VIDEO ARTEFACTS IN MOBILE IMAGING DEVICES

LOKE MEI HWAN
(B.Eng, NUS)

A THESIS SUBMITTED FOR THE DEGREE OF MASTER OF ENGINEERING

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2010

Acknowledgements

Content Page

Summary

For various lossy video compression and network transmissions systems, video artefacts are introduced during the process. As image quality is perceived by the human observer, it would be ideal if only those video artefacts that are discernable to human eyes are detected during quality evaluation. Such quality evaluation requires much computational power and careful understanding of the human visual sensitivity towards these video artefacts.

This work involves a study of the human visual sensitivity towards video artefacts on mobile imaging devices. In our experiments, we evaluate the sensitivity of fifteen users towards some common video artefacts using a database of test video sequences recorded off the screen of a PDA device.

Our results show that the human eye is very sensitive to spatial content loss and its sensitivity towards "blockiness" is dependent on video content.

List of Tables

List of Figures

List of Symbols

Symbols   Definition

H	Height of the image on the display screen
D_1	Discriminant value used in the freeze video artefact detection algorithm
Max	Maximum value
Abs	Absolute value
I	Frame index number
f_i	Current frame of index i
f_{i-1}	Previous frame of index $i-1$
A	Averaging filter
D_2	Discriminant value used in the loss video artefact detection algorithm
D	Percentage of data loss between the consecutive frames
G	Number of pixels which have a large difference of more than 20 grey levels between consecutive frames
g	Subset of G; number of pixels which have a difference of 20 grey levels between consecutive frames and exhibits low grey level values below 40 in the current frame
$f_i(x,y)$	The pixel value of the current frame at position (x, y)
$f_{i-1}(x,y)$	The pixel value of the previous frame at position (x, y)
n	The horizontal length of the frame
m	The vertical length of the frame
p	Number of frames with artefacts in a test video sequence
r	Number of frames with false alarms selected by the system
q	Number of frames with artefacts correctly pick up by the system

1 Introduction

The field of multimedia is constantly growing and there is an ever-increasing number of new imaging displays and devices. Commonly used displays include those in mobile devices such as Personal Digital Assistants (PDA) and mobile hand phones and most of these are able to handle low bit rate videos. However, such compact imaging devices are not able to render the displayed images in high quality due to various limitations (of the hardware & etc). As such a tool for evaluating the quality of images produced by these mobile devices would be useful for the manufacturers of such devices.

This evaluation of the video/image quality of the mobile devices would have to be based on the hardware specifications of the device and displayed video clips/images. This is accomplished traditionally by displaying a reference video clip in the device, and manually examining the displayed output for the presence of any video artefacts, which are the undesirable distortions or defects of the video sequences [1] [2]. In this work, we design a system to quantify the sensitivity of the human visual system with respect to each video artefact.

Test video

Source

Channel simulator

Analysis & Measurement
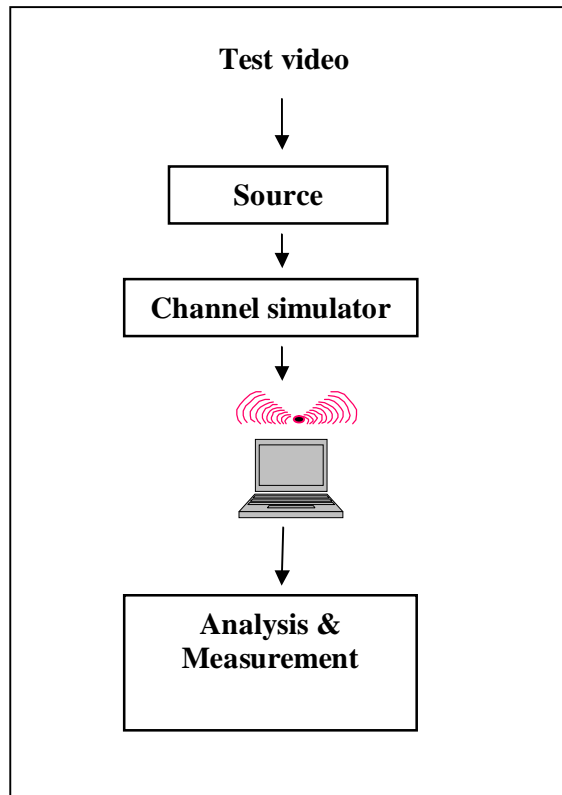
Figure 1. The Video Quality Evaluation Flow Chart

Figure 1 shows an example of a typical video quality evaluation arrangement where first, a reference video sequence is source coded to compress it into an encoded low bit-rate form. At the next stage, a channel simulator which simulates the behaviour of a network sends the encoded data to the imaging device (presented in the form of a monitor screen in Figure 1) which displays the received images. The video displayed on the imaging device display is then subjected to visual analysis and measurement. This workflow results in the imaging device displaying video artefacts.

Video artefacts are the undesirable distortions or defects of the video sequences, and are usually the results of hardware or software defect. It is therefore useful to be able to detect the presence of video artefacts. However, the combination

of both the elements of hardware and software can produce variations of video artefacts which are difficult to detect. The hardware elements that could contribute to the appearance of the artefacts are the image capture card or screen defects. Software designs that could contribute to the appearance of the artefacts include coding and quantization.

A major consideration in designing an automated system is the human visual system which has a range of sensitivities that makes it more attentive to certain details as compared to others. For example, it would be a waste of resources to remove video artefacts that the human viewer is not able to perceive. Therefore, a good understanding of the human visual sensitivity to different video artefacts is needed to design a system for artefact detection. Human visual sensitivity is discussed in detail in Section 2.1.

1.1 <u>Previous Works</u>

The issue of detecting video artefacts is closely related with the field of video quality measurement which has been widely studied. For video quality metrics, the most important task is to quantify the quality of the video, or alternatively to quantify the impact of distortion within the video based on an understanding of the human visual system. The goal of this work is to quantify the sensitivity of the human visual system with respect to each video artefact.

Many video quality metrics in the research field use a standard defined by the International Telecommunications Union (ITU) technical paper "*Methodology for the*

subjective assessment of the quality of television pictures" [3]. This work conducted a series of subjective tests which tabulated the mean opinion scores (MOS) against a database of videos. The performance of several video quality metrics is then compared against the results from the subjective tests. The results from the subjective tests serve as a valuable benchmark for the output of video quality metrics in research, as well as provide the environmental conditions required for a subjective test. This thesis will often make reference to the ITU work for the design of the subjective tests. Out of several video metrics created [4][5][6][7][8], one of the better performing metric was the National Telecommunications and Information Administration (NTIA) video quality metric (VQM) [4], which scores relatively better over a wide range of videos. The VQM metric used a set of weighted-parameters on several components such as image blurriness, colour, and presence of blockiness. These parameters were determined through intensive subject testing and studies by NTIA. However, the performances of these video quality metrics are poor when tested upon a set of videos with a limited bit rate range. In another work [9], the results showed that video quality metrics in general did not perform well when restricted to the videos with low bit ranges.

Although there is research on the effect of the video artefacts toward the overall video quality, there has been limited research on the individual artefacts itself. A previous work by Qi was done as a subjective test which measured the effect of frame freezing and frame skipping on the video quality [10]. In this work, the freeze artefacts and loss artefacts are inserted randomly into parts of the sequences. However, the results of the experiment still aimed at determining the overall video quality, instead of the individual artefacts. The methods for evaluating the subjective

tests and the video sets were based on the Subjective Assessment Methodology for Video Quality (SAMVIQ), which focused on the use of streamlined videos from a network [11]. An important point demonstrated by this work is that research in human vision studies had paid lesser attention to the temporal aspects of video as compared to the spatial aspect. In another artefact work by Lu, the effect of frame-dropping and blurriness on the overall video quality is measured, to examine the relative strength of each artefact to each other [12]. The various factors that contributed to the perceived blur effect included the motion, image contrast and orientation of the distorted videos. The targeted range of videos covered was that of the low bit-rate videos.

Among the various video artefacts, the blockiness artefact is the most studied artefact in the field of image processing. While many metrics and studies aim at investigating the effects of blockiness artefacts on the overall quality of the video sequence, there are relatively few tests trying to quantify the presence of the blockiness artefact itself [13] - [21]. Most of these works are related to the video processing field, which try to reduce the effects of blockiness present, and cannot be used to detect the blockiness that is induced through hardware defects.

To our knowledge, there have been industrial products that are supposed to measure these artefacts, but these systems are intensive in computations, expensive and are only used for the measurement of the processed videos against referenced videos. These systems are not usable for a video quality pipeline which considers the quality of the video as viewed from the device's display. Most of the targeted videos

in hardware applications are stream-lined videos from a network with no reference videos.

1.2 *Thesis Overview*

In this work, we conduct a study aimed at evaluating the sensitivity of the human visual system sensitivity towards 3 common video artefacts, namely: 'freeze', 'frame loss', and 'blockiness'. Video artefacts are inserted into the test videos to simulate the post-effects of hardware defects.

A good understanding of the nature of the video artefacts is needed before the features/parameters that need to be extracted/ measured can be identified. These features/parameters will be measured by conducting a series of subjective tests to measure the human visual sensitivity to each of the artefacts.

In order to test the validity of the subjective results, the extracted parameters are applied onto another set of video sequences with different video content. Much of the work done in this field focuses on quantifying the overall video quality rather than quantifying the threshold of the individual video artefacts.

1.3 *Thesis Overview*

The next chapter provides details of the human visual system, video artefacts and developments in the field of video quality analysis. In Chapter 3, we discuss

details of the video artefacts examined in this work. Algorithms for detecting the

video artefacts are also described here. Chapter 4 describes the materials and

environment of the subjective test while in chapter 5 we describe the subjective test

procedures. In Chapter 6, the results of the study is presented and further examined

while in Chapter 7 we conclude the thesis with discussions and possible future works.

2 Literature Review

2.1 _Human Visual Sensitivity_

In the field of video processing, the quality of an image is traditionally evaluated using methods such as the Peak Signal-to-Noise Ratio (PSNR) or the Mean Squared Error (MSE) method. However, these methods pose several problems that make it difficult for both practical integration and implementation into a video pipeline. The first feature of these methods is the initial requirement for a reference image. This reference image with no distortion is then computed against the distorted counterpart to determine the amount of distortion [22]. Based on this issue, these methods cannot be employed in the use of an environment where no reference image is available. In a quality analysis pipeline, it is often the case that a reference image is not readily available. Placing the reference image through an imaging device would result in a blurring effect when viewed on its display screen, which is what the human eye will see as the end result. Since different types of hardware devices with varying display surfaces is used in the testing process, it is not ideal to keep creating reference images that must be placed and viewed through the various device displays. In this thesis work, the video artefacts are simulated as the defects of the hardware imaging device.

The second issue with the PSNR/ MSE method is that the sensitivity of the human eye is not considered into its computations. While this makes the computations relatively fast and elegant, it is not a completely accurate interpretation

of what the human visual sensitivity notices. The human visual system is structured such that the eye is only sensitive to a certain range of artefacts or differences. This means that a significant amount of details could be removed before the perceptual difference is noticed by the human subject. It is then possible to either reduce or compress the data transmitted without compromising the perceptual quality of the video. In many video processing applications, this perceptual insensitivity is commonly used in the stage of video compression. This is used where a reduction of the bit rate is desirable. By making use of the human eyes insensitivity to details, minimal information is required for the user to appreciate same level of video quality. In many perceptual quality works, the term 'just-noticeable-difference' of an image refers to the threshold that determines the amount of the distortion that must occur between the original images and the distorted images before the subjective viewer notices the difference between the images [23].

Another human-visual related field is the topic of visual attention, where a person's attention is most focused on an area of interest on the screen. During a visual search, the human eye uses a saccadic eye movement which is rapid and jumpy in order to perform quick searches. When focused on a point of interest, the human eye changes its movement to a fixation, where it focuses on the object of interest. The spatial focus is narrowed on the stimulus. The viewer is then likely to be most sensitive to changes made on the area within the eyes' *fixation* focuswhich is a point of interest to the viewer (e.g., a human face). Several contributing factors that will determine the focus of interest include the colour, contrast sensitivity and movement of objects within the video scene [24].

To design an automated process for video quality analysis, it is necessary to understand some characteristics of the human visual system with relation to video artefact. This will allow for a design which is more coherent with the human perception of the image quality. The first reason to consider the human visual system in the process design includes the fact that the human eye is the ultimate end-process evaluator of the image.

2.2 *Video Artefacts*

Video artefacts constitute the undesirable distortions of a video sequences, which renders the video unpleasant to the viewers' eyes. There are several types of video artefacts, ranging from blurriness, blockiness, and ringing. Most works aim at reducing the presence of these artefacts at the software level, but not at the detection of these artefacts.

In the research done on the evaluation of image artefacts by A. Punchihewa [1], objective quality measures were used to evaluate the video quality with relation to the *blockiness, edge-blur, ringing* and *contouring* artefacts. In another work about video artefacts [2], he outlined the various components of a video pipeline and the artefacts mitigation in these pipelines . Most artefacts come about due to a trade-off between the limited bandwidth and optimizing the video quality and so there is a need to better understand the processes in which video artefacts are introduced   to aid in the development of a suitable workflow for proper evaluation of the video quality and the artefacts that arise through the process.

A complication which undermines the study of video artefacts is the spatial-temporal relationship present. Most works evaluate the final quality of the video sequence with relation to the video artefacts added to it, such as the work by Qi [10]. Another type of work which is done in the video processing field is to create a workflow to reduce the number of artefacts in a video sequence [16].

In this thesis work, the number of artefact occurrences is measured through the detection by a real-time system such as a mobile device [25].

3 Common Video Artefacts

In this work, the three video artefacts evaluated are the freeze, frame loss, and blockiness artefacts. These are artefacts which are commonly seen in transmitted videos, such as those in wireless networks. The relation of these video artefacts to visual perception is a key area of examination in this work. By studying the cause and characteristics of these video artefacts, suitable threshold parameters are chosen for measurements during the subjective experiments.

3.1 *Frame Freeze Artefacts*

The freeze video artefact is a video artefact which appears to have no visible change in content during a consecutive sequence of video frames. This freeze effect creates a discontinuous disparity in the video playback, which is perceived as unpleasing to the viewer's eyes.

The presence of this artefact is caused by the slow capturing rate of the camera device, or by the inability of the handheld device to process and display the imaging data at its optimum frame rate. For a network transmission, the freeze video artefact occurs when insufficient data packets are transmitted to form the consecutive frame, and the display algorithm duplicates the previous frame for display. The occurrence of the freeze video artefact is usually followed by an abrupt motion jerk within the video sequence. Due to these characteristics, the freeze artefact affects both the temporal and spatial aspects of the video sequence.

Previous frame to next consecutive frame
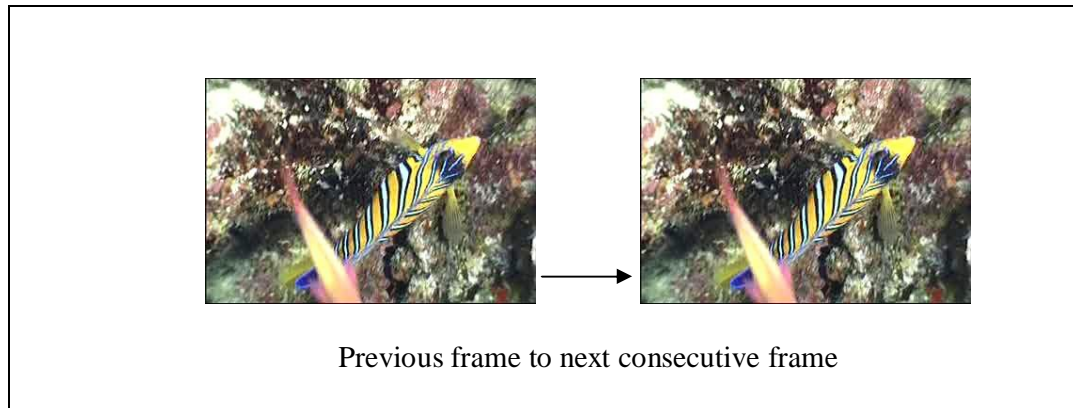
Figure 2. A Pair of Frames with a Potential Freeze Artefact

The images in Figure 2 show an example of a potential freeze video artefact occurrence. The two consecutive frames (previous and current frames) appear to exhibit none or minimal noticeable changes. The term 'noticeable' is the keyword here since the grey level differences between the two video frames cannot be detected by the human eyes, and therefore appears to have no content change. Even if there are differences in pixel values, the viewer will deem the lack of content change as a potential freeze artefact.

Based on the understanding of its characteristics, detecting the freeze video artefact requires 2 components to be measured during the subjective experiments: the spatial and the temporal aspects of the artefact's occurrence. The spatial component refers to the amount of content change between 2 consecutive frames. As mentioned, the human viewer considers a potential freeze artefact only if there is no noticeable content changes. The spatial variable is measured as the minimal change of grey values of the pixels between consecutive frames. The grey value channel consists of

the luminance of the video, and consists of the majority of the information about the video frame. For the temporal component, it is seen that the freeze artefact affects the temporal continuity of the video. Not only must there be a lack of noticeable content change, this occurrence must last at least a specific length of time. This length of time duration of the artefact is the threshold that needs to be measured in the experiment later. This threshold is expressed in the subjective experiments as the number of frames, and is determined under the situation of 30 frames per seconds (fps).

Designing an automatic method for the detection of the freeze artefact is made complicated by a trade off between the measured thresholds and the presence of noise within the video. Noisy artefacts in the video sequence are caused by either software defects such as corruption of the image during transmission or hardware defects. Faulty display of the imaging device, screen reflectance and other external hardware defects such as camera resolutions reduce the chance of gaining the original pixel values of the video sequences.

As measuring the pixel grey values is an important component of content change measurements in this work, it is found that a large amount of noise present in the environment affects the detection of the freeze video artefact. Therefore, the threshold of content change could be adjusted along with the consideration of noise tolerance included. Under the presence of noise, this work will determine the spatial and temporal thresholds in which the human eye will detect the freeze artefact. This is based on the understanding that human eye will detect a freeze artefact only if the conditions of time duration and a lack of content changes are fulfilled.

With the understanding above, we perform the subjective experiments in Section 5, and aim to emulate the results achieved from these experiments. The detection algorithm makes use of the characteristics of the freeze artefact occurrence as mentioned. The 2 conditions for the freeze video artefact are:

1. The content change between 2 consecutive frames must not be perceptually visible.

2. The freeze artefact must occur for a significant period of time.

The threshold results from the subjective experiment are used with the conditions for detecting the freeze video artefact. A perceptual threshold is determined for noticing change in details between consecutive frames. If the amount of grey level changes between consecutive frames is below this threshold, the human eye does not see the details. For the experiments, the freeze artefact was simulated by repeating the frames in-between. The human eye is most sensitive to the luminance value of the frame, with the grey level values ranging from 0 to 255.

The first condition requires the detection of these 'freeze frames'; the video frames without any visible content. The second condition requires the time duration of the freeze frames to be at least of a minimum threshold. Therefore, the main task in a detection algorithm is to firstly determine the presence of freeze frames, and secondly measure the duration of their occurrences. The methods taken to detect the freeze video artefact is described in the following paragraphs. The flowchart and details of the program for this algorithm is presented in the Appendix A.

To determine if the current frame is a freeze frame, the change in content between consecutive frames is measured. This change in content is represented by a discriminant value D_1, which is computed by using the highest absolute difference between 2 consecutive frames.

At frame f_i, the discriminant value D_1 is computed as:

$$D_1 = max(\ abs(\ f_i - f_{i-1}\ ) * A\ )$$ **(1)**

Where:
      D_1 is the discriminant value computed,
      i is the number index of the current frame being analysed,
      f_i is the current frame being analysed,
      f_{i-1} is the previous frame being analysed,
      A is an averaging filter.

An averaging filter A is applied to the recorded image sequence to reduce the external environmental noises that influence the readings. The averaging filter A is a 3 x 3 matrix given as:

$$A = \frac{1}{9} \times \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$ **(2)**

Discriminant value D_1 is reflective of the content change between consecutive frames. When this discriminate value is smaller than a specific threshold, there is insufficient noticeable content change between consecutive frames. From the subjective experiments, the threshold for the discriminate value D_1 was found to be 16.5. The values of the discriminate value was determined by examining the

subjective videos in which participants had noticed the artefacts and measuring the change between the frames based on Equation 1 above. In the presence of noise, this threshold could be given a higher value to enable a small percentage of noise to be tolerated. In lighting and camera situations with higher noise levels, where the original threshold is deemed to be too sensitive, it is found that the threshold value for D_1 can be adjusted to 19.5.

After a freeze frame is identified, the time duration of this freeze frame occurrence has to be measured. The result of time threshold comes from the results of the subjective experiments detailed in Section 5, and was stated to be the duration of 3 frames. During the detection process, the system tracks the number of consecutive freeze frames that had occurred.

Once the threshold (i.e. 3 frames) has been reached, this sequence of frames is identified as a single occurrence of freeze artefact. Any freeze frame which occurs after these 3 frames also constitutes as the same freeze artefact. If a non-freeze frame (a video frame that consists of a change of image content) is present thereafter, this signals that the current instance of freeze artefact has ended. The detailed diagram of the freeze detection algorithm is shown in Appendix A.

3.2 _Frame Loss Artefacts_

The frame loss artefact is a video artefact which appears as a sudden loss of video data or frames. This is commonly noted by a discontinuity in the content of the

image sequence. The affected video sequence would appear to have a momentary flicker on the screen if the loss artefact occurred briefly. Otherwise, it would be displayed as a sudden blank screen. The loss video artefact affects both the spatial and temporal aspect of the video sequence, creating an unpleasant flickering effect. The effects of the different loss artefacts (full-loss and half-loss frame types) can be seen in consecutive images in the following Figure 3. Video flickering caused by the loss video artefact is unpleasant to the user viewing the imaging device. Loss of visual content is a very critical issue in video processing and network applications.

Normal Frame
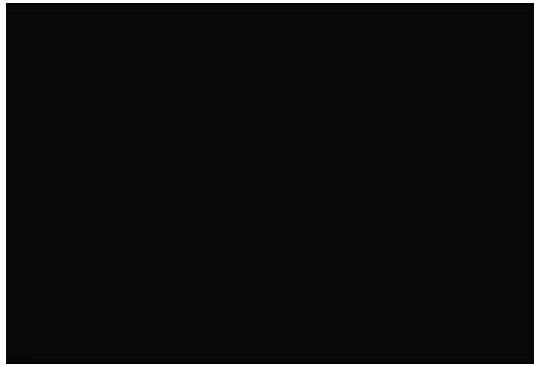
Full Loss Frame (Lossy)

Half Loss Frame (Lossy)

Figure 3. Comparison of a Normal Frame and Lossy Frames

In this work, the loss video artefact is categorized into the two types shown in Figure 3: a *full-loss* frame, and a *half-loss* frame. The presence of a full-loss and half-loss frame will bring about the effect of a screen flicker or blank screen. The presence of loss video artefacts in a video sequence is due to loss of data packets during the network transmission. When data packets are lost and the imaging device still attempts to continue displaying the transmitted video frames, the lost packet components form the blank parts in the frame loss. As a result, the receiving display will display video frames that are either completely blank (full-loss frames) or incomplete (half-loss frames).

The video loss artefact is characterized with the sudden loss of data, with the following consecutive frames not expressing any useful data for the viewer. Similar to the freeze video artefact, the loss video artefact affects both the spatial and temporal component of the video. Loss of video content severely affects both the spatial component and the temporal continuity of the video sequence. Therefore, 2 thresholds parameters need to be measured from the subjective experiments: firstly, the threshold of distortion within the video frame, and secondly, the threshold of the time duration of the artefact. The threshold of distortion within the video frame is a numerical value derived from change of pixels grey levels within consecutive video frames. The threshold of time duration is measured as number of consecutive frames occurrence, under the imaging device's play-rate of 30 fps.

Difficulties of designing an automatic method of detecting loss artefact involve the false alarms of selecting frames with the fade-out effect and sudden scene

changes. The fade-out effect is a typical video effect which makes the scene darken to a blank screen and is typically used in film production for transition to another scene. The method should be designed with consideration of minimizing the chance of false alarm detections.

The detection algorithm for the loss video artefact considers both the spatial and temporal aspects of the video. The 2 conditions of a loss video artefact are defined by the following:

1. The content change between 2 consecutive frames must be abrupt and significant.
2. The content change must be viewed as a loss of data, where the changed pixels become pixels of low grey level value.

Based on the two conditions, it is necessary to keep the knowledge of the previous and current frame status, which requires knowing whether they are considered as loss frames. In this work, we consider three possible types of loss frame statuses that are based on the percentage of data loss : *Full*, *Half*, and *Normal*. The *Full* and *Half* types are considered as contributors to the frame loss artefacts.

Using the first condition, the first task is to detect sudden and significant content change between consecutive frames. This content change is represented by a disciminant value D_2, which is computed as the absolute change in the mean pixel grey levels. If this discriminant value is larger than the perceptual threshold, there is said to be sufficient content change between the frames.

For each video frame, the system computes the discriminant value, until it encounters a video frame with a disciminant value larger than the perceptual threshold. This perceptual threshold found from the videos used during the subjective experiments is 9.5. Thereafter this video frame can be evaluated for its image content to determine its frame status with respect to the loss artefact. Any later consecutive video frame that does not differ largely in disciminant value is likely to be of the same frame status.

The equation for the discriminant value D_2 is given to be:

$$D_2 = abs\left(\frac{1}{n \cdot m} \sum_{x=0}^{x=n-1} \sum_{y=0}^{y=m-1} f_i(x, y) - \frac{1}{n \cdot m} \sum_{x=0}^{x=n-1} \sum_{y=0}^{y=m-1} f_{i-1}(x, y)\right) \qquad (3)$$

Where:
> D_2 is the discriminant value,
> i is the number index of the current frame,
> $f_i(x,y)$ is the pixel value of the current frame at position (x, y),
> $f_{i-1}(x,y)$ is the pixel value of the previous frame at position (x,y),
> n is the horizontal length of the frame,
> m is the vertical length of the frame.

Upon finding the first frame that exhibits a significant change in content, the next condition is to identify whether it is a loss frame and measure the duration of the occurrence. In order to identify the status of the frame, the percentage of data loss between the previous and current frame is measured. Based on the knowledge of the previous frame and the amount of data loss, the current given frame is determined to be a *Full* or *Half* loss frame, or a *Normal* frame. In this work, the *Half* frame loss refers to any frame with 50 – 85% data loss. A higher data loss (more than 85%) indicates a *Full* frame loss, whilst lower data loss (lesser than 50%) indicates a

Normal frame. The percentage of data loss chosen for the *Normal* frame were placed at higher value of 50% as this reduces the chance of false alarms through gradual scene change.

Two different measurements are used based on the previous frame. The first case is when the previous frame state is a *Normal* or *Half* frame, while the second situation is when the previous frame state is a *Full* loss frame. This is because of the possible frame state transitions when there is content change between the consecutive frames.

In the first scenario where the previous frame state is a *Normal* or *Half* frame, the data loss is determined by the following:

$$D = \left(\frac{g}{G} \right)$$
(4)

Where:
      D is the ratio of data loss,
      G is the number of pixels which have a difference of more than 20 grey
      levels between consecutive frames
      g is the subset of G which also exhibit grey level values lower than 40

For the second scenario where the previous frame state is a *Full* loss frame, the amount of data loss is determined by:

$$D = \left(\frac{g}{n \cdot m} \right)$$
(5)

Where:
      D is the ratio of data loss,
      G is the number of pixels which have a difference of more than 20 grey
      levels between consecutive frames
      n is the horizontal length of the frame
      m is the vertical length of the frame

The computation of data loss is dependent on the number of pixels which have experienced a change in grey levels and the ratio of these pixels which had became low grey values.

After identifying a loss frame, the algorithm will determine the duration of the loss artefact. Using the results from the subjective experiments in Section 5, it was found that the number of frames required for a loss artefact to be noticed is 1. This means that the occurrence of a single loss frame is sufficient for this to be a frame loss artefact. This is due to the human visual system being sensitive to sudden changes in spatial content. A consecutive sequence of loss frames is considered to be a single occurrence of a loss artefact. When a *Normal* frame is encountered after a sequence of loss frames, this is considered to be the end of a loss artefact occurrence.

This algorithm workflow prevents fade-out effects from being detected as false alarms. The fade-out effect is a common transition scene used in movie clips. As the fade-out effect usually progresses over a significant number of frames, the human eye does not pick this up as a loss artefact. This implemented workflow will also prevent picking the scene change as a false alarm as the next scene consists of image information. The detailed diagram and parameters table for the loss artefact detection algorithm is found in the Appendix B, whilst Section 6 describes the implementation of the subjective test results.

3.3 Frame Blockiness Artefacts

The blockiness video artefact embeds discontinuous edges of blocks into the video image, making it discomforting to the viewer's eyes. The blockiness artefact is commonly seen together with the other two video artefacts in video transmission. The presence of this artefact is also often found together with many other kinds of image-related artefacts such as blurring and ringing.

The following Figure 4 shows an example of the blockiness video artefact. The presence of the blockiness artefact is mostly introduced during video compression processes with block-transforming techniques, such as the MPEG compression. Such methods make use of lossy quantization processes in order to maximize the compression of the video to low bit rates. For networks, blockiness artefacts tend to appear along side with loss video artefacts when there is a loss in data packets during a video transmission.
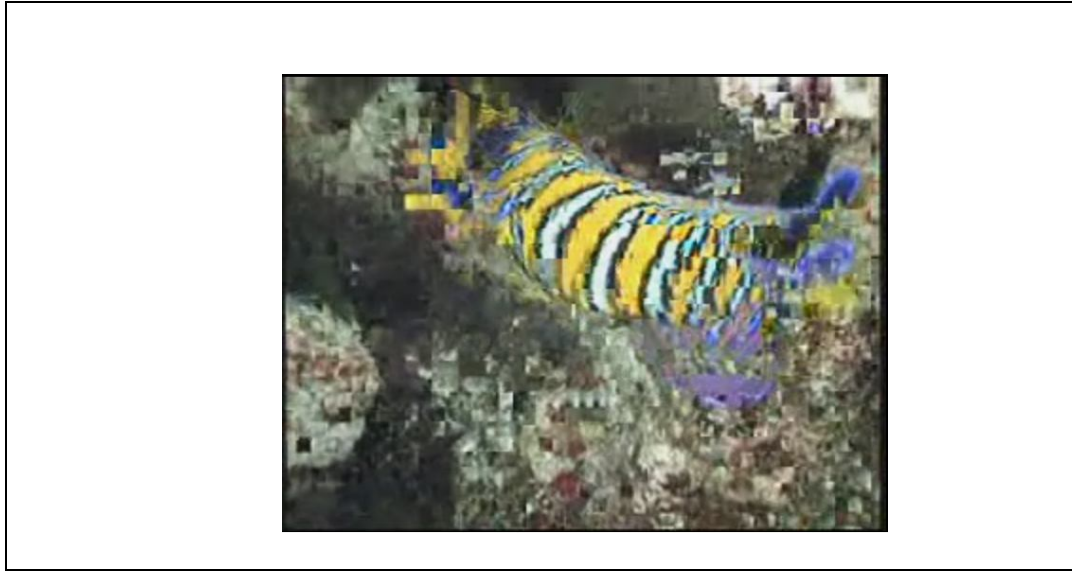
Figure 4. A Blocky Video Artefact

Among the known imaging artefacts, the blockiness artefact is a frequently studied artefact in research. There had been several research papers written on the effects of the blocking artefacts on the overall quality of the video sequences, but several issues within these works have not been addressed [13] - [21]. Firstly, these works do not measure the quantity of blockiness artefact alone, but instead relate the blockiness quantity with the overall video quality. Secondly, most of the existing related-works still use the mean square error as the main method for measuring the severity of distortion, which does not accurately reflect the sensitivity of the human visual system.

As it is often seen in the presence of other artefacts, the detection of the blockiness video artefact alone presents a difficulty. The work is to determine the conditions where the subjective viewer will start to notice the blockiness artefact.

Other details of interest include the characteristic of the videos where the blockiness artefact occurs.

The blockiness artefact affects the spatial aspect of the video. The parameter that is considered is the rate of compression done on the video and the content characteristics of the videos. The procedures for subjective experiments are further described in Section 5.2.

4 Designing Subjective Experiments

In the experimental procedures for video artefact detection, the main driving factors behind the designs are the human visual system and the video quality pipeline. The video quality pipeline is aimed at detecting the video artefacts on a mobile imaging device using a non-reference method.

Figure 5 shows the proposed pipeline which takes into consideration human visual system:
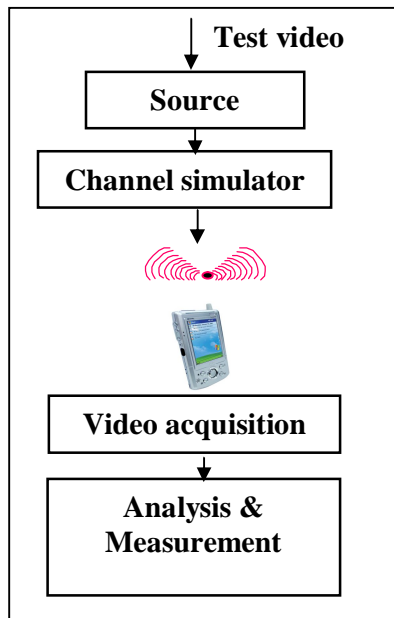
Figure 5. Proposed Video Quality Evaluation Pipeline

The proposed video quality evaluation pipeline setup is similar to that in Figure 1. The concept behind the pipeline is as follows: if a video sequence with no

distortion was placed into an imaging device (such as a PDA), the system could perform quality evaluation based on the hardware defects of the imaging device. During the playback of the video sequence in the imaging device, the screen of the imaging device is recorded. Analyzing this recorded playback off the device screen will allow for the testing of the artefact based on the hardware defects, although this method assumes that the recording device has minimal errors introduced.

However, it is difficult to create and control the amount of hardware artefacts in quantity. Therefore, the situation in Figure 5 is simulated using another method. First, video sequences with added and controlled quantities of artefacts are generated. These distorted sequences are then placed into the imaging device. The imaging device in this case, is a PDA device. The final output on the imaging device display will appear to the viewer in a similar output as a hardware artefact. This displayed image is recorded by a camera system, which can pass the captured video frames to the computer for video quality analysis. The camera device has to be adjusted to obtain a clear image of the imaging device, and its parameters are fixed between the experiments. In this work, the captured video frames are used as the control group for the subjective experiments in Section 5. The new workflow using the distorted video sequences with quality loss is shown in Figure 6.

```
┌─────────────────────────────────────────────┐
│           ┌─────────────────────┐            │
│           │  Video Sequence with │            │
│           │      Artefacts       │            │
│           └─────────────────────┘            │
│                      │                        │
│                      ▼                        │
│              ┌──────────────┐                 │
│              │  PDA Screen  │                 │
│              └──────────────┘                 │
│                      │                        │
│                      ▼                        │
│              ┌──────────────┐                 │
│              │ Camera System │                 │
│              └──────────────┘                 │
│                      │                        │
│                      ▼                        │
│              ┌──────────────┐                 │
│              │ Captured Video│                 │
│              │    Images     │                 │
│              └──────────────┘                 │
└─────────────────────────────────────────────┘
```
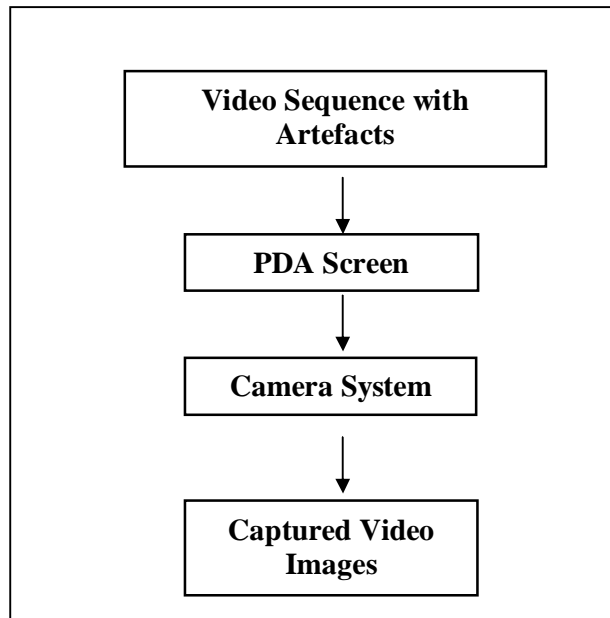
Figure 6. Flowchart for Obtaining the Image for Evaluation

The pipeline shown in Figure 6 will produce output images from the device screen that will be analysed. In a typical video quality analysis, these images will be processed by the computer.

The experimental study carried out in this work will determine the following factors for each video artefact:

1) The characteristic of each video artefact

2) The thresholds and parameters that should be measured with respect to the human visual system

3) Determining the validity of the threshold parameters obtained in the subjective experiments

The characteristic of each video artefact was described in Section 3. With respect to each of the video artefact, the subjective experiments are carried out in different stages to determine each of the factors. These thresholds are determined with relation to the human visual sensitivity. After achieving the threshold parameters, following the workflow in Figure 6, the results can be validified with experimental programs. The experiment program reads the output video images from the camera, and is expected to give similar results to the subjective experiments.

The experiment is dependent on the environmental design and setup. In this work, the camera is used to capture the image of the video playing on the PDA screen and do automatic detection of the video artefact in real time. As the camera needs to record the video image off the imaging device screen, external environment factors such as lighting and the camera focus become factors that can affect the results of the experiment. The camera focus and resolution is adjusted to obtain optimum sharpness where the details of the image can be seen without the presence of electrostatic lines. In a video quality evaluation, this pipeline process will allow for the system to pick up a video artefact based on the hardware. In an automatic detection case, this will allow for a detection of the video artefact due to the hardware defect, assuming minimal defect in the camera device.

4.1 _Camera Setup_

The camera setup is shown in Figure 7. The camera used for the image capture process was a CV-M9 CL model JAI camera [26] which is a progressive scan RGB colour CCD camera with a maximum resolution of 1024 x 768 pixels.
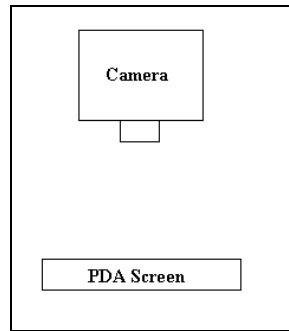
Figure 7. Camera and System Physical Setup

The camera records the image of the PDA screen which will be analysed by the computer in real time. For the subjective tests, the captured video sequences recorded by the camera are used for the subjective experiments. In this work, the imaging device under investigation is a D810 Dopod model PDA. The distance of the camera to the screen is adjusted so that the captured area of interest would be about the size of a typical VGA video frame (640 x 480 pixels). By default, the maximum resolution of the camera is bigger (1024 x 768 pixels) but the smaller VGA frame size is used because it is more commonly used, especially for PC based processing and the smaller VGA frame size allows for faster computations. An Intel Pentium-4 PC with clock speed of 3.0GHz, 1 GB RAM, and SCSI hard disk of speed 10,000 rpm was used for the processing of the captured video sequences.

Due to the physical setup, there are some problems with the captured images. Firstly is the presence of electro-static lines distortions that appears on the captured frames. To overcome this issue, the camera lens focus had to be re-adjusted to a trade off so that the details of the resultant image could be seen, along with a reduction in

the presence of electrostatic lines. The second issue deals with the surrounding environment and has a strong influence on the captured image of the PDA screen. Excessive light thrown onto the PDA screen will result in an output image which has lower contrast and thus, makes the image content harder to view and process. In this work, the PDA screen is adjusted to be brightly lit, while the surrounding room environment is kept dark.

4.2 *Subjective Videos Setup*

Other than the camera, the PDA, and the surrounding environment, test videos were also required for the subjective experiments. For this work, test videos were prepared for the subjective experiments and loaded into the PDA, which in turn displays the videos that are then captured by the camera at a resolution of 640 x 480 (VGA). The processed video sequences used for the subjective tests were of progressive type (non-interlaced); of video size 352 x 288 (CIF) pixels, and YUV 4:2:0 formats.

Five sets of video sequences were selected from the Moving Pictures Experts Group (MPEG) video dataset [27]. These video sets are commonly used in video compression research and in this work. These videos were used as the main reference videos. Using these reference videos, video sequences with varying quantity of artefacts were generated for the subjective experiments. The following Table 1 lists the video sequences with some brief description of the contents.

Video Sequence	Description of Video Sequence Content
Foreman	A man with a headgear talks in the foreground, with a various number of facial expressions. In the beginning of the video, the background scene is static, while the end of the video shows the background shifting very quickly.
Tempete	A group of plants is shown. There are many flying leaves which are quickly falling. The camera slowly pans out as the video progresses.
Mobile	A toy train moves slowly in a room. The room environment consists of many objects such as a spinning metal piece, and a calendar on the wall. The background scene shifts slowly.
News	This video consists of 2 news-announcers in the foreground and a television screen in the background. The news-announcers are sitting in stationary poses and talking. The television shows a ballerina moving vigorously.
Hall	This static scene consists of a corridor within an office. Two men start at different angles of the office and walk in opposite direction past each other.

Table 1. Video Sequences with Descriptions

Video Sequence	Speed of Foreground object	Speed of Background objects
Foreman	**Medium - high**	**Stationary - high**
Tempete	**High**	**Medium**
Mobile	**Low**	**Low**
News	**Low**	**Stationary**
Hall	**Medium**	**Stationary**

Table 2. Content Characteristics of Video Sequences

Table 2 provides information of the speed of the moving objects in the video sequences. The video artefacts added to each video sequences are as follows:

- *Freeze artefacts*: For every set of m video frames in the sequence, remove n normal frames and insert n freeze frames. Freeze frames were produced by replicating the first frame of the consecutive set before the removal. By duplicating the frames, there is no content change within that time period, hence simulating the 'freeze' effect. In this work, the value of m is fixed at 12, whilst the value of n was used for producing the artefacts in the subjective experiments ranged from 1 to 5.

- *Loss artefacts*: The operation of adding loss artefacts is similar to that of freeze artefacts. For every set of m number of video frames in the original video sequence, n number of video frames are removed, and replaced with n number of black screens. The black screens in this work are blank empty frames with no significant content in them. The lossy frame is defined as a complete black screen which abruptly occurs. In this work, the value of m is 10, and the value of n ranges from 1 to 3. The videos used in the subjective experiments used the full-loss artefacts.

- ***Blockiness artefacts***: To generate the artefacts for these videos, the reference videos were passed through an MPEG-2 encoder for compression. They were then decompressed again to obtain the video at different bit rates. The blockiness effect was obtained through the quantization process. The bit rates used were 128, 256, 384, 512, 768 and 1024 kbits per second.

Using the original 5 reference videos, the total number of test videos generated with artefacts was 70 videos. The videos were generated under controlled conditions. Each video consisted of 250 frames. 25 of the videos were added with the freeze artefact, 30 sequences were with the blockiness artefacts, and 15 sequences consisted of the loss frames artefacts. These 3 simulated artefacts are considered as the video artefacts that are commonly introduced through the hardware process.

5 The Subjective Experiments

The subjective experiments are carried out to quantify the sensitivity of the human visual system to the video artefacts. The subjective experiments were conducted using the videos sequences created (see Chapter 4). The experiments consisted of 15 participants (subjects) watching the video sequences (i.e, the recorded images of the PDA screen) in randomised order on a monitor.

5.1 *Setup of Subjective Experiment*

The video sequences were placed and viewed by the subjects on a Tobii 1750 eyetracker LCD monitor, of which its specifications are shown in Table 3.

Monitor size	17 inch, 1280x1024pixels
Monitor Resolution	1280x1024pixels
Resolution of Video Frame	CIF size
Monitor Refresh rate	75

Table 3. Hardware Specifications of Monitor

Similar to the video recording process, the room in which the subjective experiments were carried out was kept dark. This enabled the details on the monitor to be viewed clearly without interference from other light sources on the monitor screen. This setup enables the experiment participants to focus fully on the information displayed on the monitor which is considered as the standard way for conducting subjective experiments. [3]

During the subject experiment, the test participant is seated in front of a LCD monitor (the Tobii Eyetracker system). The viewing distance between the participant and the LCD monitor is $4H$, where H is the length of the image when displayed on the monitor. In this work, the participant sits at a distance of 60cm from the monitor screen.

5.2 *Procedure of Subjective Experiment*

Most subjective experiments require the participants to quantify the overall video quality. In this work, the subjective experiment is conducted to determine the participants' visual sensitivity to the individual video artefacts. The objective is to determine the human perceptual sensitivity to each artefact which is later validated with further experimentation. The subjective experiment is conducted by displaying a series of videos with varying quantities of added artefacts. Thereafter, the participant is asked to rate the individual artefacts based on the severity of its presence. If the presence of the artefact in the video is visually annoying, the participant gives a higher rating for the artefact's severity. While individual opinion may differ, these results generally reflect how sensitive the human eye is to the presence of video artefacts.

The procedure of the subjective experiments is similar to the ITU work [3], but with some modifications. We used a modified version of the Double Stimulus Impairment Scale Method (DSIS) where part of the method was demonstrated and shown in Lu's work [12]. The original DSIS method is complex and requires the participant to do a high amount of manual entry during the experiment. The selected

DSIS variant II method reduces the effort needed on the part of the participants, and differs on the way participant answer questions during the experiments.

The original DSIS procedure is as follows; first the reference video sequence is shown, followed by the processed video sequence. The participant rates the video quality thereafter. The reference video sequence is the original uncompressed sequence, while the processed video sequence consists of added video artefacts. The DSIS variant II procedure in Lu's work [12] differs in that it repeats the original sequence and processed sequence twice as shown in Figure 8. This additional procedure gives the participant more opportunity to compare the videos and identify the artefacts. Another difference between the DSIS variant II procedure and its original is in the way the questions are posed to the participants and the way participants answer these questions. In the original DSIS method, 5 varying levels of descriptions are used to express the range of video quality. These levels consists of 'Bad', 'Poor', 'Fair', 'Good', and 'Excellent'. The levels are respectively represented by a score range of 1 to 5, with 'Bad' having the lowest score of 1 and 'Excellent' having the highest score of 5.

However, our experiments focus on the sensitivity to the presence of artefacts while the original method was intended to quantify the severity of distortion on the overall video quality. Therefore, the questions posed in the subjective experiment have been changed. The scoring range is replaced with a 'yes'/ 'no' type question and the assessor was asked if he/she noted the presence of any video artefacts in the processed video. By randomizing the videos and tabulating the results, the threshold for each parameter described in Section 3 can be estimated.

Instructions for the experiments were described in English. All participants come from an English speaking background and therefore have a reasonable level of understanding to participate in the experiment. During the experiments, an experiment conductor was on-site to help explain the experiments procedures, answer any questions and control the pace of the experiments.
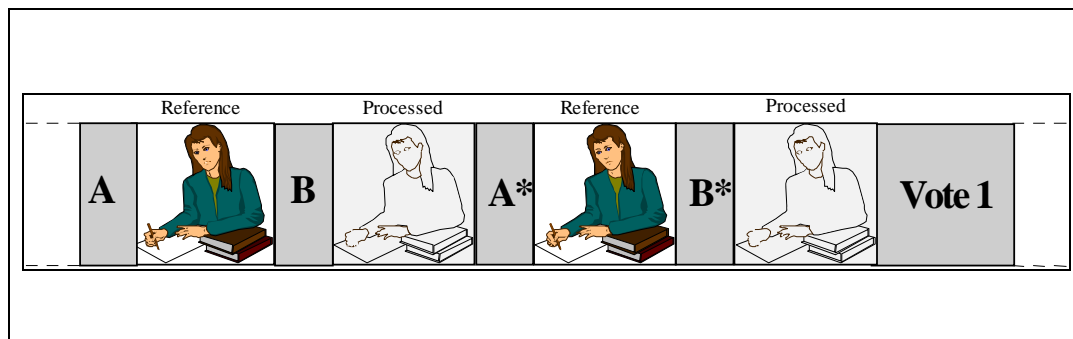
Figure 8. DSIS Variant II Basic Test Cell Process

Figure 8 shows the procedure of the DSIS variant II basic test cell. A fixed set of display operations is performed for each processed sequence. Screen messages are inserted between the displays of each sequence; this will assist in keeping the pace of the experiment and signalling to the assessor of the upcoming video status. The screen messages were shown briefly on the screen for about 2 seconds. They consist of the alphabets 'A', 'B', and the word 'Vote'. Display of the letter 'A' meant that the upcoming video was a reference video, 'B' meant that the video was a processed video, whilst having a '*' symbol next to the corresponding letter indicated that the upcoming video was being displayed for a second time. The 'Vote' screen indicated the time period where the participant could give his answer and comments.

Figure 9. Screen Messages used during Subjective Tests

Figure 9 shows the individual screen messages of 'A', 'B', and 'Vote', which were displayed for each set of the DSIS cell. The procedure is as shown in Figure 8.

The message 'Vote' remains on the screen until the participant provides his/her input about the presence of the video artefact. The participant must determine if a video artefact (freeze, loss or blockiness) is present in the processed video sequence. These video artefacts are the simulation of hardware artefacts.

5.3 *Preparations for the Experiments*

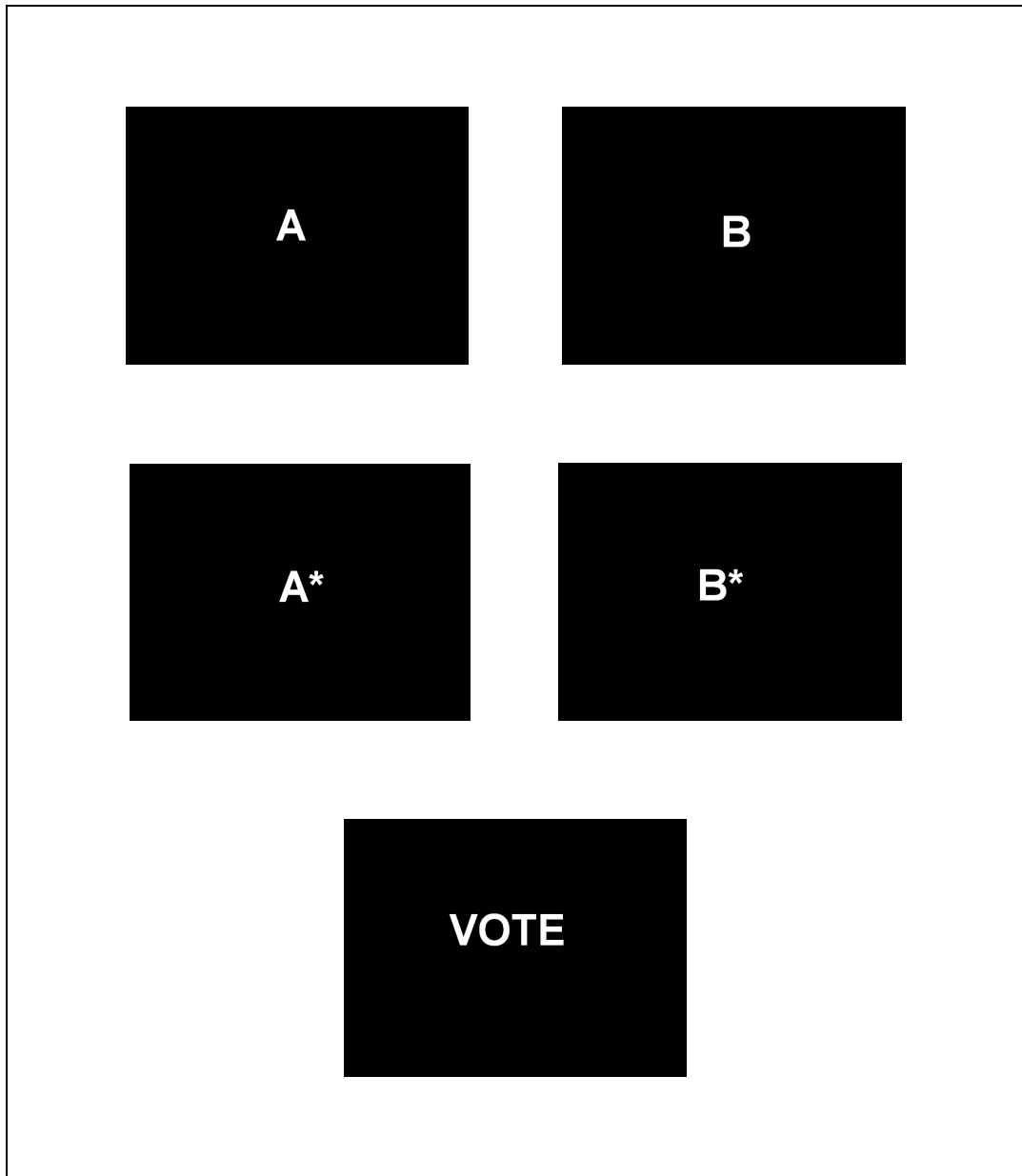Each experiment participant is trained before the actual subjective experiment. This provides the participants with some confidence and understanding before the actual experiment. Before the actual experiment, the experiment procedure was explained carefully to the participant. Following the explanation, examples of each video artefact were shown. These video artefacts were generated with a MPEG reference video set called 'hill', which is the sequence used for training the participant. This video consists of a progressively moving camera which shows scenery consisting of flowers and a house. The objects in the video content are moving at a moderate speed. This training exercise was done in the same order of the DSIS cell shown in Figure 8.

For subjective experiments, the threshold parameters were obtained using 15 participants as this would help to develop a sufficient large database. 15 participants were chosen as a sufficient large sampling database. They are a mixture of male & female experts from the image processing field and non-experts, and some wore spectacles. The participants' ages range from 20 to 50 years old. Each participant was scheduled for the experiment at individual timings. The summary of the subjective experiment participants topology are listed in the following Table 4.

Subject	Gender	Spectacles
1	Male	No
2	Male	Yes
3	Male	Yes
4	Male	Yes
5	Female	No
6	Female	Yes
7	Female	No
8	Male	Yes
9	Male	Yes
10	Male	Yes
11	Female	Yes
12	Male	No
13	Female	Yes
14	Male	Yes
15	Male	No

Table 4: Overall Subject statistics

Throughout the experiment, the participants' responses were recorded and tabulated by the experiment conductor who was there to control the pace and timing of the experiment.

6 Experimental Results

The results of the subjective experiments are collected and tabulated. The results for the freeze frames subjective experiment presented in Table 5 shows the number of freeze frames occurrences required before each assessor considers the occurring artefact as a freeze frame artefact. There is a set of outlier values present in user 12, for the News and Hall video sequences, but its impact is reduced due to averaging between the participants.

Subject	Fore_ freeze	Temp_ freeze	Mobile_ freeze	News_ freeze	Hall_ freeze
1	3	3	3	2	2
2	2	3	2	2	2
3	3	3	3	2	2
4	3	3	4	3	3
5	3	2	3	2	3
6	3	2	2	2	2
7	3	3	4	2	2
8	2	3	3	3	2
9	2	2	2	2	2
10	2	4	3	2	2
11	4	3	4	3	3
12	2	3	2	6	6
13	2	3	2	5	3
14	3	3	3	3	2
15	2	3	2	2	2
Mean	2.6	2.8667	2.8	2.7333	2.5333

Table 5: Results of Freeze Subjective Test

Similarly, the results for the loss video artefact are tabulated in Table 6. The recorded values are the number of loss frames occurrence required for the assessor to consider it as a loss video artefact.

Subject	Fore_ loss	Temp_ loss	Mobile_ loss	News_ loss	Hall_ loss
1	1	1	1	1	1
2	1	1	1	1	1
3	1	1	1	1	1
4	1	1	1	1	1
5	1	1	1	1	1
6	1	1	1	1	1
7	1	1	1	1	1
8	1	1	1	1	1
9	1	1	1	1	1
10	1	1	1	1	1
11	1	1	1	1	1
12	1	1	1	1	1
13	1	1	1	1	1
14	1	1	1	1	1
15	1	1	1	1	1
Mean	1	1	1	1	1

Table 6: Results of Loss Subjective Test

Table 7 summarizes the results from Table 5 and Table 6, about the average number of artefact frames occurrences required for each video sequence. The left column of Table 7 indicates that it takes at least more than 2 consecutive freeze frames to notice a freeze artefact for all video contents. For the practical execution of the verification program, the frame number input is required to be an integer value, so the threshold value for the freeze frames occurrence is rounded up to 3. The right column of Table 7 shows that it takes only one loss frame for the loss artefact to be noticed.

	Average Number of Freeze Frames	Average Number of Loss Frames
Fore	**2.6**	**1**
Temp	**2.8667**	**1**
Mobile	**2.8**	**1**
News	**2.7333**	**1**
Hall	**2.5333**	**1**

Table 7: Tabulation of Overall Freeze and Loss Video Artefacts Results

Compared to freeze and loss, the blockiness video artefact is more dependent on video content. The factor identified for the subjective experiment was the amount of compression required before the assessor considers the video sequence to be a blocky one. The amount of compression is defined by the bit rate of the compressed video. Results of the blockiness subjective experiments in following Table 8 shows all of the assessors considered videos from 3 sequences (namely, Fore, News, and Hall), and consisting of compression rate of less than 768 kbits/sec to consist of blocky video artefacts. The tabulated results for the blockiness test are shown in Table 9. It can be seen that for videos consisting of the Temp and Mobile sequences, some assessors considered the videos of lower compression rate of 1024 kbits/sec to still be blocky.

Subjects		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Fore	128	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	256	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	384	Y	-	Y	Y	Y	Y	Y	Y	-	Y	Y	Y	Y	Y	Y
	512	Y	-	Y	Y	-	-	-	Y	-	Y	Y	-	Y	Y	Y
	768	Y	-	-	Y	-	-	-	-	-	Y	-	-	Y	-	-
Temp	128	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	256	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	384	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	512	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	768	-	Y	Y	Y	Y	Y	-	Y	Y	Y	-	Y	Y	Y	Y
	1024	-	Y	-	-	-	-	-	Y	Y	-	-	-	-	-	-
Mobile	128	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	256	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	384	Y	-	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	512	Y	-	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	768	Y	-	Y	Y	-	Y	Y	Y	Y	Y	Y	Y	Y	Y	-
	1024	Y	-	-	-	-	-	-	Y	Y	-	-	-	-	-	-
News	128	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	256	Y	Y	Y	Y	Y	Y	-	Y	Y	Y	Y	Y	Y	Y	Y
	384	Y	Y	-	Y	-	Y	-	-	-	Y	Y	-	Y	Y	-
	512	Y	-	-	Y	-	-	-	-	-	Y	-	-	Y	Y	-
	768	-	-	-	-	-	-	-	-	-	Y	-	-	Y	-	-
Hall	128	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	256	Y	Y	Y	Y	Y	Y	-	Y	Y	Y	Y	Y	Y	Y	Y
	384	Y	-	Y	Y	-	-	-	Y	Y	Y	-	Y	Y	Y	Y
	512	-	-	-	-	-	-	-	-	-	Y	-	Y	Y	Y	-
	768	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-

Table 8: Results of Blocking Subjective Test

Subject	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Fore	768	256	512	768	384	384	384	512	256	768	512	384	768	512	512
Temp	512	1024	768	768	768	768	512	1024	1024	768	512	768	768	768	768
Mobile	1024	256	768	768	512	768	768	1024	1024	768	768	768	768	768	512
News	512	384	256	512	256	384	128	256	256	768	384	256	768	512	256
Hall	384	256	384	384	256	256	128	384	384	512	256	512	768	512	384

Table 9: Tabulated Results of Blocking Subjective Test

The results for the freeze and loss video artefacts are consistent for the various sequences. The results from these two subjective tests are verified with algorithms in the following sections. Using another set of video sequences with known number of freeze and loss frames occurrences, the threshold retrieved from Table 7 will enable the artefacts in this new set of sequences to be detected. The physical setup for the camera, imaging device and computer system follows that of Section 4.1. The detection algorithms used for the freeze and loss video artefacts are explained in Sections 3.1 and 3.2. We examine the validity of the test results in the following section.

6.1 Examining Validity of Subjective Test Results

A C++ based software application is created to test the subjective test results. This application detects the artefacts from the mobile device screen in real time. A screenshot of this application is shown in Figure 10.

The program runs in real time on a PC, and displays the current captured camera image. Users are able to draw an area of interest around the image in the screen of the PDA device to select the area of interest for detection analysis.
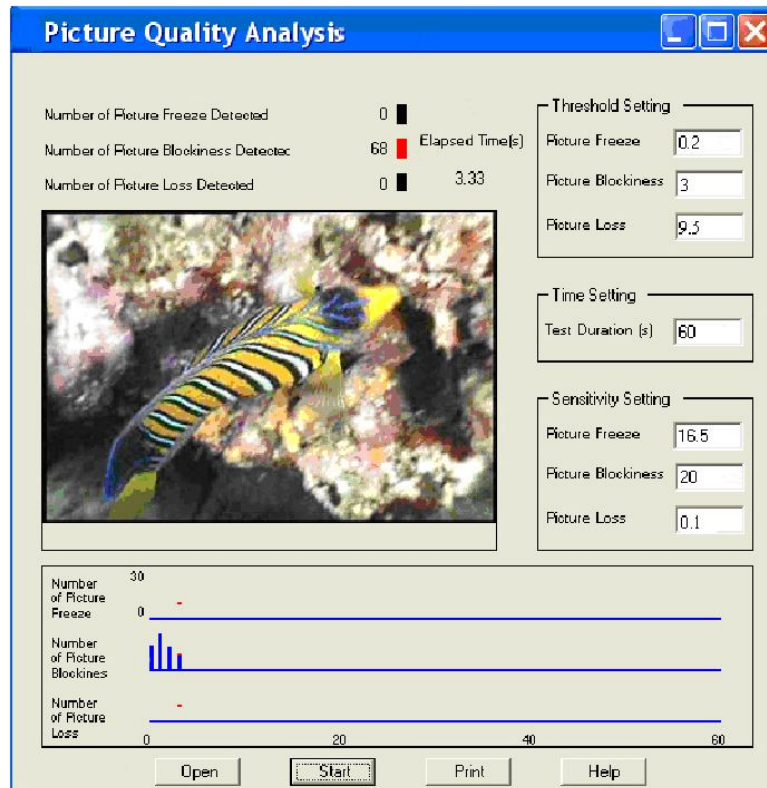

Figure 10. GUI of Artefact Detection Application

Figure 11. Area of Interest drawn around the Image

The parameters values obtained from the subjective experiments are tested using another set of video sequences. Given these values, the implemented application should detect the artefacts on this set of sequences.

The set of video sequences consists of videos from MPEG standards and the RSSCA testing video set. The video set consists of the sequences: coastal guard, dancer, group, royangle, and squirrel. The video artefacts were inserted in the method used in Section 4.2. Similarly, the physical environment follows the conditions in Section 5.1. For the MPEG video set, there were 20 occurrences of freeze artefacts and 25 occurrences of loss artefacts respectively. The RSSCA video set consisted of fewer artefacts but each artefact lasted for a longer duration. On multiple loops of the camera recording the video from the PDA screen, the implementation should be able to detect the artefacts consistently.

Given a test sequence with p artefact frames, if the system accurately detects q artefacts and picks up r artefacts of false alarm, the detection accuracy is computed as

the formula: *(q-r)/p*100%*. During the experiment, any artefact detections done by the system is recorded, and the results can be examined by viewing these recordings. The following paragraphs present the individual results for the freeze and loss artefacts.

Freeze video artefacts: In an optimum situation, the detection system was able to detect 95% of the freeze video artefacts. In the worst case, there was a case of false alarm when 2 consecutive freeze sequences were counted as a single occurrence of freeze artefact.

Loss video artefacts: The system detection rate was 97%. There were less occurrences of false alarms for the loss artefact as compared to the freeze artefact. However, from examining the video recordings, there were instances of missed detections where a *Half* loss frame is labelled as a *Normal* frame instead. This is due to the percentage of data loss between consecutive frames not being significant enough to trigger detection within the system. These *Half* loss frames consist of data loss and therefore would be considered as lossy frames to the viewer.

6.2 *Discussion*

In following section, we raise several issues about this work with respect to the subjective experiments and the characteristics of the individual artefacts. The frame loss and blockiness video artefacts exhibit specific cases which may affect the effectiveness of the algorithm.

For the frame loss artefact, the assumption used in the detection workflow is that frames with loss artefacts consist of low grey level values, as shown in Figure 3. However, if this definition is expanded to loss frame with high grey level values, the algorithm used would need to be corrected. For this situation, the mean variance of the image frame could be used as the means for detection instead.

The frame loss detection implemented in Section 3.2 would fail in cases of a scene transition to a dark scene. Although this is not a frequent case, this situation must be noted. It would be picked up as a false alarm.

During the subjective experiments, the blockiness artefact appears to have less consistent results between assessors as seen in Table 9 due to its dependence on its content and the assessors' varying opinions. Although the detection algorithm was not implemented to detect the blockiness video artefact, several characteristics of the blockiness artefact can be observed during the experiments. Firstly, video sequences with fast moving objects have blocky artefacts that are spotted by the assessors even at higher bit rates. One sample sequence is the tempete video with fast falling leaves. Another type of video with such noticeable artefacts is the image sequence with complicated textures: mobile sequence. This is due to the quantization process in compression creating the blocky regions at areas of the image which are moving or consist of complicated textures.

Besides the individual video artefacts, one of the topics that are not covered is how the video content contributes to the human visual sensitivity. For instance, the

human visual system has a heightened sensitivity to artefacts found on the human face

in the foreman video.

7 Conclusion

In this work, subjective experiments are conducted to obtain threshold parameters for the human perceptual sensitivity for three common video artefacts: the frame freeze, frame loss, and blockiness artefacts. The workflow is designed with the purpose of performing automated video quality analysis of mobile devices. This setup is used to pick up on hardware faults through detection of the video artefacts.

Video sequences are recorded off the PDA screen, using a camera in real time, in a dark room environment. Through the subjective experiments, the parameters and related variables were determined. From the subjective tests, it is shown that it takes an average of 3 freeze frames to occur before an assessor notices a freeze artefact, and it takes 1 loss frame for the assessor to notice the loss artefact. For the blockiness artefact, it is seen that for videos with fast moving content or complicated textures, the assessors detects the artefacts at roughly 1024 kbits/sec. For other set of videos, the blockiness artefact was detected at 768 kbits/sec.

The parameters were dependent on the spatial and temporal properties of the video artefacts. The results of the experiment are then passed through a software implementation of the detection algorithm, using a second set of video sequences. The implemented software tests have a range of over 90% detection rate for the validation stages.

7.1 *Future Works*

The results of the subjective tests could be used for further investigations in the area of quality analysis. While this work provides a way to automatically detect the artefacts, the environment is still heavily controlled and uses several assumptions in its implementation. By understanding the limits of the human visual, it brings further foresight towards future developments in the field of video quality metrics or quality evaluation.

Another possible area of future research is how each video artefact dominates in the presence of each other. In broadcasted video sequences, there are often cases where the freeze and blockiness artefacts occur at the same time. A subjective experiment could be designed such that controlled pairs of artefacts are placed into the viewed video sequences. Hence, the point of interest is which artefact is more likely to be noted by the assessor.

Bibliography

[1]   A. Punchihewa, D.G. Bailey, R.M. Hodgson, "Benchmarking image codes by assessment of coded test images: the development of test images and new objective quality", Journal of Telecommunications and Information Technology, 2006, pp. 11-16

[2]   A. Punchihewa, D.G. Bailey, "Artefacts in Image and Video Systems: Classification and Mitigation", Proceedings of Image and Vision Computing New Zealand 2002, pp.197-202, 2002

[3]   ITU-R BT 500, "Methodology for the subjective assessment of the quality of television pictures", June 2002.

[4]   Stephen Wolf, Margaret Pinson, "Video quality measurement techniques", NTIA Report, NTIA/ITS, June 2002.

[5]   Feng Xiao, "DCT-based Video Quality Evaluation", MSU Graphics and Media Lab (Video Group), Winter 2000.

[6]   Zhou Wang, Alan Conrad Bovik, Ligang Lu, "Video Quality Assessment Based on Structural Distortion Measurement", IEEE Signal Processing: Image Communication, Vol 19, No 2. pp. 121-132, February 2004.

[7]   A.B.Watson, "Towards a perceptual video quality metric", Human Vision, Visual Processing, and Digital Display VIII, 3299, 139-147.

[8]   A.B.Watson, James Hu, John F McGowan III, "DVQ: A digital video quality metric based on human vision", Journal of Electronic Imaging, Vol. 10(1), 20-29.

[9]   M.H. Loke, E.P.Ong, W.S Lin, Z.K. Lu, S.S. Yao, "Comparison of Video Quality Metrics on Multimedia Videos", IEEE ICIP 2006, pp 457-460, 8-11 Oct 2006

[10] Y. Qi, and M. Dai, "Effect of freezing and frame skipping on video quality", International Conference on Intelligent Information Hiding and Multimedia, 2006, pp. 423-426.

[11] F. Kozamernik, "Subjective quality of internet video codecs using SAMVIQ", 2005, http://www.ebu.ch/trev_301-samviq.pdf.

[12] Z. Lu, W. Lin, E.P. Ong, S. Yao, S. Wu, B.C. Seng, S. Kato, "Content-based quality evaluation on frame-dropped and blurred video", *IEEE International Conference on Multimedia and Expo*, pp. 1455-1458, July 2007.

[13] H. R. Wu, and M. Yuen, "A generalized block-edge impairment metric for video coding," IEEE Signal Processing Letters, vol. 4, No. 11, pp. 317-320, Nov. 1997.

[14] H. S. Malvar, and D. H. Staelin, "The LOT: Transform coding without blocking effects," IEEE Transactionson Acoustic, Speech, and Signal Processing, vol. 37, pp. 553-559, 1989.

[15] H. Paek, R.-C. Kim, and S.-U. Lee, "On the POCS-based post-processing technique to reduce the blocking artifacts in transform coded images," IEEE Transactions on Circuits and Systems for Video Technology, vol.8, no. 3, pp. 358-367, June 1998.

[16] Z. Wang, and D. Zhang, "A novel approach for the reduction of blocking effects in low-bit-rate image compression," IEEE Transactions on Communications, vol. 46, no. 6, pp. 732-734, June 1998.

[17] N. C. Kim, I. H. Jang, D. H. Kim, and W. H. Hong, "Reduction of blocking artifacts in block-coded images using wavelet transform," IEEE Transactions on Circuits and Systems for Video Technology, vol. 8, no. 3, pp.253-257, June 1998.

[18] Y. Yang, and N. P. Galatsanos, "Removal of compression artifacts using projections onto convex sets and line process modeling," IEEE Transactions on Image Processing, vol. 6, no. 10, pp. 1345-1357, Oct. 1997.

[19]   G. A. Triantafyllidis, D. Tzovaras, and M. G. Strintzis, "Blocking artifact detection and reduction in compressed data," IEEE Trans. Circuits Syst. Video Techn. Vol.12, no.10, pp. 12(10): 877- 890, 2002.

[20]   Zhou Wang, Alan C. Bovik, and Brian L. Evans, "Blind measurement of blocking artifacts in images," in Proc. IEEE Int. Conf. Image Processing 2000.

[21]   A. Petrovski, T. Kartalov, Z. Ivanovski, L. Panovski, "Blind measurement and reduction of blocking artifacts," in Proc. 48th Int. Symp. ELMAR-2006, Croatia, 2006.

[22]   B. Girod, "What's wrong with mean-squared error," Digital Images and Human Vision, A. B. Watson, ed., Chapter 15, pp. 207-220, the MIT press, 1993.

[23]   X. K. Yang, W.S. Lin, Z. K. Lu, E. P. Ong, S. S. Yao, "Just Noticeable Distortion Model and its Applications in Video Coding", *Signal Processing: Image Communication*, European Association for Signal Processing, Vol. 20, Issue 7, pp. 662-680, August 2005.

[24]   S, Shioiri, T. Inoue, K. Matsumura, H. Yaguchi, "Movement of Visual Attention", *Proceedings of IEEE Int Conference of System, Man, and Cybernetics*, Vol. 2, pp. II-5-II-9, 1999.

[25]   E. P. Ong, S. Wu, M. H. Loke, S. Rahardja, J. Tay, C. K. Tan, L. Huang, "Video quality monitoring of streamed videos", *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1153-1156, April 2009.

[26]   JAI – Industrial CCD/CMOS cameras and video imaging for traffic management, 2007, http://www.jai.com/EN/Pages/home.aspx

[27]   The MPEG HomePage, *http://mpeg.chiariglione.org/*

Appendix A

This appendix consists of all the diagrams and parameters which were used in the freeze algorithm.

Freeze Artefact Detection Algorithm

Figure 12 shows the diagram of the workflow used for detecting the freeze artefact as explained in Section 3.1.
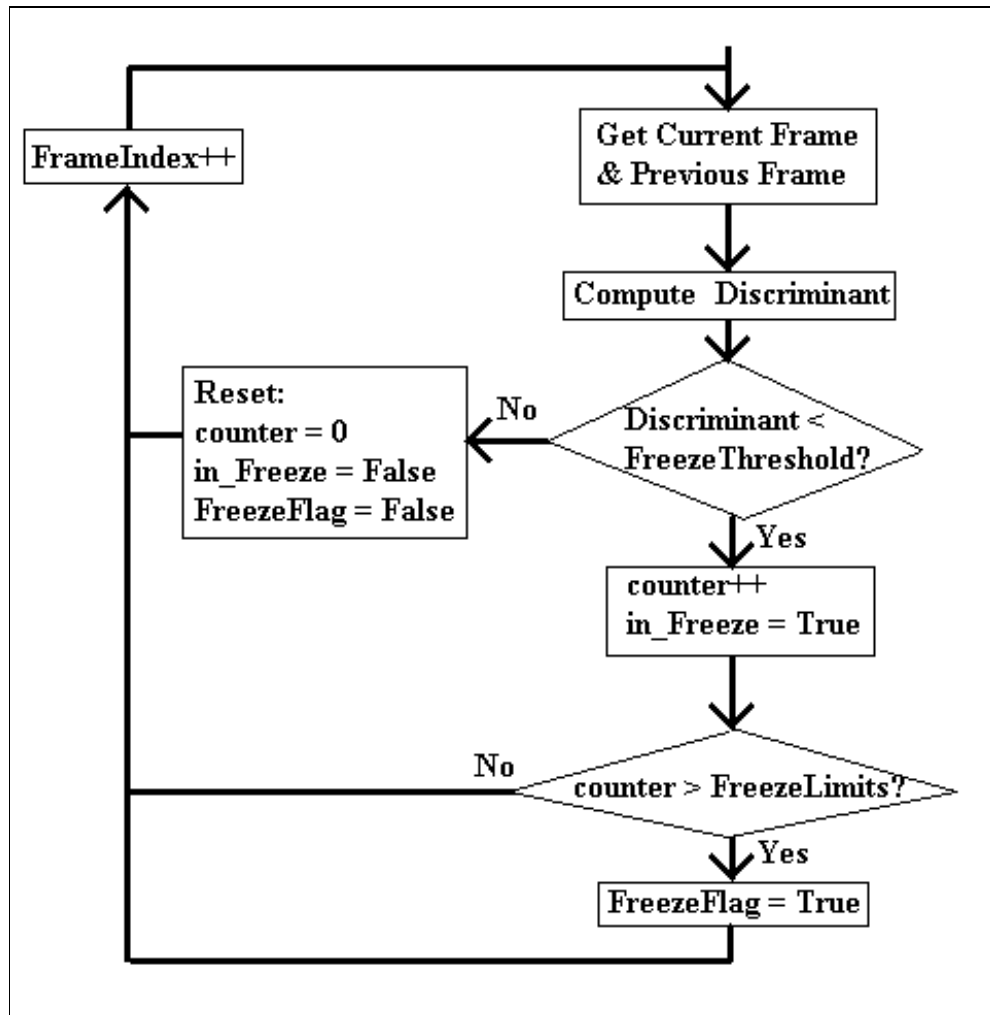

Figure 12. Flowchart for the Detection of the Freeze Video Artefact

For the freeze artefact algorithm in Figure 12, the following terms and parameters in Table 10 are used:

Parameters	Detail
counter	The integer variable records the number of freeze frames (frames with no visible content change) that has occurred in consecutive frame sequence. Its default value is 0. If the next frame is detected as a non-freeze frame, the *counter* parameter resets to its default value.
in_Freeze	This is a TRUE/FALSE flag variable that indicates that the algorithm is in the process of evaluating a possible sequence of consecutive freeze frames. Its default value is FALSE. The *in_Freeze* variable is TRUE when at least 1 freeze frame is detected. When the encountered consecutive frame is a freeze frame, the value remains as TRUE, otherwise, the value resets.
FreezeThreshold	This is a float threshold parameter which is used to determine the frame status. Another discriminant value is first computed from two consecutive frames to determine the amount of content change. This discriminent value is checked against the *FreezeThreshold* parameter; if the discriminant value is lower than the threshold, this means there is a lack of content change between the frames, and therefore the current frame is considered as a 'freeze frame' and *in_Freeze* value is changed to TRUE. The higher the value of FreezeThreshold the more noise and motion is tolerated before determining the frame to be a freeze frame.

FreezeLimits	This is an integer threshold parameter. Before a consecutive series of freeze frames is considered to be a single occurrence of a freeze video artefact, the minimum number of freeze frames required must be known. The value for this threshold parameter was found through the subjective experiments in Section 5. The value for this was 3 frames.
FreezeFlag	This is a TRUE/FALSE flag variable. The default value is FALSE. When one or more freeze video artefacts are detected, the value is TRUE, otherwise, the value resets to default.

Table 10: List of Parameters used in Freeze Artefact Detection

Appendix B

This appendix consists of all the diagrams and parameters which were used in the freeze algorithm.

Loss Artefact Detection Algorithm

Figure 13 shows the diagram of the workflow used for detecting the frame loss artefact as explained in Section 3.2. The step '*Update FrameState*' involves another sub-process which is further illustrated in Figure 14.
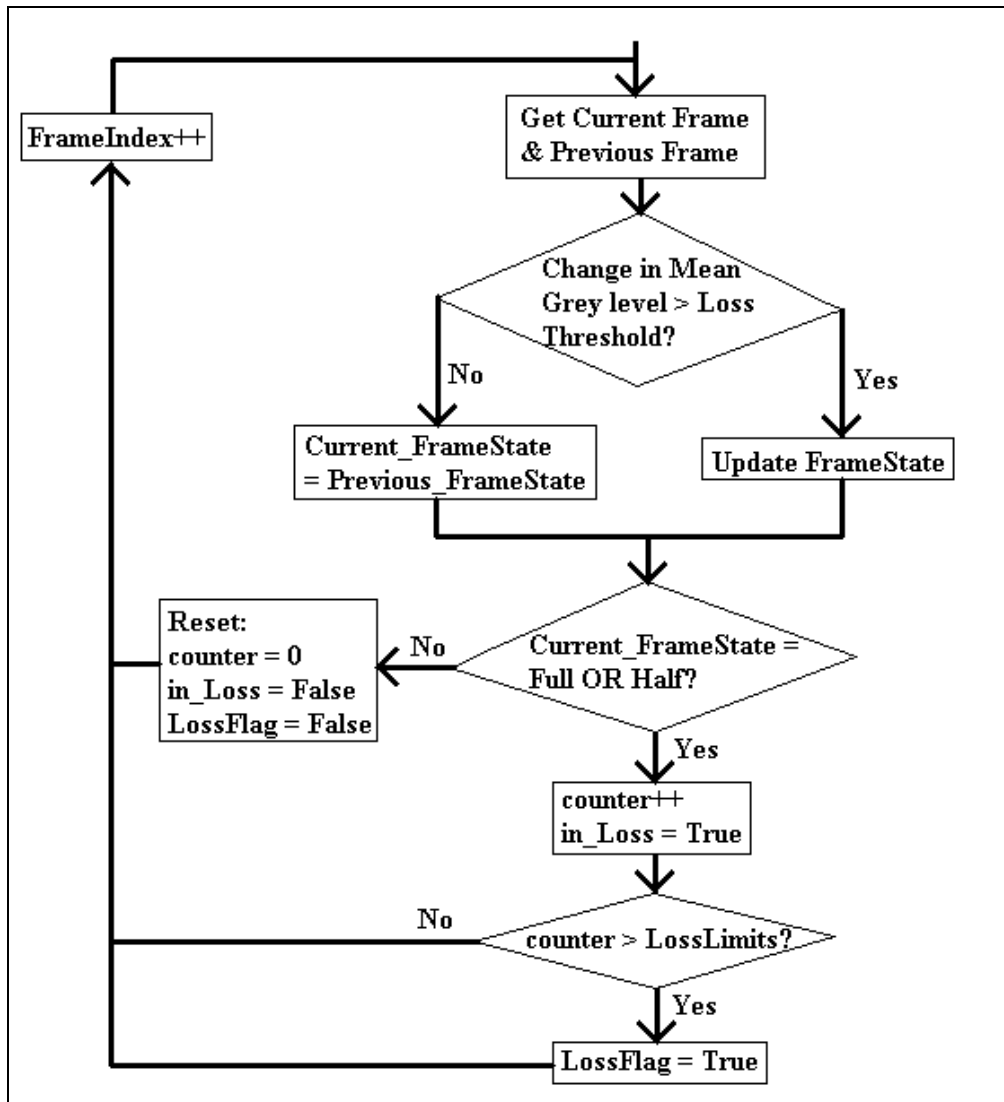
Figure 13. Flowchart for the Detection of the Loss Video Artefact

For the loss artefact detection algorithm in Figure 13, the following terms and parameters in Table 11 are used:

Parameters	Detail
Current_FrameState	This variable records the current frame state. 3 possible values for this parameter are *Normal*, *Half*, and *Full*. The initial default for this variable is *Normal*. This variable is updated through the sub-process *UpdateFrameState*.
counter	This is an integer variable which records the number of consecutive loss frames occurrences. The default value is 0. This variable is incremented for each occurrence of frames with loss artefacts. When a normal frame is encountered again, this counter is reset to its default.
in_Loss	This flag variable indicates the process of evaluating a loss artefact. When at least one loss frame is detected, the flag variable is TRUE. The default value is FALSE. When the system encounters a normal frame again, the *in_Loss* parameter will return to its default value.
LossThreshold	This is the threshold parameter that determines the maximum amount of pixel change that must occur before a frame status (*CurrentFrameState*) check is done. Its value must be a float. If the discriminant value D_2 is larger than the *LossThreshold*, it means there was a significant change in the image content. Otherwise, there was no significant change, and the next pair of consecutive frames can be retrieved. To check the frame

	status, the sub-process *UpdateFrameState* is called.
LossLimits	This is the integer threshold parameter that determines the minimum number of loss frames that must occur before it is considered as a loss artefact. For this range of frames, *in_Loss* value must remain TRUE. The value for this parameter was found through the subjective experiments in Section 5.
LossFlag	This is a flag variable which value depends on the detection of the loss artefact. Its default value is FALSE. When a loss artefact is detected, the flag variable value is TRUE. Otherwise, the occurrence of a non-loss frame will reset the value of variable.

Table 11: List of Parameters used in Loss Artefact Detection

Figure 14 shows the process used for determining the current frame status as described in Section 3.2.
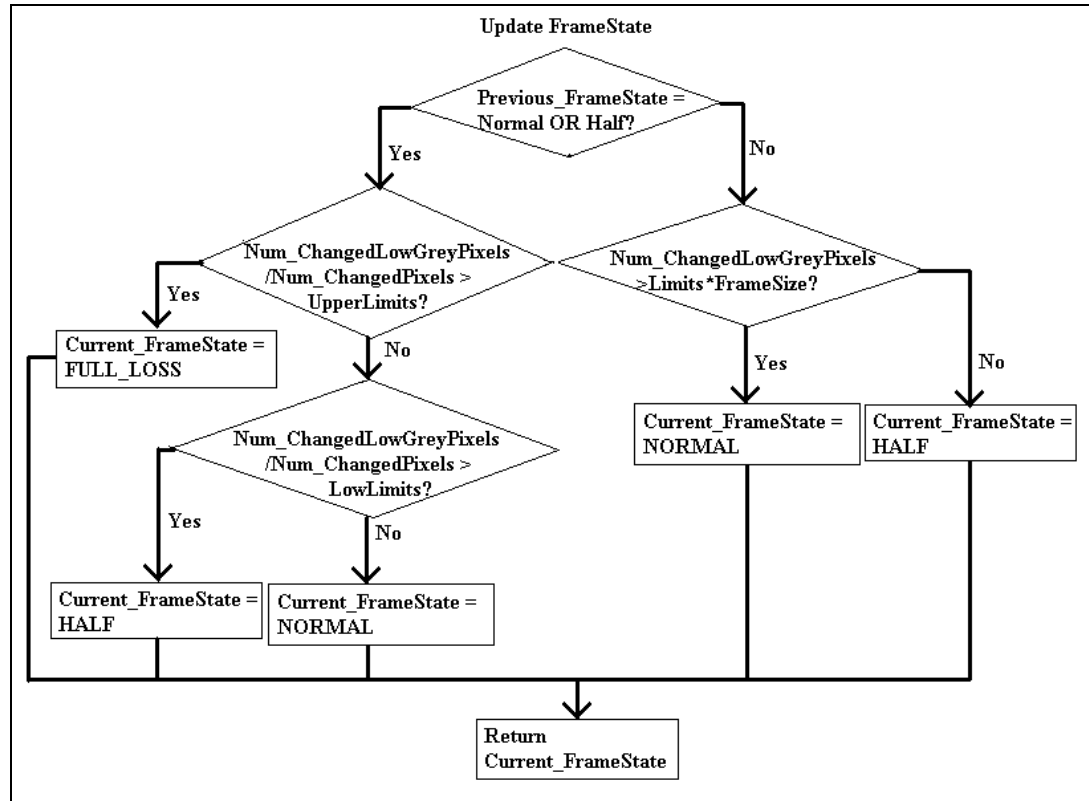

Figure 14. Sub-Process of the UpdateFrameState found in Loss Detection

For the sub-process of determining the frame state in diagram Figure 14, the following terms and parameters in Table 12 are used:

Parameters	Detail
Update FrameState	The sub-process to update the loss status of the current frame.
Previous_FrameState	This variable stores the previous frame state. The values could be *Full*, *Half* or *Normal*.
Num_ChangedPixels	This variable represents the number of pixels whose grey level values have changed significantly between the previous frame and current frame due to possible data loss. For this work, the value of change required between grey levels is 20.
Num_ChangedLowGreyPixels	This variable stores the number of pixels within variable *Num_ChangedPixels* which are of low grey level values. Pixels of interest are those with grey level values of 40 and lower.
UpperLimits	This parameter sets a threshold which is used to determine if the current frame is a *Full* loss frame. This process is used when the previous frame state is *Normal* or *Half*. The parameter *UpperLimits* is set at a percentage of 85%.
LowLimits	Similar to the *UpperLimits* parameter, this sets the threshold to determine if the current frame state was

	Half or *Normal*. This process is used when the previous frame state is *Normal* or *Half*. The parameter *LowLimits* is set at a percentage of 50%.
Limits	When the previous frame state was a *Full* loss frame, this parameter represents a threshold percentage needed to determine if the current frame state is *Normal* or *Half*. This parameter is used if the previous frame state was a *Full* loss frame. The value of this parameter is set at a percentage value of 75%.
FrameSize	This variable is the total number of pixels within the frame, and is a product of the width and length of the video frame.

Table 12: List of Parameters used in the Sub-Process UpdateFrameState