CHAPTER SIX

Compressing Video Based on Region of Interest

Hussein Muzahim Aziz, Markus Fiedler, Håkan Grahn, and Lars Lundberg

Abstract

Real-time video streaming suffer from bandwidth limitation that are unable to handle the high amount of video data. To reduce the amount of data to be streamed, we propose an adaptive technique to crop the important part of the video frames, and drop the part that are outside the important part; this part is called the Region Of Interest (ROI). The Sum of Absolute Differences (SAD) is computed to the consecutive video frames on the server side to identify and extract the ROI. The ROI are extracted from the frames that are between reference frames based on three scenarios. The scenarios been designed to position the reference frames in the video frames sequence. Linear interpolation is performed from the reference frames to reconstruct the part that are outside the ROI on the mobile side. We evaluate the proposed approach for the three scenarios by looking at the size of the compressed videos and measure the quality of the videos by using the Mean Opinion Score (MOS). The results show that our technique significantly reduces the amount of data to be streamed over wireless networks with acceptable video quality are provided to the mobile viewers.

Keywords

Sum of Absolute Differences, Region of Interest, Reference Frames, Video Compression, Mean Opinion Score


6.1 Introduction

Bandwidth is one of the most critical resources in wireless networks, and the available bandwidth of wireless networks should be managed efficiently [10]. Therefore, the size of a video stream should be adapted according to the network bandwidth [5],[6]. Network adaptation refers to how much network resources (e.g., bandwidth) a video stream should be utilize i.e., it is an adaptive streaming mechanism for video transmission [4].

Video Coding [8], like H.264, is developed to encode/decode the video frames with respect to the specific rate required by a certain application. In a low bit-rate video coding, such as on mobile video streaming, the video could skip some temporal/spatial levels to meet the bandwidth limitations.

The main feature of H.264/SVC [11] is to provide bandwidth-optimized transmission for video streaming by observing current network conditions. H.264/SVC provides three types of enhancements for optimized bandwidth transmission. First, it can support temporal enhancements by changing the frame rate. Second, it can support spatial enhancements through resolution, and finally it can support enhancements of the quality through a signal-noise-rate.

The basic element of H.264/AVC video sequence is slicing, where each frame can be divided into several slices [7] and each slice contains a group of macroblocks (MBs) [3],[12]. H.264/AVC introduces Flexible Macroblock Ordering (FMO) as a useful error resilient tool, where H.264/AVC defines seven different types of FMO modes: Interleaved, Dispersed, Fore-ground with left-over, Box-out, Raster-scan, Wipe, and Explicit [16]. The H.264/SVC encodes the video in the way that can be selectively transmitted according to the type option; contents and network condition by using a bit stream extractor [11].

In this paper, we present a video adaptation technique to reduce the amount of data to be streamed over wireless network. The streaming server will identify and extract the high motion slice region (ROI) from the frames that are between reference frames and drop the less motion slice region (non-ROI). Four different ROIs for three different
scenarios are proposed to study the effect of the compression size on the video streaming.

After the mobile device has received the video stream, linear interpolation between reference frames will be performed to reconstruct the pixels that are outside the ROI (non-ROI). Mean Opinion Score (MOS) measurement is used to evaluate the quality of the reconstructed videos.

6.2 Background and Related Work

Several techniques have been proposed for spatial adaptation for slicing the video frames. Mavlankar et al. [2] examine how to determine the slice sizes for streaming the ROI. In their work, the server will adapt and stream according to the regions size of the video content that is desired at the client’s side. They study the trade-off in the choice of slice size. The optimal slice size achieves the best trade-off to minimize the expected number of bits that are transmitted to the client per frame. The output of their work is to predict the optimal slice size regions, which depends on the signal as well as the display resolution for the mobile screen.

Moiron et al. [15], proposed a slicing scheme for enhanced error resiliency. The proposed scheme used FMO without introducing any increases in the bit rate or computational complexity. The frame level priority is provided based on the slice position within the frame. The frame was sliced into three distinct regions. The slice structuring in the encoder was modified to accommodate a new set of rules for video slicing. These rules prevent macroblocks from different regions from being packed into the same slice. They also define when the current slice should be terminated.

Wang et al. [17] applied a cropping technique to perform spatial adaptation to the video stream to overcome the display constraints by producing the ROI and according to the mobile screen resolution. The ROI is determined by using an attention based modelling method. The ROI is automatically detected and crop the informative region in each frame to generate a smooth video sequence.
6.3 The Proposed Technique

In the related work section, the researchers identify the ROI as the most attractive region to the viewers. The ROI is extracted from the video frames to cope with the bandwidth limitation and the mobile screen resolution. In this study, the Sum of Absolute Differences (SAD) metric is used to identify the motion region position which we considered it as the ROI. The ROI will be extracted from the video frames in the server and stream it over the network. On the receiver side, the pixels that are outside the ROI (non-ROI) will be reconstructed from the reference frames by using linear interpolation [18].

6.3.1 Identifying the ROI

The SAD technique is a commonly used technique for motion estimation in various video encoding standards like H.264 [1]. The idea is to take the absolute differences between consecutive video frames. The SAD value will be zero except for the changes induced by the objects moving between the video frames. If there is a lot of motion in one region of the frames, the SAD value in this region will be relatively high, and if there is no motion then the SAD value in this region will be zero. The SAD is computed to identify the position of the ROI. The idea is to place the ROI where there is a lot of motion in the video scene and drop the less motion region (non-ROI). We assume that we could have different ROIs for different sequences in a video.

Two ways of scanning the consecutive video frames are considered to identify the highest intra-slice differences, as shown in Figure 1 and Figure 2, respectively:

- The first way is to scan the consecutive video frames from top-to-bottom, as shown in Figure 1.
Chapter Six

\[
SAD_H(k) = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} \sum_{x=0}^{L-1} |F_x(i, j + k) - F_{x-1}(i, j + k)| \quad (1)
\]

- The second way is to scan the consecutive video frames from left-to-right, as shown in Figure 2.

\[
SAD_V(k) = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} \sum_{x=0}^{L-1} |F_x(i + k, j) - F_{x-1}(i + k, j)| \quad (2)
\]

where \( L \) is the length of the frame sequences, \( N \times M \), is the height and width for the ROI, and \( k \) is a fixed region.

The test videos used in this work were the samples of video sequences Highway, Akiyo, Foreman, News, and Waterfall, with a resolution of 176 x 144 pixels [19]. The chosen videos are well known as professional test videos that have different characteristics.
For Highway video, the $\text{SAD}_H$ value is the largest when the ROI is close to the bottom of the frames as the highest differences among the intra-slices as shown in Figure 3(a). The $\text{SAD}_V$ value for Highway video is the highest when the ROI is close to the sides of the frames as shown in Figure 3(b).

For Akiyo and News videos, the $\text{SAD}_H$ and $\text{SAD}_V$ values are the highest in the middle of the frames as shown in Figure 4 and Figure 6, respectively.

For Foreman video, the $\text{SAD}_H$ value is the highest at the bottom of the frames and the $\text{SAD}_V$ value is the highest in the middle of the frames as shown in Figure 5. It is been notice from that, the $\text{SAD}_H$ and $\text{SAD}_V$ values are approximately within the same range as the Foreman video is shaking all the times.

For Waterfall video, the $\text{SAD}_H$ value is increasing dramatically as the video is zooming out all the times, where the highest value is in the bottom of the frame, while the $\text{SAD}_V$ value it is the highest on the left side of the frames as shown in Figure 7.

From the result that we are obtained from calculating the SAD values; four different ROIs cases are proposed in this study as shown in Figure 8:

- ROI case A corresponds to the highest value of the $\text{SAD}_H$; see Figure 3(a), Figure 4(a), Figure 5(a), Figure 6(a) and Figure 7(a).
- ROI case B corresponds to the highest value of the $\text{SAD}_V$; see Figure 3(b), Figure 4(b), Figure 5(b), Figure 6(b) and Figure 7(b).
- ROI case C is chosen statically in the middle of the frames as shown in Figure 8 (c) and Table 1.
- ROI case D is actually two regions. The reason for investigating this ROI is that some video, e.g., Highway (see Figure 3(b)), has a lot of motion on the right and left side of the video frames as shown in Figure 8 (d) and Table 1.
Chapter Six

Figure 3: Highway video.

a. The SADH \((k)\)

b. The SADV \((k)\)
a. The SADH \((k)\)

b. The SADV \((k)\)

Figure 4: Akiyo video.
Chapter Six

Figure 5: Foreman video.

a. The SADh \((k)\)

b. The SADV \((k)\)
Chapter Six

Figure 6: News video.

- Figure 6a: The SADh $(k)$
- Figure 6b: The SADV $(k)$
Chapter Six

Figure 7: Waterfall video.

a. The SADh (k)

b. The SADv (k)
Figure 8: The proposed ROI.
a. The reference frames are every 3rd frame

b. The reference frames are every 4th frame

c. The reference frames are every 5th frame

Figure 9: The proposed scenarios.

Table 1: The SAD values (Mbit) for case C and D.

<table>
<thead>
<tr>
<th>Videos</th>
<th>Highway</th>
<th>Akiyo</th>
<th>Foreman</th>
<th>News</th>
<th>Waterfall</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Case C</strong></td>
<td>2018.591</td>
<td>17.789</td>
<td>67.018</td>
<td>29.901</td>
<td>22.718</td>
</tr>
<tr>
<td><strong>Case D</strong></td>
<td>247.313</td>
<td>3.636</td>
<td>60.173</td>
<td>4.391</td>
<td>27.739</td>
</tr>
</tbody>
</table>
6.3.2 The Proposed Streaming Scenarios

The streaming video is divided into two videos: one video containing a complete version of every $n^{th}$ frames in the original video sequence, which we call it the reference frames and one video containing the ROI that are between reference frames. In this study we consider three scenarios of $n$ ($n = 3, 4$, and $5$), as shown in Figure 9.

In the first scenario where the reference frames are every 3rd frame, e.g., 0, 3, 6, 9,..., in the second scenario where the reference frames are every 4th frame, e.g., 0, 4, 8, 12,..., in the third scenario where the reference frames are every 5th frame, e.g., 0, 5, 10, 15,...

The two videos (one with every $n^{th}$ frames and one with the ROI) are encoded and transmitted in the normal way by using H.264 [20].

At the mobile side, the receiving video frames are been reordered to it is original sequence position, while linear interpolation is performed between reference frames to reconstruct the part of the region that is outside the ROI (non-ROI). This means that the value of a pixel $(x,y)$ outside of the ROI for frames $k*n+i$ ($k = 0, 1, 2, 3$ and $i = 0, ..., n-1$) is obtained by using the following formula $((n-i)F_{kn}(x,y) + iF(k+1)n(x,y))/n$, where $F_{kn}(x,y)$ is the pixel value for position $(x,y)$ in frame number $kn$ in the original video sequence.

6.4 The Effect of the Proposed Technique on the Streaming Size

Compressing the video frames according to the proposed technique by using H.264 ffmpeg codec [20] will affect the amount of redundancy information to be removed from the video frames. Since we will drop the pixels that are outside the ROI (non-ROI), we expect that the compression size of the videos will be affected.

The proposed scenarios are applied for different cases (ROIs) to the videos Highway, Akiyo, Foreman, News, and Waterfall [19], with a coding rate of 30 frames per second. The chosen videos have different characteristics and motions level. Therefore, the compression size of
the videos is expected to be different from one video to another and from one scenario to another.

The compression size of the videos is decreases when the proposed ROI method is applied, as the amount of decrease is different from one scenario to another and from one ROI to another. For Highway videos, ROI case A shows the smallest decrease as shown in Table 2 (a). The reason for that, the ROI case A has the highest SAD value and the non-ROI that are related to the ROI case A is more static compared to the non-ROI that are dropped from the other ROIs, as shown in Figure 3 and Table 1.

Table 2: The compressed video size (bytes).

a. Highway

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Case A</th>
<th>Case B</th>
<th>Case C</th>
<th>Case D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario one</td>
<td>25891300</td>
<td>23470760</td>
<td>24326420</td>
<td>25241910</td>
</tr>
<tr>
<td>No. of packets</td>
<td>17589</td>
<td>15945</td>
<td>16526</td>
<td>17148</td>
</tr>
<tr>
<td>Decrease</td>
<td>34.5%</td>
<td>40.6%</td>
<td>38.5%</td>
<td>36.1%</td>
</tr>
<tr>
<td>Scenario two</td>
<td>24626600</td>
<td>21835484</td>
<td>22809872</td>
<td>23834016</td>
</tr>
<tr>
<td>No. of packets</td>
<td>16730</td>
<td>14834</td>
<td>15496</td>
<td>16196</td>
</tr>
<tr>
<td>Decrease</td>
<td>37.7%</td>
<td>44.8%</td>
<td>42.3%</td>
<td>39.7%</td>
</tr>
<tr>
<td>Scenario three</td>
<td>23685770</td>
<td>20741686</td>
<td>21772332</td>
<td>22887146</td>
</tr>
<tr>
<td>No. of packets</td>
<td>16091</td>
<td>14091</td>
<td>14791</td>
<td>15548</td>
</tr>
<tr>
<td>Decrease</td>
<td>40.1%</td>
<td>47.5%</td>
<td>44.9%</td>
<td>42.1%</td>
</tr>
</tbody>
</table>

b. Akiyo

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Case A</th>
<th>Case B</th>
<th>Case C</th>
<th>Case D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario one</td>
<td>12679446</td>
<td>15669696</td>
<td>15418140</td>
<td>9685038</td>
</tr>
<tr>
<td>No. of packets</td>
<td>8614</td>
<td>10645</td>
<td>10474</td>
<td>6580</td>
</tr>
<tr>
<td>Decrease</td>
<td>30.7%</td>
<td>14.3%</td>
<td>15.7%</td>
<td>47.0%</td>
</tr>
<tr>
<td>Scenario two</td>
<td>11739910</td>
<td>15044034</td>
<td>14787634</td>
<td>8414166</td>
</tr>
<tr>
<td>No. of packets</td>
<td>7975</td>
<td>10220</td>
<td>10046</td>
<td>5716</td>
</tr>
<tr>
<td>Decrease</td>
<td>35.8%</td>
<td>17.8%</td>
<td>19.2%</td>
<td>54.0%</td>
</tr>
<tr>
<td>Scenario three</td>
<td>11120176</td>
<td>14569282</td>
<td>14286604</td>
<td>7600286</td>
</tr>
<tr>
<td>No. of packets</td>
<td>7554</td>
<td>9898</td>
<td>9706</td>
<td>5163</td>
</tr>
<tr>
<td>Decrease</td>
<td>39.2%</td>
<td>20.3%</td>
<td>21.9%</td>
<td>58.4%</td>
</tr>
</tbody>
</table>
### Chapter Six

#### c. Foreman

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Case A</th>
<th>Case B</th>
<th>Case C</th>
<th>Case D</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scenario one</strong></td>
<td>29306084</td>
<td>29182400</td>
<td>29005722</td>
<td>29542048</td>
</tr>
<tr>
<td><strong>No. of packets</strong></td>
<td>19909</td>
<td>19825</td>
<td>19705</td>
<td>20069</td>
</tr>
<tr>
<td><strong>Decrease</strong></td>
<td><strong>36.6%</strong></td>
<td><strong>36.9%</strong></td>
<td><strong>37.3%</strong></td>
<td><strong>36.1%</strong></td>
</tr>
<tr>
<td><strong>Scenario two</strong></td>
<td>26949156</td>
<td>26750004</td>
<td>26565964</td>
<td>27182634</td>
</tr>
<tr>
<td><strong>No. of packets</strong></td>
<td>18308</td>
<td>18173</td>
<td>18046</td>
<td>18466</td>
</tr>
<tr>
<td><strong>Decrease</strong></td>
<td><strong>41.7%</strong></td>
<td><strong>42.2%</strong></td>
<td><strong>42.6%</strong></td>
<td><strong>41.2%</strong></td>
</tr>
<tr>
<td><strong>Scenario three</strong></td>
<td>25331482</td>
<td>25165654</td>
<td>24951476</td>
<td>25598484</td>
</tr>
<tr>
<td><strong>No. of packets</strong></td>
<td>17209</td>
<td>17096</td>
<td>16951</td>
<td>17390</td>
</tr>
<tr>
<td><strong>Decrease</strong></td>
<td><strong>45.2%</strong></td>
<td><strong>45.6%</strong></td>
<td><strong>46.0%</strong></td>
<td><strong>44.6%</strong></td>
</tr>
</tbody>
</table>

#### d. News

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Case A</th>
<th>Case B</th>
<th>Case C</th>
<th>Case D</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scenario one</strong></td>
<td>19146848</td>
<td>19093548</td>
<td>20964438</td>
<td>15158016</td>
</tr>
<tr>
<td><strong>No. of packets</strong></td>
<td>13007</td>
<td>12971</td>
<td>14242</td>
<td>10298</td>
</tr>
<tr>
<td><strong>Decrease</strong></td>
<td><strong>26.8%</strong></td>
<td><strong>27.0%</strong></td>
<td><strong>19.8%</strong></td>
<td><strong>42.0%</strong></td>
</tr>
<tr>
<td><strong>Scenario two</strong></td>
<td>17809400</td>
<td>17799214</td>
<td>19882266</td>
<td>13443716</td>
</tr>
<tr>
<td><strong>No. of packets</strong></td>
<td>12099</td>
<td>12092</td>
<td>13507</td>
<td>9133</td>
</tr>
<tr>
<td><strong>Decrease</strong></td>
<td><strong>31.9%</strong></td>
<td><strong>31.9%</strong></td>
<td><strong>24.0%</strong></td>
<td><strong>48.6%</strong></td>
</tr>
<tr>
<td><strong>Scenario three</strong></td>
<td>16762434</td>
<td>16574180</td>
<td>18756994</td>
<td>12303162</td>
</tr>
<tr>
<td><strong>No. of packets</strong></td>
<td>11388</td>
<td>11260</td>
<td>12743</td>
<td>8358</td>
</tr>
<tr>
<td><strong>Decrease</strong></td>
<td><strong>35.9%</strong></td>
<td><strong>36.6%</strong></td>
<td><strong>28.3%</strong></td>
<td><strong>52.9%</strong></td>
</tr>
</tbody>
</table>

#### e. Waterfall

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Case A</th>
<th>Case B</th>
<th>Case C</th>
<th>Case D</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scenario one</strong></td>
<td>30666804</td>
<td>29842876</td>
<td>30307382</td>
<td>31064252</td>
</tr>
<tr>
<td><strong>No. of packets</strong></td>
<td>20833</td>
<td>20274</td>
<td>20589</td>
<td>21103</td>
</tr>
<tr>
<td><strong>Decrease</strong></td>
<td><strong>34.6%</strong></td>
<td><strong>36.3%</strong></td>
<td><strong>35.3%</strong></td>
<td><strong>33.7%</strong></td>
</tr>
<tr>
<td><strong>Scenario two</strong></td>
<td>28110148</td>
<td>27215516</td>
<td>27722092</td>
<td>28525000</td>
</tr>
<tr>
<td><strong>No. of packets</strong></td>
<td>19097</td>
<td>18489</td>
<td>18833</td>
<td>19378</td>
</tr>
<tr>
<td><strong>Decrease</strong></td>
<td><strong>40.0%</strong></td>
<td><strong>41.9%</strong></td>
<td><strong>40.8%</strong></td>
<td><strong>39.1%</strong></td>
</tr>
<tr>
<td><strong>Scenario three</strong></td>
<td>26546070</td>
<td>25609170</td>
<td>26142780</td>
<td>26972772</td>
</tr>
<tr>
<td><strong>No. of packets</strong></td>
<td>18034</td>
<td>17398</td>
<td>17760</td>
<td>18324</td>
</tr>
<tr>
<td><strong>Decrease</strong></td>
<td><strong>43.3%</strong></td>
<td><strong>45.3%</strong></td>
<td><strong>44.2%</strong></td>
<td><strong>42.4%</strong></td>
</tr>
</tbody>
</table>
For Akiyo videos, the compression size is decreases, as shown in Table 2 (b). The ROI for cases B and C shows the lowest decreases. These are also the ROIs with the largest SAD value, as shown in Figure 4 and Table 1.

For Foreman videos, the decreases of the compression size for all ROIs are approximately within the same range as the video is shaking all the time, as shown in Table 2 (c). The SAD value is also approximately within the same range for all cases as shown in Figure 5 and Table 1.

For News videos, the ROI case C shows the lowest decreases in the compression size, as shown in Table 2 (d). This is also the ROI with the largest SAD value, as shown in Figure 6 and Table 1.

For Waterfall videos, the amount of compression size is within the same range for all ROIs, since the video is zooming out. However, ROI case D has somewhat lower compression size as shown in Table 2 (e). Looking at Figure 7 and Table 1, we see that the SAD value for ROI case D is the highest.

It can be summarized from the compression Tables, the size of the compressed videos are related to the SAD value to the corresponding ROI.

6.5 Subjective Viewing Test

6.5.1 Testing Materials and Environments

The videos are displayed on a 17 inch FlexScan S2201W LCD computer display monitor of type EIZO with a native resolution of 1680 x 1050 pixels. The videos are displayed with resolution of 176 x 144 pixels in the centre of the screen with a black background with a duration of 66 seconds for Highway video and 10 seconds for Akiyo, Foreman, News and Waterfall videos.
6.5.2 Testing Methods

It is well known that Peak Signal-to-Noise Ratio (PSNR) does not always rank quality of an image or video sequence in the same way as a human being. There are many other factors are considered by the human visual system and the brain [13]. One of the most reliable ways of assessing the quality of a video is subjective evaluation of the Mean Opinion Score (MOS) measurement, MOS is a subjective quality metric obtained from a panel of human observers. It has been regarded for many years as the most reliable form of quality measurement technique [14]. The MOS measurements are used to evaluate the videos quality in this study and based on the guidelines outlined in the BT.500-11 recommendation of the radio communication sector of the International Telecommunication Union (ITU-R). We use a lab with controlled lighting and set-up according to the ITU-R recommendation. The score grades in this methods range from 0 to 100. These ratings are mapped to a 5-grade discrete category scale labelled with Excellent, Good, Fair, Poor and Bad [9].

The subjective experiment has been conducted at Blekinge Institute of Technology in Sweden. The users observed the proposed scenarios in two groups, where the participated of thirty non-expert test subjects for each group were all university students. The participated of the first group are 25 males and 5 females for evaluating the Highway videos and their age’s range of 20 to 41. The participated for the second group are 27 males and 3 females for evaluating Akiyo, Foreman, News, and Waterfall videos and their age’s range of 20 to 35.

The amount of data is gathered from the subjective experiments with respect to the opinion scores that were given by the individual viewers. Concise representation of this data is achieved by calculating the conventional statistics such as the mean score and 95% confidence interval.

6.6 Experimental Results

For Highway videos, the first scenario (every 3rd frame), as shown in Figure 9 (a) and Figure 10 (a). The observers evaluate the videos after
it been reconstructed by using linear interpolation. The scores are
within the same range for the four ROI cases, as the MOS is larger than
3 and lower than 4. However, the MOS for ROI case A is slightly
higher than the rest. The second scenario (every 4\textsuperscript{th} frame), as shown in
Figure 9 (b) and Figure 10 (b). The MOS vary from one case of ROI to
another and the ROI case D shows the lowest score, while the ROI case
A shows the better score than the rest. For the third scenario (every 5\textsuperscript{th}
frame), as shown in Figure 9 (c) and Figure 10 (c), the MOS for ROI
case A shows the highest score as ROI case A has the highest SAD
value as well, as shown in Figure 3 and Table 1.

The MOS for Akiyo videos as shown in Figure 11. The ROI case D
shows the lowest score for the three scenarios. From Table I, the SAD
value for ROI case D is significantly smaller than other ROIs. As the
ROI case D is been defined as a two side regions as shown in Figure 8
d, where the highest motion region is in the middle of the frames as
shown in Figure 4. Therefore, after the videos have been reconstructed
by interpolation, the motion region is been highly affected and for this
reason the MOS is low for ROI case D. The differences in both MOS
and SAD value for ROIs A, B and C is rather small for Akiyo video.

The MOS for Foreman videos as shown in Figure 12. The scores are
below 2.5 for the four ROI cases, as the Foreman video frames are
shaking all the time, which is negatively been affected by the
interpolation.

The MOS for News videos as shown in Figure 13. The Figure shows
that the ROI case C gets the highest MOS, as the highest motion is in
the middle of the frames, while ROI case D gets the lowest MOS, as
the motion region been affected by interpolation. From Figure 6, Figure
8 and Table I, we observed from that the ROI case C has the highest
SAD value and ROI case D has the lowest SAD value.

The MOS for Waterfall videos as shown in Figure 14. The scores are
greater than 2.5 for the four ROI cases. ROI case D gets the lowest
MOS but ROI actually has the highest SAD value, even all the ROIs
have approximately similar SADs values. This is an exception to what
we have been seen in other videos.
Chapter Six

Figure 10: The MOS for the four proposed slice cases for Highway.
Figure 11: The MOS for the four proposed slice cases for Akiyo.
Chapter Six

Figure 12: The MOS for the four proposed slice cases for Foreman.
Figure 13: The MOS for the four proposed slice cases for News.
Figure 14: The MOS for the four proposed slice cases for Waterfall.
In general, it seems that a high value of SAD results in a high MOS. However, it seems that if the SAD value does not differ so much between different ROIs, e.g., Foreman and Waterfall videos, then ROIs of case B and C seem to be promising.

Figures 10 - 14 shows that by sending the reference frames more often, e.g., every 3rd frame instead of every 4th and 5th frame will increase the MOS in most cases.

6.7 Conclusion

To reduce the amount of data that are streamed over wireless networks, a Region Of Interest (ROI) compression scheme is proposed in this study. The scheme is based on identifying and extracting the motion ROI by computing the Sum of Absolute Differences (SAD) for video streaming.

Four different ROI cases for three different scenarios are proposed to evaluated each video, corresponding to send the full frame information (reference frames) every 3rd, 4th, and 5th frame, respectively.

The proposed adaption scheme is compressed by H.264 codec to study the effects on the video size, by sending the reference frames more often, e.g., every 3rd frame instead of every 4th and 5th frame; it will increase both the MOS and the file size.

Our experiment shows, and quantifies, the trade-off between high compression and high MOS. How often we will send the reference frames is an engineering decision that depends on the available bandwidth and the need for compression level. However, since it is probably more important to keep the MOS on a high level, we would argue that one should select ROIs with the highest SADs values.

The SAD can easily been calculated for different parts (scenes) of a video. We could then use this for obtaining an appropriate ROI for that particular part of the video, i.e., we do not need to use the same ROI for the entire video.
REFERENCES


