Master Thesis
Electrical Engineering

Performance analysis of no-reference video quality assessment methods for frame freeze and frame drop detection

AUTHOR: Bharat Reddy Ramancha

School of Engineering
Blekinge Institute of Technology
37179 Karlskrona
Sweden

This thesis is submitted to the School of Engineering at Blekinge Institute of Technology in partial fulfillment of the requirements for the degree of Master of Science in Electrical Engineering. The thesis is equivalent to 20 weeks of full time studies.

Contact Information

Author:
Bharat Reddy Ramancha, E-mail:bharat.reddy0@gmail.com

University advisor: Muhammad Shahid
University Examiner: Dr. Benny Lövström

School of Engineering
Blekinge Institute of Technology
37179 Karlskrona
Sweden

Internet: www.bth.se/ing
Phone: +46 455 385000
SWEDEN

Abstract

Data transmission through a network has become a challenging task considering the loss of data factor. This is similar in the case of video transmission through a network. Error prone channels present in the network are responsible for the quality degradation of decoded video. There must be some metric monitoring the quality of this decoded video. Lot of research work is done in this field and many metrics are designed for video quality estimation considering various factors. Most of the metrics are designed considering only few factors thereby limiting their scope for few videos. In this research we investigate few metrics and propose the best metric by implementing them on different videos. Best metric proposed gives the video quality estimation independent of resolution and artifact in a video. Subjective analysis is conducted for validation of our results. Metrics that are implemented belong to NO-REFERENCE video quality metric group.

Keywords: Video quality, No-reference, frame freezing, frame dropping, objective measurement, Subjective measurement.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Digital transmission of video is very important and challenging task. Due to increase in technology and modernization many new methods are developed for transmission of videos digitally. These digital video systems have replaced many analog video systems as there are many advantages compared to analog video systems.

Digital video transmission system chain consists of four major blocks acquisition, compression, transmission and reconstruction. Acquisition refers to source or library of video, where videos are generated. In compression block the acquired video is compressed with available encoders in the market to limit the bandwidth. In this process of limitation of bandwidth many parameters like frame rate, bit rate etc can be altered. The quality degradation due to compression of video can be categorized into two types [2] spatial and temporal. Next stage is the transmission of video through a channel. Transmission of these video through error porne channels causes packet loss, band width fluctuation, jitter and delay etc. Video transmitted through these channels is then reconstructed and displayed. Video transmitted in this process gets degraded and loses its quality.

Video quality assessment of reconstructed videos is very important to judge the efficiency of the transmission system. This help monitoring the digital video transmission system to maintain quality levels during display. There are many factors that effect the video quality and many research works are done exploring these factors and methods to assess the video quality.

In this thesis report two research questions have been answered. Firstly, need for the performance analysis of metrics. Secondly, selecting the best metric among the studied metrics. Performance analysis of video quality assessment methods has its advantages like scope to improve the existing metric and selecting the better metric for video quality assessment. We

have compared different metrics that have been proposed to assess video quality and suggest the best method among them. No- reference model was our main criteria in this work. We have discussed mainly the impact of frame dropping and frame freezing on video quality.

This report is organized in five chapters. Chapter 2 gives the background work done on factors effecting the video quality and highlighting the important factors. Chapter 3 gives brief description about the quantification of video quality and different methods and models used for it. Chapter 4 describes the different metrics used for comparing, this is main aim of our thesis. Chapter 5 reveals about the results. Last in the conclusion part, we conclude our results and suggest the best metric for video quality assessment in the scenarios of frame freeze and frame drop.

Chapter 2

Background

2.1 Overview of Video Quality Assessment

Video quality is a measure of perceived video degradation passed through a channel and evaluation of video quality based on certain assumptions is referred as video quality assessment. No channels are ideal and these channels introduce some distortions affecting the video quality. There are many reasons behind video quality degradation during transmission through a channel. There are many video based applications developed now a days and there is a necessity for the prediction of video quality to maintain standards of applications [3].

2.1.1 Packet loss

When a video is transmitted through an IP network, it is done in the form of packets. Due to errors in the network some of these packets are not transmitted successfully. This results in packet loss. Hardware misconfiguration, limitation of band width and duplex mismatching are some of the reasons for packet loss while transmitting a video. Packet loss is considered as one major error in digital transmission of videos. Packet loss has huge impact on perceived video quality. Packet loss can be a random loss or it can occur in blocks. Both have their own affect on video quality. Impact of packet loss on video quality may differ with type of packet or content in it. Different packet types with different loss distance and loss frequencies may lead to different video quality [2]. Packet loss impact on video quality can be detected when it is identified. Research works are done on the visibility of individual packet losses in MPEG video [4].
Two effects of packet loss mainly discussed in this report are

- Frame freezing
- Frame dropping

Frame freezing

Frame freezing is defined as freezing of a frame for certain duration of time. When packet loss occurs some frames are lost, in order to avoid discontinuity in display the same frame is repeated for certain duration until a new frame is available. Perceptual impact of frame freezing is mainly content based. Frame freezing results in irregular display of video. Identification of these frozen frames is very important task in video quality assessment techniques. Multiple freezes also have major impact on video quality. Relation between probability of detection and duration of dropped frames was proposed [5], considering this relation metric has been proposed detecting the effect of multiple freezes [6].

Many research works are done on identification of frozen frames and predicting the quality based on frozen frames. But, predicting video quality based on number of frozen frames is not an appropriate way of video quality assessment as there are other factors that effect the quality of the video along with frozen frames. Most basic method for identification of a frozen frame is by computing motion energy between two frames. If a frame is considered to be frozen then motion energy between such two frames is zero as shown in Fig 2.1. In real time applications a single but long interval frame freezing has less impact than multiple short frequent freezes [7].

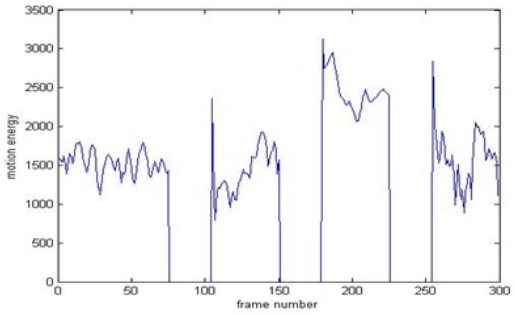

Figure 2.1: Shows frame freezing at three locations

Frame Dropping

Frame dropping is defined as the dropping of frames at certain locations. This is done to reduce frame rate of video to maintain overall rate of playback with audio. Frame dropping has a negative impact on perceptual visual quality of a video [8]. Detecting dropped frames and determining video quality from it is explained in Chapter 3. Effect of single frame dropped location on video quality differs with effect of multiple frame dropped locations. Regular frame dropping (i.e reducing frames alone) has less impact

on video quality when compared to irregular frame dropping [5]. It can also be said that several number of small duration of frame dropping has more impact when compared to single large duration.

Frame dropping is identified when there is a sudden change in motion intensity between two frame.In the Fig 2.2 it is seen that at two locations there is sudden drop of motion intensity indicating frame drop at those locations. In real time applications a single but long interval frame dropping has less impact rather than multiple short frequent dropping.

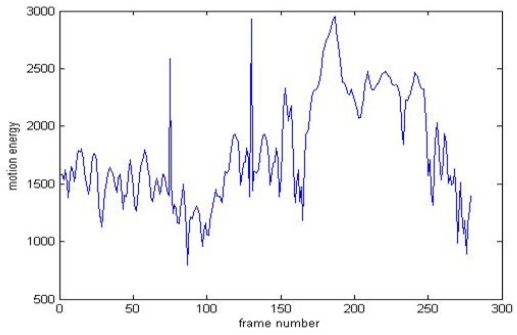

Figure 2.2: Shows frame dropping at two locations i.e sudden change in motion energy

2.1.2 Spatial Distortion and Temporal Distortion

Spatial distortion occurs at the image acquisition or at the image compression stage where the video is up sampled or down sampled with respect to the frame size. This degrades the perceptual quality of video as extra pixels will be added when up sampled where as few pixels are removed when down sampled. Generally blockiness,blurring ringing effect are referred to spatial distortions. Blockiness refers to extended block effect and false contouring effect. Blurring in a video is caused mainly due to incorrect focus of a frame or due to irregular movement of subject. Blurring has very bad impact on perceptual quality and human eye. Ringing effect is also considered as spatial distortion as it effects the pixel values because of ringing phenomena.

Temporal distortion occurs mainly in transmission of video through the error prone network channel. When a video is transmitted through an error prone network channel parameters such as frame rate, bit rate etc will be effected and this is observed at the receiver end when video is displayed.

Some of the important forms of temporal distortions are jerkiness, delay, flickering, freezing and frame dropping. Frame freezing, frame dropping and their impact on video quality are explained in previous section.There are metric proposed considering parameters obtained from both spatial approach and temporal approach involving motion activity density of a video as controlling factor [9]

2.1.3 Related work and Motivation

There are many no-reference video quality assessment metrics proposed by many researchers to provide better video quality assessment for all videos. But, not all of these proposed metrics were successful in all cases. It is observed that few metrics proposed are purely based on certain artifacts i.e only certain artifacts are assumed in a video. But in real time there are many artifacts that degrade the video quality. For example [10] [11] have been proposed by considering the frame freezing and frame dropping artifacts. Though they produce better results in identifying the frame freezing and frame droppings but video quality estimation basing upon only one artifact is not enough.

Many video quality metrics are available in real world. But, not all of these metrics predict better video quality for all types of videos. Performance analysis of such metrics has to be done to provide better video quality assessment. Finding the best metric that can provide precise video quality rating is the motto of this work. Many metrics have been studied and implemented. Metrics that provided better results are lead to comparison of their performance. The metrics used in this work are applied to different videos with different artifacts.

2.2 Methods to quantify Video quality

Two basic methods for quantifying the video quality are

- Subjective Video Quality Measurements
- Objective Video Quality Measurements

2.2.1 Subjective Video Quality Measurements

Subjective video quality assessment is one of the most accurate method to estimate video quality rating [12]. In subjective testing, a number of videos are displayed to viewers and they are rated.Generally the ratings are given on the scale of 1 to 5 , where 5 for good quality video and 1 for low quality video.Mean Opinion Score (MOS) is calculated by taking average over all viewers.Mean opinion score is most widely accepted subjective method [13].

Subjective video quality assessments is carried on recommendations specified by International Telecommunication Union (ITU) such as ITU-R Rec. BT.500-11 [14].

There are many methods to conduct these subjective measurements as specified in ITU recommendations. Some of them are Double Stimulus Continuous Quality Scaling (DSCQS), Single Stimulus Continuous Quality Evaluation (SSCQE) etc. In our work SSCQE is used, following the methodologies used in [15].

Although this method estimates the accurate result, there are few problems while measuring video quality. Firstly, number of viewers required is not predefined and in the case large number of viewers it is not possible to conduct the testing in a short time. Secondly, high preparation cost and time involved in conducting in subject test.

2.2.2 Objective Video Quality Measurements

Objective video quality measurement is another method to quantify video quality [15]. Generally, this is done by designing some mathematical model involving the parameters that effect the video quality. This mathematical model predicts the video quality and can be automatically evaluated by computer program. Video quality values obtained from subjective testing are compared with values obtained from objective testing to know reliability of the metric.

There are many methods in objective quality measurements and they differ on the availability of reference video in predicting the video quality. They are

- Full-Reference (FR)
- Reduced Reference (RR)
- No-Reference (NR)

Full reference

Full Reference method refers to availability of original video for video quality estimation of distorted video. In this method pixel wise comparison is done for each frame for original and distorted video. The original video which is referred as reference video, is an uncompressed version.

This method is restricted only to the offline video quality measurement as in real time applications in case of online or live broadcasting, availability of original video is not possible.
In such cases immediate result is required to enhance the video quality that cannot be provided by FR method. But, a precise analysis of video quality can be done using this method for offline videos. Full reference metric can estimate the video quality of compressed videos effectively [16].

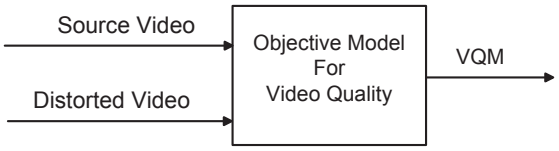

Figure 2.3: Block diagram for FR method

Reduced Reference

Reduced reference is similar to FR method but only few parameters of original video and distorted video i.e partial information of videos are taken into consideration and compared to estimate video quality of distorted video. This reduces the operational time when compared to FR method. Availability of original video is not essential. This holds as major advantage for RR over FR method.

RR method is useful in many application such as tracking image quality degradation and controlling the steaming resources. Results obtained from RR method can be comparable to FR method. Similar to FR metric RR metric is also practically limited. For example the case discussed in FR metric

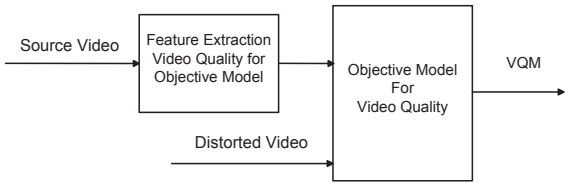

Figure 2.4: Block diagram representation of RR Method

No Reference

The main goal of this method is to develop a computational model that predicts the video quality without any prior reference videos. NR metrics can be used for both offline and online video quality assessment. These are more suitable for in service quality monitoring. In real time applications information about the type of distortion is not available. No additional network is required for reference data to assess video quality. This feature of no-reference makes it more applicable for situations like real time streaming, broadcasting etc. Implementation of NR metric is not as simple when compared to FR and RR metrics.

Metrics designed in this method may or may not have assumptions based on type of distortions. But, while evaluating, it does require some examples with similar distortion for training. There are few metrics that consider few important factors before designing metric in this method making the computation easier. For example model proposed in [17] uses quantization parameter, the motion and bit allocation factor and also few characteristics of HVS. With increase in technology in video handling devices there is a need for no-reference video quality prediction based on the available encoding parameters and providing them as an input to the artificial neural network [18]. Computational complexity in neural network based method increases with increase input space. To overcome this feature of complexity a machine learning based model has been proposed which uses support vector machine (SVM) and extracted visual quality bitstream parameters [19]. Many FR metric are predicted using this model and produced good results.

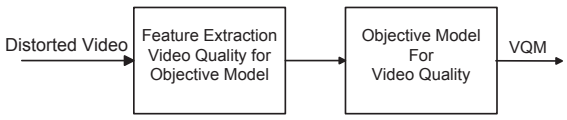

Figure 2.5: Block diagram NR metric analysis

Mean Squared Error (MSE) Peak Signal to Noise Ratio (PSNR)

Both MSE and PSNR are simple and easy methods making them most commonly used methods in objective video quality assessment. They are fast to compute. However, these methods only predict the approximate value of perceived video quality. They are based on pixel to pixel comparison of video frame irrespective of the video content. For example when we consider motion of objects in a video i.e fast motion videos and slow motion videos have different perceptual quality but this is not considered in MSE and PSNR calculations. MSE and PSNR also ignore the spatial relationship of

the pixels and they cannot accurately predict video quality experienced by human. There are metrics that estimate video quality based on MSE and video content [15] PSNR and MSE are calculated as follows:

$$PSNR = 10 \log \frac{{MAX_t}^2}{MSE(n)} \tag{2.1}$$

MAX_t is maximum pixel value and MSE is average of square of difference between luminance values of corresponding pixels between two frames.

$$MSE = \frac{1}{UV} \sum_{u=1}^{U} \sum_{v=1}^{V} [I_R(u,v) - I_D(u,v)]^2 \tag{2.2}$$

$I_R(u,v)$ is intensity value of reference video frame at pixel location (u, v) and $I_D(u,v)$ is intensity value of distorted video frame at pixel location (u, v). U and V are number of rows and columns in a video frame.

PSNR is a widely used quality metric for measuring the video quality of lossy compression

PSNR is calculated for entire sequence of video of length N

$$PSNR = \frac{1}{N} \sum_{n=1}^{N} PSNR(n) \tag{2.3}$$

SSIM: Structural Similarity Index is quality metric that predicts the structural similarity among two frames. SSIM is considered as full reference metric as it measure the image or frame quality based on uncompressed or distortion free image or frame. SSIM can be used as an alternative for measuring the perceptual video quality. SSIM considers frame degradation as perceptual change in a video i.e the pixels have strong dependencies when the y are spatially close. SSIM is calculated as follows.

$$SSIM(n) = \frac{[2\mu_{I_R}(n)\mu_{I_D}(n) + C_1][2\sigma_{I_R I_D}(n) + C_2]}{[{\mu_{I_R}}^2(n) + {\mu_{I_D}}^2(n) + C_1][{\sigma_{I_R}}^2(n) + {\sigma_{I_D}}^2 + C_2]} \tag{2.4}$$

$\mu_{(I_R)}(n), \mu_{(I_D)}(n)$ are mean intensity of nth frame of reference (I_R) and distorted (I_D) video sequence, $\sigma_{(I_R)}(n)$ and $\sigma_{(I_D)}(n)$ are contrast of nth frame of reference (I_R) and distorted (I_D) video sequence. C_1, C_2 are constants used in order to evade any instabilities in the structural similarity comparison.
SSIM is calculated for entire sequence of video of length N

$$SSIM = \frac{1}{N} \sum_{n=1}^{N} SSIM(n) \tag{2.5}$$

Perceptual Evaluation Video Quality(PEVQ):

PEVQ provides Mean Opinion score on the scale 1-5 for a particular video sequence. This is used for predicting the video quality for videos resulting from various applications like TV, video call, streaming video etc. Quality of video is predicted based on modeling the behavior of HVS characteristics. PEVQ is accepted by ITU-T recommendation in VQEG [20].
We have used PEVQ and SSIM for all the videos used in this research work. The obtained results are analyzed with subjective readings and the outlier ratio is calculated to predict the consistency of the metric.

Analysis of Objective Results with Subjective Results:

Validation of the objective results are done by analyzing them with subjective results. Correlation between objective and subjective results shows the validation of results. Pearson correlation co-efficient (PCC) and Spearman correlation co-efficient (SCC) are most commonly used in this case. PCC is defined as co-variance of two variables divided by product of standard deviation of the variables. PCC is value ranging from -1 to 1with 1 resembling the perfect relation between variables. SCC is defined as PCC between ranked variables.SCC is a non-parametric measure of two variables that are statistically dependent. Similar to PCC, SCC also range from -1 to 1.

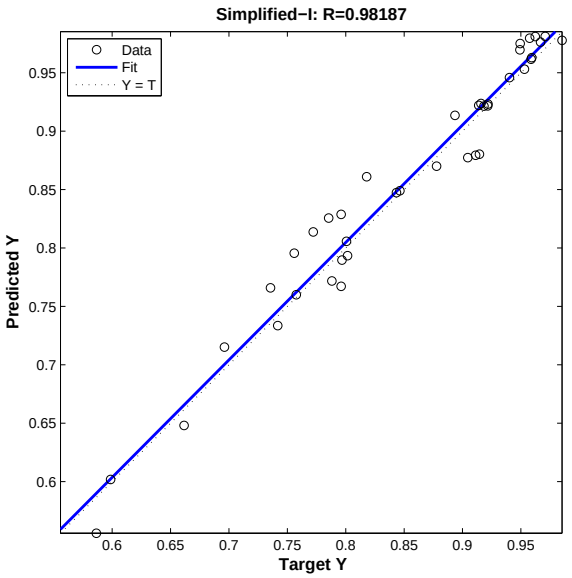

Figure 2.6: Regression plot of subjective and objective scores

The relation between subjective and objective results is shown in Fig 2.6. Both subjective and objective values are plotted within the range [0-1].

Subjective testing for the videos in our research work is based on Single Stimulus Quality Evaluation (SSQE) [14].

In this chapter three metrics of no-reference video quality assessment are discussed. They are

- A model of jerkiness for temporal impairments in video transmission[1]
- No-reference temporal quality metric for video impaired by frame freezing artefacts[10]
- A No Reference (NR) and Reduced Reference (RR)Metric for Detecting Dropped Video Frames[11]

Performance of these metrics with reference to MOS will be analyzed.

2.3 Model of jerkiness for temporal impairments in video Transmission by S. Borer[1]

Many research works are done in designing a metric for computing amount of jerkiness and evaluating video quality. These metric's differ with each other with their mathematical forms and evaluation methodologies. Similarly, a model of jerkiness has been proposed for temporal impairments with different evaluation methodology and mathematical form. Compared to previous research works this metric is not limited to few resolutions and fixed frame rate. This metric can be applied to videos of resolution varying from QCIF to HD. Mathematical form of this video is more general and can be applied for variable frame rate video sequences. Amount of jerkiness for the given video sequence is the final outcome of the metric.

2.3.1 Parameter description:

Two basic parameters considered in this metric are time stamping and motion intensity. Time stampings is defined as display time of each frame. Generally a video sequence is in the form of

$$v = (f_i, t_i) i = 1, ...n \tag{2.6}$$

Where f_i denotes frame number that is display at time t_i. Display time can be calculated as

$$\Delta t_i = (t_{(i+1)} - t_{(i)}) \tag{2.7}$$

Another important parameter is motion intensity that defines the motion of objects in a video. Motion intensity is the inter frame distance between

each frame.

$$m_{(i+1)}(v) = \sqrt{\sum_{x}(f_{(i+1)}(x) - f_i(x))^2)} \tag{2.8}$$

where $f_i(x)$ denotes the Y (luminance value) components of frame i at location x.

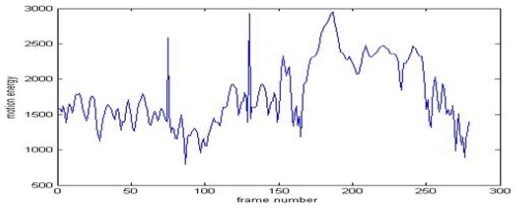

Figure 2.7: Shows the motion intensity of sample video

2.3.2 Jerkiness evaluation:

Input to this model is video sequence with no frame repetitions. Consider two video sequences one with no frame freezing and other with frame freezing at isolated location. Observing the motion intensity in both the video sequences, in second video sequence we have zero motion intensity at some location indicating no motion and there is sudden jump in motion intensity at the end of freeze interval.

Following considerations are made in design of measure for jerkiness. First, the motion intensity function should be a monotone function as increase in motion intensity increases jerkiness. Second, zero motion intensity resembles zero jerkiness value. Third, for larger motion intensities jerkiness value might saturate and for small motion intensity perceptual impact of jerkiness might be small. Thus it is defined as parameterized S-shape function. Fourth, similar to motion intensity jerkiness increases with increase in time stampings. Therefore time stamping function is also considered to be monotone function.
Jerkiness is the sum of the product of relative display time and monotone function of display time and monotone function of motion intensity

$$J(v) = \frac{1}{T}\sum \Delta t_i T_\alpha(\Delta t_i) u(m_{(i+1)}(v)) \tag{2.9}$$

Here T_α and μ are the S-shape function or monotone function of display time and motion intensity respectively.
S-shape function is defined as

$$f(x,y) = \begin{cases} a.x^b & \text{if } (x \le p_x) \\ \frac{d}{1+exp(-c*(x-p_x))}1 - d & \text{else} \end{cases}$$

Where $a = \frac{p_y}{\frac{qp_x}{p_y}}$
p_x
$b = \frac{qp_x}{p_y}$
$c = \frac{4q}{d}$
$d = 2(1 - p_y)$

S-shape function used is shown in Fig 3.3. The parameters (p_x, p_y, q) are different for motion intensity and display time. For motion intensity the values are given by $(p_x, p_y, q) = (5, 0.5, 0.25)$ and for display times time $(p_x, p_y, q) = (0.12/c, 0.05, 1.5.c)$. Value c is different for different resolutions. c is 1 for QCIF, 1.36 for CIF, 1.08 for QVGA and 1.18 for VGA resolution
In previous works jerkiness is measured only along high motion intensity or

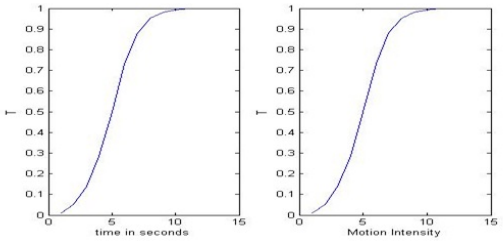

Figure 2.8: S-Shape function used for both display time and motion intensity

high bit rate or long display time. But, in this metric it is measured along both high motion intensity and long display time. lower the frame rate more is the jerkiness as there will be high motion intensity at the end of the frame drop.

2.4 Identification of frozen frames:

Metric [10] deals with frame freezing impairments on video quality and is applicable to both no-reference and full reference video quality assessment models. Our research is limited to only no-reference quality assessment model. In no-reference approach, frozen frames are identified only on the processed sequence.

Basic idea behind identifying frozen frames is calculating Mean Square Error (MSE) between current frame and previous frame. MSE between two frames is calculated according to the equations 3.5, 3.6 and 3.7 where W and H are frame width and height respectively in pixels and I represents current frame. Input video is converted to YUV color space.

$$YM_1(i) = \frac{1}{(W * H)} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} (Y(x, y, i) - Y(x, y, i-1))^2 \tag{2.10}$$

$$UM_1(i) = \frac{1}{(W * H)} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} (U(x, y, i) - U(x, y, i-1))^2 \tag{2.11}$$

$$VM_1(i) = \frac{1}{(W * H)} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} (V(x, y, i) - V(x, y, i-1))^2 \tag{2.12}$$

If MSE between current and previous frames is zero then the frame is considered as frozen frame. But, in real time, video capturing involves slight signal noise in the signal or slight change in the luminance values of the pixels and this can lead to false detection of frozen frames. To avoid this false detection of frozen frames a non-zero threshold for MSE is used. On other hand, there may be case for video with low motion video content where MSE is a non-zero low value. This may also lead to false detection of frozen frames . Following steps are taken to identify frozen frames.
Step1: Firstly, potential frozen frames are identified by computing MSE between current frame and previous frame based on equations 3.5 3.6 and 3.7.
Step 2: Potential frozen frame is also checked against first frame of the freeze event i.e MSE is calculated between potential frozen frame and first frame of the freeze event.

$$YM_2(i) = \frac{1}{(W * H)} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} (Y(x, y, i) - Y(x, y, i-k))^2 \tag{2.13}$$

$$UM_2(i) = \frac{1}{(W * H)} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} (U(x, y, i) - U(x, y, i-k))^2 \tag{2.14}$$

$$VM_2(i) = \frac{1}{(W * H)} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} (V(x, y, i) - V(x, y, i-k))^2 \tag{2.15}$$

where (i-k) is the temporal index of the first frame of the freeze event in which frame i might be.
Frame i is said to be frozen when MSE between current and previous frames but also first frame of the freeze event is less than a non-zero threshold value.

$$FreezeFlag(i) = \begin{cases} 1 & \text{if } (YM, UM, VM)_{1,2} < T \\ 0 & \text{otherwise} \end{cases}$$

In this algorithm T is empirically set as 1.

Now, a metric is designed based on the histogram of occurrences and histogram of duration for freeze events. Variables FrDur and FrTotDur are defined. These variables represent the duration of individual freeze event (in ms) and total duration obtained by multiplying number of occurrences by its associated duration (in ms) respectively. Values for these variables are obtained from the histogram of occurrences and histogram of durations.

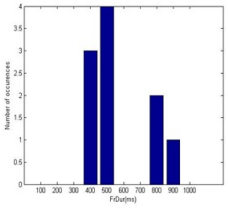

Figure 2.9: Number of occurrences of freeze events

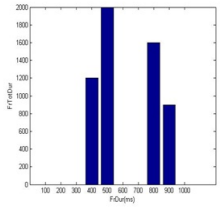

Figure 2.10: Total duration of freeze events

Next, these values are normalized with respect to total time duration of video (in ms) for each bin in histogram as shown in equations

$$FDP(b) = \frac{FrDur(b)}{TotDur * 100} \tag{2.16}$$

$$FTDP(b) = FrTotDur(b)TotDur * 100 \tag{2.17}$$

Where b represents bin in histogram and TotDur represents total duration of video (in ms). FDP(b) and FTDP(b) are assigned as the variables for the mapping function to provide good correlation with subjective data. The mapping function is defined as:

$$T1(b) = \frac{1}{(f2(FTDP(b)) * f1(FDP(b)) + f3(FTDP(b)))} \tag{2.18}$$

$$f1(x) = a1 + b1 * log(c1 * x + d1) \tag{2.19}$$

$$f2(x) = a2 * x^2 + b2 \tag{2.20}$$

$$f3(x) = a3 * x^2 + b3 \tag{2.21}$$

The constants in equations are derived from subjective data based on least square regression. $a1 = 5.767127, b1 = -0.580342, c1 = 3.442218, d1 = 3.772878, a2 = -0.00007, b2 = -0.088499, a3 = 0.000328, b3 = 0.637424$ Each T1(b) obtained is bounded in the range of [1 5].

$$\acute{T1}(b) = min(max(T1(b), 1), 5) \tag{2.22}$$

Minimum value obtained from the set of $\acute{T1}(b)$ is the temporal quality of the video as shown in equation. The output value ranges from 1 to 5, where 1 being poor and 5 being excellent quality. If the temporal quality obtained is near to 5 then video has less frame freezing impairments and if the temporal quality obtained is near to 1 then video has more frame freezing impairments.

2.5 Identification of dropped frames

The detection of the dropped frames is based on an algorithm developed by Wolf [5].
The NR algorithm developed estimates the number of dropped frames in a video sequence and their temporal locations.
NR algorithm developed is mainly divided into two sections:

- Computing motion energy time history
- Examining motion energy time history to identify dropped frames.

Computation of motion energy time history: Computing motion energy gives us clear view about the motion of video and temporal location of sudden change in motion.

Step1 Compute the Temporal Information (TI) difference sequence given by

$$TI(i,j,t) = Y(i,j,t) - Y(i,j,t-1)t = 2,3,..N \tag{2.23}$$

This is pixel wise luminance difference between frames.

Step2 A threshold is set for the image motion (i.e M_{image}). All the values below this are set to zero. This is done to eliminate low level noise. M_{image} can be adjusted for higher values.

$$TI(i,j,t) = \begin{cases} TI(i,j,t) & \text{if } abs(TI(i,j,t)) > M_{image} \\ 0 & \text{otherwise} \end{cases}$$

Step3 Square TI to convert from amplitude to energy and compute the mean of each video frame. Here the resulting time history frame- by -frame values that contains the motion energy will be represented as TI2 and computed as

$$TI2(t) = MeanTI(i,j,t)^2 \tag{2.24}$$

This TI2 waveform is examined to locate dropped frames.

Examining motion energy time history to locate dropped frames: In a video sequence scenes differ with the amount of motion they carry (for example: a pedestrian video or football video). TI2 determines amount of motion of one frame over the previous frame as it is the average motion energy of temporal information differences in pixels.

A dynamic threshold value is generated to determine the dropped frames. This threshold can be varied with motion i.e it can be increased when motion increases and vice versa. This concludes that even though dropped frames contain more residual motion they can still be perceived as dropped frames.

Step4 Compute the average TI2 value $TI2_{ave}$ for the scenes in the whole video.

$$TI2_{ave} = meanTI2_sort(k) \tag{2.25}$$

TI2 values are sorted from low to high
Scene cut problem may arise in few cases. Hence while computing $TI2_{ave}$ influence of scene cut must be considered as scene cuts may lead to high TI2 values affecting $TI2_{ave}$ values.

$$ceil(F_{cut} * (N-1)) \leq k \leq floor(1 - F_{cut}) * (N-1) \tag{2.26}$$

F_{cut} is the percentage of scenes cut to eliminate before computing the average.

Step5 Compute the dynamic factor dfact as :

$$dfact = a + b * log(TI2_{ave}) if (dfact < c) \tag{2.27}$$

then set dfact=c. Here a, b and c are positive constants as mentioned in table 3.1. This equation certifies that perception of frame drops is linearly dependent upon the log of the average motion energy

Step6 Compute the Boolean variable drop (equal to 1 when frame is drop is detected , otherwise equal to 0)

$$drops(t) = \begin{cases} 1 & \text{if } TI2(t) \leq dfact * M_{drop} \\ 0 & \text{otherwise} \end{cases}$$

Step7 Compute the Boolean variable dips (equal to 1 when a frame dip is detected, otherwise equal to 0) as

$$dips(t) = \begin{cases} 1 & \text{if } TI2(t) \leq dfact * M_{dip} \text{ and } dips_{mag}(t) * A_{dip} \\ 0 & \text{otherwise} \end{cases}$$

where $dips_{mag}$ is a function that finds the magnitude of the dips and is given by

$$dips_{mag}(t) = minTI_2(t-1) - TI_2(t), TI_2(t+1) - TI_2(t) \tag{2.28}$$

if $dips_{mag}(t) < 0$ then $dips_{mag}(t) = 0$ The endpoints (i.e., t =2 and t=N) of the dips are set equal to 0 since the $dips_{mag}$ function is undefined for these two data points.

Step8 Some frames are detected as both dips and drops. Therefore fraction of dropped frames FDF can be calculated by computing logical or between dips vector and drops vector and dividing it number of samples.

$$FDF = \sum(dropsORdips)/(N-1) \tag{2.29}$$

Table 2.1: Constants mentioned in metric

Parameter	Value
M_{image}	30
F_{cut}	0.02
a	2.5
b	1.25
c	0.1
M_{drop}	0.015
M_{dip}	1.0
A_{dip}	3.0

Chapter 3

Experimental analysis

3.1 Subjective Test

Subjective video quality assessment is a traditional method of video quality. It is considered as the most appropriate method for video quality rating. Video quality rating depends on average of ratings obtained from all the viewers. These ratings are based on standards mentioned in ITU-R Rec. BT.500-12 (2009) [14].

In our work subjective analysis is conducted on test videos from [21] and were comprised of four different resolutions namely QCIF, CIF, QVGA, VGA. Each resolution has 3 sample videos with different frame rate (30fps, 15fps,7.5fps) are used. Different amount of frame freezing at different locations is introduced in these videos using Matlab. Length of freezing (seconds) and location of freezing are shown in Table 4.1. A total of 144 videos have been used for video quality assessment. Subjective testing is conducted on 20 observers with all viewing parameters set according to ITU-R Rec. BT.500-12 (2009) [14]. In our work Single Stimulus Continuous Quality Evaluation

Table 3.1: Freezing length and freezing locations for videos used in subjective analysis with frame rates

Frame rate(fps)	No. of freezing locations	Freezing location	Freezing length
30	One	1st quarter	1 sec
30	Two	1st and 2nd quarter	1 sec each
30	Three	1st,2nd and 3rd quarter	1 sec each
15	One	1st quarter	1 sec
15	Two	1st and 2nd quarter	1 sec each
15	Three	1st,2nd and 3rd quarter	1 sec each
7.5	One	1st quarter	1 sec
7.5	Two	1st and 2nd quarter	1 sec each
7.5	Three	1st,2nd and 3rd quarter	1 sec each

(SSCQE) method is used. This method is suitable as the observers only rate the impaired videos.

3.1.1 Grading Scale

Subjective video quality assessment of a video is done on a scale of 0-100 with 0 being poor video and 100 being excellent video whereas the metric predicts the rate on scale of 0-5. So, there is a need to map the Subjective MOS to prediction MOS. This mapping is done based on the equation 4.1.

$$MOS_{0-5} = \frac{MOS_{(0-100)}}{20} \tag{3.1}$$

3.2 Predicted Results

Results obtained from the metrics for all the videos are tabulated and analyzed. The prediction is done on the scale 0-5.

3.3 Evaluation

Accuracy of the predicted results is calculated by computing Pearson linear correlation co-efficient and Spearman's rank correlation coefficient. These coefficients are calculated by processing the predicted results and subjective test results to predefined Matlab function corr( ) by mentioning the type of coefficient required.
The performance of the metric is estimated by calculating the outlier ratio. It is defined as the ratio of number of false scores to total number scores, where an outlier is a point for which

$$|Perror(i)| > 2.07\sigma(MOS(i)). \tag{3.2}$$

Here $P_error(i)$ is difference between subjective MOS and predicted MOS of metric.

3.4 Analysis:

Results obtained from three metrics are compared and analyzed with results obtained from subjective testing. Subjective testing is conducted as per ITU-R Rec BT.500-12 (2009). Spearman coefficient, Pearson coefficient with subjective scores is calculated. S.Borer metric predicted better results compared to other metrics. Motion intensity and time stampings are the main factors involved in this metric and these factors make this metric more reliable compared to other metrics. Spearman co-efficient (.98) and Pearson co-efficient (.96) suggests correlation of the results obtained with PEVQ

scores. M.Ghanbari's (2009)[9](section 3.2) video quality assessment is done based only on freezing artifact. This metric suggests that short multiple frequent frame freezing whose length on the whole predicts same video quality as single frame freezing of similar length. But practically this is incorrect as multiple frame freezing events has more impact compared to single frame freeze event.The major drawback for this metric is, it considers only one artifact (frame freezing) bounding it to only few videos. But, identification of frozen frames can be justified from this metric. The objective and PEVQ results are correlated and it is found that correlation value is quite low for this metric. S.Wolf(2008)[10] is another metric that is considered in our research work. Frame drops and frame dips are the main factors involved in this metric. Identification of frame droppings and frame dips plays major role in this metric design but prediction of video quality based on these factors alone is not suggestible.

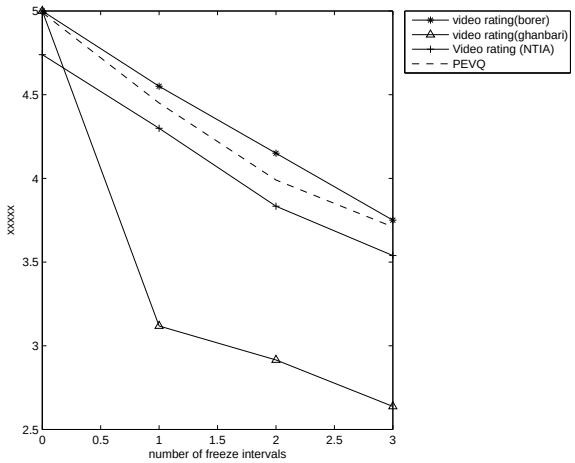

Figure 3.1: Metric behavior along with PEVQ MOS for CIF resolution videos

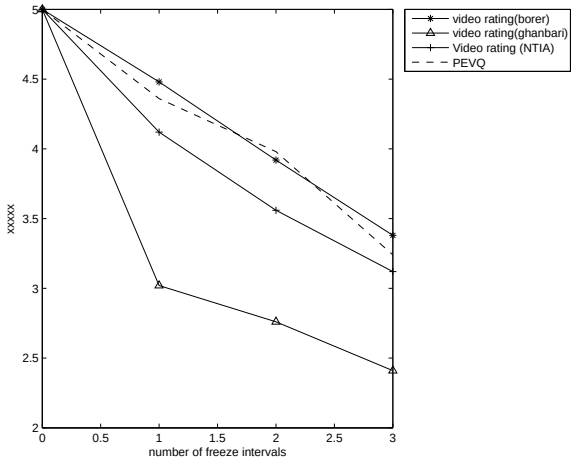

Figure 3.2: Metric behavior along with PEVQ MOS for QCIF resolution videos

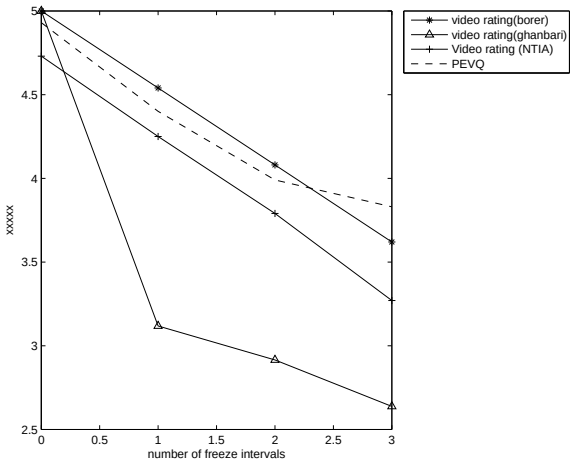

Figure 3.3: Metric behavior along with PEVQ MOS for QVGA resolution videos

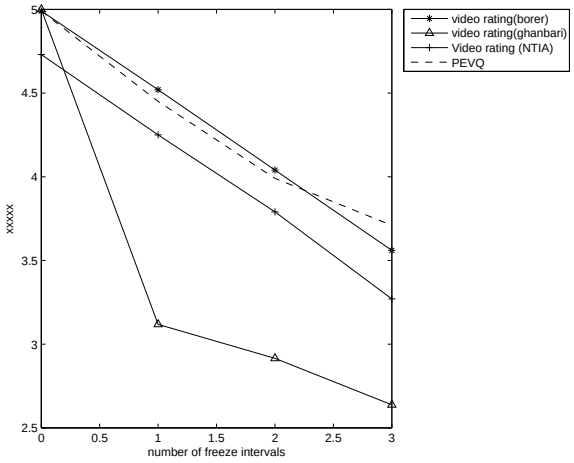

Figure 3.4: Metric behavior along with PEVQ MOS for VGA resolution videos

Chapter 4

Conclusions

This research work addresses the necessity for the performance analysis of existing metrics. Our research starts with study on factors responsible for video quality degradation and then the performance analysis of no-reference video quality assessment methods and suggesting the best metric. We have compared three models of No-reference video quality estimation methods for frame freezing and frame dropping. Metrics used in this research work are studied and implemented on different type of videos involving different parameters responsible for quality degradation. Prediction ability of these metrics is tested using PSNR, PEVQ and subjective MOS. Though all these metrics are predicted based on motion intensity Borer metric has considered many other factors like time stamping, resolution, and bitrate making it more reliable compared to other metrics. Video quality prediction using Ghanbari and S.Wolf's is based only on frame freezing and frame dropping respectively, bounding these metrics only to few videos. Behavior of these metrics can be studied from the graphs and correlation coefficients obtained. Performance of Borer metric is high compared to other metrics. The research work carried out in this thesis is based on the videos which are already captured. It would be an interesting task to carry out this work on live streaming videos there by making few modifications if necessary and predicting a best metric for future.

Bibliography

[1] S. Borer. A model of jerkiness for temporal impairments in video transmission. In *Quality of Multimedia Experience (QoMEX), 2010 Second International Workshop on*, pages 218 –223, june 2010.

[2] M.Yuen, H.Wu, and K.Rao. Coding artifacts and visual distortions. In *Digital Video Image Quality and Perceptual Coding CRC Press*, pages 87–122, 2005.

[3] M.M Farias, M.C.Q.and Carvalho, Kussaba H.T.M, and B.H.A Noronha. A hybrid metric for digital video quality assessment. In *In Broadband Multi Media Systems and Broadcasting (BMSB)*, volume 1, pages 1–6, 2011.

[4] Kanumuri S., Subramanian S.G., Cosman P.C., and Reibman A.R. Predicting H.264 packet loss visibility using a generalized linear model. In *IEEE International Conference on Image Processing*, pages 2245 –2248, oct. 2006.

[5] Ricardo R. Pastrana-Vidal, Jean Charles Gicquel, Catherine Colomes, and Hocine Cherifi. Sporadic frame dropping impact on quality perception. In *Proc. SPIE 5292, Human Vision and Electronic Imaging IX, 182*, june 7 2004.

[6] Huynh-Thu Q. and Ghanbari M. No-reference temporal quality metric for video impaired by frame freezing artefacts. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 2221 – 2224, nov. 2009.

[7] S.van Kesterand, T.Xiao, R.E Kooijand, K. Brunnstrom, and O.K.Ahmed. Estimating the impact of simple and multiple freezes on video quality. In *in proceeding of SPIE*, volume 7865, page 1, 2011.

[8] Zhon Kang Lu, Weis Lin, B.C Seng, and X.K. Yang. Measuring the negative imapct of frame dropping on perceptual video quality. In *in proceeding of SPIE*, volume 5292, 2004.

[9] S.A Amirsahi and M.C Larabi. Spatial-temporal video quality metric based on estimation of qoe. In *Third International Workshop on Quality of Multimedia Experience*, 2011.

[10] Quan Huynh-Thu and M. Ghanbari. No-reference temporal quality metric for video impaired by frame freezing artefacts. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 2221 –2224, nov. 2009.

[11] S.Wolf. Fourth interbational workshop on video processing and quality metrics for consumer electronics(vpqm-09). In *A No Reference (NR) and Reduced Reference (RR)Metric for Detecting Dropped Video Frames*, 2009.

[12] A. Khan, L. Sun, J. Farjardo, I. Taboado, F Liberal, and E. Ifeachor. Impact of end devices on subjective video quality assessment for qcif video sequences. In *in Quality of Multimedia Experience QoMEX, Third Intenational Workshop*, volume 1, pages 177–182, 2011.

[13] J. Xing Xu, A. L. Perkis, and Y. Jiang. On the properties of mean opinion scores for quality of experience managemnet. *In Multimedia ISM 2011 IEEE Symposium*, pages 500–505, 2011.

[14] ITU-R BT.500-12,radio communication sector of itu, 2009. http://www.itu.int/.

[15] S. Chikkeur, V. Sundarm, M. Reisslien, and L.J. Karam. Objective video quality assessment methods: Aclassification review and performance analysis. *Broadcasting, IEEE Transactions on*, 57(2):165–182, 2011.

[16] A. Bhat, S. Kannangara, Yafan Zhao, and I. Richardson. A full reference quality metric for compressed video based on mean squared error and video content. *Circuits and Systems for Video Technology, IEEE Transactions on*, 22(2):165 –173, feb. 2012.

[17] Xiangyu Lin, Hanjie Ma, Lei Luo, and Yaowu Chen. No-reference video quality assessment in the compressed domain. *Consumer Electronics, IEEE Transactions on*, 58(2):505 –512, may 2012.

[18] Muhammad Shahid, Andreas Rossholm, and Benny Lovstrom. A reduced complexity no-reference artificial neural network based video quality predictor. In *4th International Congress on Image and signal Processing CISP*, volume 1, pages 517–521, Oct 2011.

[19] Muhammad Shahid, Andreas Rossholm, and Benny Lovstrom. A no-reference machine learning based video quality predictor. In *Fifth In-*

ternational Workshop on Quality of Multimedia Experience, pages 176 –181, july 2013.

[20] Opticoms pevq, 2012. http:www.pevq.org.

[21] Test video sequences, 2012. http://http://media.xiph.org/video/derf/.