description of PVQM can be found in [1]. PVQM is a full reference algorithm to estimate the quality of the coded video sequences. This estimation is achieved by processing the distortions, which viewers could perceive in the coded video sequences. PVQM processes the given pair of video sequences by employing the following steps sequentially.

- Alignment
- Modelling
- Cognitive modelling

Overview of involved steps

In the alignment phase, the software does the spatio-temporal and luminance chrominance alignment which paves the way for a better video quality. This phase is necessary to avoid the errors which could occur when comparing the original and the degraded images. As a first process, histogram matching is done and is explained in the later section. The spatio-temporal alignment is done to find the corresponding best matching original to all the distorted frames. The spatio-temporal alignment is done in two different phases namely, probing and matching phases. Probing phase is done in a wide range of original frames whereas, matching phase is done in correspondence to the result of the probing phase so as to cut down the search range. Finally, alignment phase gives out spatial and temporal information which were used in the extraction of perceptual measures for different features of the impaired sequences. In probing and matching phases the best matching original is found by using block matching technique. The three main features used in the extraction measures are edginess of the luminance, temporal decorrelation between present and previous frames and the third one is, the normalized color error. These three indicators are explained in the later part of the chapter. A linear cognitive model is finally used to give out the final video quality. The various parameters that are generated based on the difference between the original and the distorted frames are listed in Appendix-A with a block diagram. The edginess and the temporal parameters are computed in the luminance domain as, Human Visual System (HVS) is highly sensitive to the changes occuring in this domain.

2.2 Analysis of PVQM

The basic components of PVQM are described in the following steps.

Histogram Matching

Histogram matching is carried out to compensate the linear or nonlinear transformations that occurred in the Y, Cb, Cr components due to the digital processing systems. These transformations are mostly induced by the codecs to improve the perceived quality. To reduce this overall gain, luminance-chrominance values of the impaired sequence are aligned based on the histogram of the original sequence. A correction curve extracted from the cumulative histograms of the original and the distorted sequences is used to compensate this induced gain. Some of the typical problems which occurs in the implementation of histogram matching are discussed and explained in reference [1]. The original sequence (O) and the distorted sequence (D) are subjected to histogram matching in order to obtain the corrected distorted sequences (D_c). Based on this the following phases are computed.

Parameter based on Edginess

To estimate the edges in an image, a sharpness measure of the original (O_s) and the distorted (D_s) are computed by

$$D_s(i,j) = \sqrt{\left(\frac{D_c(i,j+2)+D_c(i,j+1)-D_c(i,j-1)-D_c(i,j-2)}{2}\right)^2 + (D_c(i+1,j)-D_c(i-1,j))^2} \tag{2.1}$$

$$O_s(i,j) = \sqrt{\left(\frac{O(i,j+2)+O(i,j+1)-O(i,j-1)-O(i,j-2)}{2}\right)^2 + (O(i+1,j)-O(i-1,j))^2} \tag{2.2}$$

In order to provide the importance to sharpness components, the process of dilation is performed. Dilation imposes weightage even to single pixel error in the sharpness component. Dilation is carried out in both original and distorted sequences. The dilated sharpness measure of the two sequences is obtained by:

$$D'_c(i,j) = \max\{D_s(i',j')|i' = i-1, i, i+1; j' = j-1, j, j+1\} \tag{2.3}$$

$$O'_c(i,j) = \max\{O_s(i',j')|i' = i-1, i, i+1; j' = j-1, j, j+1\} \tag{2.4}$$

Based on these two sequences, omitted sharpness (e) is computed. The omitted sharpness measure is the difference between the dilated corrected distorted image and the dilated original sharpness image normed by deviation. Omitted sharpness is calculated as,

$$e(i,j) = 100 \cdot \frac{D'_c(i,j) - O'(i,j)}{|O'(i,j)| + 80 + |d(i,j)|} \tag{2.5}$$

e(i,j) is clipped between -40 *and* 40

Deviation is a measure extracted to reflect the background illumination, which is usually grey, in the subjective test. It is calculated by

$$d = \max\{|D_c(i,j) - 100|, |O(i,j) - 100|\} \tag{2.6}$$

Upon the estimation of edges, PVQM generates different parameters by applying different Lebesque weights. In computing all the parameters, a window function [1] is implemented in order to give more weightage to the center part of the image than the periphery, as the distortions in former part will have more impact on the viewers quality estimation.

A set of parameters namely, E1, E2, E3, E4 are generated by using only the negative part of the sharpness measure. They are calculated as:

$$e_{neg} = |\min(e, 0)| \tag{2.7}$$

$$E_1 = \frac{\sum\limits_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right| \cdot |e_{neg}[i,j]|}{\sum\limits_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right|} \tag{2.8}$$

$$E_2 = \sqrt{\frac{\sum\limits_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right| \cdot |e_{neg}[i,j]|^2}{\sum\limits_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right|}} \tag{2.9}$$

$$E_3 = \sqrt[5]{\frac{\sum\limits_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right| \cdot |e_{neg}[i,j]|^5}{\sum\limits_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right|}} \tag{2.10}$$

$$E_4 = \sqrt[7]{\frac{\sum\limits_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right| \cdot |e_{neg}[i,j]|^7}{\sum\limits_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right|}} \tag{2.11}$$

Another set of parameters G is generated by using both the positive and the negative parts of the sharpness measure.

$$G_1 = \frac{\sum\limits_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right| \cdot |e[i,j]|}{\sum\limits_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right|} \tag{2.12}$$

$$G_2 = \sqrt{\frac{\sum\limits_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right| \cdot |e[i,j]|^2}{\sum\limits_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right|}} \tag{2.13}$$

$$G_3 = \sqrt[5]{\frac{\sum_{i,j} \left|\sin\frac{i\pi}{W} \cdot \sin\frac{2j\pi}{H}\right| \cdot |e[i,j]|^5}{\sum_{i,j} \left|\sin\frac{i\pi}{W} \cdot \sin\frac{2j\pi}{H}\right|}} \tag{2.14}$$

$$G_4 = \sqrt[7]{\frac{\sum_{i,j} \left|\sin\frac{i\pi}{W} \cdot \sin\frac{2j\pi}{H}\right| \cdot |e[i,j]|^7}{\sum_{i,j} \left|\sin\frac{i\pi}{W} \cdot \sin\frac{2j\pi}{H}\right|}} \tag{2.15}$$

Blockiness

Most of the codecs introduce blockiness artifacts within the degraded video sequence. An estimation for this blockiness in the corresponding sequence is obtained by:

$$B_{cd}(i,j) = \sqrt{\left(\frac{\sum_{x=-4}^{3} D_c(j+x,i) - D_c(j+x,i+1)}{8}\right)^2 + \left(\frac{\sum_{x=-4}^{3} D_c(j+1,i+x) - D_c(j,i+x)}{8}\right)^2} \tag{2.16}$$

$$B_o(i,j) = \sqrt{\left(\frac{\sum_{x=-4}^{3} O(j+x,i) - D_c(j+x,i+1)}{8}\right)^2 + \left(\frac{\sum_{x=-4}^{3} O(j+1,i+x) - D_c(j,i+x)}{8}\right)^2} \tag{2.17}$$

Then the introduced artifacts are extracted by norming this blockiness measure with deviation. The introduced artifacts i.e. the introduced sharpness measure is computed by:

$$I_s = \frac{100(B_{cd}(i,j)) - 2(B_o(i,j))}{|B_{cd}(i,j)| + 10 + |0.25(d(i,j))|} \tag{2.18}$$

I_s(i,j) is clipped to a maximum of 40

The positive part of the introduced sharpness is used to generate the parameters with different Lebesque weights. They are computed as:

$$I_s = |\max(I_s, 0)| \tag{2.19}$$

$$F_1 = \frac{\sum_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right| \cdot |I_s[i,j]|}{\sum_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right|} \tag{2.20}$$

$$F_2 = \sqrt{\frac{\sum_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right| \cdot |I_s[i,j]|^2}{\sum_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right|}} \tag{2.21}$$

$$F_3 = \sqrt[5]{\frac{\sum_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right| \cdot |I_s[i,j]|^5}{\sum_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right|}} \tag{2.22}$$

$$F_4 = \sqrt[7]{\frac{\sum_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right| \cdot |I_s[i,j]|^7}{\sum_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right|}} \tag{2.23}$$

Normalized Difference parameters

Apart from these sharpness measures and edginess measures there are few more parameters that are generated based on the normalized difference measure(L). It is computed based on the following equation:

$$L(i,j) = 100 \cdot \frac{(D_c(i,j) - O(i,j))}{80 + |0.7(d(i,j))|} \tag{2.24}$$

The same procedure is repeated for the generation of other parameters,

$$L_1 = \frac{\sum_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right| \cdot |L[i,j]|}{\sum_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right|} \tag{2.25}$$

$$L_2 = \sqrt{\frac{\sum_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right| \cdot |L[i,j]|^2}{\sum_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right|}} \tag{2.26}$$

$$L_3 = \sqrt[5]{\frac{\sum_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right| \cdot |L[i,j]|^5}{\sum_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right|}} \tag{2.27}$$

$$L_4 = \sqrt[7]{\frac{\sum_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right| \cdot |L[i,j]|^7}{\sum_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right|}} \tag{2.28}$$

Temporal measure

In quality assessment of video, apart from the spatial measure one of the most important issues is to exploit the temporal correlation between the consecutive pictures within the sequence. This temporal measure is generated as the decorrelation between the present and the previous images of the original sequence. Temporal measure generated here is to compensate the spatial measure and the edginess parameters. This measure is also the information perceived by the subject at the edges of an image as a factor of motion present in the sequence. Luminance being the most informative part of an image, decorrelation measure is also extracted using Luminance. PVQM algorithm first, generates measures which estimate the temporal correlation using the changes in pixel by pixel comparison between the present and the previous frames of the corrected distorted sequence and the original sequence. The change in pixel by pixel comparison is given by:

$$C_d = |D_c - D_{cp}| \tag{2.29}$$

$$C_o = |O - O_p| \tag{2.30}$$

and the difference between them can be formulated as,

$$C = (C_d - C_o) \tag{2.31}$$

The negative part and the positive part of this changed measure C are handled separately in the same manner to generate four parameters from either of those measures. The measures generated by using the positive parameters are:

$$C_{pos} = |\max(C, 0)| \tag{2.32}$$

$$I_4 = \left(\frac{\sum\limits_{i,j} \sqrt{|C_{pos}(i,j)|}}{(W \cdot H) + 1} \right)^2 \tag{2.33}$$

$$I_5 = \frac{\sum\limits_{i,j} |C_{pos}(i,j)|}{(W \cdot H) + 1} \tag{2.34}$$

$$I_6 = \sqrt{\frac{\sum\limits_{i,j} (|C_{pos}(i,j)|)^2}{(W \cdot H) + 1}} \tag{2.35}$$

$$I_7 = \sqrt[5]{\frac{\sum\limits_{i,j} (|C_{pos}(i,j)|)^5}{(W \cdot H) + 1}} \tag{2.36}$$

the measure generated by using the negative parameters are,

$$C_{neg} = |\min(100 \cdot C, 0)| \tag{2.37}$$

$$I_0 = \left(\frac{\sum\limits_{i,j} \sqrt{|C_{neg}(i,j)|}}{(W \cdot H) + 1} \right)^2 \tag{2.38}$$

$$I_1 = \frac{\sum\limits_{i,j} |C_{neg}(i,j)|}{(W \cdot H) + 1} \tag{2.39}$$

$$I_2 = \sqrt{\frac{\sum\limits_{i,j} (|C_{neg}(i,j)|)^2}{(W \cdot H) + 1}} \tag{2.40}$$

$$I_3 = \sqrt[5]{\frac{\sum\limits_{i,j} (|C_{neg}(i,j)|)^5}{(W \cdot H) + 1}} \tag{2.41}$$

Along with the previous measures some more measures are generated by using root mean square value of distorted and original sequence. The root mean square (RMS) is calculated by using the changes that occurred in the sequences. The RMS of the sequences can be formulated as:

$$Rms(C_d) = \sqrt{\frac{\sum\limits_{i,j} C_d^2(i,j)}{W \cdot H}} \tag{2.42}$$

$$Rms(C_o) = \sqrt{\frac{\sum\limits_{i,j} C_o^2(i,j)}{W \cdot H}} \tag{2.43}$$

A parameter generated by using RMS measures is calculated by

$$I_8 \;=\; (Rms(C_d) - Rms(C_O)) \cdot 100 \tag{2.44}$$

Correlations of the frames in the sequences are calculated by

$$Corr_o = \frac{\frac{\sum\limits_{i,j} O(i,j) \cdot O_p(i,j)}{W \cdot H}}{\sqrt{\frac{\sum\limits_{i,j} O^2(i,j)}{W \cdot H} \cdot \frac{\sum\limits_{i,j} O_p^2(i,j)}{W \cdot H}}} \tag{2.45}$$

$$Corr_d = \frac{\frac{\sum\limits_{i,j} D_c(i,j) \cdot D_{cp}(i,j)}{W \cdot H}}{\sqrt{\frac{\sum\limits_{i,j} D_c^2(i,j)}{W \cdot H} \cdot \frac{\sum\limits_{i,j} D_{cp}^2(i,j)}{W \cdot H}}} \tag{2.46}$$

Parameters generated by using correlation measure are calculated by:

$$I_9 \;=\; (1 - Corr_o) \cdot 100 \tag{2.47}$$

$$I_{10} \;=\; (1 - Corr_d) \cdot 100 \tag{2.48}$$

$$I_{11} \;=\; (I_{10} - I_9) \cdot 100 \tag{2.49}$$

Chrominance Measures

Video signals are a composition of chrominance (color portion) and luminance. The chrominance gives the color portion to be displayed. Higher chrominance levels produce more powerful or stronger colors. HVS is more sensitive to Cr measures than Cb. Also to reflect the fact that the HVS is less sensitive to errors in the color saturated areas, chrominance is normed by the saturation measure S(n,i,j) [1].

PVQM computes two kinds of measures L and G in chrominance domain. L is the one generated by the difference between the distorted and the original image and it is given by:

$$L(i,j) = \frac{100 \cdot (D_c(i,j) - O(i,j))}{25 + |0.3 \cdot S(n,i,j)|} \tag{2.50}$$

Parameters generated by using the above difference measure are:

$$L_1 = \frac{\sum_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right| \cdot |L[i,j]|}{\sum_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right|} \tag{2.51}$$

$$L_2 = \sqrt{\frac{\sum_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right| \cdot |L[i,j]|^2}{\sum_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right|}} \tag{2.52}$$

$$L_3 = \sqrt[5]{\frac{\sum_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right| \cdot |L[i,j]|^5}{\sum_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right|}} \tag{2.53}$$

$$L_4 = \sqrt[7]{\frac{\sum_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right| \cdot |L[i,j]|^7}{\sum_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right|}} \tag{2.54}$$

G measures are generated by the difference in the dilated corrected distorted sequence and the dilated original sequence. This difference can be found by:

$$S_{oc}(i,j) = \frac{100 \cdot (D'_c(i,j) - O'(i,j))}{|O'(i,j)| + 40 + |0.8 \cdot S(n,i,j)|} \tag{2.55}$$

Parameters generated by using the difference measure (S_{oc}) of chrominance are:

$$G_1 = \frac{\sum_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right| \cdot |S_{oc}[i,j]|}{\sum_{i,j} \left|\sin \frac{i\pi}{W} \cdot \sin \frac{2j\pi}{H}\right|} \tag{2.56}$$

$$G_2 = \sqrt{\frac{\sum_{i,j}\left|\sin\frac{i\pi}{W}\cdot\sin\frac{2j\pi}{H}\right|\cdot|S_{oc}[i,j]|^2}{\sum_{i,j}\left|\sin\frac{i\pi}{W}\cdot\sin\frac{2j\pi}{H}\right|}} \quad (2.57)$$

$$G_3 = \sqrt[5]{\frac{\sum_{i,j}\left|\sin\frac{i\pi}{W}\cdot\sin\frac{2j\pi}{H}\right|\cdot|S_{oc}[i,j]|^5}{\sum_{i,j}\left|\sin\frac{i\pi}{W}\cdot\sin\frac{2j\pi}{H}\right|}} \quad (2.58)$$

$$G_4 = \sqrt[7]{\frac{\sum_{i,j}\left|\sin\frac{i\pi}{W}\cdot\sin\frac{2j\pi}{H}\right|\cdot|S_{oc}[i,j]|^7}{\sum_{i,j}\left|\sin\frac{i\pi}{W}\cdot\sin\frac{2j\pi}{H}\right|}} \quad (2.59)$$

A traditional parameter PSNR is also computed by having the mean square error and is computed as

$$MSE = \frac{\sum_{i,j}(D_c(i,j) - O(i,j))^2}{W \cdot H} \quad (2.60)$$

$$PSNR = \begin{cases} 100 & \text{if } MSE \leq 0, \\ 20 \cdot log_{10}\frac{255}{\sqrt{MSE}} & \text{otherwise} \end{cases} \quad (2.61)$$

Chapter 3

Subjective and Objective Analysis

This chapter deals with the analysis of the objective scores obtained from the PVQM. The aim of every objective model is to match with the result of the subjective scores. To achieve this aim it is essential to analyze the behavior of every objective model for every kind of distortion in the distorted sequence and also to investigate the various features of the original sequence. Visualization of the analysis is done by making a scatter plot with the available data set. A scatter plot depicts the scores of the objective and the subjective test as shown in the figure 3.1. If the quality of the video sequence is not estimated properly by the objective model then they fall on the critical areas. There are two cases to achieve critical results. They can be classified as over estimation and under estimation. Both the cases are represented in the figure 3.2.

In addition, to have a deeper understanding of the behavior of the algorithm, the objective scores are also plotted with reference to the hypothetical reference circuit(HRC), degraded sequence and source sequence(SRC). The characteristics of algorithms for different features present in the original video is obtained by plotting the objective scores with respect to the SRC. Figure 3.3 shows the plot based on SRC. The plot based on HRC is shown in figure 3.4 where different artifacts generated by the digital processing system could be significantly noted. Some of the features of the original sequence are contrast, motion content, etc. Some of the artificats in the impaired sequences are tiling, blockiness, edge busyness, etc. This analysis also reveals the kind of parameters that have to be optimized based on the perceptual feature of

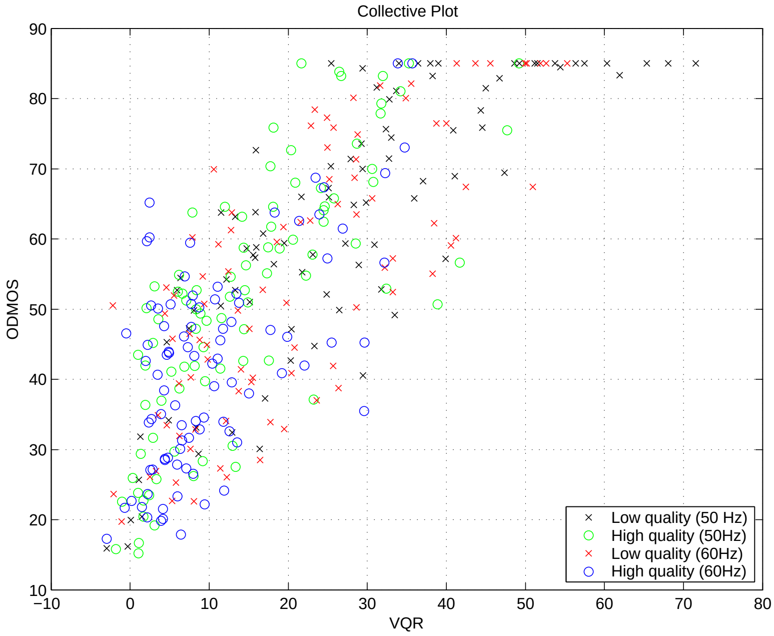

Figure 3.1: Scatter plot

HVS. Hence, the characteristic of PVQM is studied based on this analysis.

3.1 Variance Weighted Regression Correlation

VQEG being the standardization organization in the area of video quality, came up with the proposal to find the correlation, named as Variance Weighted Regression Correlation. In this text Variance Weighted Regression Correlation is abbreviated as VWRC for convenience. By finding the correlation, VQEG compared different competing models which were submitted to them. To perform the process of VWRC, VQEG test plan permitted a pre-processing on the results of each objective model. This pre-processing was released as metric-2. In this section, monotonic cubic regression used in metric-2 is discussed and a possible method to perform the regression is also explained.

Degree of agreement between the subjective and the objective scores

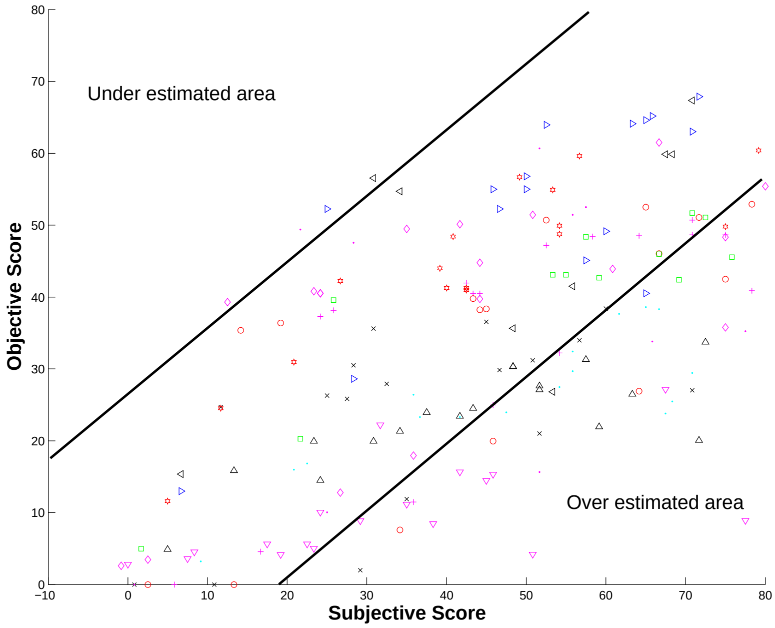

Figure 3.2: Over and Under estimated areas

should be found. Regression of Differential Mean Opinion Score (DMOS) against objective model might not adequately represent the relative degree of agreement of the subjective scores[2]. Therefore, VQEG came up with metric-2. By applying metric -2 to the objective data of a model, the predicted objective score of that model is found and is used for the calculation of the correlation coefficient. Metric-2 instructs that *"Recognizing the potential non-linear mapping of the objective model outputs to the subjective quality ratings, the objective test plan provided for fitting each proponent's model output with a non-linear function prior to computation of the correlation coefficients. As the nature of the non-linearities was not well known beforehand, it was decided that two different functional forms would be regressed for each model and the one with the best fit (in a least squares sense) would be used for that model. The functional forms used were a 3^{rd} order polynomial and a four-parameter logistic curve [3]. The regressions were performed with the constraint that the functions remain monotonic over the full range of the data. For the polynomial function, this constraint was implemented using the procedure outlined in reference [4]"*. Weighted regression

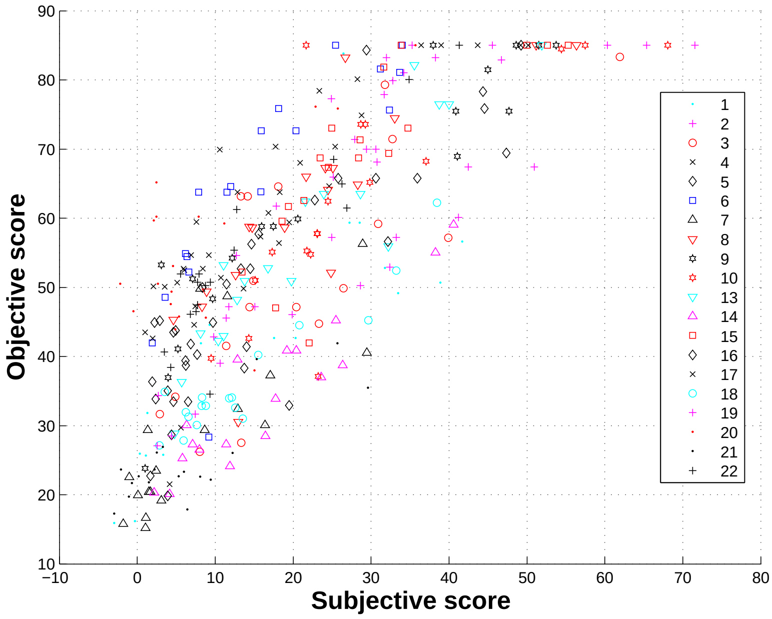

Figure 3.3: Scatter plot based on SRC

calculation is explained further.

3.1.1 Monotonic Cubic Regression (MCR)

The process of regression is done with the help of a cubic polynomial which is monotonic in nature and is known as Monotonic Cubic Regression. MCR was performed by the method suggested in [4]. A cubic polynomial is said to be monotonic, if it satisfies a non-linear inequality constraints as shown below:

$$x_1(x_2 + 2x_2 + 3x_3) \geq 0 \quad (3.1)$$

$$\sqrt{(x_2^2 + 3x_2 \cdot x_3)^2 + (2x_1 + 2x_2 + 3x_3)^2 \cdot (3x_1 \cdot x_3^2 - x_2^2 \cdot x_3)^2}$$

$$+x_2^2 + 3x_2x_3 + (2x_1 + 2x_2 + 3x_3) \cdot (3x_1 \cdot x_3^2 - x_2^2 \cdot x_3) \quad \geq \quad 0 \quad (3.2)$$

A third order polynomial which satisfies the non-linearity is used as the regression function to find the predicted DMOS based on the objective measure

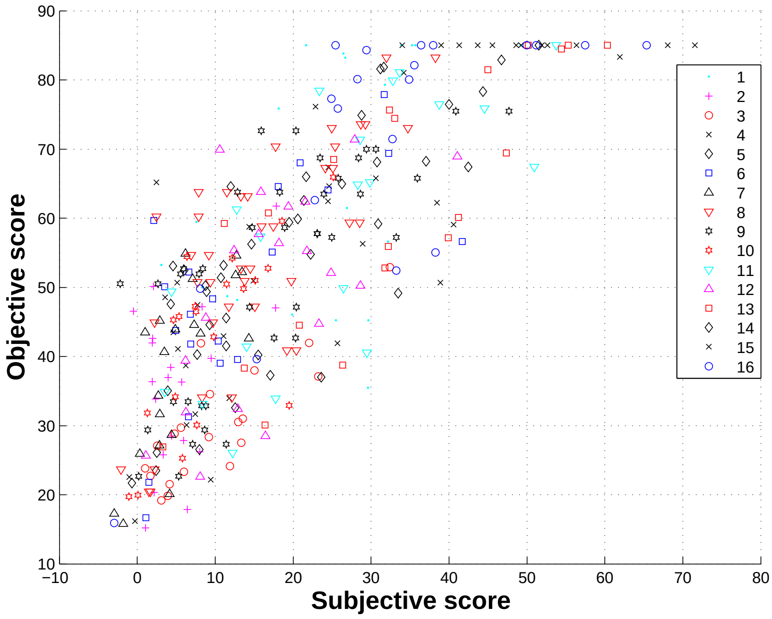

Figure 3.4: Scatter plot based on HRC

and the subjective DMOS. A cubic polynomial could be:

$$P(t) = x_0 + x_1 \cdot t + x_2 \cdot t^2 + x_3 \cdot t^3 \tag{3.3}$$

A typical problem in here is to determine the coefficients of the cubic polynomial in equation 3.3. The coefficients of the cubic polynomial could be found by reduction of the error function. The error function here is the difference between the predicted value and the corresponding original objective value. This could be formulated as:

$$e(i) = x_0 + x_1 \cdot t_i + x_2 \cdot t_i^2 + x_3 \cdot t_i^3 - y_i \tag{3.4}$$

The above equation 3.4 could be written in a least square sense as:

$$F(x_0, x_1, x_2, x_3) = \sum (x_0 + x_1 \cdot t_i + x_2 \cdot t_i^2 + x_3 \cdot t_i^3 - y_i)^2 \tag{3.5}$$

The optimal coefficients which have a minimum error can be found using MATLAB. MATLAB provides a function *fmincon* which performs the constrained optimization. This function accepts various kinds of inputs. In this work, three among those are used.

The inputs given to *fmincon*

- Initial guess of the coefficients
- Function which gives the least square error
- Function that computes the gradient of the non-linear constraint.

Based on these inputs, *fmincon* performs the optimization and produces the coefficients. An initial value of zero was chosen for each coefficients and were given to the function *fmincon*. The equation 3.5 was written in the function which returns the least square error for every value of predicted coefficients starting from the given initial values. The function which provides the gradient for every iteration was written using the equations 3.1 and 3.2. Later on, by using these coefficients, predicted DMOS is found and processed to find the correlation with the subjective scores. A fit for a given set of subjective and objective scores is shown in figure 3.5.

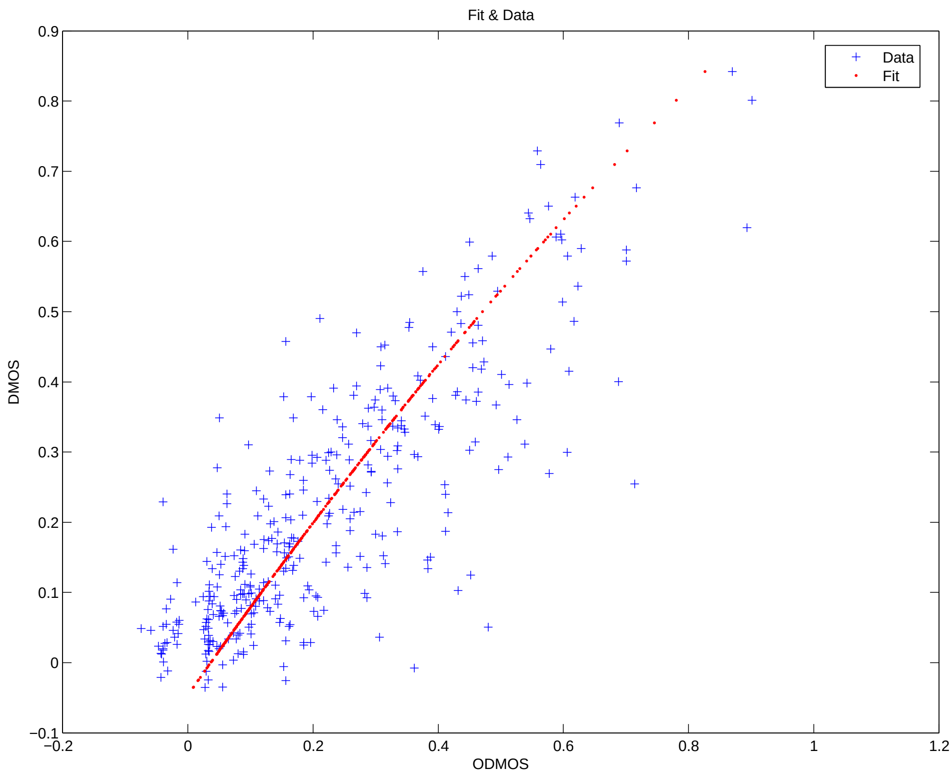

Figure 3.5: Data and Fit

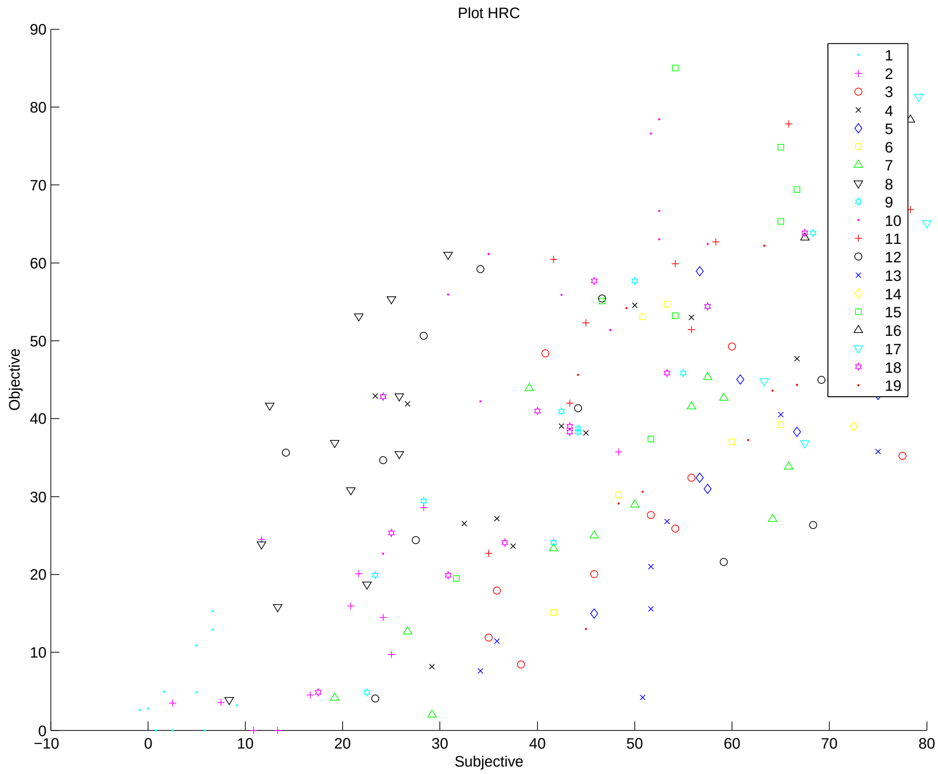

Figure 3.6: Scatter plot with Objective scores of CIF sequences

3.1.2 Adaptations for PVQM

The PVQM software initially processed television material. To handle low bit rate sequences, i.e. multimedia signals, PVQM is adapted to CIF and QCIF image sizes. Evaluation of CIF video sequences were done and verified across the results obtained from the locally conducted subjective tests. The scatter plot of objective and subjective scores for the multimedia material is shown in figure 3.6. The parameters used in the reference [1] are employed in the generation of the objective score depicted in the figure 3.6.

Parameters used in the final linear combinations were chosen by a tool OPTIMAP which selects the measures which have the higher degree of correlation with the subjective scores.

Chapter 4

Temporal Measure

The analysis of the PVQM and its behaviour brought few things to limelight. Analysis revealed that PVQM finds the best matching original frame for every frame in the impaired sequence. The distortion measures are extracted by finding the difference between the distorted frame and the corresponding matching original. Whereas, none of these measures concentrated in the time alignment between the frames present in the distorted sequences. Also, the present temporal decorrelation measure is entirely dependent on the original sequence. In addition to this, PVQM is designed to evaluate TV materials which are of good quality when, compared to multimedia signals. Whereas, multimedia signals are the one which are subjected to higher level of temporal distortions. The above issues paved way for a new temporal measure.

To estimate the temporal distortions in an impaired sequence an algorithm is taken as reference [5]. This reference algorithm is explained in the former part of this chapter and the implemented algorithm in the latter part.

4.1 Reference Algorithm

For every frame a corresponding temporal indicator is generated and used for further computations. The temporal indicator is generated by finding out the pixel by pixel difference between the present and the previous frame in the luminance domain. Later on, standard deviation (SD) is computed for the resulting difference image. This process is carried out separately for the original and the distorted one. The measure for the two sequences are plotted across the corresponding frame as shown in figure 4.1

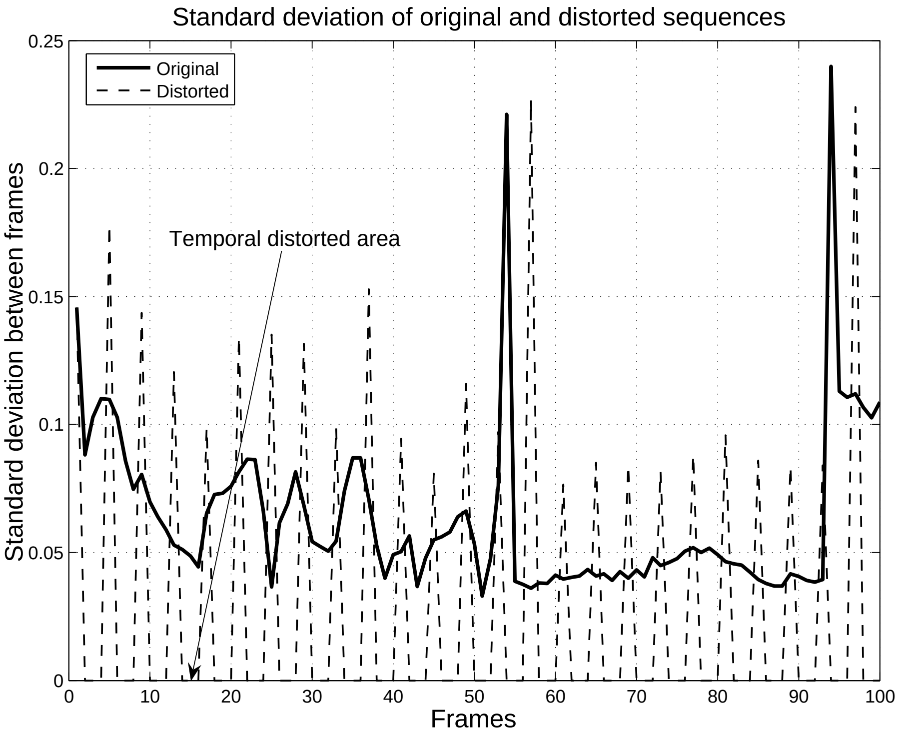

Figure 4.1: Standard deviation between frames

4.1.1 Characteristics of the Plotted Curve

The original sequence is the one which has no repeated frames and hence, results with at least a minimum measurable difference in the temporal indicator for every frame. Whereas, the standard deviation is zero in the case of frame repetition. This characteristic behavior is reflected in the curve with a variation in every frame. Unlike the original sequence, distorted sequence is subjected to various temporal distortions like frame repeat, frame skips, etc. These are clearly noticeable either with very high standard deviation or with zero standard deviation in the places where the distortions have occurred.

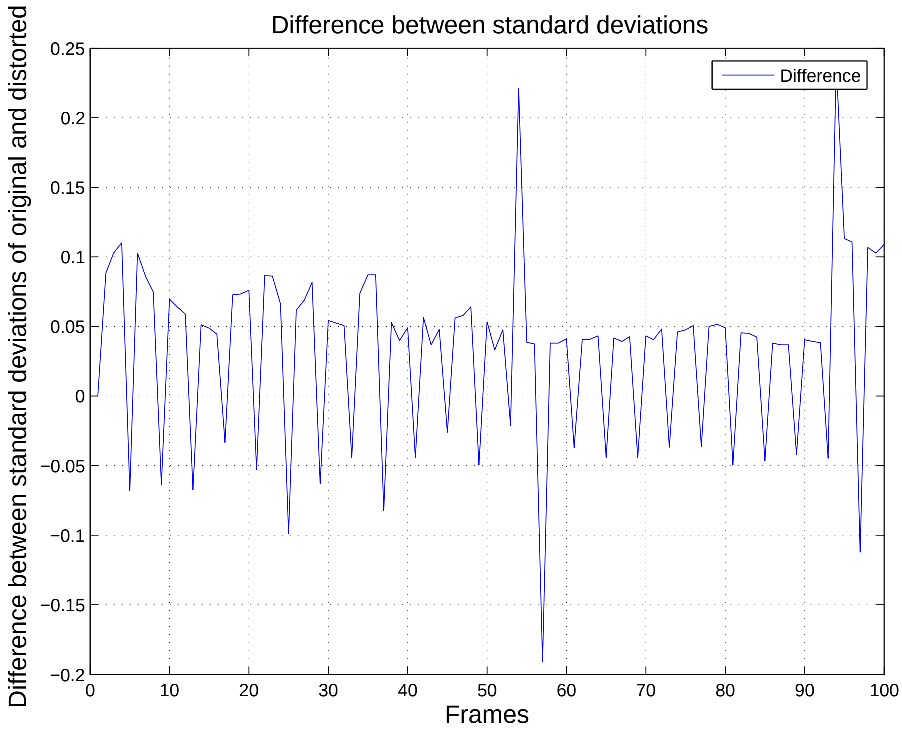

Figure 4.2: Difference between standard deviations

4.2 Temporal Score

The difference in the temporal indicator measures are found, frame by frame, between the original and the distorted sequences. This difference is shown in figure 4.2. The authors of the reference [5] extracts two kinds of parameters from the obtained difference between the standard deviations of the original and the distorted sequences. One with the positive part of the difference and the other with the negative part. The positive and the negative part of the difference are depicted in figure 4.3 and figure 4.4 respectively. In the positive part of the difference, the curve with the largest area is found and this area is multiplied with the width of the respective curve. The assumptions made in the above-mentioned algorithm is that 'the frame next to the temporal distortion is the frame to be displayed at that time stamp'. Whereas, in practical situations, different kinds of possibilities might occur. They are discussed in appendix-B.

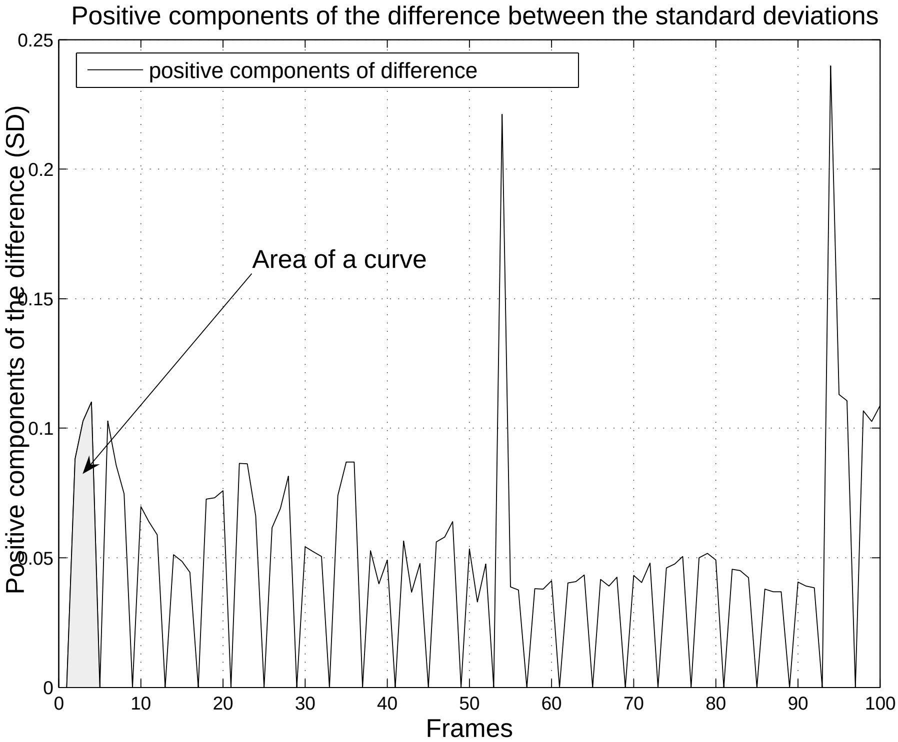

Figure 4.3: Positive part of the difference between standard deviations

4.2.1 Adaptations

The aim of the adaptation is to generate a temporal measure which considers the conditions given in appendix-B. The adaptations were made in such a way that the temporal measure increases with the increase in the amount of temporal distortions. The idea of finding the area and multiplying that with the corresponding frame repeat is used in a different way. The adaptations are explained using the figure 4.5. The figures 4.5a and 4.5b depict the temporal indicator of an original sequence and the corresponding distorted sequence. The frame A in the figure 4.5b is the frame which precedes the temporal distortion. The matching original frame C for the frame A is found. Similarly, a matching original D is found for the frame B, the frame which follows the temporal distortions. The corresponding area for the temporal distortion that occurred between the frames A and B is found as the next

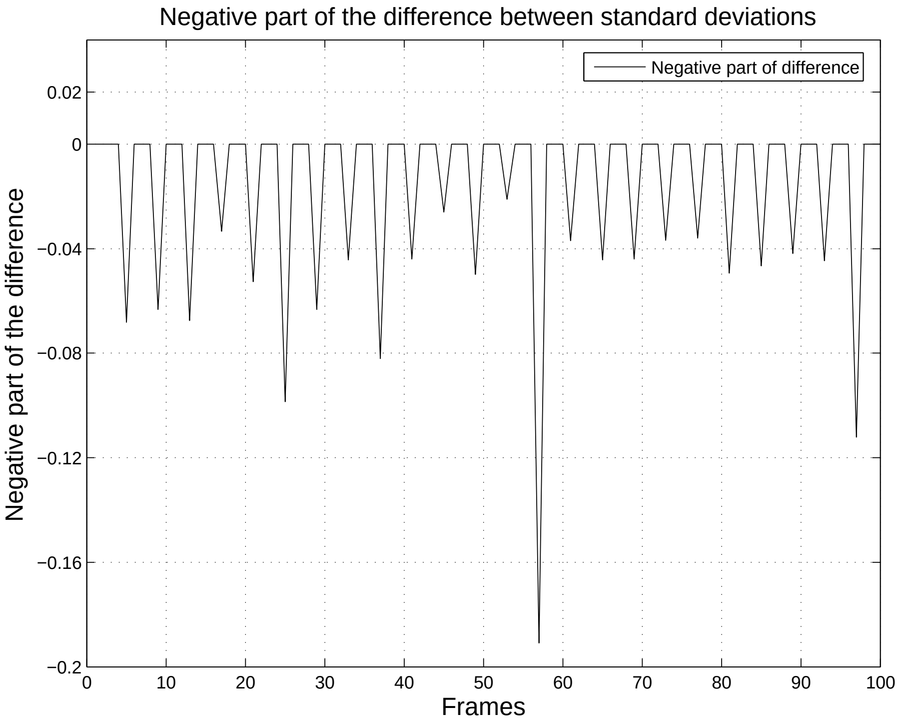

Figure 4.4: Negative part of the difference between standard deviations

step. The area under the curve formed by the matching originals C and D is computed. This area is then multiplied by the number of frame repeats occurred between the frames A and B in the impaired sequence. In case of no frame repeats a value of 1 is multiplied with the area. This procedure is applied to all the instances of temporal distortions that occurred within the impaired sequence. Finally, by choosing the highest of all the temporal scores, a temporal measure for the given impaired sequence is extracted.

This paragraph discusses the logic behind the mathematical calculations involved in the generations of the temporal measure. In case of a temporal distortion, there may be a frame skip, frame repeat or frame freeze. The viewer gets annoyed by any of these distortions stated above. The level of annoyance varies depending upon the amount of motion present in the sequence and the number of frames involved in the temporal distortion. The area of the curve is computed to have an estimate of change that a viewer

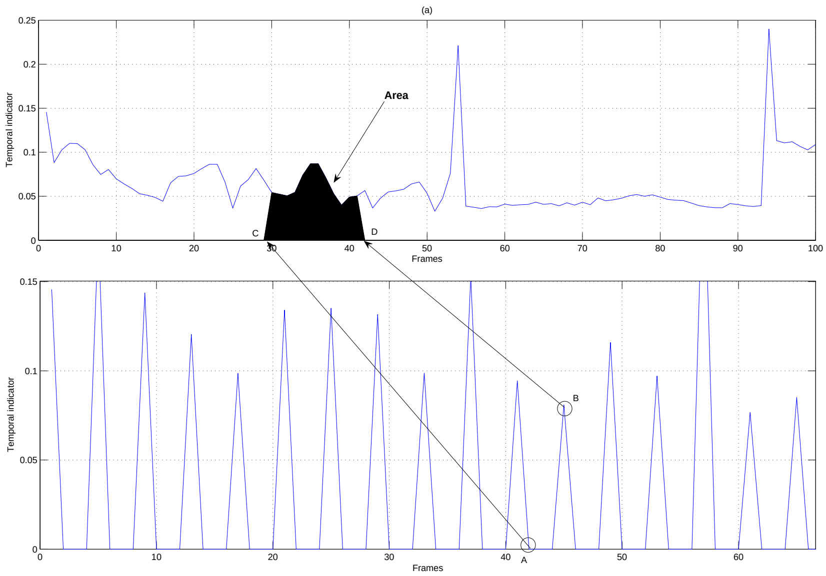

Figure 4.5: Temporal indicator of an original sequence and its corresponding distorted sequence

experiences after the temporal distortion. The multiplication is to consider the amount of annoyance caused by frame repeat or frame freeze. The highest among the temporal scores generated for a sequence is calculated to have the estimate of the most annoying part in the impaired sequence.

To analyze the behavior of this model, sequences with known temporal distortions are evaluated. The nature of temporal distortions are described in appendix-C. The figures 4.6 and 4.7 shows the behavior of the temporal measure for different kinds of temporal distortions. From those figures it is evident that the temporal score increases with the increase in the distortion.

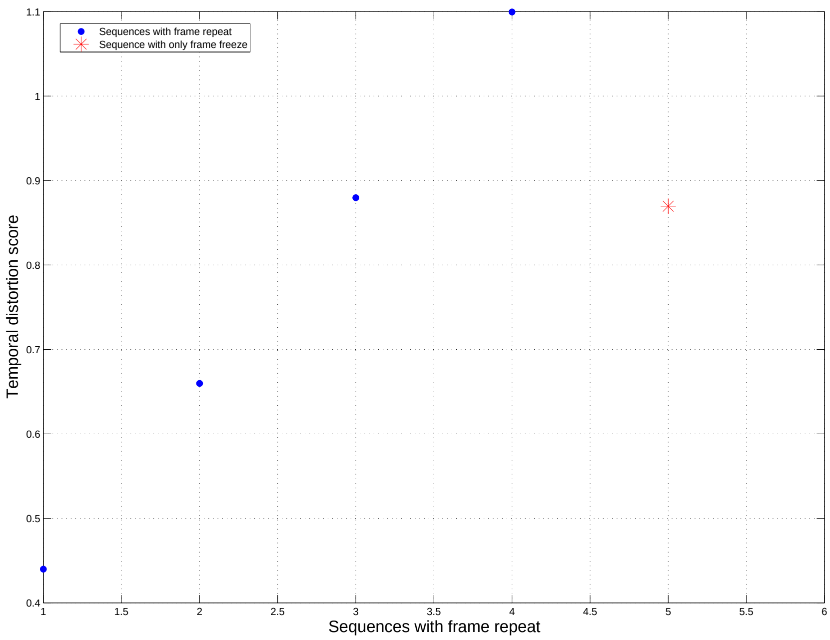

Figure 4.6: Impaired sequences with frame repeat

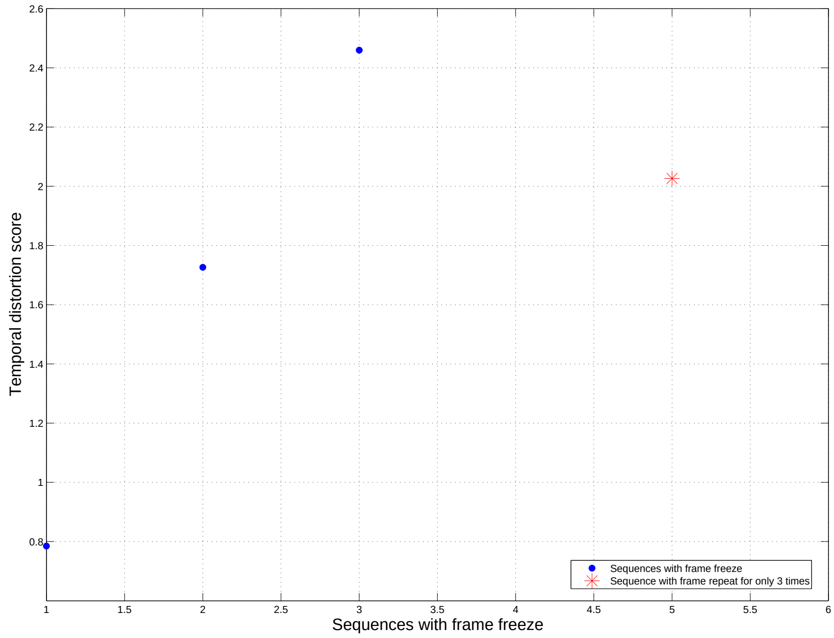

Figure 4.7: Impaired sequences with frame freeze

Chapter 5

Results and conclusion

5.1 Temporal score Results

The results and the conclusion of the work done in this master thesis is dealt in this chapter.The goodness of the new temporal measure is achieved by the comparison of the temporal measure with the results of the subjective test. The results of the sequences with temporal distortions were chosen for the comparison. Temporal effect is one among the other features which contribute to the video quality. Hence, the temporal measure alone cannot result in a higher correlation with the subjective scores. The temporal measure generated here is combined with the other parameters to obtain the final objective scores. The objective score is now a combination of the parameters used in the reference[1]. This objective score plotted across the respective subjective scores for the temporally impaired sequences is shown in figure 5.1. The figure 5.1 also depicts the fit for the plotted data set. The fit provides the optimal curve of the agreement between the parameters on its respective coordinates. In this case the diagonal fit shows a reasonable agreement between the objective and the subjective scores. Computing the correlation between the data set gives a better understanding about the level of agreement in numerical values. The data set in the figure 5.1 showed a Pearson correlation of 0.64. Owing to the fact that a mere addition of temporal measure will not result in the improvement of correlation a linear combination is performed. To perform this linear combination the temporal measures are first compared with the difference between the objective and the subjective score as depicted in the figure 5.1. This comparison is shown in the figure

5.2. The linear fit of the data in figure 5.2 gives the required information for the linear combination of objective scores and the temporal measure. Figure 5.3 shows the data set and its fit for the new objective and the subjective scores. Though the figures 5.1 and 5.3 appear to be similar they result in different correlations. The addition of temporal measure achieves a pearson's correlation of 0.70. The video quality is a measure which rely on HVS. HVS is a system with a nonlinear characteristics. Thus, any distortion beyond a level of threshold will produce a negligible amount of variation in the video quality. To adapt this behaviour of HVS, a simple nonlinear treatment of taking square root is implied to the temporal measure. The data set using the temporal measure after nonlinear treatment is shown in figure 5.4. This further improved the correlation to 0.75.

5.2 Conclusion

The purpose of this thesis is to adapt the PVQM to low bitrate multimedia applications. As the first step the given software, which evaluated the TV material, is analyzed. The parameter generated to achieve the final quality is formulated. The impact of important parameters is analyzed. Much time was not spent in investigating the nature of the parameters as the PVQM is developed based on a published paper [1]. The work then continued with the adaptation to low bit rate video sequences. These analyzations lead the way to a temporal measure. Thesis work proceeded in generating the temporal measure which concentrates only in the temporal distortions. A suitable algorithm [5] is adapted to generate the temporal measure. These adaptations were made to consider the real world temporal distortions. Finally, the result part introduces a possible way to combine the temporal measure which could result in higher correlation with the subjective scores. The ideas which were thought at different instances during this thesis is summed up at the end of this thesis.

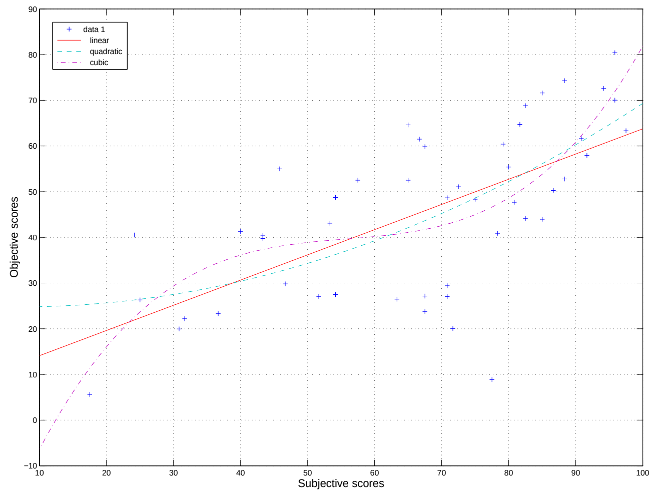

Figure 5.1: Objective and subjective scores for temporally distorted sequences

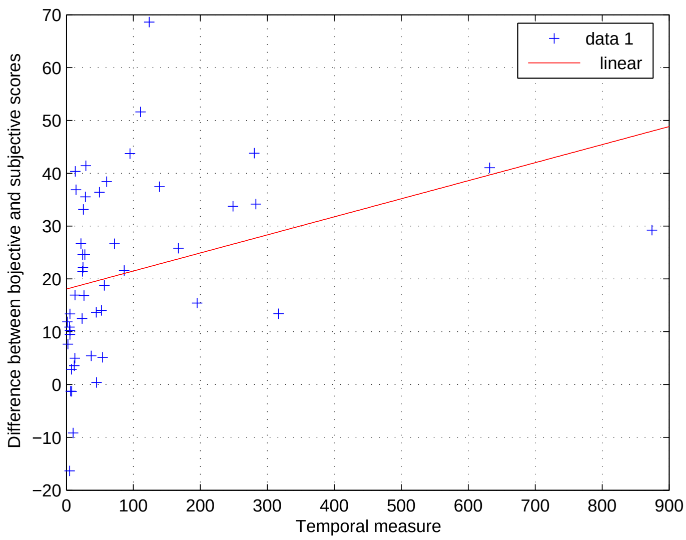

Figure 5.2: Temporal measure and difference between objective and subjective scores

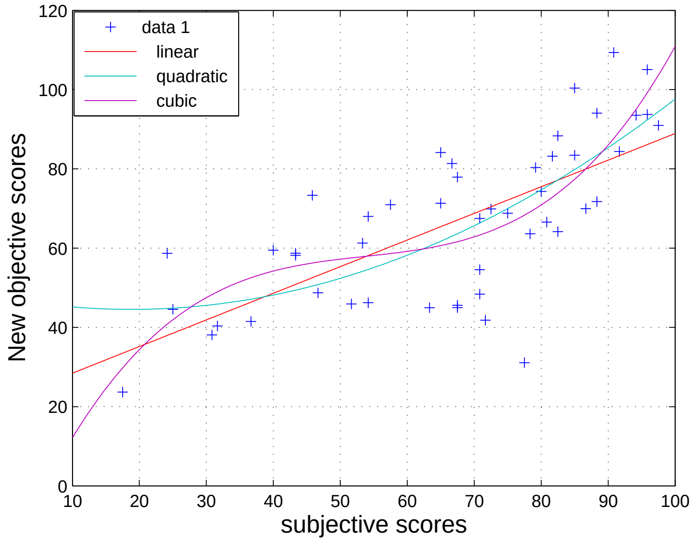

Figure 5.3: Objective scores including temporal measures

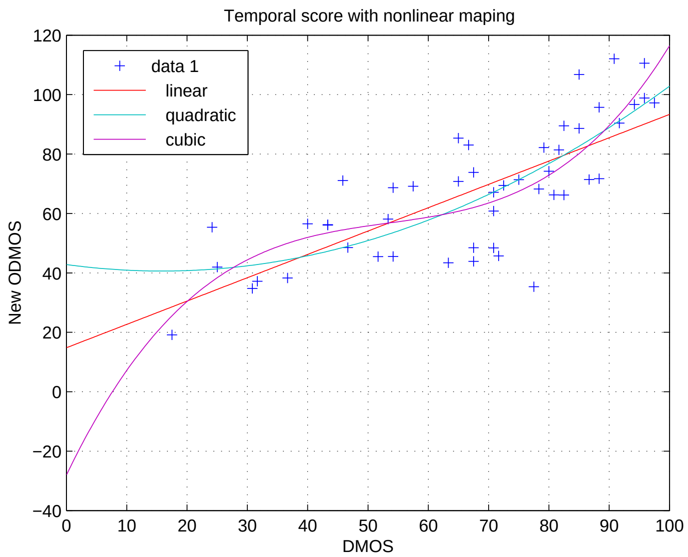

Figure 5.4: Objective scores with nonlinearly mapped temporal measure

Chapter 6

Future Proceedings

During the tenure of this master thesis, apart from the temporal measure described in the previous chapters, at different points of time different kinds of ideas were tried. One such model which was found worthy is described in this chapter. This model is not implemented in this thesis whereas it is suggested for future proceedings.

In the initial stages of developing a temporal model, thesis work concentrated in finding a temporal score based on frame rate. From reference [6] it is evident that the nature of the original sequence plays an important role in deciding the required framerate. Also, the result from reference [6] shows the need for higher framerate for the sports sequences. The sport sequences are the one with more motion content. There are many algorithms in the recent developments to have an estimate of motion content. To restrict the work towards the generation of temporal measure, much time was not spent on motion estimation algorithm. The frame rate not only depends on the amount of motion content, but also on the perceived motion of the subject. The *Department of Psychology, Stanford University* published a paper on *Perceived speed of colored stimuli* [7]. Based on reference [7] the color content in the image, could influence the perceived motion, which in turn shows its impact on the frame rate. The proposed model is represented as a block diagram in figure 6.1. Though this is an abstract model with small amount of information about the different blocks, it could give a lead to the person who wishes to concentrate in this area.

In the above model, color estimate could be done either by edge based or by area based algorithm. Though HVS is sensitive to changes occuring within the area in the color domain, contrast at the edges also contributes

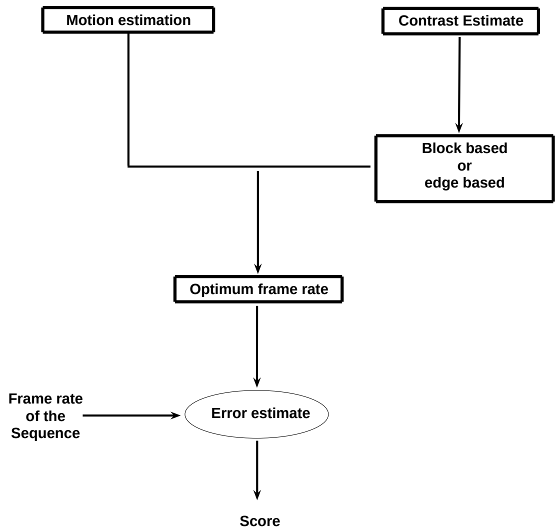

Figure 6.1: Block diagram of proposed model

in perceived motion. The framerate of an impaired video could be found in many ways. A method to find the frame rate of the degraded video sequence is described in this chapter and has been implemented within the algorithm. The frames which are repeated within the impaired sequences are found. A histogram is obtained for the number of times each frame has been repeated. Then by considering a threshold of 90%, the number of repeats that has occurred often is found. Furthermore, a weighted mean of the obtained number of repetitions above threshold of 90% is calculated. Finally, the framerate of the impaired sequence is obtained by dividing the framerate of the original by this weighted mean. Thus, by comparing the framerate of the impaired sequence and the required framerate, an error estimate could be achieved. The temporal measure could be generated using this error estimation.

Appendix A

Block diagram of the generated parameters

There are different kinds of parameters generated in PVQM to achieve the final video quality estimate. This block diagram gives an abstract view of the parameters generated.

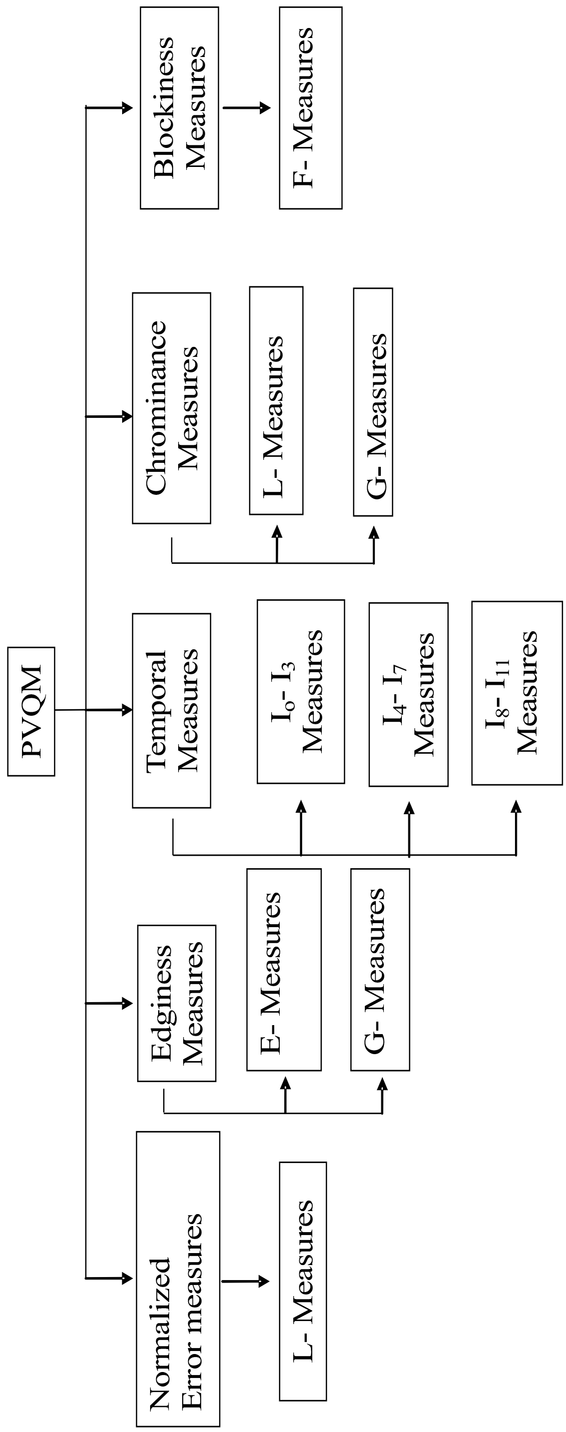

Figure A.1: Measures in PVQM

Appendix B

Real time temporal distortions

The nature of few among the different temporal distortions which occur in the real time are discussed in this session. The figure B.1 shows the frame alignment of an original sequence. When a sequence is subjected to temporal distortions this alignment of the frames shown in figure B.1 is disturbed. The nature of the distortions are depicted in the other figures B.2 to B.4. The main three kind of temporal distortions are

- Frame skip
- Frame repeat
- Frame freeze

Frame skip is depicted in the figure B.2. Here the frames after the 5^{th} frame are skipped and the latter frames follow the 5^{th} frame. In case of this distortion, a viewer experiences a jerk in the flow of the video. This jerk depends upon the number of frames skipped. Frame repeat is shown in the figure B.3. Viewers experience a slow sequence, when compared to the original sequence. Frame freeze is shown in the figure B.4. In this case, a single frame is repeated for many times within the distorted sequence. The viewer experience no movement in the video until a new frame appears. The movement that a viewer can see after a frame freeze depends on the frame, which follows the frame freeze.

1	2	3	4	5	6	7	8	9	10	11	12	13

Figure B.1: Frame alignment of an original sequence

1	2	3	4	5	9	10	11	12	13

Figure B.2: Frame alignment of a sequence with frame skip

1	1	2	2	3	4	4	5	6	7	7	8	9

Figure B.3: Frame alignment of a sequence with frame repeat

1	2	3	4	4	4	4	4	4	4	4	4	13

Figure B.4: Frame alignment of a sequence with frame freeze

Appendix C

Nature of the test sequences

Two kinds of temporal distortions are employed in the comparison of the temporal measure generated in this thesis. The nature of the sequences are shown in the figures C.1 and C.2. Few more sequences from either of the distortions shown in the figures were generated to investigate the nature of the temporal measure.

Nature of sequences based on frame repeat :

- Sequence 1 : All the frames are repeated for 1 time (figure C.1)
- Sequence 2 : All the frames are repeated for 2 times
- Sequence 3 : All the frames are repeated for 3 times
- Sequence 4 : All the frames are repeated for 4 times
- Sequence 5 : Repetition of a frame for many number of times

Nature of sequences based on frame freeze :

- Sequence 1 : All the frames are repeated for 1 time (figure C.2)
- Sequence 2 : All the frames are repeated for 2 times
- Sequence 3 : All the frames are repeated for 3 times
- Sequence 5 : Frame repetition of 4 for 3 times

The results of the comparisons of temporal measure for each of the above sequences are shown in the chapter 4.

1	1	2	2	3	3	4	4	5	5	6	6	7

Figure C.1: Sequence with frame repeat

1	1	3	3	5	5	7	7	9	9	11	11	13

Figure C.2: Sequence with frame freeze

Bibliography

[1] A.P. Hekstra, J.G. Beerends, et.al.: *PVQM - A perceptual video quality measure,* Signal processing: Image communication, 17 (2002) 781-798

[2] VQEG: *Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment* Document: COM 9-80-E, March 2000

[3] ITU-R Recommendation BT.500-8: *Methodology for the subjective assessment of the quality of television pictures,* September 1998.

[4] C. Fenimore, J. Libert and M.H. Brill: *Monotonic cubic regression using standard software for constrained optimization*, NIST/ITS internal report, November 1999.

[5] Arthur Webster, Steve Wolf: *Spatial and Temporal Measures for Video Quality*, ANSI contribution number T1Q1.5/92-113, January 22, 1992

[6] Patrick Brun, Gert Hauske, and Thomas Stockhammer: *Subjective Assesment of H.264/AVC Video for Low-Bitrate Multimedia Messaging Services* IEEE International Conference on Image Processing 2004, Singapore, October 2004

[7] Robert F. Dougherty, William A. Press, and Brian A. Wandell: *Perceived Speed of Coloured Stimuli*, Neuron, Vol. 24, 893-899, December, 1999.

[8] E. Bruce Goldstein: *Sensation and Perception*, Wadsworth Publishing Company, 6 edition (August 2001), ISBN: 0534539645

[9] Hultquist, Paul F: *Numerical Methods for Engineers and Computer Scients*, Publisher: Benjamin-Cummings Pub Co (February 1, 1988), ISBN: 080534652X

[10] Video Quality Experts Group, homepage: www.vqeg.org

[11] Website about statistics, homepage: http://davidmlane.com/hyperstat/

[12] Visual Science, homepage: http://cvs.anu.edu.au/

[13] Color conversion, homepage: www.easyrgb.com/math.html

[14] Perceived motion, homepage: www.cis.rit.edu/people/faculty/montag/