Performance Analysis of Various Key Frame Extraction Methods for Surveillance Applications

S. Ariffa Begum¹, Dr. A. Askarunisa²

¹Full-Time Ph.D., Scholar, ²Professor & Head, Department of CSE, Vickram College of Engineering, Enathi, Tamilnadu, India

Abstract— With the large amounts of video data available, it has become increasingly important to have the ability to quickly search through and browse through these videos. The key frame extraction is an essential technology of video retrieval and has a great impact on retrieval efficiency by reducing the redundant information. In this paper we analyses the performance of simple Edge Change Ratio (ECR) method, color histogram method and distance based methods for key frame extraction in surveillance applications.

Keywords— Video Surveillance, Key frame Extraction, Color Histogram, Edge Change Ratio.

I. INTRODUCTION

Nowadays, video surveillance systems are installed worldwide in many different sites such as airports, hospitals, banks, railway stations and even at home. This leads to large collection of video data. Processing this large amount of data is a challenging problem. So, it is necessary to have the technique which reduces the amount of data to be processed. This is achieved by eliminating the redundant frames of the video. The frames which contain meaningful information about the content of the video are called as key frames. The researchers have attempted to exploit various features for the extraction of key frames in videos. Some of the low level features which are commonly used include color histogram, frame correlation, motion information and edge histogram etc. In this paper, we compare the performance of color histogram method, Edge Change Ratio (ECR) method and distance based methods.

A video can be segmented into different units, such as frames, shots, or scenes. The complete moving picture in a video can be discretized to a finite image sequence, i.e., many still images. Each still image is called a “frame” [1], which is the basic unit of the video. The image sequence is naturally indexed by the frame number.

A video shot is defined as a series of interrelated consecutive frames taken contiguously by a single camera and representing a continuous action in time and space [1]. In general, shots are joined together in a process called editing to produce a video. The unbroken image sequence in a shot usually has consistent content. While scene is a more semantic notion, which is essentially a story unit.

II. LITERATURE REVIEW

Key-frame extraction techniques have been used extensively in video retrieval, which is used for video browsing and content-based retrieval applications. In the past decade, there has been significant work done in the area of video processing to partition a given video into separate shots. The basic idea of most of the techniques is to measure and compare the similarities between consecutive frames.

Among these techniques, pair wise pixel comparison [1, 2, 3] is a straightforward and simplest way, in which the number of pixels changed from a frame to the next is counted.
When the total percentage of the pixels has changed, a shot is detected. The weakness of this algorithm is it is more sensitive to noise.

Likelihood ratio [4, 5] is a region-based technique. It solves the false detection due to small camera motions. Instead of comparing individual pixel, it compares the statistical characteristic, the so-called likelihood ratio, of the corresponding regions (i.e. blocks) in two successive frames. If the likelihood ratio is larger than a preset threshold, the region is regarded as being changed. A shot boundary is found if more than a certain number of blocks have changed. It is less sensitive to noise, camera and object motion.

Histogram differences [6] are most widely used in shot detection. This technique can be used to find the frame whose histogram varies significantly from another frame histogram. Thus we can detect shot boundaries. This approach is less sensitive to motion. A shot boundary is found if more than the threshold block has changed. One major problem of histogram differences is that two images can have exactly the same histograms while the shown content differs extremely. In addition, Region based histogram differences are adapted. Each frame is divided into several blocks. The gray-scale histogram is computed for each block.

In the algorithms above, a fixed threshold was set previously. In order to enhance the performance, two thresholds and adaptive threshold [7] rather than one fixed global threshold are employed. In the two thresholds method, a higher threshold is used for hard cut detection, and a low threshold is set for special effects. While in the adaptive threshold method, the local thresholds consider various features and adapt the thresholds within different windows. Like in the previous case, if the feature is higher than the corresponding threshold a shot can be declared.

Unlike the difference comparison methods, the edge change ratio (ECR) attempts to compare the actual content of two consecutive frames. It transforms both frames to edge pictures, and compares them using dilatation to compute a probability that the second frame contains the same objects as the first frame. This approach is very sensitive to hard cuts. The basic idea of edge change ratio (ECR) comparison method is summarized as following:

**Step 1: Pre-processing**
- Convert the color frames into gray-scale frames.
- Detect edges in the frames using sobel method
- Find the inverted frame
- Apply dilation to the frame

**Step 2: Compute ECR**
- Detect two contiguous frames Fi and Fi+1.
- Count the number of edge pixels in the frames Fi (δn) and Fi+1(δn+1).
- Define the entering and exiting edge pixels En+1 and En.
- The Edge Change Ratio (ECR) can be calculated as
  \[ \text{ECR} (n, n+1) = \max \left( \frac{E_{n+1}}{\delta_{n+1}} , \frac{E_n}{\delta_n} \right) \]

**Step 3: Key Frame Extraction**
- If the edge change ratio is larger than a predefined threshold, we consider the two as a cut.
- After going through the whole video, the hard cuts can be detected, the video can be broken into shots and the key frames are extracted.

IV. COLOR HISTOGRAM METHOD

The color histograms have been commonly used for key frame extraction in frame difference based techniques. This is because the color is one of the most important visual features to describe an image. Color histograms are easy to compute and are robust in case of small camera motions. The idea behind histogram based approaches is that two frames with unchanging background and unchanging (although moving) objects will have little difference in their histograms.

The steps involved in color histogram method are as follows:
1. Convert the RGB color model to HSV color model.
2. Find the color histogram for each frame.
3. Normalize the color histogram of each frame
4. The distance between two consecutive frames is calculated.
5. Find the mean value of HSV color histograms and this is taken as dissimilarity between the frames.
6. If the dissimilarity is larger than a predefined threshold, we consider the two as a cut.
7. After going through the whole video, the hard cuts can be detected, the video can be broken into shots and the key frames are extracted.

V. DISTANCE BASED METHOD
The similarity of frames is also measured by the distance between two frames. The frames are similar if the distance between them is too small. So, we can eliminate the frames which are having smaller distance among them. It is important to select a suitable similarity measurement method which makes the distance of feature match with the similarity of frames possibly.

In this paper to find the difference between frames; Euclidian and Correlation based methods had been taken for analysis.

VI. EXPERIMENTAL RESULTS AND DISCUSSIONS
In order to analyse the performance of ECR and color histogram method, a number of surveillance video samples have been tested. These videos are real time CCTV footages taken from various locations like shopping malls, railway junctions and many more.

The following table shows the number of frames extracted as key frames for the above said algorithms:

<table>
<thead>
<tr>
<th>Test Videos</th>
<th>Total Number of Key frames</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ECR</td>
</tr>
<tr>
<td>RobWoman</td>
<td>1412</td>
</tr>
<tr>
<td>TrolleyTrack</td>
<td>888</td>
</tr>
<tr>
<td>JewelTheif</td>
<td>1610</td>
</tr>
<tr>
<td>ManAttack</td>
<td>1740</td>
</tr>
<tr>
<td>GunMan</td>
<td>1380</td>
</tr>
<tr>
<td>ChildTheft</td>
<td>1860</td>
</tr>
</tbody>
</table>

Key frame extraction aims to reduce the amount of video data, and the frame sequence must preserve the overall contents of the original video. From the results obtained from the above methods the color histogram method produces better compression ratio for all videos. But each of the methods is having their own pros and cons. Even though the histogram method yields better compression ratio the results produced by it is not best suitable for further analysis such as left object and untoward incidents identification.

Nearly ten cctv video footages of different categories of misbehaviours like theft at railway junction, jewellery shop, pushing trolley on track etc., had been taken for analysis.

In the above case when the resultant key frames are analysed the correlation method produces best key frames collection for theft at jewellery shop. The Euclidian distance method produces best key frames collection for trolley on track.

ECR method detects the edge changes in the frames and adaptive threshold value is used for edge changes in the frames. Threshold value is dynamic and will be changing during the course of each shot so that key frames can be retrieved efficiently. This produces best key frame collection for GunMan video footage.

The ECR method focus on edge feature, histogram method focus on colors, the Euclidian and Correlation methods focus on distance between the frames.

VII. CONCLUSIONS AND FUTURE ENHANCEMENTS
In this paper, the basic algorithms used for key frame extraction had been analysed for different surveillance videos. And the detailed discussion is presented in the previous section. Our aim is to extract the key frames that can be used to represent the video as a whole and summarizes the salient content of the video. From this analysis it is evident that the extracted key frames depend on the input video content and the threshold values used in the above methods.

This analysis shows the need for common (new or a hybrid) key frame extraction method which is computationally simple, independent of input with high accuracy and low error rate. So, in future our aim is to develop an algorithm that satisfies above requirements.
REFERENCES


