# VIDEO QUALITY ASSESSMENT USING SUBJECTIVE AND OBJECTIVE METRICS

RAHUL GAURAV

B.Tech., Bachelor of Technology, India, 2009

Submitted to the Graduate School of Science and Engineering

in partial fulfillment of the requirements for the degree of

Master of Science

in

Electronics Engineering

ISIK UNIVERSITY

2013

ISIK UNIVERSITY

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

VIDEO QUALITY ASSESSMENT USING SUBJECTIVE AND OBJECTIVE METRICS

RAHUL GAURAV

APPROVED BY:

Assoc. Prof. Hasan F. Ates, Isık University ______________________

(Thesis Supervisor)

Prof. Ercan Solak , Isık University ______________________

Assist. Prof. Umit Guz Isık University ______________________

APPROVAL DATE: ..../..../....

# VIDEO QUALITY ASSESSMENT USING SUBJECTIVE AND OBJECTIVE METRICS

## Abstract

Since past few years, quality assessment of digital videos has acquired a lot of attention in the video processing community, leading to the growth of different 2D and 3D objective metrics for images and videos. Some of these metrics correlate quite well with the Human Visual System (HVS), while some do not. Most of the popular objective metrics do not perfectly correlate with the HVS. This correlation can be measured by the opinion scores based on observations of human subjects.

In this thesis, we extensively study about Video Quality Assessment and complexity of HVS by proposing a method which helps us to find the best correlation between HVS and the objective metrics. The motive behind this work is to to introduce an objective metric that is adequate to predict the MOS (Mean Opinion Score) of distorted video sequences based on the FR (Full Reference) method, keeping in mind the benchmark set by the Video Quality Expert Group (VQEG). Subjective quality evaluation experiments using the human observers were performed, that is, a group of non-experts viewing the video sequences in original and distorted forms, as it is the human perception method which is evidently the most reliable one. Thereafter, objective test results, including the introduced metric, called PQM (Perceptual Quality Metric), and several other state-of-art metrics, are compared with MOS provided by human subjects. These comparisons elucidate the level of correlation between tested objective metrics and MOS. As an ultimate objective, we show the accuracy and monotonicity of various quality metrics, including PQM. The results obtained show that the PQM is better correlated with the human subjective judgement.

# ÖZNEL VE NESNEL METRİKLER KULLANARAK VİDEO KALİTE DEĞERLENDİRMESİ


## Özet

Son birkaç yılda dijital videoların kalite değerlendirmesi video işleme araştırmacılarının çokça ilgisini çekmiş, ve bu da video ve resimler için farklı 2-boyutlu ve 3-boyutlu nesnel metriklerin geliştirilmesine yol açmıştır. Bu metriklerden bazıları İnsan Görsel Sistemi (İGS) ile oldukça ilintili iken, bazıları ise değildir. Aslında birçok popüler nesnel metrik İGS ile yüksek bir ilintiye sahip değildir. Bu ilintinin seviyesi, insan deneklerin gözlemlerine dayanan fikir skorları ile ölçülebilir.

Bu tezde, Video Kalite Değerlendirmesi ve İGS'nin karmaşıklığı üzerinde durulmuş ve İGS ile eniyi ilintiyi verecek bir nesnel metrik hesaplama yöntemi önerilmiştir. Bu çalışmanın arkasındaki esas amaç bozulmuş videoların OFS (Ortalama Fikir Skoru) değerlerini kestirebilecek bir nesnel metrik geliştirilmesidir. OFS değerleri TR (Tam Referanslı) karşılaştırma yöntemine göre Video Quality Expert Group (VQEG) tarafından ortaya konan test şartlarına uygun olarak belirlenmiştir. Bu öznel kalite değerlendirme deneyleri uzman olmayan bir grup deneğe video dizileri orijinal ve bozulmuş durumda gösterilerek gerçeklenmiştir. Böylelikle görsel video kalitesini en güvenilir şekilde ölçen OFS değerleri bulunmuştur. Daha sonra önerilen AKM (Algısal Kalite Metriği) ve diğer en gelişmiş nesnel metrikler insan deneklerinden elde edilen OFS değerleri ile karşılaştırılmıştır. Bu karşılaştırmalar sonucu test edilen nesnel metriklerle OFS arasındaki ilinti seviyeleri ortaya konmuştur. Böylelikle AKM de dahil farklı metriklerin doğruluk ve tekdüzelilik seviyeleri ölçülmüştür. Elde edilen sonuçlar AKM'nin insan öznel değerlendirmeleri ile daha yüksek ilintiye sahip olduğunu göstermektedir.

## Acknowledgments

First and foremost, I would like to express my sincere gratitude to my advisor Dr Hasan Fehmi Ates. I am grateful that he shared his immense knowledge and great ideas for improving my research skills and diligence. His guidance and support helped me throughout my graduate study and writing of this thesis. I have been amazingly fortunate to have an advisor who gave me the freedom to explore on my own, and at the same time the guidance to recover when my steps faltered. I should also thank him not only for the time and attention he dedicated to my work but the time he spent during these years listening patiently to me. Undoubtedly, I could not have imagined having a better advisor for my graduate study.

I am extremely thankful to the people of Turkey from the bottom of my heart. Despite being a foreigner, they made my two years in Turkey marvellous and never made me feel like an outsider. They have been more than kind and helpful to me, especially my friends at ISIK University. I greatly value their friendship and I deeply appreciate their belief in me. Without their support, I wouldn’t have adjusted in a foreign country.

My special thanks go to the ISS security people of our university. Their kindness and support cannot be expressed and forgotten. They are the greatest people on earth.

Most importantly, I owe a lot of thankfulness to all my friends throughout the globe especially Ceren, Utkristha, Serkan, and Clementine. Finally, I would like to express my heart-felt gratitude to my mother and my genius brother. Thanks for trusting me and supporting me in my endeavours.

This study was supported by The Scientific and Technological Research Council of Turkey (TÜBİTAK) Grant No: 108E201.

Dedicated to all those poor children throughout the globe

who cannot even afford to go to school ...

# Table of Contents

## List of Tables

# List of Figures

# List of Abbreviations

| | |
|---|---|
| ACR | Absolute Category Rating |
| CIF | Common Intermediate Format |
| CSF | Contrast Sensitivity Function |
| DCR | Degradation Category Rating |
| DSCQS | Double Stimulus Continuous Quality Scale |
| DSIS | Double Stimulus Impairment Scale |
| FR | Full Reference |
| HDTV | High Definition Television |
| HVS | Human Visual System |
| IEEE | Institute of Electrical and Electronics Engineers |
| IHS | Information Handling Services |
| IQA | Image Quality Assessment |
| ISO | International Standards Organization |
| ITU | International Telecommunications Union |
| LGN | Lateral Geniculate Nucleus |
| MPQM | Moving Pictures Quality Metric |
| MOS | Mean Opinion Score |
| MOVIE | MOtion-based Video Integrity Evaluation index |
| MSE | Mean Squared Error |
| MSSIM | Multi-scale Structural SIMilarity |
| MSU | Moscow State University |
| NR | No Reference |
| PC | Pair Comparison Method |
| PDM | Perceptual Distortion Metric |
| PLCC | Pearson Linear Correlation Co efficient |
| PQM | Perceptual Quality Metric |

| | |
|---|---|
| PSNR | Peak Signal-to-Noise Ratio |
| PVQM | Perceptual Video Quality Metric |
| QCIF | Quarter Common Intermediate Format |
| QoS | Quality of Service |
| QoE | Quality of Experience |
| RMS | Root Mean Square Error |
| RR | Reduced Reference |
| SAD | Sum of Absolute Error |
| SDSCE | Simultaneous Double Stimulus for Continuous Evaluation |
| SIF | Source Input Format |
| SROCC | Spearman Rank Order Correlation Co efficient |
| SS | Single Stimulus Method |
| SSCQE | Single Stimulus Continuous Quality Evaluation |
| SSIM | Structural Similarity index |
| VOD | Video On Demand |
| VQEG | Video Quality Experts Group |
| VQA | Video Quality Assessment |
| VQM | Video Quality Metric |

# Chapter 1

# Introduction

## 1.1 Background

Over the last few years, digital video technology has seen drastic changes and is growing vigorously in an unimaginable way. Sometimes it might seem that this technology has almost reached its pinnacle but indeed it is spreading out its proverbial wings.

Sony was the first to introduce digital video in 1986 and since then digital video technology has been gaining constant admiration due to its fast progress especially if we focus on applications such as Video on Demand (VoD), Internet Video, Digital Cinema, Video Phones, Video Surveillance, HDTV and several other video based applications [21].

We have travelled a long way since the 1980s when the very first digital video format was formulated. Despite having a low maximum resolution, this was a breakthrough which set the stage for digital video. Since then, the world of video technology has shown no signs of slowing down and because of innovations that started that were introduced at that time; we are now able to stream very high quality lengthy videos over the internet in high definition. The extent of this study lies in the fact that today we have reached a level that we almost watch more streaming video than the physical media. According to a new study by Information Handling Services (IHS) screen digest, American people watched more videos streaming over the internet than they did on DVD or Blu-ray combined [7]. That's the awe-inspiring breakthrough of digital video technology we are going to discuss.

In the recent past, Digital Video Communication has shown a significant progress especially the transmission and compression aspects with an aim to render a very good quality video to the ultimate user. But still the visual content is one of the most thought-provoking content types not only because of the bandwidth but also due to the ultimate user quality perceptual experience. The technological advancements in this field need better ways of quantification of the performances of the communication system. Eventually in most of these applications the end results are prepared for usage by human observers therefore humans are the ones who decide the fate of such operations.

Undoubtedly, this is one of the most crucial and decisive factors so human subjects should always be taken into consideration. Otherwise predicting the real success of the transmission or extent of improvement in quality would not be accurate.

According to the International Standards Organization (ISO), “Quality is the degree to which a set of inherent characteristics of a product fulfills customer requirements” [16]. Despite quoting this straightforward definition, past research has proved that quality estimation is highly subjective and relatively hard to model within the proper framework.

Up until now, the most precise and valued way of assessment of the quality of a video is the evaluation using subjects in the form of human participants (expert or non-experts based on the target) [26] But these psychophysical experiments are not as simple as they sound. As a matter of fact, they are quite expensive, pretty difficult and comprising of lengthy tasks. Also, it is not pragmatic to involve human subjects in most of such applications [12]. This leads to a need of a highly robust system which is able to assess the video quality in an effective way without introducing any human observers. So the experts need to keep in mind about the proper algorithm which would yield objective evaluation of the video quality under consideration.

Although some well-known objective metrics like Peak Signal to Noise Ratio (PSNR) and Sum of Absolute Error (SAD) have made a mark in recent years, the irony lies in the fact that these metrics have a major drawback as they do not correlate with the human subjective judgments very well.

## 1.2 Quality of Service (QOS) and Quality of Experience (QOE)

There has been a substantial amount of work in the field of Image Quality Assessment for several decades. But these days, a slight shift has been viewed as researchers have started to focus on the assessment of video quality as well. Sometimes it seems hard to fathom the way VQA based research is gaining a strong momentum. Few things can easily be deduced from literature reviews in which the focus has been on spatial distortion rather than the temporal ones and in general on the Quality of Service (QOS) rather than the Quality of Experience (QOE).

When we start talking about these two terms, we realise QOS is a quite well known term compared to QOE. It is a well-known fact that the former tries to objectively quantify the services handed over by the vendor. In fact, it has nothing to do with the view point of the customer, but its more relevant to the media. While the latter speaks about the subjective measure of a customer's experience. But before we introduce the term QOE in a broader way let us throw some light on the need for QOE systems. As the emerging digital video communication technologies gain industry and consumer acceptance at a progressively rapid pace, the increase in the quality of the video is the first striking thing visible to a naive user. Presently it is not only a question of services provided but also how nicely the characteristics are handled because that has a long lasting impact on the ultimate user. The predominance of user-targeted applications are already being witnessed these days where the role of quality is the most crucial. In our research, quality is mostly referred to as the signals meant for consumption and interaction by ultimate users. But from the acquisition phase until the end, these signals may suffer different kinds of artefacts during the processing which may lower the quality of experience for the perceiver. There are also some other factors, such as delays in communication and other context related issues that may affect the experience by an ultimate user substantially. Before the introduction of QOE, service providers used to address these matters with respect to the QOS factors from the provider’s point of view.

QOE offers a new trend to look at the quality related issues that are not dealt with the QOS metrics. QOE focusses on motives, behaviours, properness, context, usability and observer components based on the final content rather than just focusing on the subjective perceptual experiences. This gradual inclination from QOS to QOE is based on the notion of how a particular observer rates a given service or product and this phenomenon is more or less related to factors like usability and perceptual experience [19].

As International Telecommunication Unit (ITU) clearly states: “The overall acceptability of an application or service, as perceived subjectively by the end-user” [20]. So QOE works inside the user’s creative mind which consequently makes it more qualitative than quantitative. With the rapid unbeatable growth in the entire digital video communication system, there is a need of such a finely chiseled QOE which should be broader in scope comprising of a proper formulation of subjective quality assessment methods and also the objective ones that model the perceptual experience of observers in a very close, careful and precise manner.

In this way QOE based methods or subjective analysis makes it the most trustworthy method as we ask for human observer’s opinion and correctly termed as subjective video quality assessment. Although it’s not feasible in all the applications as we need methodological involvement of human subjects, it provides worthwhile data for evaluation of the automatic methods for quality assessment. It elucidates the obtained results of state-of-the-art video quality assessment and also helps us ameliorate the performance of Video Quality Assessment metrics.

So in order to gauge the performance of the quality assessment, Mean Opinion Score (MOS) comes into play, which is the subjective quality measurement carefully done by using human subjects as observers and helps us to correlate with the obtained objective scores.

Objective Video Quality Assessment methods, the QOS based ones are essentially based on statistical features. Normally absolute differences between the reference and degraded video series for each pixel are taken into account and thereafter transmuted into a quality evaluation score using various statistical methods, especially if we consider the popular metrics like MSE, PSNR, etc. [22]. They have a big drawback that they are not well correlated with the human subjective judgments. However recently proposed metrics such as Structural Similarity (SSIM) Index tries to address this issue by using the Human Visual System (HVS) properties (discussed in chapter 2), and have also performed better than PSNR. [23]

Video Quality Experts Group (VQEG) claims that the performance of HVS based VQA models still lacks lot of features and needs further advancement [24]. Usually when assessing the video quality we use frame by frame features rather than working on motion information. So we cannot assume such models are perfectly-correlated ones. Clearly, there is a need of a versatile QOS model which complies with the QOE in the best possible way. Our thesis proposes one solution to this issue. We intend to work on such objective metrics which performs better than the state of art objective models and mimics the human visual system.

### 1.3 Dissertation Objective and Overall Description

In this thesis, we consider both aspects of Video Quality Assessment viz. Subjective as well as the Objective Evaluation. To a great extent, our work is inspired by Perceptual Quality Metric (PQM) for dealing with 3D video datasets [47]. We do not consider the 3D aspects. As a matter of fact, we deal with the 2D video datasets. A robust objective algorithm has been proposed namely Perceptual Quality Metric for 2D (PQM2D) using the ideas from the above mentioned work. The aim of this work is to show better results for 2D video quality assessments and outperform the various popular state-of-art metrics like Structural Similarity (SSIM), Multi Scale Structural Similarity (MS-SSIM), Peak Signal to Noise Ratio (PSNR) etc. For the verification phase, series of subjective experiments are performed to demonstrate the level of correlation between objective metrics and the user scores obtained by Subjective Evaluation using human observers keeping in mind the standards set by International Telecommunication Union (ITU) [26].

The extremely subjective nature of our entire project, together with the complexities of Human Visual System (HVS) makes it a cutting edge research topic and an interesting problem to work on and this attempt would be evident throughout the thesis.

### 1.4 Dissertation Structure

The present thesis is structured into the following six chapters:

Chapter 2 provides a general overview of the Video Quality Assessment and complexities of Human Visual System (HVS), we discuss about standard testing conditions and review previous relevant work.

Chapter 3 presents an overall description of our Subjective Quality Evaluation and how the test was performed. Also, we discuss the features which are crucial for getting trustworthy results from subjective experiments.

Chapter 4 focusses on our full reference objective video quality assessment method and evaluation. We also discuss the state-of-art metrics and study the proposed objective model and the algorithm involved.

Chapter 5 analyses the various experimental results obtained by subjective and objective methods statistically. We talk about accuracy and monotonicity of our results by examining the correlation using various kinds of graphical representations and discuss the superiority of our proposed algorithm. And in Chapter 6, we conclude by giving final remarks on our entire research work and point out some important future direction.

# Chapter 2
# Video Quality Assessment

## 2.1 Introduction

We have already briefly shown that there are two general methods of assessing video quality, the subjective and objective quality metrics. In this chapter, we will see the overview of these methods. We will focus on some literature work and state of art metrics. We will also see the crucial role of recommendations by International Telecommunication Union (ITU).

## 2.2 Human Visual System

The first stage of the Human Visual System (HVS) is the eye in which the visual stimulus goes through the physical properties of the eye and thereafter towards the photoreceptors behind the eye. There are so many typical features demonstrated by the eye comprising phenomenon such as abnormalities based on lens like lens aberrations which are difficult to model in the HVS based Video Quality Assessment algorithms. The working of the eye is similar to a low pass filter because of the optics of eye which is band limited. This is the reason why point spread functions are used for modeling HVS based quality evaluation systems. [30][31].

The photoreceptors are further arranged into rods and cons. The rods are responsible for the working of vision system under the scotopic (i.e. night vision) conditions while the cons are responsible for vision under the photopic conditions. They are also responsible for encoding colour information. In fact the spreading of rods and cones in the eye is not uniform. The numbers of photoreceptors are prominent at a region called the fovea and there is a noticeable deterioration as one makes motion away from the fovea. This concept is important for quality assessment of an image or video because the human eye does not absorb the total visual stimulus at the same resolution. The component of the stimulus which is imaged on the fovea has the highest resolution and the others have lower resolution. For the sake of taking in the complete stimulus, the human glances over the image with the help of a set of fixations followed by quick movements of eye known as saccades. During the process of saccade, very little information is collected [32].

Furthermore, the information from the photoreceptors is then worked on by the retinal ganglion cells and then passed onto the Lateral Geniculate Nucleus (LGN) which acts as a relay station supposedly [32]. And it is the first location along the visual pathway where the information from the left and the right eye merges. The LGN receives not only the feed-forward information from the retinal cells, but also feed-back information from the next stage of processing - the primary visual cortex (area V1). The amount of feedback received leads one to believe that the LGN may not be just a relay station in the visual pathway [34]. Further, recent discoveries show that the LGN may perform certain normalisation computations [35].

A significant amount of neural activity is devoted to motion based works. As a matter of fact, Human Visual System is susceptible to motion, that is why objective assessment of video quality takes motion into careful consideration. Despite the fact we have made very little progress in understanding the HVS, there is still a lot of research to be done in this field. Each of the areas above discussed areas is an active discipline of research for getting an engineering view to understanding and analysing the Human Visual System. Figure 2.1 shows approximation of the absorption spectra of the three cone types [30].

### 2.3 Complexities of Human Visual System

There are some distinguishing characteristics of the HVS. They influence the Image/Video Quality Assessment as most of these characteristics govern the discernibility of distortions and eventually of the quality. The Contrast Sensitivity Functions (CSFs) also play an important role for decreasing sensitivity of the Human Visual System as there is an increase in spatial frequencies. Human Visual System also demonstrates changing sensitivity to the temporal frequencies. Lots of quality assessment models assume that the spatial and temporal responses are capable of being dissociated [40].

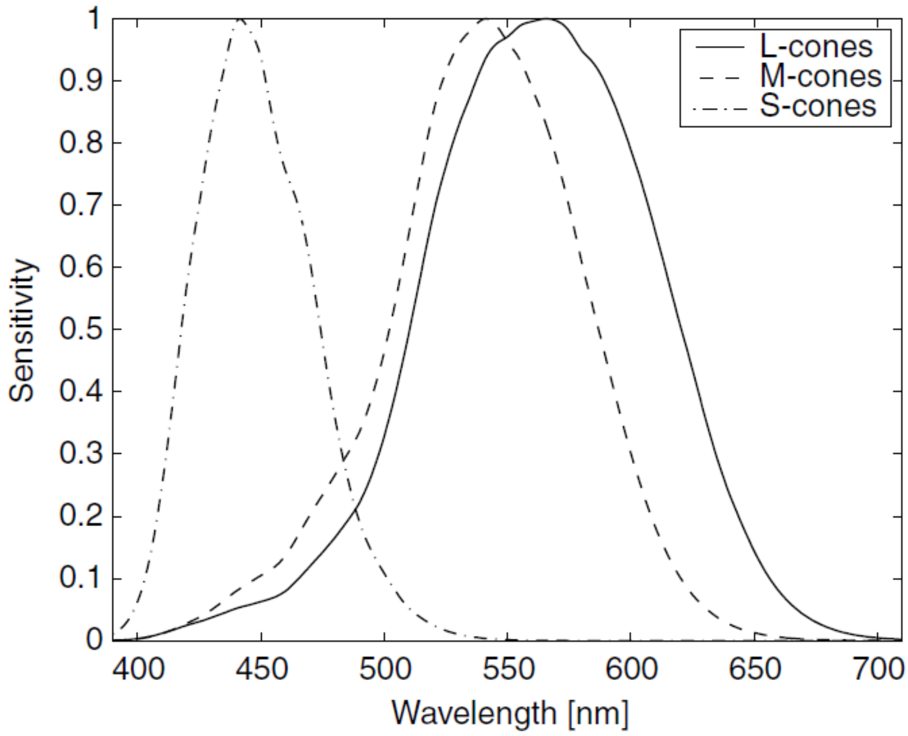


Figure 2.1: Normalized absorption spectra of the three cone types: L-cones (solid), M-cones (dashed), and S-cones (dot-dashed) (Photo taken from [30]).

### 2.4 Subjective Quality Metrics

In this section, we will see an overview about several methodologies, rules and other crucial criteria for carrying out subjective test as described and standardized in the Recommendation ITU-R BT.500 and in the Recommendation ITU-T P.910 by the International Telecommunication Union group [48]. The subjective evaluation gives the Quality of Experience (QOE) based on the information provided to the user.

#### 2.4.1 Viewing Conditions

The most important factors one should consider are the lighting, the ambience noise and the quality and calibration of display. The table shown below shows the requirements for setting up proper viewing condition for subjective test [48].

Table 2.1: Some details of the recommendations on viewing conditions

| Viewing Conditions | |
|---|---|
| **Parameters** | **Settings** |
| Viewing distance | 1-8H |
| Background Room Illumination | <20lux |
| Peak Luminance of the Screen | 100-200 cd/m |
| Ratio of luminance of inactive screen to peak luminance | <0.05 |
| Ratio of luminance of background behind the display to peak of luminance | <0.2 |

### 2.4.2 Selection of Test Materials

The subjective VQA results are largely affected by the test video scene or sequence taken into account because the observers view those contents so it should be selected carefully. The wider the range of test sequence, the more meaningful results we get. Therefore videos with detailed backgrounds are useful for yielding sensible results [48].

### 2.4.3 Observers Selection

As per the recommendation booklet discussed above, the minimum number of observers should be 15 so that we can get more reliable results. The higher the number of subjects, the more trustworthy the output becomes. Also, non-experts should be preferred over experts in video/image quality area.

### 2.4.4 Video evaluation session

The session plays an important role because the observers should not become tired or confused. That’s why the duration of a normal session should not last more than half an hour. The test session should comprise of a warm up session as well before the actual test starts. We stabilise the users’ opinion by showing around five “dummy presentations” and the obtained data should not be used in the final results of the test. If we are carrying out more than one session then three dummy presentations are enough at the beginning of the actual test. Essentially, the warm up sessions allows the users get acquainted with the real test session [48].

Figure 2.2 shows a sample of videos ordered for warm up and real test session. It is a series of video sequences taken from our subjective experiment depicting the warm up session, the data which is not counted but is just for making the user for the actual test and the main test session, the session where data is taken into consideration for the final results. Furthermore, we should use a random order for the presentations and some of the presentations can be performed repeatedly from session to session to check cohesiveness.

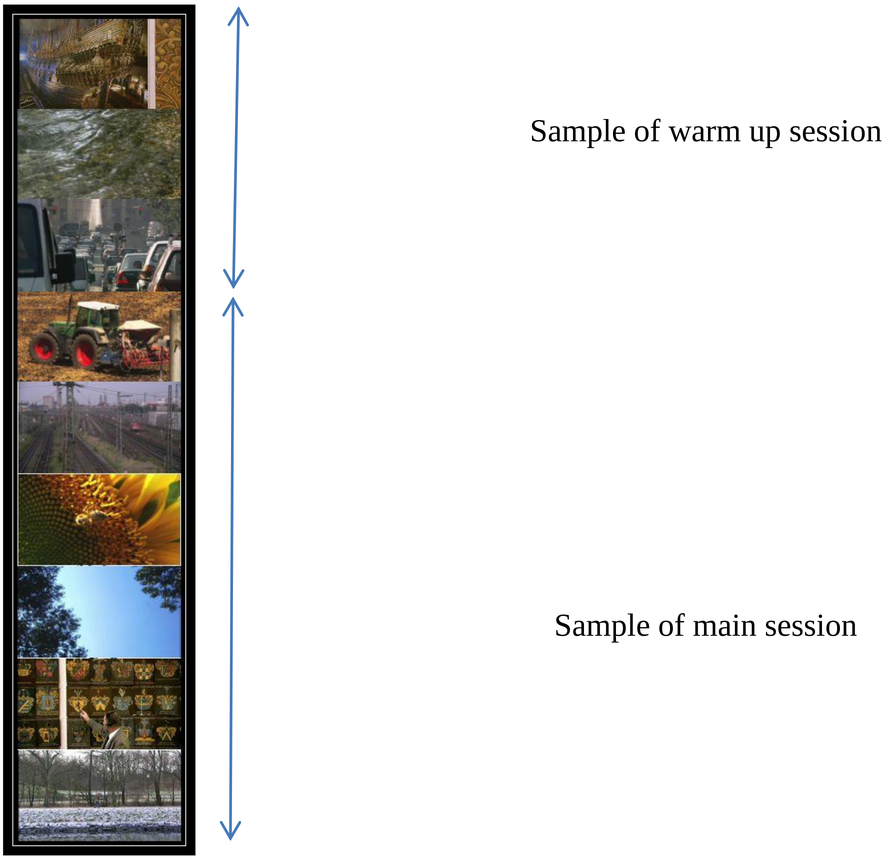


Figure 2.2: Test Session structure

### 2.4.5 Useful information for the assessment

In this section, the user should be introduced to the test, the grading scale and the sequence timing so that there is no room for misunderstanding when the test starts. We have already explained in the video evaluation session how warm-up sessions help us to give a demonstration of the real test session (2.4.4). It is important that this information is provided to the users so that they can assess the videos diligently.

### 2.4.6 Video quality assessment methods

There are a number of proposed methodologies, but few of them have been standardized so far. Some of them are discussed below.

#### 2.4.6.1 The Double Stimulus Impairment Scale (DSIS) or Degradation Category Rating (DCR)

The Double Stimulus Impairment Scale (DSIS) is a tool applicable for such cases where we can easily see the distortions produced by the encoding process like blockiness, ringing and blurring. It is the fundamental methodology for assessing television pictures. This method is useful for assuring the similarities of test condition with respect to the reference condition and it is effective for high quality system evaluation in the multimedia communications [26].

For the trial of DSIS method, the observer is shown the video sequences in pairs where the first one is called the reference sequence and the second one is called the test or impaired sequence. The reference is the original, undistorted source sequence whereas the impaired sequence is a distorted version of the reference video.

#### 2.4.6.2 The Comparison Scale Method (CS) or Pair Comparison method (PC)

It is a method which performs a direct one on one comparison between two systems in order to judge which one is better and how much better it is than the other one. It is also called Pair Comparison method in the case of multimedia applications. In this method, the reference and test sequences are not displayed in a specific order.

#### 2.4.6.3 The Single Stimulus Method (SS) or Absolute Category Rating (ACR)

The Single Stimulus method is such a method where the test sequences are demonstrated one at a time and are graded individually on a category scale. It is also referred to as Absolute Category Rating. It is a fast and easy method as it allows users to improve their time efficiency. The presentations are shown in a random order for each subject.

#### 2.4.6.4 The Single Stimulus Continuous Quality Evaluation (SSCQE)

Basically, the SSCQE method is made to assess the digital television systems. It measures the quality of a video over time; hence the subjects are continuously grading the video quality on a linear scale. It is one of those methods where a series of video sequences are shown only once to the subject. In comparison to the various other methods that also use continuous rating scales, this method allows the observers to assess both audio and video in video-conferencing applications. It tries to maintain a high level of focus and attention from the subjects in order to reduce fatigue, the SSCQE method advices the introduction of breaks during each test session.

#### 2.4.6.5 The Double Stimulus Continuous Quality Scale (DSCQS)

The Double Stimulus Continuous Quality Scale (DSCQS) method recently gained popularity for being used in performance evaluation of the digital HDTV (High Definition Television) Grand Alliance System, which was the foundation for the North American standards for digital TV broadcasting [26]. Essentially, this method is useful when it is not possible to span the full range of quality stimulus [48]. In the DSCQS method, the order of the reference and the test sequences are randomised. Thus, the observer does not know which presentation is the reference or which is the test sequence. Thereafter, the observer is asked to grade both the sequences.

#### 2.4.6.6 The Simultaneous Double Stimulus for Continuous Evaluation (SDSCE)

This SDSCE method is one of the internationally accepted and standardised methods for image and video quality assessment like the aforementioned methods. Specifically, the SDSCE can be appropriately employed to all those cases where fidelity of visual information, influenced by time-varying degradation, has to be taken into account [48].

The SDSCE method comprises of assessing two sequences simultaneously and continuously. Both the clips are displayed in parallel positions using one or two displays.

The user needs to shift the attention between right and left presentations over and over again because of the simultaneous display of both the clips so this factor is considered a crucial drawback for such a method. The subjects continuously judge the fidelity of the video of the impaired sequence with respect to the reference, by moving a slider on a simple voting device. In this method, users know very well as to which sequence is the reference and which sequence is the impaired one.

In the upcoming chapter on subjective quality evaluation, we have shown the method followed in our experiment.  We have employed the Double Stimulus Impairment Scale (DSIS), also called Degradation Category Rating (DCR) method where the videos are shown pairwise. The first stimulus is the reference video and the second one is always the impaired version. The voting time for the observers are supposed to be less than or equal to ten seconds. It is a well-suited method for our experiments, especially when we deal with full reference metrics.

## 2.5 Objective quality metrics

The objective quality metrics plays a key role in the VQA/IQA because it helps us in predicting the quality score that human observers would give to a particular video sequence. These metrics can give us the Quality of Service (QoS) in digital video communications.

### 2.5.1 Classification of Objective VQA

On the basis of information we have, it is divided into three kinds.

• Full Reference metrics (FR):

These metrics need the original video and the impaired video. We compare the quality of the impaired version of the original video in a pixel by pixel manner. It is the most commonly used metrics type. In the upcoming chapters we will see how we implement our proposed FR metrics and compare it with various other FR metrics.

• Reduced Reference metrics (RR):

These metrics require the description of some of the parameters from the original video and the impaired video. We use such metrics when there is lack of availability of original video. For example, in case of transmission with a limited bandwidth, we use RR metrics type.

• No Reference metrics (NR):

As the name suggests, such metrics just has the access to the impaired video, with no access, whatsoever, to the original video. It is one of the hardest types of objective methods.

### 2.5.2 Related Work

From the mathematical point of view, elementary error indices like the Mean Squared Error (MSE) are frequently used to evaluate video quality because of their simplicity. It is a well-known fact that the MSE does not correlate well with the Human Visual System (HVS). That is the reason why VQA research community has been studying full reference VQA methods in a profound way [57]. Different weighted versions of MSE have been suggested by some research groups. In order to overcome the drawbacks of the aforesaid methods, such approaches were considered which could predict the quality of an image or video in the best way using systems similar to the HVS. Therefore significant effort has been made on using models of the HVS to develop quality indices, which in other words we generically term as HVS-based indices. Typical HVS-based indices use linear transforms with the possibility of separation in the spatial and temporal dimensions to decompose the reference and test videos into various channels. The main idea is to model based on the errors from different channels at each pixel and aggregate to get space-varying map that predicts the probability that a human percipient will be able to observe any dissimilarity between the two images. Some of the well-known HVS-based video quality indices are the Moving Pictures Quality Metric (MPQM), Perceptual Distortion Metric (PDM) and the Sarnoff JND vision model [30].

In the recent past, there has been a shift toward VQA methods that seek to characterize features that the human eye relates with loss of quality. More precisely, distortions like blur, blocking artefacts, colour information and so on. One of the obvious reasons for this shift has clearly been the complexity and imperfectness of the various HVS models. Seemingly, some other factors also play a decisive role in the above mentioned shift. For example such HVS based models which generally model the sensitivity of the HVS to various features evaluated at the threshold of perception [30]. There are some other popular algorithms which have originated from industries and they attempt to detect the distortion and measure the level of annoyance. Such algorithms are Video Quality Metric (VQM) from NTIA [55], the SSIM index for video [51], Perceptual Video Quality Measure (PVQM) and so on [30]. Nonetheless, these frameworks extensively capture spatial distortions but do not deal with the temporal distortions in the video so much. For example, VQM deals with the 3-D spatio-temporal blocks of video in finding some features but in the case of temporal component it just treats frame differences [55]. The extended version of the SSIM index estimates local spatial SSIM indices at each frame and employs motion information to blend into a single quality score for the full video.

In order to deal with the mentioned issues, LIVE group [57] singled out spatial and temporal characteristics. In the beginning, the reference and test videos are disintegrated into spatio-temporal bandpass channels with the use of a Gabor filter. In a wide way, spatial components are dealt with using the basic ideas of SSIM index [51]. Thereafter, temporal component is quantified with the help of available motion information obtained from the reference video. Eventually, both the spatial and temporal quality scores are combined diligently and we get the overall video quality score called MOtion-based Video Integrity Evaluation index (MOVIE) [57].

# Chapter 3

# Subjective Quality Evaluation

## 3.1 Introduction

The subjective quality assessment methods are essentially used to gauge the performance of multimedia or television systems with the help of responses obtained from observers who view the system under test. As a matter of fact, it is evident from the previous chapter that it may not be possible to fully describe the performance of a system using objective experiments. As a result, there is a distinct need of well-configured subjective experiment with a proper set up following the benchmark of International Telecommunication Union (ITU) as much as possible. Though it is very difficult to accomplish the subjective task following the ITU standards completely, in our experiments we have tried to follow the instructions as far as we can in order to have more meaningful results. In order to carry out such experiments, it is important to choose from the range of options accessible to us which are well suited for the objective test and for the consideration of the assessment task we have.

It is well known that, subjective quality assessment is a method based on user's perceptual experience involving well-structured experimental designs and human participants. So in this chapter, we will discuss our subjective experiment and how cautiously we performed it. The target of the subjects involved is to judge the video quality of the presented series of video sequences. With the help of this experiment, we are able to find the Mean Opinion Score (MOS) of the various video sequences under consideration. We will see that firstly computations are made by constituting all the observers involved in the test and later statistical analysis is employed in order to substantiate the observer's opinion (chapter 5).

### 3.2 General Viewing Conditions

The laboratory environment was set up as per the guidelines of the ITU. Our attempt was to follow the guidelines as much as possible but some conditions were really hard to implement precisely. Some details are mentioned in the table 3.1. The maximum duration was strictly kept within 25 minutes and not lesser than 2 subjects in a session which eventually comprised of 16 subjects in total, including all sessions. Figure 3.1 shows the diagram of a human observer in a testing room and the viewing distance has also been shown.

Table 3.1: Laboratory Condition and Display details

| **Parameters** | **Settings** |
|---|---|
| Peak luminance of the screen | 150 cd/m2 |
| Other room illumination | quite low |
| Height of image on screen (H) | 11 cm |
| Viewing Distance | 88 cm |

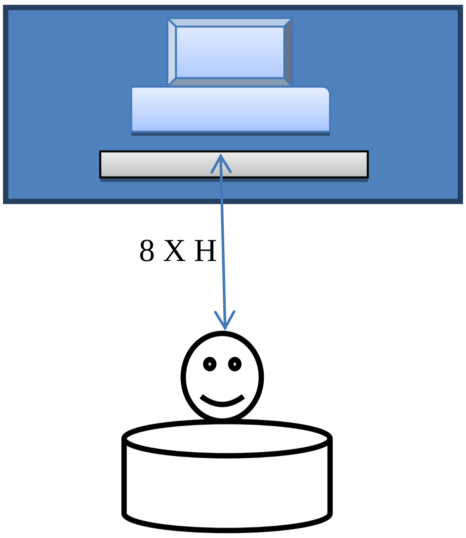


Figure 3.1: Diagram depicting a subject in the testing room

### 3.3 Source Sequences

The videos were obtained from the Laboratory for Image and Video Engineering (LIVE) at The University of Texas at Austin. They are carefully and professionally filmed video datasets using high end equipment and thereafter converted into digital format with extreme care and attention. This makes them distortion free samples (from now on these videos are referred to as "reference videos"). These videos have no sound component and are recorded in High Definition (HD) with YUV 4:2:0 formats. In order to keep the aspect ratio of HD videos, a standard resolution of 768X432 pixels has been used which would also denigrate visual distortions involved [28].

In our subjective quality assessment experiments, we used nine reference videos in the test session. Figure 3.2 shows one frame from each of the original video sequences used in our experiment. Figure 3.3(a) shows the screenshot of the original video sequence called "park run" and in fig 3.3(b), 3.3 (c), 3.3 (d) and 3.3 (e), the same video is distorted by wireless, IP, H264 and MPEG2 respectively. Figure 3.4 shows the histogram of PSNR variations in the selected set of source sequences. Test cases were carefully selected so that the maximum range of PSNR is covered. It is clearly demonstrated that varying videos were selected, on the basis of its PSNR so that we get more reliable and better results from a huge range of video datasets.

A short summary of the videos used in the first session is provided below (Fig 3.2).

- Rush Hour: In this scene, still camera shows the rush hour traffic on a street.
- Sunflower: In this clip, a still camera shows an insect (bee) which is moving over a sunflower and it's a close-up view.
- Station: In this clip, a still camera shows a railway track where a train is visible and few people are walking across the track.
- Tractor: In this case, there is a tractor which moves across the field (camera pan).
- Mobile and Calendar: Here, there is a moving toy train in a horizontal manner with a calendar moving vertically in the background (camera pan).
- Blue Sky: A blue sky is being shown in this clip with some trees and it involves circular camera motion.
- River Bed: In this clip, a still camera shows a river bed which contains some small smooth rounded rocks and water.

- Park run: A person is running across a park in this scene (camera pan).
- Shields: In this video, in the beginning, the camera pans then becomes still and further zooms in. A person is shown who is walking across the display and points to it.

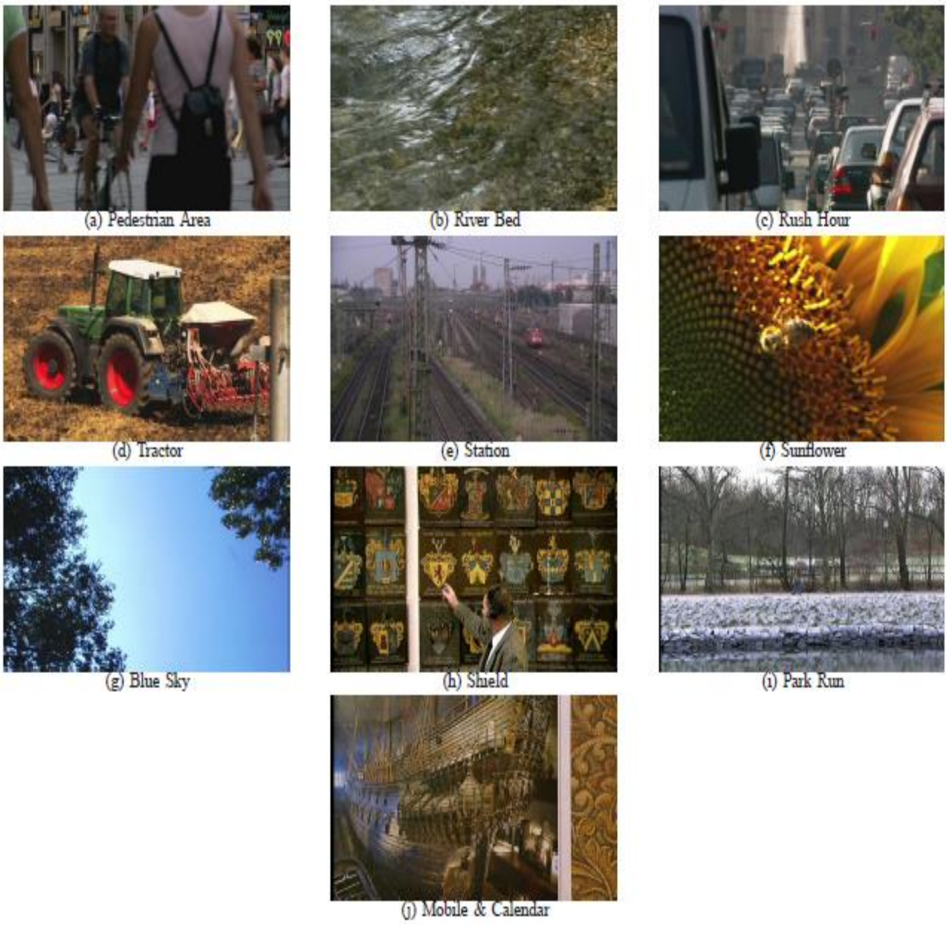


Figure 3.2: One frame from original video

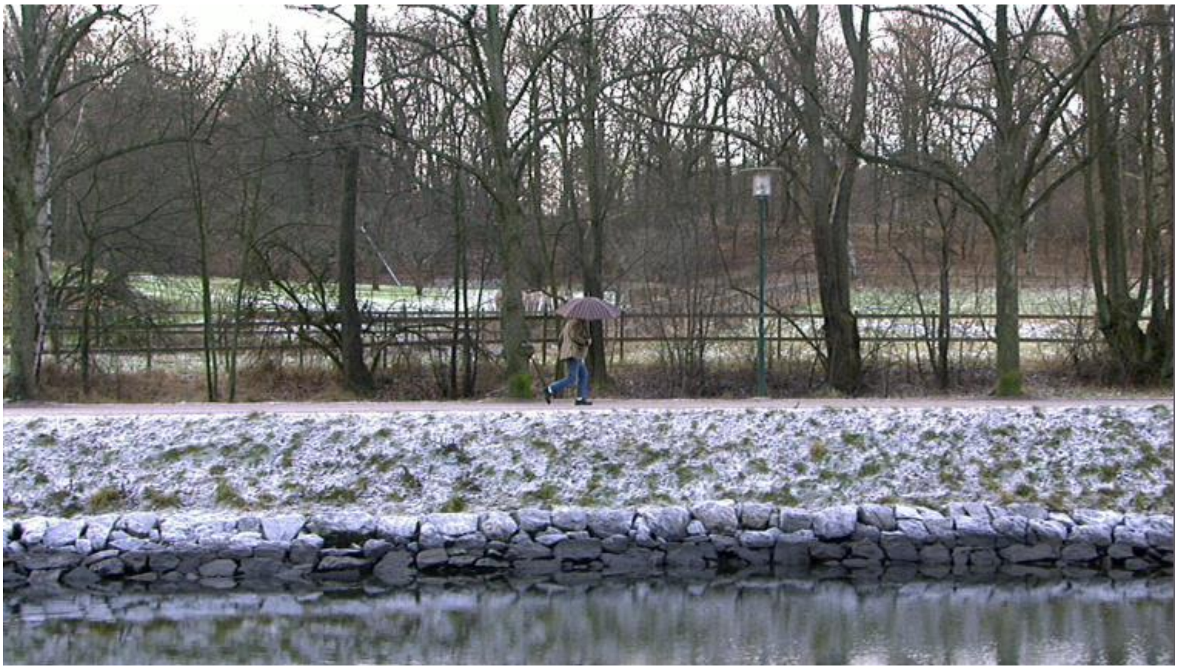

Figure 3.3(a): screen shot of reference video “park run”

Figure 3.3(b): screen shot of “park run” distorted by wireless transfer

Figure 3.3(c): screen shot of “park run” distorted by IP transfer

Figure 3.3(d): screen shot of “park run” distorted by H264

Figure 3.3(e): screen shot of “park run” distorted by MPEG2

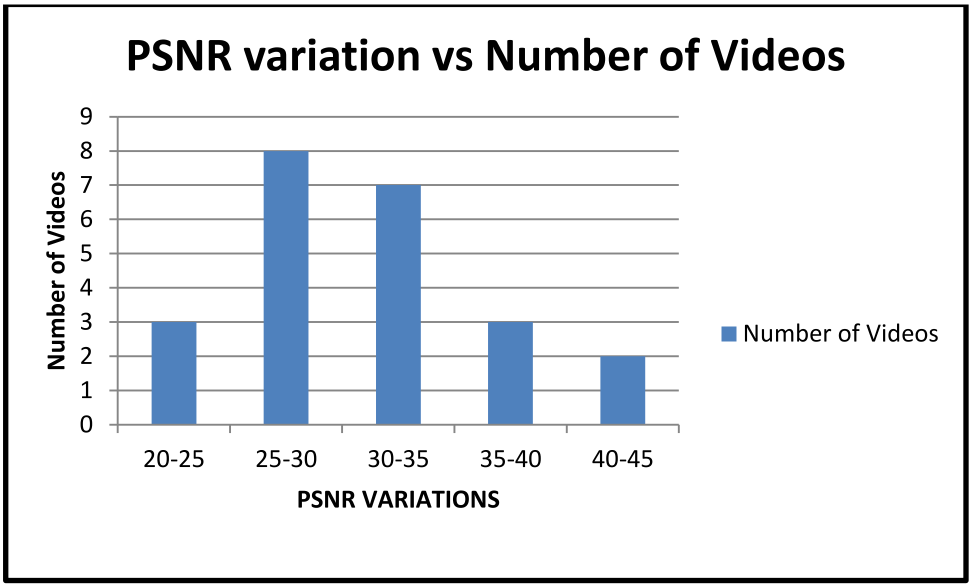


Figure 3.4: PSNR variations of selected video sequences

### 3.4 Test Sequences

For test sequences cases, we used four types of distortions mentioned below:

- Wireless distortion

- IP network distortion
- H.264 compression
- MPEG-2 compression

We used distorted samples from the LIVE Video Quality Database [27] [28]. It is discernible that different sources of distortions cause different types of artefacts and degradations. Especially if we talk about H.264 and MPEG-2 compressed sequences, they give us consistent impairments like common artefacts including blur, ringing, blocking around the edges for both the components, spatial and temporal ones. If we compare the test sequences obtained from lossy transmission, there are prominent errors due to block losses in large areas covering the frames for wireless network distortions. In case of IP distortions, block errors in large regions of the frame are noticeable. The artefacts in the wireless transfer appear different than the IP network artefacts. There are likely chances of packet loss in case of wireless transfer because it is more vulnerable to bit errors due to attenuation, shadowing, fading etc.

The benefits of using LIVE Video Quality Database lies in the fact that most of the distortions are not just spatial ones but are spatio-temporally localized distortions. Still it's a matter of concern that most of the database for VQA research is not free to use as they have copyright issues or are sold for business purpose.

As we discussed about the selection of test material for our subjective based experiments, there are some other factors which impact the Human Visual System (HVS) the most. Apparently the spatial as well the temporal activity plays a decisive factor in calculating the human perceptual experience for a video sequence [53].

### 3.5 The Subjective Assessment

#### 3.5.1 Subjective Testing Design

In chapter 2, we discussed the pros and cons of various test methodologies. So the test methodology used in our experiments is known as Double Stimulus Impairment Scale (DSIS) or the Degradation Category Rating (DCR). These kinds of methods are popular in the full reference video quality assessment where we want to measure the robustness of the system [26]. A carefully selected playlist was prepared by the author. Entire playlist comprised of 24 videos in total, that is, 9 reference videos in total with various kinds of distorted counterparts.

As the subjective experiments are reasonably lengthy and there are chances of viewer's fatigue, we broke the session into various parts. Latency and frame drops were also dealt with utter seriousness so that the user does not get bothered by any kind of other issues apart from video distortion based issues. We tried our best to strictly follow the recommendations of International Telecommunication (ITU).

### 3.5.2 Observer Selection and Training

Most of the subjects who took part in this subjective quality assessment research were non-expert undergraduate students from the department of psychology of ISIK University, Turkey. Each video was rated by 16 subjects in total. Each subject was told about the aim of the experiment one by one and was carefully instructed about the methods of grading the video sequences. Subjects were given a warm up session in order to familiarise themselves with the environment and experiments before they started the main session. They were also asked to provide ratings for the warm up session to help them understand the testing procedure, these warm up ratings were not used in finding the final Mean Opinion Score (MOS). Figure 3.5 shows the test session presentation structure.

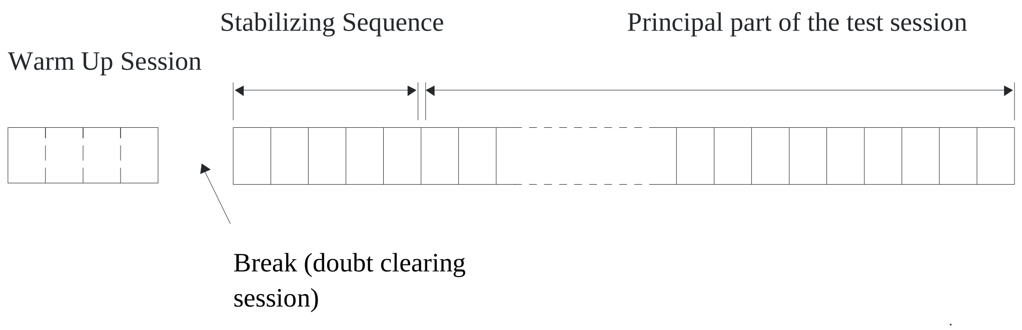


Figure 3.5 Test session presentation structure

### 3.5.3 Grading Scales

Using the DSIS methodology, the subjects were asked to rank the impairment in a five grade impairment scale (Fig 3.6):

5: imperceptible
4: perceptible, but not annoying
3: slightly annoying
2: annoying
1: very annoying

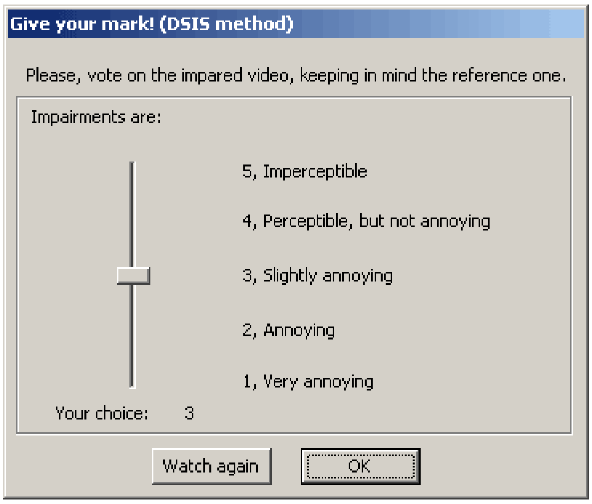


Figure 3.6: A screenshot from the MSU program showing the grading scale for DSIS method

## 3.6 Video Quality Evaluation Program Interface

In our subjective experiments, the program used for carrying out the various tasks such as test presentations, results storage etc. was the MSU (Moscow State University) perceptual video quality player formulated by the Graphics and Media Lab Video Group. Figure 3.7 shows the program interface. It is an efficient program where the user starts the session by pressing the start button and eventually the reference and thereafter impaired video is displayed. It has some other proficient features such as a time bar in order to give the user an idea of time duration of a video sequence. Finally, a five point rating scale appears after the display of reference and impaired video which asks the user to rate the impairment level. In this way, the user can record their opinion easily.

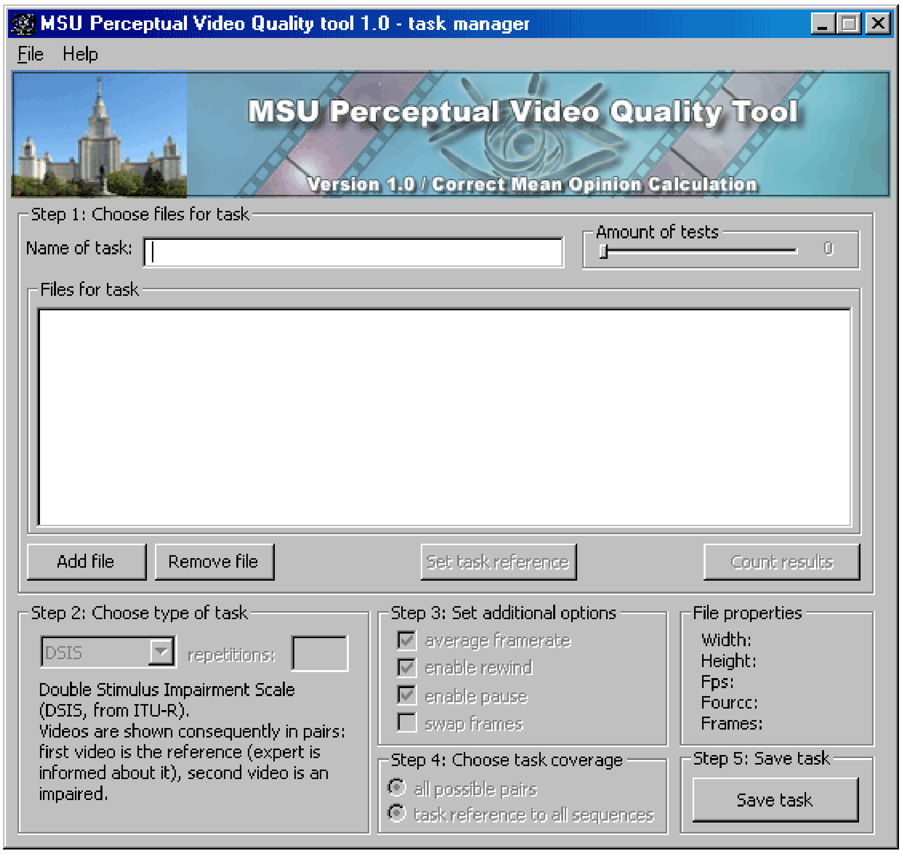


Figure 3.7: Task Manager of the Video Quality Evaluation Program Interface

# Chapter 4

# Objective Quality Evaluation

## 4.1 Introduction

We have already discussed the importance of subjective tests in previous chapters. They give us the best means of measuring quality as it is perceived by the audience. But still when it comes to monitoring the entire video data quality, the subjective methods are not so worthy of being depended on. That’s why we need automatic method which can assess the quality of a broadcast video. Nonetheless, for substantiating the performance of such automatic methods, subjective testing should be employed. This chapter proposes a new objective video quality assessment metrics which is more reliable than the available popular state of art metrics as we will see in the upcoming chapter.

We carried out the full reference objective quality assessment test for our selected set of videos from the LIVE database mentioned in the previous chapter [28]. We used various algorithms for this purpose, especially the ones which are popular in the VQA field of study. Essentially, the algorithms used in our research are the ones which are free to use for academic and research purpose. In fig 4.1 we can see a basic structure of how our state of art metrics and proposed metrics work in full reference quality evaluation manner. In this chapter, we will probe the objective evaluation of various available metrics in full reference form and then we will also see the idea behind proposed metrics for the mean opinion score prediction and its significance.


Figure 4.1: Basic structure of a full reference objective model

### 4.2 State-of-Art Objective VQA

The test comprised of the usage of the following available objective video quality assessment metrics.

#### 4.2.1 Peak Signal to Noise Ratio (PSNR)

It is one of the simplest and popular VQA algorithms which is an unsophisticated function of the Mean Square Error and gives us some classical VQA results. Although it has some obvious drawbacks such as not correlating well with the human visual system, it has been popular in the IQA and VQA community since long. But its ease lies in the fact that it uses logarithmic scale which is more convenient.

$$PSNR = 10\log_{10}\frac{255^2}{MSE} \tag{4.1}$$

where *MSE* is the Mean Square Error.

#### 4.2.2 Structural SIMilarity Gaussian (SSIM G)

It is a method for measuring change in the information between the reference and distorted video. It is known to give better results than PSNR. In order to use this algorithm precisely, frame by frame luminance component was found and SSIM index was calculated. The overall SSIM index value for the video was generated [51].

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{4.2}$$

where $\mu_x$ is the luminance of reference video and $\mu_y$ is the luminance of impaired video, $\sigma_x^2$ is the variance of reference video and $\sigma_y^2$ is of the variance of impaired video. The $C_1$ and $C_2$ are constants.

The SSIM G deals with the Gaussian blur and is also called SSIM precise. In this version, SSIM is computed per pixel using a Gaussian window around each pixel. Then SSIM for the whole frame is the average of pixel-level SSIM values.

### 4.2.3 Structural SIMilarity Box Filter (SSIM B)

The only difference between SSIM G and SSIM B is that in this case SSIM is computed on an 8x8 block basis, and the average SSIM for the whole frame is the average of block SSIMs. It is also termed as SSIM fast.

### 4.2.4 Multi-scale Structural SIMilarity Gaussian (MS-SSIM G)

The fact which distinguishes MSSIM from SSIM is that this VQA algorithm evaluates multiple SSIM values at multiple resolutions. Although it does not lay stress on the luminance component in general, nonetheless we implemented it frame by frame to the luminance part and finally average value was computed. Despite its surpassing qualities, the computations are comparatively slower for this algorithm [54]. The right scale relies upon the viewing conditions such as the display resolution and viewing distance.

In defining MS-SSIM, luminance, contrast and structure comparison measures are computed at each scale as follows:

$$l(x,y) = \frac{(2\mu_x\mu_y + C_1)}{(\mu_x^2 + \mu_y^2 + C_1)} \tag{4.3}$$

$$c(x,y) = \frac{(\sigma_x\sigma_y + C_2)}{(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{4.4}$$

$$s(x,y) = \frac{(\sigma_{xy} + C_3)}{(\sigma_x \sigma_y + C_3)} \tag{4.5}$$

where $C_1$, $C_2$ and $C_3$ are small constants given by the following relation

$$C_1 = (K_1 L)^2,\ C_2 = (K_2 L)^2 \text{ and}$$
$$C_3 = \frac{C_2}{2} \tag{4.6}$$

Furthermore, $L$=255, $K_1$<<1 and $K_2$<<1.

Using the above equations, we compute the MS-SSIM values as follows

$$MS - SSIM(x,y) = [l_M(x,y)]^{\propto_M} \prod_{j=1}^{M} [c_j(x,y)]^{\beta_j} [s_j(x,y)]^{\gamma_j}. \tag{4.7}$$

where $s_j$ is the structure comparison at the jth scale, $c_j$ is the contrast comparison at the jth scale and $l_M$ is luminance comparison only at the scale M.(Eqn. 4.3, 4.4. 4.5).

Thereafter, the results were normalized to one and utilized to find the exponents in Eqn. 4.7. The obtained parameters are $\beta_1 = \gamma_1 = 0.0448$, $\beta_2 = \gamma_2 = 0.2856$, $\beta_3 = \gamma_3 = 0.3001$, $\beta_4 = \gamma_4 = 0.0448$ and $\propto_5 = \beta_5 = \gamma_5$=0.1333 respectively.

For our computation, the level is assumed to be 5. We have used the above shown (Eqn 4.7), the product rule for our implementations. We use Gaussian smoothing window in this case and apply the sliding window approach. The main idea is to apply low pass filter iteratively and downsample the filtered image by a factor of 2.

### 4.2.5 Multi-scale Structural SIMilarity Box Filter (MS-SSIM B)

In this case, we make a slight change, that is, rather than using the Gaussian window while computing the SSIM level by level, we use the 8x8 block level SSIM computation at each resolution.

## 4.3 Proposed Perceptual Quality Objective Metric

Our work is greatly inspired by the perceptual video quality metric used for the 3D video quality assessment [47]. Rather than dealing with the 3D video components at all, our metrics would assess the quality of 2D video sequences extensively. The proposed metric is called Perceptual Quality Metrics for Two Dimensional Videos. From now we will refer to the new metrics as ‘PQM2D’.

The idea behind the formation of the new metrics is taken from the notion that the luminance value is one of the most crucial and is the essential component that determines the quality of an image. On the contrary, chrominance is basically responsible for colour in the image. In this way, we can say that luminance provides lots of information about the image especially pertaining to the structure of the image rather than the colour of the various objects in the image. Therefore, for video quality assessments, luminance value is crucial. This method is based on the idea of finding the difference between luminance values in the test and impaired frames [47]. As it is obvious that there might be variations in the structure as and when the frames become distorted, there should be prominent deviations in the luminance values. Furthermore, these luminance deviations, when considered at a specific pixel coordinate of reference as well as the impaired frames, gives us meaningful values. That means, the greater the impairment in the structure of the processed frame at a certain pixel coordinate, the greater is the luminance deviation from the reference frame, at that very point.

Let us have a look at the step by step algorithm of the PQM2D metrics.

Let $c_O$ and $c_R$ be the original and distorted frames respectively. Frames are divided into 8x8 blocks and following steps are implemented in each block:

1. Compute the pixel mean, variance and covariance of blocks $b_O$ and $b_R$; i.e $\mu(b_O), \mu(b_R), \sigma^2(b_O), \sigma^2(b_R), \sigma(b_O b_R)$.
2. Compute weighted distortion coefficient for each pixel in the block ($(m,n) \in b_R$):

$$\propto (m,n) = \begin{cases} 0, & \mu(b_o) \leq 1 and\ \mu(b_R) \leq 1 \\ 1, & \mu(b_o) \leq 1 and\ \mu(b_R) > 1 \\ min\left[1, \frac{(c_o(m,n) - c_R(m,n))^4}{\mu(b_o)^2}\right], & \text{else} \end{cases} \quad . \tag{4.8}$$

For contrast distortion in the block, define $K(b_R)$:

$$K(b_R) = 1 + \frac{(\sigma^2(b_o) - \sigma^2(b_R))^2 + 255}{(\sigma^2(b_o))^2 + (\sigma^2(b_R))^2 - 2(\sigma(b_o b_R))^2 + 255} \tag{4.9}$$

3. Perceptual distortion in the whole block is defined as:

$$PDM(b_R) = \frac{K(b_R)}{64} \sum_{(m,n)\in b_R} \propto (m,n) \tag{4.10}$$

After PDM is computed for all blocks, total perceptual distortion in the frame is equal to weighted mean of block distortions:

$$PDM(c_R) = \frac{\sum_{b_R\in c_R} w(b_R) PDM(b_R)}{\sum_{b_R\in c_R} w(b_R)} \tag{4.11}$$

Where weights are chosen inversely proportional to block means:

$$w(b_R) = \begin{cases} 1, & \mu(b_o) = 0 \\ \dfrac{255}{\mu(b_o)}, & \text{else} \end{cases} \tag{4.12}$$

Finally perceptual quality metrics (PQM) is defined as follows:

$$PQM(c_R) = \begin{cases} 0, & PQM(c_R) < 0 \\ 1 - PDM(c_R), & \text{else} \end{cases} \tag{4.13}$$

Frame level PQMs are averaged for the whole video. In order to obtain the overall objective score for a sequence, the scores of the frame level PQM are added up and divided by the number of frames in the sequence and we get the PQM2D score representing the overall quality score judging on a scale of measurement of 0 to 1 where 0 stands for the worst quality and 1 for the best.

The main idea for the PQM is based on the notion that the HVS gives the quality by first measuring the errors in the luminance which in fact comprises of the structure in an image and also is quite less sensitive to the chrominance element of an image [47]. We will study the correlation of PQM2D with the HVS obtained from the MOS scores of the subjective tests in the next chapter.

In table 4.1, we are trying to show a small set of data taken from the obtained results after simulating the above objective metrics, including the PQM2D. In the next chapter, all the considered videos taken from the LIVE database are analysed thoroughly and the results are discussed with respect to the correlation with the Mean Opinion Score (MOS) obtained by the subjective VQA.

Table 4.1: Showing a sample of objective VQA results of the few frames

| **Sequence Name** | **Distortion Type** | **PQM2D** | **PSNR** | **SSIM G** | **SSIM B** | **MS-SSIM G** | **MS-SSIM B** |
|---|---|---|---|---|---|---|---|
| River Bed | Wireless | 0,583 | 27,5 | 0,770 | 0,784 | 0,893 | 0,899 |
| Park Run | IP transfer | 0,683 | 30,53 | 0,896 | 0,912 | 0,981 | 0,984 |
| Shield | MPEG-2 | 0,605 | 28,94 | 0,842 | 0,864 | 0,969 | 0,975 |
| Sunflower | H.264 | 0,724 | 33,94 | 0,916 | 0,918 | 0,966 | 0,970 |

# Chapter 5

# Simulation Results and Discussion

## 5.1 Introduction

In the previous chapters, especially in chapters 4 and 5, we explained the methods and principles of performing subjective and objective quality assessment tests of various digital videos obtained from the LIVE database [28]. Eventually we formed a set of data bank obtained from the subjective and objective VQA. We used 16 human observers as our subjects and followed the standards of ITU for conducting the experiments [26]. Our objective VQA experiments involved 5 different types of state of art metrics and aforementioned proposed PQM2D. The experiments were performed on 9 different kinds of video scenes distorted by 4 kinds of distortions, namely Transmission over wireless networks, H.264 compression, Transmission over IP Networks and MPEG2 compression. The aim of this chapter is to examine the various experimental results obtained from all of our experiments and discuss its relevance with the Human Visual System by assessing the performance of objective metrics. The various performance criterions are monotonicity and accuracy of the objective metrics, which are determined based on correlations with MOS.

## 5.2 Processing of Subjective Quality Assessment Scores

### 5.2.1 Method of Analysis

Our subjective VQA was performed as per the recommendations of International Telecommunication Union (ITU) and we followed the DSIS procedure for the test [26], so the human subjects graded the impaired videos on the scale of 0 to 5. The main idea is to form a proper and reliable mathematical modelling so that we can see the correlation with the simulation results obtained from the objective VQA which has been discussed below.

Now we show the various results of our subjective experiment below.

Number of presentations is given as L=24,

Number of test conditions (distortion types) is given as J=4,

Number of test sequences is given as K=9,

Number of observers is given by N=16,

The number of selected presentations for each distortion type is given as follows: 7 for wireless, 5 for H264, 6 for IP and 6 for MPEG-2. The entire 24 presentations are selected in a way to provide maximum variation in video content, distortion types and level of quality.

### 5.2.2 Calculation of mean scores

Score of observer i for the presentation j is given as $u_{ji}$ and for each and every presentation carried out in the experiment, we find the mean score, given by $\bar{u}_{ji}$

$$\bar{u}_j = \frac{1}{N} \sum_{i=1}^{N} u_{ji} \qquad (5.1)$$

Using the above equation (Eqn 5.1), we form a mean opinion score databank. Furthermore, the obtained MOS is used to judge the performance of our proposed metrics and compare it with various states-of- art metrics.

Figure 5.1 shows the histogram depicting the MOS results. The MOS values have been normalised on a scale of 0 to 100. We can see the variation of scores with respect to the number of videos. Clearly, a wide range of score has been covered.

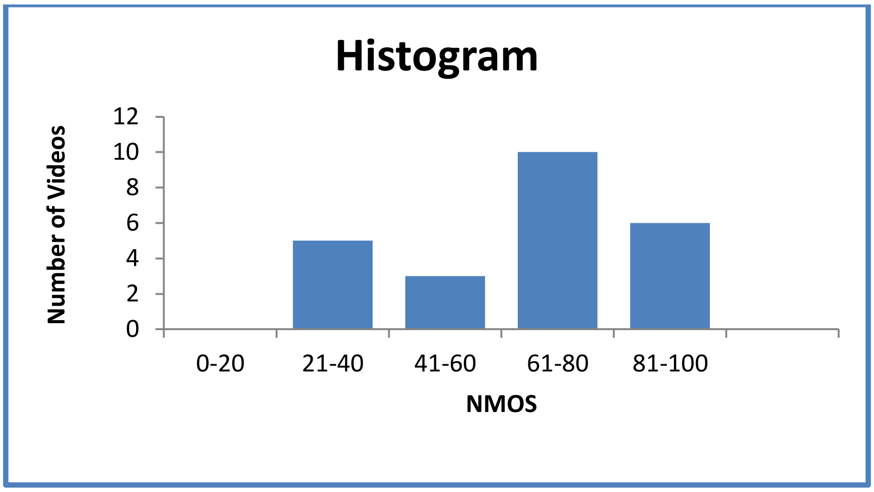


Figure 5.1: Histogram of MOS results

### 5.3 Processing of Objective Quality Assessment results

After performing various objective experiments on the video database discussed previously, we obtain the simulation results for various metrics and compare the performance with respect to the proposed metrics PQM2D.

### 5.4 Metrics Performance

After obtaining various results from subjective and objective experiments, we intend to gauge the performance of the metrics effectively. In order to do so, metrics performance estimation using correlation coefficients were used so that we get the monotonicity and accuracy of objective metrics when comparing with the subjectively obtained mean opinion scores.

#### 5.4.1 Accuracy

It is that attribute of a metric which helps to predict the subjective ratings with minimum average error. It is computed by means of Pearson Linear Correlation Co efficient (PLCC).

With the help of following algorithm, we compute the PLCC

1. MOS and the objective scores are given by length N vector pairs $x_i$ and $y_i$

2. Compute $\bar{x}$ and $\bar{y}$, the means of respective data sets, that is, the MOS and the objective score
3. Ultimately, PLCC is computed as

$$r_p = \frac{\sum(x - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}} \quad (5.2)$$

.

### 5.4.2 Monotonicity

It estimates the association of increase or decrease in the MOS value to the increase or decrease in the objective metrics value, independently of the magnitude of the increase or decrease [30]. It can be measured by the Spearman Rank Order Correlation Co efficient (SROCC). The algorithm below shows the steps to compute the SROCC.

1. MOS vector $x_i$ is sorted in descending order and the indices of the sorted vectors are assigned as rank values for the original scores. The rank vector is called $X_i$
2. Objective score vector $y_i$ is sorted in descending order and the indices of the sorted vectors are assigned as rank values for the original scores. The rank vector is called $Y_i$
3. SROCC is given by $\mathrm{r}_s$

$$r_s = \frac{\sum( X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2}\sqrt{\sum(Y_i - \bar{Y})^2}} \quad (5.3)$$

.

Table 5.1: Comparison of various VQA algorithms

**PLCC**

| | **Wireless** | **H,264** | **IP** | **MPEG2** | **ALL** |
|---|---|---|---|---|---|
| **PQM2D** | 0,916 | 0,942 | 0,924 | 0,956 | **0,883** |
| **PSNR** | 0,537 | 0,915 | 0,713 | 0,918 | **0,722** |
| **SSIM G** | 0,928 | 0,940 | 0,609 | 0,930 | **0,838** |
| **SSIM B** | 0,904 | 0,950 | 0,622 | 0,920 | **0,840** |
| **MS SSIM G** | 0,812 | 0,887 | 0,647 | 0,830 | **0,707** |
| **MS SSIM B** | 0,789 | 0,868 | 0,654 | 0,800 | **0,674** |

**SROCC**

| | **Wireless** | **H,264** | **IP** | **MPEG2** | **ALL** |
|---|---|---|---|---|---|
| **PQM2D** | 0,786 | 0,900 | 0,771 | 0,843 | **0,885** |
| **PSNR** | 0,714 | 0,900 | 0,429 | 0,929 | **0,780** |
| **SSIM G** | 0,964 | 1,000 | 0,600 | 0,929 | **0,877** |
| **SSIM B** | 0,955 | 1,000 | 0,714 | 0,929 | **0,877** |
| **MS SSIM G** | 0,893 | 1,000 | 0,829 | 0,929 | **0,746** |
| **MS SSIM B** | 0,964 | 0,900 | 0,829 | 0,814 | **0,716** |

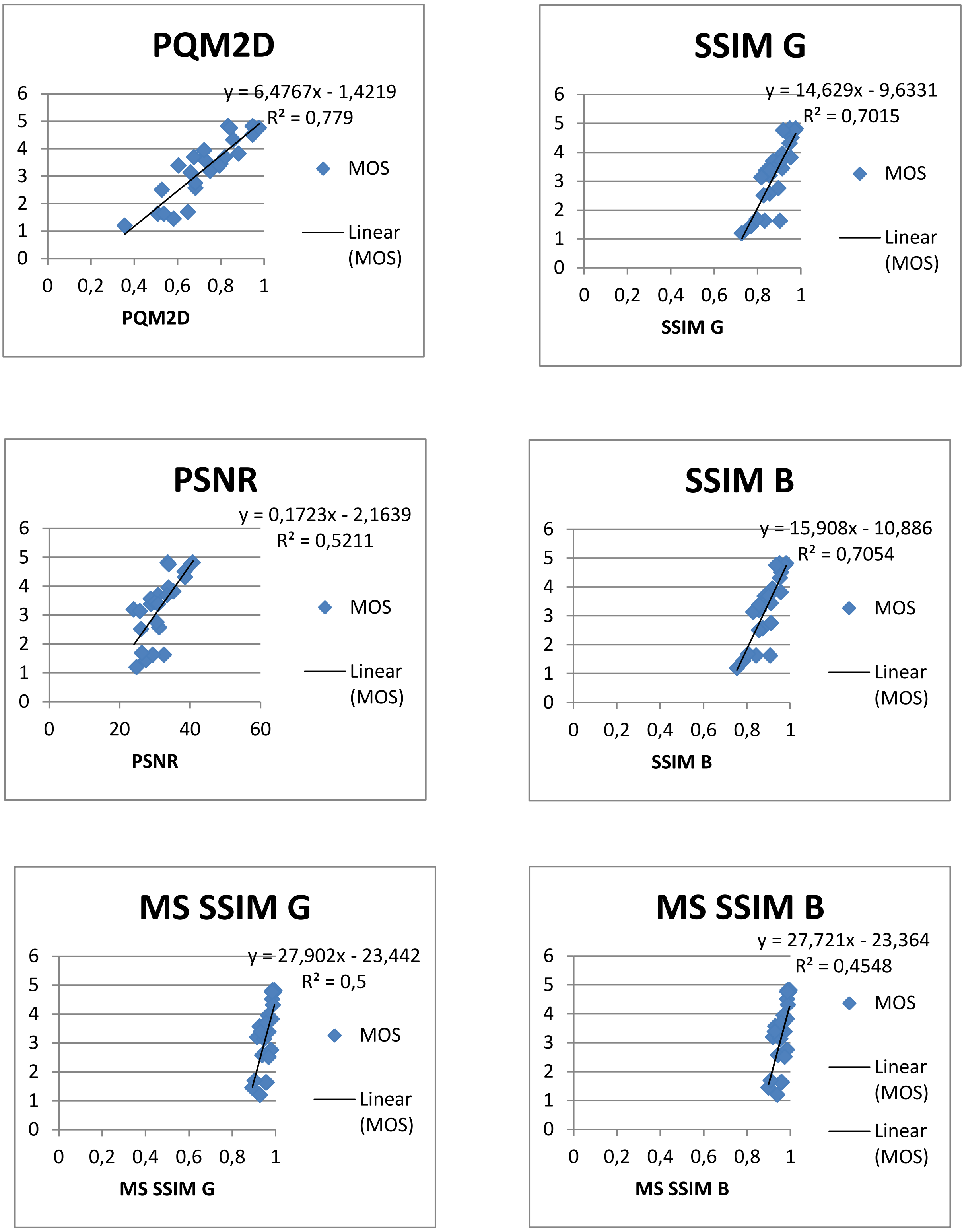


Figure 5.1: Scatter Plots of Objective vs Subjective model

### 5.5 Discussion of Results

The scatter plots have been shown for the various results discussed above. We use statistical procedures of various regression types like exponential, linear, logarithmic, power for finding the best fitting trendline in our data values in order to predict the accuracy of our results. The various equations of the best fitting trendline are also shown in the plot. On the basis of the square of correlation ($R^2$), we fit the best trendline. High value of $R^2$gives us the idea of the best fitting trendline. After comparing all the $R^2$ values for the best fitting line, we found that the linear fit is the best trendline for all our models. Clearly, the PQM2D has the highest value and both the MS-SSIM values are lowest and all the trendlines are apparently showing linear behaviour.

Tables 5.1 shows us the performance estimation of all the objective models with respect to the statistical measures of co efficients of the pearson linear correlation coefficient (PLCC) and the spearman rank order correlation coefficient (SROCC) respectively for all selected video scenes and also individually for each of the four distortion types.

It is clearly evident from the results of the metric PQM2D, with respect to SROCC and PLCC that it outperforms all the other objective models. Our tactfully organised digital video database taken from the LIVE database testifies the drawbacks of PSNR and both MS-SSIM as it is substantially lower than most of the objective models.

When we study the linear correlations based on distortion types, we see that PQM2D is mostly superior like in IP and MPEG2 distortion cases and close to the superior in case of wireless and H264 ones. Nevertheless both SSIM have shown their fairly efficient performance. For example SSIM G and SSIM B perform the best in wireless and H264 distortions respectively. However SSIM G and SSIM B perform poorly for the IP distortions, causing their overall performance to lower than the PQM2D. Likewise MS-SSIM has shown inferior performance in most of the distortion types. Therefore it can be said that PQM2D performs consistently well for all distortion types while other metrics fail for certain types of distortions.

When we study the monotonicity of the model using the SROCC results, we still see that PQM2D has the highest overall correlation score. When distortion types are individually considered, correlation values of PQM2D are fairly close to the best one except for the wireless case where it performs sub optimally. Yet, for the full data, PQM2D again has the highest SROCC scores among the tested metrics.

The higher quality of performance of our metrics PQM2D is elucidated in both the correlation results as it is always slightly larger than SSIM Gaussian and SSIM Box and also is fairly larger than MS-SSIM Gaussian and MS-SSIM Box. Nevertheless, the SSIM results are apparently comparable to the best performing algorithm.

The gist of our discussion is that PQM2D is superior in performance on the carefully formed video sequence bank obtained from LIVE database and thus gives us the perfect picture of our research theme that a robust objective algorithm, well-correlated with the human perceptual experience can provide us the best method to estimate the digital video quality. In other words, a well formed QOS can only be justified when the QOE has been obtained systematically.

# Conclusion and Future Works

In this thesis, we studied about video quality assessment and how the significance of objective model (QOS) leads to an emerging branch of equal importance, subjective model (QOE). We expressed the idea in a lucid way that for assessing the quality of a digital video both the models, QOS or QOE, are equally essential tools. It has been understood that the subjective study can be used to evaluate the influence of present generation video compression and communication technologies on the perceptual quality of digital video. That means, proper validation of QOS model is not possible without the QOE model. In our research on VQA so far, we used 24 distorted videos made from the original samples. We used four distortion types, namely wireless, IP network, H.264 and MPEG-2. The videos were evaluated by 16 subjects.

The obtained results were self-explanatory as we saw in the previous chapter. We expressed the results with the help of graphs, Pearson's Linear Correlation Coefficient and Spearman Rank Order Correlation Coefficient to make it more meaningful. The superiority of PQM2D metrics was self-evident through the substantially higher correlation values and the best fitting lines in the graphs. PQM2D can be considered quite well-correlating with the mean opinion scores obtained from the subjective experiments. Nonetheless, the subjective experiments were conducted carefully and the ITU standards were followed as much as we could. Clearly, we can say at this point that the PQM2D metrics mimics the human visual system better than available metrics such as MS-SSIM, PSNR or SSIM. The drawback of MS-SSIM was also exposed through the correlation results. Through our objective metric discussions, the sensitiveness of HVS to the luminance component is clearly visible. Evidently, we came across several artefacts in our test videos arising due to different types of distortions in our experiment. Seemingly, it is hard to fathom that a single quality evaluation metrics can deal with all kinds of artefacts. In fact different quality metrics may be required to deal with different artefacts efficiently.

The crux of the entire thesis is that complexity of the HVS is still not much known. As we solve the complexity day by day, we can have more reliable and precise results for quality assessment. For future work and in order to enhance the MOS prediction models, other features of HVS can be stressed upon. Another possible enhancement could be made while dealing with the temporal features which are not employed in most of the QOS models. Presumably, incorporating both spatial as well as the temporal component into the QOE model could lead to a rather effective prediction of the QOE.

# References


[1] Farias, Mylene CQ. "*Video Quality Metrics." Digital Video*, Book edited by: Floriano De Rango, ISBN (2010): 978-953.

[2] Amirshahi, Seyed Ali. *Towards a perceptual metric for video quality assessment,*. Masters Thesis report, Gjovik University College, Norway, 2010.

[3] Roque, Luis, *Quality Evaluation of Coded Video,* Masters Thesis, Universidade Tecnica de Lisboa, Portugal, 2009.

[4] "How Technology Has Changed Our Lives in the Last Five Years." - The Center of Technology. N.p., n.d. Web. 15 Jan. 2013.

[5] http://visual.ly/brief-history-digital-video

[6] http://www.real.com/resources/digital-video-file-formats

[7] http://www.screendigest.com/

[8] Yang, Chunting, Yang Liu, and Jing Yu. "Research on Video Quality Assessment." *Journal of Multimedia 7.1*, 49-56, 2012.

[9] Seshadrinathan, Kalpana, and Alan C. Bovik. "Motion-based perceptual quality assessment of video." IS&T/SPIE Electronic Imaging. *International Society for Optics and Photonics*, 2009.

[10] Moorthy, Anush K., Zhou Wang, and Alan C. Bovik. "Visual Perception and Quality Assessment." *Optical and Digital Image Processing: Fundamentals and Applications* (2011): 419-439.

[11] Hekstra, A. P., et al. "PVQM–A perceptual video quality measure." *Signal processing: Image communication* 17.10 (2002): 781-798.

[12] Seshadrinathan, Kalpana, et al. "Study of subjective and objective quality assessment of video." *Image Processing, IEEE Transactions on* 19.6 (2010): 1427-1441.

[13] Brandao, Tomás, and Maria Paula Queluz. "No-reference quality assessment of H. 264/AVC encoded video." *Circuits and Systems for Video Technology, IEEE Transactions on* 20.11 (2010): 1437-1447.

[14] Huang, Zixia, et al. "Towards the understanding of human perceptual quality in tele-immersive shared activity." *Proceedings of the 3rd Multimedia Systems Conference*. ACM, 2012.

[15] Wang, Zhou, and Qiang Li. "Video quality assessment using a statistical model of human visual speed perception." *JOSA A 24*.12 (2007): B61-B69.

[16] http://www.iso.org/iso/home.htm.

[17] Van Moorsel, Aad. "Metrics for the internet age: Quality of experience and quality of business." *Fifth International Workshop on Performability Modeling of Computer and Communication Systems*, Arbeitsberichte des Instituts für Informatik, Universität Erlangen-Nürnberg, Germany. Vol. 34. No. 13. 2001.

[18] Wang, Zhou, Hamid R. Sheikh, and Alan C. Bovik. "Objective video quality assessment." *The handbook of video databases: design and applications*(2003): 1041-1078.

[19] Ebrahimi, Touradj. "Quality of multimedia experience: past, present and future." *ACM international conference on Multimedia*. 2009.

[20] http://ties.itu.ch/ftp/public/itut/fgiptv/readonly/Previous_Meetings/20070122_MoutainView/il/T05-FG.IPTV-IL-0050-E.htm

[21] Yang, Chunting, Lian Zhao, and Zaiyi Liao. "Objective Quality Metric Based on Perception for video." Computer Engineering and Technology, 2009. *ICCET'09. International Conference on*. Vol. 2. IEEE, 2009.

[22] Wang, Zhou, and Alan C. Bovik. "A universal image quality index." *Signal Processing Letters, IEEE* 9.3 (2002): 81-84.

[23] Seshadrinathan, Kalpana, and Alan C. Bovik. "A structural similarity metric for video based on motion models." *Acoustics, Speech and Signal Processing, ICASSP. IEEE International Conference on*. Vol. 1. IEEE, 2007.

[24] Z. Wang, R. Hamid, and A.C. Bovik, *The handbook of video database : design and applications*, CRC Press , New York , 2003.

[25] Canadell Pulido, Josep Antoni. "*Video quality evaluation in IP videoconference between fixed and mobile devices*." (2011).

[26] International Telecommunication Union- Radiocommunication Sector, RECOMMENDATION ITU-R BT.500-13, “ M*ethodology for the subjective assessment of the quality of television pictures*”. (2012)

[27] K. Seshadrinathan, R. Soundararajan, A. C. Bovik and L. K. Cormack, "A Subjective Study to Evaluate Video Quality Assessment Algorithms", *SPIE Proceedings Human Vision and Electronic Imaging*, Jan. 2010.

[28] (2003) LIVE image quality assessment database. [Online]. Available: http://live.ece.utexas.edu/research/quality/subjective.htm

[29] "Video Quality Experts Group (VQEG)." N.p., n.d. Web. 15 Jan. 2013.

[30] Winkler, Stefan. *Digital video quality*. Wiley, 2005.

[31] Sekuler, R. and Blake, R., *Perception*, Random House USA Inc, 1988.

[32] Moorthy, A. K. and Bovik, A. C., "Visual importance pooling for image quality assessment," *IEEE Journal of Selected Topics in Signal Processing*, Issue on Visual Media Quality Assessment, 3, pp. 193–201, 2009.

[33] Hubel, D., Wensveen, J. and Wick, B., "*Eye, brain, and vision*," 1988.

[34] Mante, V., Bonin, V. and Carandini, M., "Functional mechanisms shaping lateral geniculate responses to artificial and natural stimuli," *Neuron*, 58, pp. 625–638, 2008.

[35] Born, R. and Bradley, D., "Structure and Function of Visual Area MT." *Annual Review of Neuroscience*, 28, pp. 157–189, 2005.

[36] Rust, N. C. et al., "How mt cells analyze the motion of visual patterns," *Nature Neuroscience*, 9, pp. 1421–1431, 2006.

[37] Wandell, B., *Foundations of vision*, Sinauer Associates, 1995.

[38] Olshausen, B. and Field, D., "How close are we to understanding V1?" *Neural Computation*, 17, pp. 1665–1699, 2005.

[39] Carandini, M. et al., "Do we know what the early visual system does?" *Journal of Neuroscience*, 25, pp. 10577–10597, 2005.

[40] Daly, S., "Engineering observations from spatiovelocity and spatiotemporal visual models," *Proc. of SPIE*, 3299, pp. 180–191, 1998.

[41] Simoncelli, E. and Olshausen, B., "Natural Image Statistics and Neural Representation," *Annual Review of Neuroscience*, 24, pp. 1193–1216, 2001.

[42] Chandler, D. M. and Hemami, S. S., "VSNR: A wavelet-based visual signalto-noise ratio for natural images," *IEEE Transactions on Image Processing*, 16, pp. 2284–2298, 2007.

[43] Watson, A., Hu, J. and McGowan III, J., "Digital video quality metric based on human vision," *Journal of Electronic Imaging*, 10, p. 20, 2001.

[44] Bovik, A., “Meditations on Visual Quality,” *IEEE COMPSOC ELETTER*, Technology Advances, 2009.

[45] Van den Branden Lambrecht, C. and Verscheure, O., “Perceptual quality measure using a spatiotemporal model of the human visual system,” *Proceedings of SPIE.*, pp. 450–461, 1996.

[46] Masry, M., Hemami, S. and Sermadevi, Y., “A scalable wavelet-based video distortion metric and applications,” *IEEE Transactions on circuits and systems for video technology*, 16, pp. 260–273, 2006.

[47] Joveluro, P., et al. "Perceptual Video Quality Metric for 3D video quality assessment." *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), IEEE*, 2010.

[48] ITU-T P.910, “*Subjective video quality assessment methods for multimedia applications*”, 1999.

[49] Winkler S., “Video quality and beyond,” in proc. of EUSIPCO - *European Signal Processing Conference*, Poznan, Poland, September 2007.

[50] Brandao, Tomas, and Maria Paula Queluz. "No-reference PSNR estimation algorithm for H. 264 encoded video sequences." *proc. of EUSIPCO-European Signal Processing Conference*. 2008.

[51] Wang Z., Bovik A., Sheikh H. and Simoncelli E., “Image Quality Assessment: From Error Visibility to Structural Similarity”, *IEEE T. Image Processing*, Vol. 13, No. 4, pp. 600-612, April 2004.

[52] ITU-T J.247, “*Objective perceptual multimedia video quality measurement in the presence of a full reference*”, 2008.

[53] Ries, M., Nemethova, O. and Rupp, M., ”Motion based reference-free quality estimation for H.264/AVC video streaming”, in *Proc. of IEEE Int. Symp. on Wireless Pervasive Computing (ISWPC)*, San Juan, Puerto Rico, US, Feb. 2007.

[54] Z. Wang, E. Simoncelli, A. Bovik, and M. Matthews, “Multiscale structural similarity for image quality assessment,” in *IEEE Asilomar Conference on Signals, Systems and Computers*, 2003.

[55] M. H. Pinson and S. Wolf, “A new standardized method for objectively measuring video quality,” *IEEE Trans. Broadcast.*, vol. 50, no. 3, pp. 312–322, Sep. 2004.

[56] Kalpana Seshadrinathan and A.C. Bovik, "Motion Tuned Spatio-temporal Quality Assessment of Natural Videos," vol. 19, no. 2, pp. 335-350, *IEEE Transactions on Image Processing*, Feb. 2010.

[57] B. Girod, “What’s wrong with mean-squared error,” in *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge, MA: MIT Press,1993, pp. 207–220.

# Curriculum Vitae

Rahul Gaurav was born in Patna, India. He received the Bachelor of Technology in Electrical Engineering from Biju Patnaik University of Technology, India in 2009. He worked as a Project Manager at one of the leading IT organisations of India for more than a year. Later he joined at ISIK University as a Researcher for TÜBİTAK (Scientific and Technological Research Council of Turkey) project from 2011 to 2013. His research interests include Processing and Quality Assessment of Images and Videos and Mathematical Morphology.

**Publications**

[1] H.F. Ates, Rahul Gaurav, "Decoder Side Candidate Motion Vector Selection for B-frames in H.264 Video Coding" in the proceedings of 20th Signal Processing and Communications Applications Conference at Fethiye, Muğla, Turkey, 2012.

[2] Rahul Gaurav, "Development of NanoPiezoSensor and Electrical Characterization using Impedance Spectroscopy" in the proceedings of 8th International Nanotechnology Symposium (Nanofair) at Dresden, Germany. 2010.

[3] Rahul Gaurav , "A Mathematical Morphological Perspective in the World of Images" in the proceedings of Seminar on Spatial Information Retrieval, Analysis, Reasoning and Modelling at Indian Statistical Institute, Bangalore, India, 2009.