PERFORMANCE EVALUATION OF VIDEO QUALITY ASSESSMENT METHODS BASED ON FRAME FREEZING

Muhammad Arslan Usman

This thesis is presented as part of Degree of Master of Science in Electrical Engineering

Blekinge Institute of Technology

April 2014

Blekinge Institute of Technology
School of Engineering
Department of Applied Signal Processing
Supervisor: Muhammad Shahid
Examiner: Dr. Benny Lövström
Abstract

A Digital Video Communication System consists of a video encoder, a channel through which the video is transmitted and finally a video decoder. These three phases can cause different kinds of impairments or degradations in the quality of a video. One of the degradations that takes place due to an error prone channel is frame freezing and it affects the quality of a video. Frame freezing is a phenomenon in which a particular frame is repeated (displayed) for a certain period of time until the next correct frame is received hence producing a pause or a freeze in a video sequence. Frame freezing can be single or multiple depending on the channel characteristics. Research is being conducted in order to devise a metric which can help in quantifying a video sequence on the basis of its quality. Researchers design metrics and then using their mathematical models, they compare their results with subjective measurements. Subjective measurements help in telling the preciseness and correctness of the metric that whether the quality assessment metric is close enough to subjective test measurements. In this report, three metrics used for video quality assessment have been studied and compared mathematically as well as by careful experiments. The three metrics, chosen for this study, work on No Reference (NR) method for judging the quality of a video. Subjective tests have been performed keeping in view of the recommendations of ITU (International Telecommunication Union). After the study of these three metrics, their advantages and disadvantages over each other have been discussed and the metric that performs the best under certain conditions has been proposed. Finally, a conclusion has been made for the metric that performs the best and also for the metric that performs the worse or even does not perform at all under certain conditions.
ACKNOWLEDGEMENT AND DEDICATION

First and most of all, I would like to thank my respectable thesis supervisor Muhammad Shahid and respectable Examiner Dr. Benny Lövström for their support and guidance throughout the duration of this research and for providing me an opportunity to develop my research abilities that helped me achieve the set goals for this thesis.

Also, I thank my parents, my family and my fiancé for supporting me through the hardest times and providing me the comfort in order to perform and deliver the best way I could.
# Table of Contents

1 Introduction .................................................................................................................. 1  
1.1 Outline of the thesis ................................................................................................. 2  
2 Background and literature review .................................................................................. 3  
2.1 Packet Losses ......................................................................................................... 3  
2.1.1 Frame Freezing ................................................................................................. 3  
2.2 Types of distortions ................................................................................................. 4  
2.2.1 Spatial Distortion ............................................................................................ 5  
2.2.2 Temporal Distortion ....................................................................................... 5  
2.3 Methods for the quality assessment of Videos ....................................................... 5  
2.4 Objective Quality Measurements ...................................................................... 5  
2.4.1 Full Reference metrics (FR) ........................................................................... 6  
2.4.2 Reduced Reference metrics (RR) .................................................................. 6  
2.4.3 No Reference (NR) ....................................................................................... 7  
2.5 Subjective Quality Measurements .................................................................... 7  
2.6 Metrics chosen for this study ............................................................................... 8  
2.7 A No Reference (NR) and Reduced Reference (RR) Metric for Detecting Dropped Video Frames [5] .......................................................... 8  
2.8 A Model of jerkiness for temporal impairments in video transmission [7] ........... 11  
2.8.1 Measuring the jerkiness ................................................................................ 12  
2.9 No-Reference temporal quality metric for video impaired by frame freezing artefacts [6] 14  
2.9.1 Identification of Frozen Frames .................................................................... 14  
2.9.2 Quality Metric ............................................................................................... 15  
3 Design, Implementation and Testing ......................................................................... 19  
3.1 Single Freeze Experiment .................................................................................... 19  
3.2 Results for Single Freeze Experiment ................................................................ 21  
3.3 Multiple Freeze Experiment ............................................................................... 23  
4 Performance analysis and evaluation ....................................................................... 27  
4.1 Grading Scale: .................................................................................................... 27  
4.2 Evaluation ........................................................................................................... 27  
4.2.1 Single Freeze Experiment ........................................................................... 27  
4.2.2 Multiple Freeze Experiment ........................................................................ 28
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3</td>
<td>Authentication of the Subjective tests</td>
<td>29</td>
</tr>
<tr>
<td>5</td>
<td>Results and conclusions</td>
<td>31</td>
</tr>
<tr>
<td>5.1</td>
<td>Performance of S. Borer’s Metric and S. Wolf’s Metric</td>
<td>31</td>
</tr>
<tr>
<td>5.2</td>
<td>Performance of Quan Huyn-Thy &amp; M. Ghanbari’s Metric</td>
<td>31</td>
</tr>
<tr>
<td>5.3</td>
<td>Noticeable flaws and loop holes in the Metrics</td>
<td>31</td>
</tr>
<tr>
<td>5.4</td>
<td>Conclusion</td>
<td>32</td>
</tr>
<tr>
<td>5.5</td>
<td>Future Work</td>
<td>32</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1. Block Diagram for Digital Transmission of Videos ................................................................. 1
Figure 2. Free Event in a series of regular frames .................................................................................. 4
Figure 3. Calculation of Motion Intensity of a Particular Frame [5] ......................................................... 4
Figure 4. Working of a (FR) Full Reference Metric .............................................................................. 6
Figure 5. Working of a (RR) Reduced Reference Metric ....................................................................... 6
Figure 6. Working of a (NR) No Reference Metric ............................................................................... 7
Figure 7. Example of how $T_{12}$ is computed [5] ................................................................................ 9
Figure 8. Measurement of Motion Intensity of a Particular Frame [5] ....................................................... 12
Figure 9. S-Shaped Function for the Display Time function $t_{dx}$ (Inflection Point in Zoomed part) [7] .......... 13
Figure 10. Freezing and repetition of same frames in a transmitted sequence ........................................ 14
Figure 11. Histogram for a 25 fps Video with reference to the NR Metric [6] ........................................... 16
Figure 12. Perceived Quality for Different Video Content Types (MOS vs. Duration of Freeze (Sec) 25 fps) .... 21
Figure 13. Perceived Quality for Different Video Content Types (MOS vs. Duration of Freeze (Sec) 25 fps) .... 22
Figure 14. Perceived Quality for Different Video Content Types (MOS vs. Duration of Freeze (Sec) 25 fps) .... 22
Figure 15. Perceived Quality for Different Video Content Types (MOS vs. Duration of Freeze (Sec) 25 fps) .... 23
Figure 16. Perceived Quality for Different Video Content Types : 1 freeze Instance and 3 Freeze Instance .... 25
Figure 17. Perceived Quality for Different Video Content Types : 5 freeze Instance and 8 Freeze Instance ...... 26
Figure 18. Comparison of the 3 Metrics and Subjective Test Results (Single Freeze) ............................... 27
Figure 19. Comparison of the 3 Metrics and Subjective Test Results (Multiple Freezes (25fps)) ................. 28
List of Important Tables

Table 1. Recommended Parameters for the Algorithm [5] ................................................................................... 11
Table 2. Defined parameters for the functions τ and μ [7].................................................................................. 13
Table 3. Description of the Absolute Category Rating (ACR) Scale................................................................... 19
Table 4. Duration of freeze occurrences and total duration of freeze ................................................................. 24
## List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACR</td>
<td>Absolute Category Rating</td>
</tr>
<tr>
<td>Dfact</td>
<td>Dynamic Factor</td>
</tr>
<tr>
<td>FDF</td>
<td>Fraction of Dropped Frames</td>
</tr>
<tr>
<td>FPS</td>
<td>Frames Per Second</td>
</tr>
<tr>
<td>FR</td>
<td>Full Reference</td>
</tr>
<tr>
<td>FrDur</td>
<td>Freeze Duration</td>
</tr>
<tr>
<td>FrTotDur</td>
<td>Total Duration of Freeze</td>
</tr>
<tr>
<td>ITU</td>
<td>International Telecommunication Union</td>
</tr>
<tr>
<td>MOS</td>
<td>Mean Opinion Score</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean Square Error</td>
</tr>
<tr>
<td>NR</td>
<td>No Reference</td>
</tr>
<tr>
<td>PVQA</td>
<td>Perceived Video Quality Assessment</td>
</tr>
<tr>
<td>RR</td>
<td>Reduced Reference</td>
</tr>
<tr>
<td>SROI</td>
<td>Spatial Region of Interest</td>
</tr>
<tr>
<td>TI</td>
<td>Temporal Information</td>
</tr>
</tbody>
</table>
Chapter 1

1 Introduction

Communication has become a basic commodity of life now a days and the present challenges related to communication are very important and cannot be neglected to meet high standards. Digital transmission of video is one of the names that come to our minds when we talk about communication. Digital transmission of videos is being carried on by many different methods and still there is more to be done and more to be achieved. These methods have their benefits as well as their disadvantages. Digital transmission has widely overcome analogue transmission as the former has many benefits over the latter.

The digital transmission system for videos consists of different phases which are shown in figure 1. The source is the part where the video is acquired which can be a means of generating a video. Then comes the encoding phase where the videos acquired are encoded in order to save the bandwidth as it is highly feasible if lesser bandwidth is used. Bandwidth limitation can be managed through limiting the bit rate, frame rate or frame size. This process of encoding is also known as compression as it compresses or limits the video in order to minimize the bandwidth usage. The compression can cause 2 types of quality degradations which are as follows:

- Temporal
- Spatial

These two different kinds of degradations will be discussed in later sections. After the compression part comes the transmission part, in which the compressed video is transmitted to its destination. The transmission is done through some channel and the channel can be error prone that can input errors in to the transmitted data. The error can be jittering, delaying and packet loss. After passing through the medium or channel, the transmitted video reaches its destination which is also called the receiver. The receiver performs the reconstruction of the transmitted video. The reconstruction part reconstructs the videos and then they can be judged and checked for their video quality.

![Figure 1. Block Diagram for Digital Transmission of Videos](image-url)
Video quality assessment is very important for the videos that are received at the reconstruction end as it can help in understanding the errors that the video attained while going through the transmission phase. The videos received are monitored in order to make sure that they reach the desired quality levels and they are extensively studied. So far many techniques have been devised to overcome the errors that videos attain during the transmission. There are many factors that can affect the quality of the video and here we need to perform video quality assessment methods, so the effect of these error factors can be studied and hence a solution can be devised.

This report contains a comparison of three metrics that have been proposed recently for the sake of video quality assessment and as a result the best method among them will be suggested. The metrics are of different kind but here only the metrics which are no reference techniques will be discussed only. No reference methods are basically the methods in which the reference video is not present at the receiving end which means there is no chance that the received video can be compared to the reference video and then the error can be calculated. The error factor which has been kept for this study is frame freezing. Frame Freeze happens when Frames are dropped or distorted. This factor impacts very negatively on video quality and for this sake it has been studied extensively with the help of the quality assessment metrics. Frame freezing is discussed in later chapters.

1.1 Outline of the thesis

The report contains 6 chapters in total.

**Chapter 2** contains the literature review and background of digital transmission, the factors which affect the transmission and other related important factors. Further, the methods that are used for perceiving video quality, subjective video quality testing and objective video quality testing have been discussed and finally the metrics that have been included in this comparative study are discussed in detail.

**Chapter 3** includes the design and details of the subjective test that was performed and its results. Also all the details related to the test environment have been discussed.

**Chapter 4** includes the performance evaluation and comparison for the 3 methods/metrics that have been studied and simulated on MATLAB

**Chapter 5** includes the conclusion, results and future work.
Chapter 2

2 Background and literature review

Digital Transmission has become an essential part of our daily lives, as the time passes, new methods and innovations take place and new techniques for transmitting of data are devised. As new techniques arrive in the market, so they become a remedy for some problems but at the same time they have their drawbacks which need to be overcome. When a digitally transmitted data passes through a channel it takes a lot of effects from the channel which can result in different kind of damages in the digital data. Similar is for a digitally encoded video passing through a channel, which can take error as the channels are mostly error prone channels. The videos need to be tested and observed in order to make an analysis of the errors that occur in the video. For this particular reason, subjective quality tests are conducted as they are beneficial to know the errors present in the channel and then those errors can be identified and studied so that a solution can be found in order to overcome the possibility of those errors and make digital transmission more reliable and accurate. There are many possible reasons for video degradation at the reception end during the whole transmission procedure but the ones discussed in this report are the few relevant ones. Most important and vital factor in digital transmission is that there should not be any packet loss during the transmission as it can result into dropping of frames, freezing of frames and pausing of the video.

2.1 Packet Losses

As the name tells, packet is a small proportion of the total data that is being transmitted through an error prone channel and if these small portions are lost due to the presence of errors then it is called packet loss. As a result of this inability of a few packets to reach the destination, there are different kinds of errors in the received data or more specifically in the received videos. Packet loss can be in a small proportion or can be in huge blocks which have different effect on the received videos. This problem can be dependent on many factors like the distance between the receiver and the transmitter, different transmitting frequencies, quality of the channel or the type of channel coding used etc. But in this report the factors which affect the videos or the data are not discussed, just the perceptual impacts that can happen have been discussed.

One of the major types of packet loss is frame freezing. Freezing of frames takes place when frames are dropped. So dropping of frames leads to frame freezing. Frame freezing has been studied in this report and the techniques which are used to overcome the frame freezing effect on videos have been discussed and tested.

2.1.1 Frame Freezing

This type of distortion is simply defined as the freezing of a particular frame for some time duration in a video. The reason for this type of degradation is that when packet loss occurs then there is chance that some frames in the video are lost and as a result a certain frame is repeated so that the video can be continued without a delay until a correct frame is received. This type of degradation can happen as a single event or as a number of events which is called multiple frame-freezing. This problem can be annoying for the user as in video broadcasting, the user expects a continuous video without any lag, delay or frame freeze event.
The easiest way to calculate a frame freeze is by knowing the difference between 2 consecutive frames; if the motion intensity of 2 consecutive frames is same or their difference is zero then it can mean that there is a frozen frame. The quality of a video cannot be judged by just the freeze events that occur in it but of course it is a very big factor of video degradation yet not the only one. In the following figure [7] it can be seen that when a frame freeze event occurs, the difference between the motion intensity for the consecutive frozen frames reaches zero.

![Figure 2. Free Event in a series of regular frames](image)

### 2.2 Types of distortions
The types of impairments or distortions that take place in a video sequence can be classified in 2 categories which are as follows.
2.2.1 Spatial Distortion

There are different types of spatial distortions that can take place in a video sequence. Spatial distortions can take place both at the transmission end and the receiving end of a communication channel. At transmission end it can take place during the compression of a video and at the receiving end it can take place during the reconstruction. Compression can result into blurring effect, addition of noise, colour distortions, ringing effect etc [15].

The factors behind such kind of distortions or impairments can be immature acquisition of the video which might include substandard recording device or rapid movements of the subject or out of focus video capturing. All these spatial distortions influence very badly on perceptual quality of a video and to the human eye.

2.2.2 Temporal Distortion

A temporal distortion is commonly defined as the temporal evolution, or fluctuation, of the spatial distortion on a particular area which corresponds to the image of a specific object in the scene [15]. There are different types of temporal distortions and they take place when a video sequence is passing through an error prone channel as discussed in earlier sections. Error prone channels can result into jittering, frame freezing or dropping, frame halting, jerkiness etc.

Both the temporal distortion and spatial distortion are a setback for video broadcasting but temporal impairments are likely to have worse influence as they have a much stronger impact on human perception of videos [15].

2.3 Methods for the quality assessment of Videos

Majorly, there are 2 methods that are used to quantify a video and they have been discussed separately in this chapter:

- Objective Quality Measurements
- Subjective Quality measurements

2.4 Objective Quality Measurements

- The purpose of an objective quality measurement or video quality evaluation is to automatically assess the quality of images or sequence of images (a video) or video sequences in agreement with human quality judgments which are also known subjective quality measurements [8]. Over the past few decades, image and video quality assessment has been broadly studied and many different criterions for objective quality measurements have been set. Objective quality measurements have mathematical models that automatically assess the videos and give them a rating on a specific scale [15] [16] [17].

Video quality metric can be classified into 3 different types which are described as follows:

- Full Reference metrics (FR)
- Reduced Reference metrics (RR)
- No Reference (NR)
2.4.1 Full Reference metrics (FR)
In this method a full reference of the original video is required in order to quantify the video at the basis of quality. A full comparison is done between the processed video and the original video which is basically pixel wise comparison. The original video is the uncompressed version of the processed video which has no distortions and is kept in its original form [2] [16] [17].

But this method is strictly limited as it cannot be used in live streaming or live video broadcasting such as a football match being live broadcasted. This method is successful in offline video testing. For the testing of live video broadcasting such methods are required which are very quick as a delay cannot be afforded in such broadcasts.

The following figure explains how FR metrics work though in real time communication it is not possible to have original video at the receiving end but in ideal cases (lab practice) it can be possible.

Figure 4. Working of a (FR) Full Reference Metric

2.4.2 Reduced Reference metrics (RR)
This technique also requires a reference at the receiving side which makes it a bit complex to be used for testing live video broadcasts. In this method only a specific parameters of the original video is present at the receiving side and these specific parameters are matched with the received video and thus the quality of the video is suggested [3] [5] [20].

It is on many occasions better than the FR method and is much quicker than the FR method. Following figure explains how reduced reference method works.

Figure 5. Working of a (RR) Reduced Reference Metric
2.4.3 No Reference (NR)

In this method, which is most commonly used for live video broadcasting, there is no reference present at the receiving end of the whole communication channel. There is some metric which is present at the receiving end and it predicts the quality of the video by some mathematical model [5] [9].

As compared to FR and RR method, NR method is much complex and its operating time depends on the computational and mathematical complexity of the metric. Following figure explains the NR method.

![Figure 6. Working of a (NR) No Reference Metric](image)

2.5 Subjective Quality Measurements

Subjective video quality assessment is the most accurate way of estimation of perceptual quality measurement methods. This method is in practice at present and provides the optimal results for quality assessment. In these measurements, a specific panel of audience is required which judges the quality of the video according to a scale. International Telecommunication Union (ITU) has prescribed some specific recommendations which are recommended for the researchers to take into account while performing subjective quality assessment [15] [16] [18].

A few commonly used methods that are recommended by ITU are as follows:

- Double stimulus Continuous Quality Scaling (DSCQS)
- Single stimulus Continuous Quality Evaluation (SSCQE)
- ACR (Absolute category rating)

In this report the method used is ACR. In this method the test sequences are presented one at a time and are rated independently on a category scale [18]. The method specifies that after each presentation the subjects are asked to evaluate the quality of the sequence shown.

The ACR method is further discussed later in chapter 3 and the chosen scale for the rating has also been discussed.
2.6 Metrics chosen for this study

Research is being conducted to devise different metrics in order to quantify a video on the basis of its quality but as there is no end to research, so innovations take place on daily basis. The three metrics that are discussed and compared in this report are all working on the NR method as I only took into account the case for live video broadcasting, which is majorly only possible by the NR method. The metrics have been described in 3 research papers separately and will be discussed briefly in this section of this report.

1. A No Reference (NR) and Reduced Reference (RR) Metric for Detecting Dropped Video Frames by Stephen Wolf [5]


3. No-Reference temporal quality metric for video impaired by frame freezing artefacts by Quan Huynh-Thu and M. Ghanbari [6]

2.7 A No Reference (NR) and Reduced Reference (RR) Metric for Detecting Dropped Video Frames [5]

The techniques used in this paper is the No Reference method which only involves luminance i.e. \( Y(i, j, t) \) where \( i \) and \( j \) are the rows and columns indices of the image and \( t \) is the total number of frames in the video sequence \( (t = 1, 2, 3, ..., N) \).

The technique has been explained in 2 steps which are:

- Frame-by-Frame Motion energy Time History
- Examining the Time History for Frozen Frames

Firstly, how the motion energy time history is calculated and computed, it has been explained. The motion energy time history of a video clip requires three processing steps on the sampled video images [5]

\[ TI(i,j,t) = Y(i,j,t) - Y(i,j,t-1), (i,j) \in SROI, and t= 2,3, ... N \]  

(1)

The term \( SROI \) is the Spatial Region of Interest, which can be the central portion of the image which is purposed for eliminating the image border pixels (e.g., some cameras may not fill the entire ITU-R Recommendation BT.601 frame, encoders may not transmit the entire frame) [5].

Step 2) There are zero image pixels present in an image and they can be or should be eliminated if they are lesser than a threshold called \( M_{image} \)
\[ TI(i,j,t) = \begin{cases} 
TI(i,j,t) \text{ if } \text{abs}(TI(i,j,t)) > M_{\text{image}} \\
0 & \text{Otherwise} 
\end{cases} \quad (2) \]

This helps in getting rid of the low level noise which is not needed at all and should be eliminated in order to have clear Temporal Information (TI) of the video sequence.

**Step 3)** In this step the TI calculated above is squared and named as \( TI_2 \), this is done to convert the Amplitude to Energy and then the mean of each frame is calculated as follows:

\[ TI_2(t) = \text{Mean}_{ij} \{TI(i,j,t)^2\} \quad (3) \]

It can be seen in the following figure that how the TI2 is computed.

![Figure 7. Example of how TI2 is computed [5]](image)

Now, the second step of the algorithm or method that is defined in this paper will be described. This step includes how the \( TI_2 \) (Temporal information) will be exploited in order to get the frozen frames or dropped frames. Here, there are 2 terms that need to be understood separately, one of them is called **dips** and other one is called **drops**. Dips are also frozen or dropped frames but they happen when there is very low intensity motion content present in the video sequence, its threshold is set as minimum as possible.
but not as minimum that it includes low level noise. So, at first it seems very simple but actually calculating the dropped frames or frozen frames is a bit complicated, as the video sequence can be a person who is sitting still talking with only lip movement (an anchor person). Dips have minimum motion energy and they have to be with minimum amplitude of $A_{dip}$.

After the calculation of dips and drops, the factors $A_{dip}$, $A_{image}$ and $M_{dip}$ are multiplied with a dynamic factor called $d\text{fact}$ and hence the dropped frames can be determined. $d\text{fact}$ is derived from average level of the motion intensity [5].

**Step 4** First of all the $TI2$ factor is calculated as an average and termed as $TI2\_ave$, which is as follows;

$$TI2\_ave = mean\over k \{TI2\_sort(k)\}, \text{ceil} \ (F_{cut} \cdot (N-1) \leq floor((1-F_{cut}) \cdot (N-1))) \quad (4)$$

Here $TI2\_sort$ is the $TI2$ vector from step 3 but sorted from low to high, with a new index $k$ rather than $t$. $F_{cut}$ is the fraction of the scene cuts that will be eliminated. Ceiling and Floor are rounding functions that round up to the nearest integer.

**Step 5** Computing the dynamic factor $d\text{fact}$ as follows;

$$d\text{fact} = a + b \cdot \log(TI2\_ave)$$

$$\text{if}(d\text{fact} < c) \text{ then set } d\text{fact} = c \quad (5)$$

Where $a$, $b$ and $c$ are positive constants and log is the natural logarithm with base e. To limit the factor $d\text{fact}$ to some small positive value, constant $c$ is necessary. The equation expresses that the perception of frame drops and dips is linearly dependent upon the log of the average motion energy [stefen].

**Step 6** Now, in this step calculate the Boolean variable drops (which is equal to 1 when a frame drop is detected and is equal to 0 otherwise); [5]

$$\text{drops} (t) = \begin{cases} 1 & \text{if } TI2(t) \leq d\text{fact} \cdot M_{drop} \\ 0 & \text{Otherwise} \end{cases} \quad (6)$$

**Step 7** Compute the Boolean variable dips (which is equal to 1 if a dip is detected and for other cases it is 0); [5]

$$\text{dips}(t) = \begin{cases} 1 & \text{if } TI2(t) \leq d\text{fact} \cdot M_{dip} \text{ and } \text{dips\_mag}(t) \geq d\text{fact} \cdot A_{dip} \\ 0 & \text{Otherwise} \end{cases} \quad (7)$$

Where, $\text{dips\_mag}$ is a function that finds the magnitude of the dips and is given by;

$$\text{dips\_mag}(t) = \min\{TI2(t-1) - TI2(t), \ TI(t+1) - TI2(t) \} \quad (8)$$

$$\text{if}(\text{dips\_mag}(t) < 0) \text{ then set } \text{dips\_mag}(t) = 0$$
Step 8) Some of the frames maybe classified as both a drop and a dip. Thus, here the logical function OR of the drops and dips vectors from steps 6 and 7 gives 1’s for those frames in the video clip that are detected as dropped/ repeated frames and 0’s otherwise [5].

Fraction of Dropped Frames (FDF) can be calculated by summation of the elements in the vector and dividing them by the maximum number of frames that could be detected as dropped or frozen.

\[ FDF = \frac{\text{sum}(\text{drops OR dips})}{(N-3)} \]  

The parameters which are recommended for all the above stated equations are given in the table below;

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>M image, step2</td>
<td>30</td>
</tr>
<tr>
<td>Fcut, step 4</td>
<td>0.02</td>
</tr>
<tr>
<td>a, step 5</td>
<td>2.5</td>
</tr>
<tr>
<td>b, step5</td>
<td>1.25</td>
</tr>
<tr>
<td>c, step 5</td>
<td>0.1</td>
</tr>
<tr>
<td>Mdrop, step 6</td>
<td>0.015</td>
</tr>
<tr>
<td>Mdip, step 7</td>
<td>1.0</td>
</tr>
<tr>
<td>Adip, step 7</td>
<td>3.0</td>
</tr>
</tbody>
</table>

Table 1. Recommended Parameters for the Algorithm [5]

2.8 A Model of jerkiness for temporal impairments in video transmission [7]

The technique described in this paper is also working on the No Reference (NR) Method and deals with motion energy as well. In this paper the main theme is to calculate the jerkiness factor of a video which can be defined as temporal variations known as jitter also known as freezing if it happens for a long time, these temporal degradations are perceived by the end user as Jerkiness. Here, a video sequence is defined by the following equation;

\[ v = (f_i, t_i), i=1...n \]

Where \( f_i \) denotes the frame \( i \) if the video, which will be displayed starting at the time \( t_i \) up to time \( t_{i+1} \), and \( n \) is the total number of frames present in the video sequence. The time \( t_i \) can be expressed as the equation below and will be referred to as time stamps;

\[ \Delta t_i = t_{i+1} - t_i \]

Which is called the display time of a frame \( i \). The first and foremost need is to plot the motion intensity of the whole video sequence for each frame and it is done by calculating the root mean squared inter-frame difference over the whole video.

Where \( f_i(x) \) denotes the pixel value of the Y-component of a frame \( i \) at location \( x \).
2.8.1 Measuring the jerkiness

In the following diagram the motion intensity of a video has been plotted to explain how the frame freezing takes place and how it can be identified.

Let’s consider a sample video sequence containing an isolated freezing event as shown in figure 1. Where there is zero motion corresponds to the freezing of the frames in the video sequence. As discussed earlier, the motion intensity main variables of the jerkiness measure are the display time $\Delta t_i = t_{i+1} - t_i$ (horizontal dashed line) of the freezing frame and the motion intensity $m_{i+1}$ (vertical dashed line) at the end of the freezing interval. At the end of the freezing interval there is a sudden jump in the motion intensity.

- Firstly, if the motion intensity is larger, then the jerkiness is larger. Thus, it shows that the jerkiness should be a monotone function (a function that only increases or decreases), say \( \mu \) of motion intensity.
- Secondly, in the absence of motion in the sequence then the jerkiness is 0. Thus, \( \mu (0) = 0 \) and jerkiness is a product of \( \mu \) and also a second part.
- Third, dependence of jerkiness on motion intensity for large motion intensities might saturate. Where, for very small motion intensity values the jitter values might be small. Thus, \( \mu \) is chosen to be a parameterised S-shaped function.
- Fourth, jerkiness depends on the frame display time $\Delta ti$ i.e. if the display time is larger, then the jerkiness is larger. Thus, jerkiness is monotonously dependent on display.
- Fifth, for very small display times there is no jerkiness or zero jerkiness. Thus, jerkiness is a product of a display time dependent part and the motion dependent part \( \mu \).
Last, for very large frame display times it is difficult to make useful subjective experiments. Thus, a simple approach is a close to linear relationship, e.g. if a 10sec video sequence freezes almost 10 seconds or if it freezes twice almost 5 seconds should give a similar jerkiness value [7].

The equation can be modelled as;

$$\Delta t_i \tau (\Delta t_i),$$  \hspace{1cm} (10)

So, from all the discussion above Jerkiness can be defined as the sum of the product of the relative display time times a monotone function of display time

$$J(t) = \frac{1}{T} \sum \Delta t_i \cdot \tau_a (\Delta t_i) \cdot \mu (m_i + 1 (v)).$$ \hspace{1cm} (11)

Here, the functions $\tau_a$ and $\mu$ are S-Shaped functions which are described as follows;

$$S(x) = \begin{cases} a \cdot x^b & \text{if } x \leq px \\ \frac{d}{1 + \exp(-c \cdot (x-px))} + 1 - d & \text{else,} \end{cases}$$ \hspace{1cm} (12)

Where $a = py/px(q_{px}/py)$, $b = q_{px}/py$, $c = 4q/d$, $And, d = 2(1 - p_x)$.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>(px, py, q)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Time Stampings Function) $\tau_a$</td>
<td>(0.12, 0.05, 1.5)</td>
</tr>
<tr>
<td>(Motion Intensity Function) $\mu$</td>
<td>(5, 0.5, 0.25)</td>
</tr>
</tbody>
</table>

Table 2. Defined parameters for the functions $\tau_a$ and $\mu$ [7]

![Figure 9. S-Shaped Function for the Display Time function $\tau_a$ (Inflection Point in Zoomed part) [7]](image-url)
The figure shows an S-Shaped function for the display time function ‘τ_u’, the zoomed in part shows the inflection point in the plot which is almost at 0.10 seconds and its value is 0.05 [7].

2.9 No-Reference temporal quality metric for video impaired by frame freezing artefacts [6]

A Frozen frame is defined as a frame being identical to its previous frame and if this problem persists then this means it is freeze event. The degradation of the video depends on the duration of a freeze event and total number of freeze events [6].

<table>
<thead>
<tr>
<th>Original</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degraded</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>H</td>
<td>I</td>
<td>J</td>
</tr>
</tbody>
</table>

Figure 10. Freezing and repetition of same frames in a transmitted sequence

It can be seen in the figure above that the original sequence has 7 different frames whereas the degraded sequence has the third frame being repeated for 4 times, hence discarding the D, E and F frame and causing a freeze event in the sequence.

2.9.1 Identification of Frozen Frames

For a no reference method or metric, the identification of frames has to be performed without any prior knowledge of the original transmitted sequence i.e. the quality of the sequence can only be judged by the processed sequence at the receiving end of the communication system. In a no reference approach it is not possible to distinguish between an intentionally caused pause in a sequence which seems like a freeze event and a real freeze event which is caused by an error prone channel etc.

A very common method is to simply calculate the MSE (mean square error) of 2 consecutive frames and if the MSE = 0 then it means that the frame is frozen. It can be done for longer freezes or freeze events as well. But if we talk about real time environments then this method is little bit difficult as the recorded video sequence must have been recorded by some capturing device and there is a higher chance that some noise might be introduced in the video frames hence making a slight difference between every frame, even though if it is a freeze event.

The method defined below is conducted after converting each and every frame into the YUV format which helps in calculating the MSE between the present frame and the previous frame.

\[ Y_{M1}(i) = \frac{1}{W \times H} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} (Y(x, y, i) - Y(x, y, i - 1))^2 \]  
\[ U_{M1}(i) = \frac{1}{W \times H} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} (U(x, y, i) - U(x, y, i - 1))^2 \]  
\[ V_{M1}(i) = \frac{1}{W \times H} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} (V(x, y, i) - V(x, y, i - 1))^2 \]
Where, in the above equations W and H are width and height of every frame respectively and N is the total number of frames in the video sequence.

If there is very low motion content in the video then there is a chance that those frames will be considered as frozen frames which is not favourable. So to deal with this problem another set of mathematical equations is used in which the frames which are in YUV format already, the MSE is calculated between the frozen frame and the very first frame of the freeze event, so that it is clear that either the frame is potentially frozen or it is a frame with just very low motion content. Let YM(i), UM(i) and VM(i) be the MSE between the present frozen frame and the first frame of the freeze event.

\[
YM(i) = \frac{1}{W \times H} \sigma \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} (Y(x, y, i) - Y(x, y, i - k))^2
\]  

\[
UM(i) = \frac{1}{W \times H} \sigma \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} (U(x, y, i) - U(x, y, i - k))^2
\]  

\[
VM(i) = \frac{1}{W \times H} \sigma \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} (V(x, y, i) - V(x, y, i - k))^2
\]  

So, a frame \( i \) will be considered a frozen frame if the MSE between this frame and the previous frame and the MSE between this frame and the first frame of the freeze event is below a certain threshold which is given as follows; [6]

\[
Freeze\ Flag(i) = \begin{cases} 
1 & \text{if } \{YM, UM, VM\}_{1,2}(i) < T \\
0 & \text{Otherwise}
\end{cases}
\]  

Where, the value of \( T \) in this metric has been empirically set to 1.

2.9.2 Quality Metric

A cumulative histogram is shown in the figure below which describes the distribution of freezes in the video sequence and it is built from the information calculated from above mentioned formulae. Each bin in this histogram named \( FrDur \) represents a freeze event. The number of occurrences of a bin specifies the number of freeze events. The histogram illustrates an example of a 25fps video having 5 freezes in it and each freeze is 400ms and one is of 800ms. But the cumulative value \( FrTotDur \) can be known by multiplying the number of occurrences of a freeze event or bin with that specific bin’s total duration [6].
Now, all the duration values are normalised in relation to the total duration of the received video which has been decoded. For each bin;

\[ FDF (b) = \frac{FrDur(b)}{TotDur} \times 100 \tag{20} \]

\[ FTDP (b) = \frac{FrTotDur(b)}{TotDur} \times 100 \tag{21} \]

Where \( b \) is the index corresponding to the bins present in the video sequence and \( TotDur \) is the total duration of the video sequence.

Then finally the following mapping function is devised which is said to be much near to the subjective data collected;

\[ T1(b) = \frac{1}{f2(FTDP(b)) + f1(FDP(b)) + f3(FTDP(b))} \tag{22} \]

Where;

\[ f1(x) = a1 + b1 \times \log(c1 \times x + d1) \tag{23} \]

\[ f2(x) = a2 \times x^2 + b2 \tag{24} \]

And;

\[ f3(x) = a3 \times x^2 + b3 \tag{25} \]
The constants in the above Equations were empirically determined using least-square regression on the subjective data: $a_1 = 5.767127$, $b_1 = -0.580342$, $c_1 = 3.442218$, $d_1 = 3.772878$, $a_2 = -0.00007$, $b_2 = -0.088499$, $a_3 = 0.000328$, $b_3 = 0.637424$ [6].

Each $T_1(b)$ value is then bound-limited in the range [1, 5]:

$$T_1'(b) = \min (\max(T_1(b),1),5) \quad (26)$$

Finally, the temporal video quality metric is set to:

$$T_1 = \min (T_1'(b)) \quad (27)$$

Where, 5 represents excellent quality video and 1 represents poor quality [6].
Chapter 3

3 Design, Implementation and Testing

In order to test the 3 metrics, it was needed to perform subjective tests in order to get the results from the human end and then compare the performance of the subjective test results and the results produced by the 3 metrics.

There were 2 types of experiments or tests performed for the subjective testing and they are as follows.

- Single Freeze Experiment
- Multiple Freeze Experiment

3.1 Single Freeze Experiment

For the Single freeze experiment 18 test subjects were carefully selected. The group of test subjects included 7 female subjects and 11 male subjects with ages ranging from 20 to 30 years. The subjects were from youth, expert and non expert category. None of the test subjects were paid for their services but a few were presented with Coffee and cake as a token of gratitude for their service. The scale chosen to rate the quality of the videos was Absolute Category Rating (ACR), which can be understood from the table below [12][13].

<table>
<thead>
<tr>
<th>MOS</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Excellent</td>
</tr>
<tr>
<td>4</td>
<td>Good</td>
</tr>
<tr>
<td>3</td>
<td>Fair</td>
</tr>
<tr>
<td>2</td>
<td>Poor</td>
</tr>
<tr>
<td>1</td>
<td>Bad</td>
</tr>
</tbody>
</table>

Table 3. Description of the Absolute Category Rating (ACR) Scale

The videos used for the single freeze experiment were obtained as follows. 5 videos with different content were selected which included

1. A Walking Space Shuttle Crew
2. Crowd doing Ice Skating
3. A city view through a Helicopter
4. Football practice
5. Routine video of a harbor

All of the videos were in AVI format as required for the test. The Videos had a frame rate of 25 frames per second (fps) and had no audio. As audio is considered to be a distraction for the test subjects in Subjective Video Testing, so the videos containing audio were made audio less by the help of software named Virtual Dub. Virtual Dub deals a video frame by frame, so the video quality was not harmed during the removal of audio content. The videos were of a duration of 12 seconds. Normally it is suggested to use videos with a greater time lapse but as it is not feasible to prolong a test more than 30 minutes, so the time laping of a video was carefully selected.

The temporal impairment which was introduced in the videos was a single freeze which had different freeze durations. The freeze was introduced in the video on a particular interval of time and it was done by copying a frame at approximately middle of the video (6 seconds) and pasting it over the specified interval of the freeze. In this way a specific number of frames were dropped and an artificial scenario of frame freeze/ frame drop was created in each video. The freeze time intervals selected for the videos are as follows:

1. 0.12 seconds
2. 0.2 seconds
3. 0.52 seconds
4. 1 second
5. 2 seconds
6. 3 seconds

The compression was not required for the videos as the videos were already obtained in a very small size (lesser than 300mb) and a very good quality.

For the freeze experiment a interface developed by a team of students from Blekinge Institute of Technology was used, known as Perceived Video Quality Assessment Tool (PVQA tool) [19]. The interface was developed in Java and had admin mode as well as the test subject mode. The admin mode required a password so that the changes needed to be done for a particular test could only be examined/ altered by the admin. As the interface used for the test was developed in Java so there was hardly a chance for network error or any kind of disruption.

The Hardware used for the test was also carefully chosen which is as follows; Compaq laptop CQ610 with a processor 2.2 GHz Core2Duo technology, 2 GB of RAM, Windows 7 Home premium and the native resolution of the screen i.e. 1366 x 768.

After generating all the videos with the help of MATLAB and Virtual Dub, the total number of videos became 35 (30 videos with frame freezes and 5 original videos). As there were 6 freeze durations and total 5 videos, so it resulted in 35 videos (5 Videos * 6 Freeze Durations = 30 + 5 Original Videos). The problem raised over here was that the total number of videos were too many and the test could not be completed in a single go i.e. 30 minutes as recommended by International telecommunication Union (ITU). So, after careful observation a SiTI plot was developed with the help of MATLAB and 4 videos were finally selected out of the 5 original ones. This resulted into drop of the following video:

- Football Practice
After dropping 1 original video, now the total number of videos which were to be used in the test became 28 (24 videos with frozen frames and 4 original videos).

A time duration of 10 seconds was given to assess the quality of the video through ACR scale [18] and a pause of 5 seconds was also given in order to start the next video. In the pause time a grey screen was shown to the subject in order to nullify the effect of color depth effects on the eyes of the subjects.

3.2 Results for Single Freeze Experiment

The 18 test subjects who took part in the test were carefully trained how to take the test as they were given a very short introduction of why and how the test is performed. The interface used for the test was user friendly so there was not a chance that the subjects would face any difficulty. The test was performed in a meeting room, with white background and white light.

![Figure 12. Perceived Quality for Different Video Content Types (MOS vs. Duration of Freeze (Sec) 25 fps)](image)
Figure 13. Perceived Quality for Different Video Content Types (MOS vs. Duration of Freeze (Sec) 25 fps)

Figure 14. Perceived Quality for Different Video Content Types (MOS vs. Duration of Freeze (Sec) 25 fps)
The subjects were familiarized with the test before starting the actual test, the test was explained on a white board and its purpose was also explained. A short test session comprising of 3 videos was also performed in order to make sure that the subject has understood the procedure perfectly. The subject was left alone in the room to perform the test so he/she might not be diverted towards anything. One of the test subjects could not perform the test properly and he was requested to take the test again and he performed it perfectly well in the second time. It is evident from figure1 that the graphs show a negative relation between the freeze duration and the video quality (MOS) as it was expected. But the most abrupt changes are for the Walking Crew video and the Ice skating video, as there is a group of people being shown in the video with a high motion content and a frame freeze causes a bigger impact on the subject. Similar is for the cases of other 2 videos but it is not so much abrupt/sudden as the video did not contain a human movement but it was a moving scenery.

3.3 Multiple Freeze Experiment

For the Multiple freeze experiment same as the single freeze experiment, 18 test subjects were carefully selected. The group of test subjects included 7 female subjects and 11 male subjects with ages ranging from 20 to 30 years (same subjects that were selected for single freeze experiment). The subjects were from youth, expert and non expert category. Once again, the scale chosen to rate the quality of the videos was Absolute Category Rating (ACR) [12] [13].

The videos used for the multiple freeze experiment were the same as used for the single freeze experiment. This time the temporal impairment which was introduced in the videos was a multiple freeze occurrence which had different freeze durations on different intervals. The freeze was introduced in the video on a particular interval of time and it was done by copying a frame random points and pasting it over the specified interval of the freeze. In this way a specific number of frames were dropped and a artificial scenario of multiple frame freeze/ frame drop was created in each video. The multiple freeze time intervals selected for the videos are as follows.

![Figure 15. Perceived Quality for Different Video Content Types (MOS vs. Duration of Freeze (Sec) 25 fps)](image-url)
It can be seen that there are 5 different freeze occurrences (1, 2, 3, 5 and 8 freezes), which have different duration of time.

The following figures shows the comparison of single freeze and 3 freezes

<table>
<thead>
<tr>
<th>Number of Freeze Occurrences</th>
<th>Duration of each freeze occurrence(s)/ total freeze Occurrence (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.067/0.067 0.133/0.133 0.4/0.4 0.8/0.8 1.6/1.6 3.2/3.2</td>
</tr>
<tr>
<td>2</td>
<td>0.067/0.133 0.133/0.267 0.2/0.4 0.4/0.8</td>
</tr>
<tr>
<td>3</td>
<td>0.067/0.2 0.133/0.4 0.267/0.8 0.533/1.6</td>
</tr>
<tr>
<td>5</td>
<td>0.067/0.333 0.133/0.667 0.333/1.667 0.667/3.333</td>
</tr>
<tr>
<td>8</td>
<td>0.067/0.533 0.133/1.067 0.2/1600 0.4/3200</td>
</tr>
</tbody>
</table>

Table 4. Duration of freeze occurrences and total duration of freeze
It can be seen that the graph for 3 freeze videos is showing lesser ACR than the graph for 1 freeze videos, which was expected. The following graphs show the perceived video quality for 5 freeze videos and 8 freeze videos.

Figure 16. Perceived Quality for Different Video Content Types: 1 freeze Instance and 3 Freeze Instance
Figure 17. Perceived Quality for Different Video Content Types: 5 freeze Instance and 8 Freeze Instance

It can be seen clearly that the 2 graphs are showing differences from the graphs that had lesser freeze occurrences and even the graph with 8 freeze occurrences has the least ACR.
4 Performance analysis and evaluation

The three metrics which have been taken into account for this study have been compared; firstly they have been implemented on MATLAB and then tested on the same videos that were used for the subjective tests. The subjective test experiment was done in 2 phases i.e. single freeze experiment and multiple freeze experiment.

4.1 Grading Scale:
The grading scale chosen for this comparison is the same which was used for the subjective tests which is the normal 0-5 grading scale also known as ITU-R five point quality scale. Subjective video quality assessment is done by means of other scales as well such as the 0-100 scale where as 0 represents a poor video and 100 represents an excellent video. But as the subjective tests were performed with the same grading method as are the metrics, so there is no need for any conversions.

4.2 Evaluation

The evaluation for both of the experiments has been discussed below and through MATLAB the following graphs have been generated.

4.2.1 Single Freeze Experiment

![Figure 18. Comparison of the 3 Metrics and Subjective Test Results (Single Freeze)](image)
The results for the Videos containing single freeze event for different time durations has been plotted in the above figure and it can be seen clearly that there is a nominal difference between the subjective test results and the 3 metrics’ results. The 2 metrics: one of them proposed by S.Borer and other one by Stephen Wolf; their performance is almost the same in predicting the effect of frozen frames on the videos. Both of these metrics have shown almost similar behaviour and follow the same curve in the graph but their difference is noticeable compared with the subjective test results. If the Metric proposed by Quan Huyn-Thy & M. Ghanbari [6] is taken into account then it can be seen that this Metric does not notice the frame freeze for freeze event less than 0.5 (0.47 approx) seconds for both 25fps videos and 30 fps videos. But for further freeze durations, this metric’s performance is much closer to the subjective test results. But as it is a very serious drawback for this metric that it cannot detect a frame freeze even less than 0.5 (0.47 approx) seconds, so this metric’s behaviour is considered to be least appreciable.

Now, the results for the multiple freeze experiment have been discussed as follows

### 4.2.2 Multiple Freeze Experiment

![Figure 19. Comparison of the 3 Metrics and Subjective Test Results (Multiple Freezes (25fps))](image)

The videos mentioned earlier were used for the subjective test experiment and then the recorded results were compared with the results of the 3 metrics. As it can be seen in the graph above that the 3 metrics have a nominal difference in performance as compared to each other but they have a slightly higher difference if compared to the results of the subjective tests. It was checked for every video having a freeze
duration lesser than 0.5 seconds or 0.47 seconds to be more precise, that the metric proposed by Quan Huyn-Thy & M. Ghanbari [6] does not work properly. Here the results for the metric proposed by Stephen Wolf show better results than the metric proposed by S. Borer. From the above results, these 2 metrics start showing similar attitude for more than 5 freeze events. As it can be seen that the MOS for 8 freeze events is almost the same as for 5 freeze events. For a video of 10 seconds, 8 freezes of similar time duration are a major impairment that is the reason that the subjective test results show the MOS for 5 and 8 freeze events much realistic than the metrics.

It should be noticed that 2 of the videos (Football practice and Ice skating) with a frame rate of 25 had very high motion intensity and the other 2 (Helicopter view of city and Harbour) had normal motion intensity.

4.3 Authentication of the Subjective tests
The authenticity of the subjective tests and their results has been verified as a white paper was followed to generate all the videos for the tests and the parameters used to generate the videos, also the recommendations of ITU-P. 190 have been followed strictly in the tests.
Chapter 5

5 Results and conclusions

5.1 Performance of S. Borer's Metric and S. Wolf's Metric
After comparison of the 3 metrics for video quality assessment of the videos which were degraded by introducing frame freezes of different durations and had a frame rate of 25 fps, it was noticed that the Metric proposed by S. Wolf based on jerkiness of the video outperforms the other 2 metrics. Though it is evident that the 2 metrics proposed by Borer and S. Wolf have a close competition as Borer's metric is showing MOS much closer to the MOS of S. Wolf's metric.

For the single freeze experiment, the video had a single freeze which is normally not the case in a real time video transmission but still it was clear from the figure 20 that S. Wolf's metric can detect the frozen frames more accurately, hence perceiving the quality of the video much better than of Borer's metric. Similar was the case for multiple freeze experiment which is more closer to the scenario that occurs in real time video transmission. Especially, when we notice that as the number of freezes are increased, the metric starts showing much better performance compared to that of S. Borer's metric. Though there is a big difference between the subjective test results but still S. Wolf's metric is noticeably a better choice regarding video quality assessment.

5.2 Performance of Quan Huyn-Thy & M. Ghanbari's Metric
In this research another noticeable result is that the metric proposed by Quan Huyn-Thy & M. Ghanbari has a very major flaw in it i.e. it does not detect frame freezes which are lower than the duration of 0.48 seconds. Though such a freeze is not easily detectable and even human eye can miss it but still it affects the quality of the video. The metric was tested for both Multiple freezes and single freezes but where ever there was a freeze lower than 0.48 seconds, it was noticed that the Metric does not detect it and resultantly gives a MOS of 5 to the video. Even for the video that has very high motion intensity, the Metric fails to detect the frozen frames.

5.3 Noticeable flaws and loop holes in the Metrics
There are a few flaws and loop holes in all the 3 metrics that were noticed during this research. The major flaw for the metric proposed by Quan Huyn-Thy & M. Ghanbari [6] has already been discussed in section 5.2.

As for the other 2 metrics, they lack the ability to assess the quality of a video which has multiple freezes, as it should be noticed that a video which a shorter total duration and larger frozen frames or freeze occurance, it should have a lesser MOS and vice versa. But these Metrics do not consider the total duration of the video in their video perception methods. For example, it is understandable that a football
match of 90 minutes having total freeze duration of 5 minutes cannot have the same MOS as of a movie trailer of 10 minutes having same total freeze duration (for multiple freeze case).

5.4 Conclusion
From the careful experimentation and comparison of the metrics with each other and with the subjective tests. It can be seen that the video quality assessment of a video sequence can be best performed with S. Wolf's metric. Borer's metric is good at detection of frozen frames as it can detect even a slightest freeze of 0.067 seconds but it does not perceive the quality of the video as better as S. Wolf's metric.

Also for multiple freeze, S. Wolf's metric shows a better performance, as it can be seen that as the number of freeze occurrences increase, the video quality decreases and MOS calculated by this metric decreases as well but in case of Borer's metric it decreases in a very negligible manner.

Finally, Quan Huyn-Thy & M. Ghanbari's metric is not good for real time video quality perception practices as it is not able to detect a freeze which is lower than that duration of 0.48 seconds.

5.5 Future Work
With the help of this performance evaluation and comparative analysis, these metrics can be studied and some mathematical changes can be done in their respective algorithms in order to improve their performance. These metrics need improvement in terms of Video quality assessment for multiple freeze videos and also some improvement needs to be done regarding total duration of a video versus the freeze duration in that video.

Also, the metric proposed by Quan Huyn-Thy & M. Ghanbari can be studied thoroughly and some mathematical changes can be done in its algorithm in order to detect the freezes lower than the duration of 0.48 seconds. This metric performs the best for all the freezes which are greater than 0.48 seconds but it is least useful when the duration is smaller.

There are many video quality assessment metrics available which can be included in this research in order to make the scope of this research much broader and a better evaluation can be done. Also, videos with greater frames per second (fps) can be included in this research to study the performance of these metrics in depth.
References


