

Key frame extraction using color histogram method.

Miss.A.V.Kumthekar*, Prof.Mrs.J.K.Patil**

*(Department of E&TC, Shivaji University,Kolhapur.

Email: aartikumthekar@gmail.com)

** (Department of E&TC, Shivaji University,Kolhapur.

Email: jaymala.p@rediffmail.com)

ABSTRACT:-

Key frame extraction is an essential part in video analysis and management, providing a suitable video summarization for video indexing, browsing and retrieval. General video is rich in content and consists of 20 to 30 frames per second. Hence a one hour video would contain around 20x60x60 to 30x60x60 frames. Most of these frames contain redundant information and thus key frame extraction is essential. The use of key frames reduces the amount of data required in video indexing and provides the frame work for dealing with the video content. Therefore, this paper proposes a method for video key frame extraction based on color histogram and edge detection, the purpose is to remove the redundant frames, reduce the computational complexity and improve recognition efficiency. The compression ratio is 98%.

Keywords: color histogram, edge detection, frame difference, key frame, shot detection.

1. INTRODUCTION:

Recent years have witnessed an enormous increase in video data on the internet. This rapid increase demands efficient techniques for management and storage of video data. Video summarization is one of the commonly used mechanisms to build an efficient video archiving system. The video summarization methods generate summaries of the videos which are the sequences of stationary or moving images. Key frame extraction is widely used for

video summarization. The key frames are the characteristic frames of the video which render limited, but meaningful information about the content of the video. 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. used the color histogram difference between the current frame and the last extracted key frame to draw out key frames from the video. In this project we are going to deal with the frame difference measures such as color histogram, frames correlation and edge orientation histogram for the extraction of key frame. To extract valid information from video, to process video data efficiently, and to reduce the transfer stress of network, more and more attention is being paid to the video processing technology. The amount of data in video processing is significantly reduced by using video segmentation and key-frame extraction using histogram difference matching. To reduce the transfer stress in network and invalid information transmission, the transmission, storage and management techniques of video information become more and more important.

Video segmentation is a fundamental process and the first step in automatic digital video analysis. It is of great importance in many applications, such as video databases, video compression and transmission, video retrieval and browsing, and so on. 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. All the frames in one video have a same size and the time between each two frames is equal, typically 1/25 or 1/30 seconds. 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. Different shots are joined together in the editing stage of video. There are several editing operations, with or without the use of the transition effect. Transitions between shots can be either ‘abrupt’ or ‘gradual’ [2].

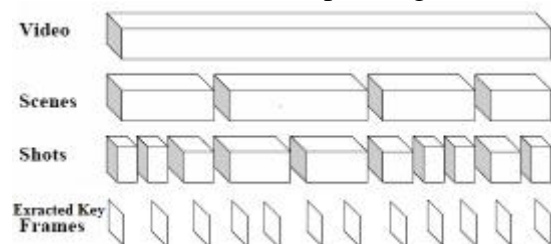


Fig 1: Structural hierarchy of a video

The main aim is to present an approach to segment a given video into different shots, find the beginning and end of each shot, and then extract a single frame representing the main content of each shot. Such a frame is so called key frame. The key frame can also be applied to the video content-based retrieval [7], content analysis, video summarization [6], editing and so on. Key-frame based video representation views video abstraction as a problem of mapping

an entire segment (both static and moving contents) to some small number of representative images. The features used for key frame extraction include colors (particularly the color histogram), edges, shapes, optical flow, camera activity, and features derived from image variations caused by camera motion. The extraction of key frames must be automatic and content based so that they maintain the salient content of the video while avoiding all redundancy. The basic requirement of video segmentation and indexing is to partition a video into shots. A shot is an unbroken sequence of frames and also a basic meaningful unit in a video. It usually represents a continuous action or a single camera operation. Generally, the purpose of a shot segmentation algorithm is to accurately find all the shot boundaries. There are two kinds of shot boundaries (shot changes) namely, abrupt changes and gradual transitions. Abrupt changes usually result from camera breaks, while gradual transitions are produced with artificial editing effects, such as fading, dissolve and wipe. A number of methods to video shot transition detection have been proposed in the past decade. Most of the methods provide metrics to quantify the difference between successive frames in order to locate shot boundaries. Among all the video shot detection methods, we roughly classify them into two categories: one uses original video data and the other uses only the compression information.

2. LITERATURE REVIEW:-

Video summarization is an active research field in computer vision for video browsing and content based video retrieval. The summarized video decreases the required storage area and increases the computational efficiency in extracting the key frames from the given video. The key frames are nothing but the meaningful part

of the entire video eliminating the redundant frames of the frames.

In the past decade, there has been significant work done in the area of video processing to partition a given video into separate shots. Shot segmentation is the first step of the key frame extraction, which mainly refers to detecting the transition between successive shots. The basic idea of most of the techniques is to measure and compare the similarities between consecutive frames. The following are some of the technique-

1. Pair wise pixel comparison

Pair wise pixel comparison [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. In this algorithm individual pixels from frames are compared to find out frame difference. Pair-wise comparison evaluates the differences in intensity or color values of corresponding pixels in two successive frames. In this algorithm the pixel-wise difference algorithm gives quite acceptable results with adaptive thresholding. By considering difference between the difference signal values of adjacent frames is a worthwhile approach. In practice, it is observed that it is useful to reduce the difference signal with a threshold derived from the maximum and minimum difference signals over a small aperture. Even with the adaptive thresholding, the algorithm produces false alarms, if the shot before/after the shot boundary includes high motion activity. The reason can be explained as follows: The weakness of the pixel based features is the high sensitivity to the video content. It is difficult for this algorithm to understand whether the change in the continuity signal is due to shot boundary or due to disturbances/motion. In order to enhance the algorithm, adaptive

thresholding can be used. However, the high level of activity in the images around shot boundary produces a larger difference signal than expected. As a result adaptively obtained threshold is larger. A threshold that is larger than expected results in missed shot boundary. The main disadvantage of this method is its inability to distinguish between a large change in a small area and a small change in a large area. It is observed that cuts are falsely detected. The other disadvantage of this method is that it is quite sensitive to fast object movements and the camera motion - fast camera panning or zooming. Also it is sensitive to noise.

2. Likelihood ratio:

Likelihood ratio [3] is a region-based technique. It is a typical statistical difference method, which can be regarded as an extension to pixel difference. It can solve the problem of 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 can be declared if a certain number of regions have changed. A shot boundary is found if more than a certain number of blocks have changed. It is less sensitive to camera and object motion and noise.

3. Motion Capture Data by Curve Saliency

In this paper, Eyuphan Bulut & Tolga Capin [8] proposes a new approach to find key frames in a motion captured sequence. Treat the input motion sequence as a curve, and find the most salient parts of this curve which are crucial in the representation of the motion behavior. We apply the idea of saliency to motion curves in the first part of

our algorithm. Then in the second part, we apply key frame reduction techniques in order to obtain the most important key frames of the motion. This method is effective to a certain extent.

3. PROPOSED METHODOLOGY

Efficient key frame extraction enables efficient cataloguing and retrieval with large video collections. Video is rich in content and it results in a tremendous amount of data to process. This can be made easier by only processing some frames, such as the key frames of video. In general, a key frame extraction technique must be fully automated in nature and must use the contents of the video to generate summary. Theoretically, key frames must be extracted using high level features such as objects, actions and events. However, the key frame extraction based on high level features is mostly specific to certain applications and usually low level features have been employed. Some of the commonly used examples of the low features are color histogram, correlation, moments, edges and motion features. These low level features can then be employed to derive high level features to generate domain specific applications. The common methodology is to compare consecutive frames based on some low level Frame difference measure and extract a key frame

The low level features used in this project are

- (1)Color histogram
- (2)Frame correlation
- (3)Edge detection

To start the extraction process, the first frame is declared as a key frame. Instead of computing one histogram for the entire image, we divide the image into total of T_s sections each of size $m \times m$. This is to effectively measure the level of difference between the two frames. Then the frame difference is computed between the current

frame and the last extracted key frame. This frame difference is computed by using color histogram, correlation, edge detection. Then the obtained frame difference is compared with certain threshold, if the difference satisfies with the threshold condition then the current frame is selected as a key frame. By continuously repeating the procedure for all frames we can extract the key frames.

The basic block diagram of key frames extraction consists of Extraction of frames, color histogram difference of block and frame and edge detection. Extraction of frames module extract all the frames from the given input video and the key frames are identified based on color histogram.

Key frame is the frame which can represent the salient content and distinct information as compared to the previous frame. Key frame extraction is a widely used method for video summarization that is the Key frames extracted will summarize the characteristics of the video. Video summarization is a method to generate succinct version of a video by eliminating the redundant frames

The Basic Framework of the Key Frame Extraction Algorithm is as shown

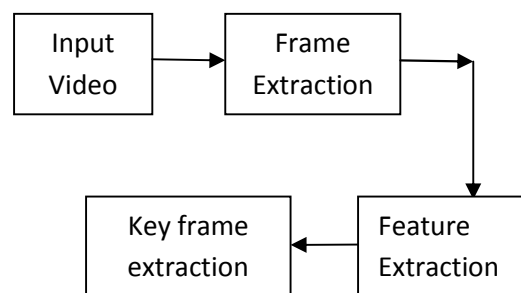


Fig 2: Basic block diagram of key frame extraction

- Input Video

The video can be in format of avi (Audio Video Interleave). To process this video,

frames have to be extracted. The AVI format was developed by Microsoft. The AVI format is supported by all computers running Windows, and by the entire most popular web browser.

- **Frame Extraction**

As video consist of number of frames depend upon size video. These frames occupy large space in memory. Frame rate is features. The low level features such as color histogram, frame correlation and edge histogram are obtained using certain frame difference measures. Then the frame difference values are calculated for all extracted frames for different videos.

a) Color histogram

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 color histogram difference $d(I_i, I_j)$ between two consecutive frames I_i and I_j can be calculated as follows:

$$D(i, i+1) = \sum_{j=1}^n \frac{|H_i(j) - H_{i+1}(j)|^2}{H_{i+1}(j)}$$

Where H_i and H_{i+1} stand for the histogram of I_i and I_{i+1} , respectively.

A shot transition occurs when $d(I_i, I_{i+1})$ is bigger than a given threshold. The experiment result illustrates that good effect can be achieved. Selecting an appropriate threshold is the key to the method.

b) Edge detection

Edge detection is one of the most commonly used operations in image analysis. An edge is the boundary between an object and the

about 20 to 30 frames per second. The video taken as input is divided into frames in this section. To do this task we have used mmreader and extracted frames. The input to mmreader can be any of the video.

- **Feature extraction**

The features of the extracted key frames can be color, edge or motion

background, and indicates the boundary between overlapping objects. This means that if the edges in an image can be identified accurately, all of the objects can be located and basic properties such as area, perimeter, and shape can be measured. Edges define the boundaries between regions in an image, which helps with segmentation and object recognition. The edges images have no differences in the distribution of the gray value. The edge matching rate is used to match the edges of adjacent frames to eliminate redundant frames. The formula for calculating the edge matching rate is as follows:

$$P(f_i, f_{i+1}) = s/n$$

$$\text{Where, } n = \text{Max}(f_i, f_{i+1})$$

$$s = \sum_i^m \sum_j^n h(i, j)$$

Where m and n indicate the height and the width of the image, f_i and f_{i+1} represent the consecutive frame and $h(I, j)$ represents the difference between the f_i and f_{i+1} .

c) Block correlation

In block motion compensation (BMC), the frames are partitioned in blocks of pixels (e.g. macro blocks of 16×16 pixels in MPEG). Each block is predicted from a block of equal size in the reference frame. The blocks are not transformed in any way apart from being shifted to the position of the predicted block. This shift is represented by a motion vector.

To exploit the redundancy between neighboring block vectors [9], (e.g. for a single moving object covered by multiple blocks) it is common to encode only the difference between the current and previous motion vector in the bit-stream. The result of this differencing process is mathematically equivalent to global motion compensation capable of panning. It is possible to shift a block by a non-integer number of pixels, which is called sub-pixel precision. The in-between pixels are generated by interpolating neighboring pixels. Commonly, half-pixel or quarter pixel precision (used by H.264 and MPEG-4/ASP) is used. The computational expense of sub-pixel precision is much higher due to the extra processing required for interpolation and on the encoder side, a much greater number of potential source blocks to be evaluated.

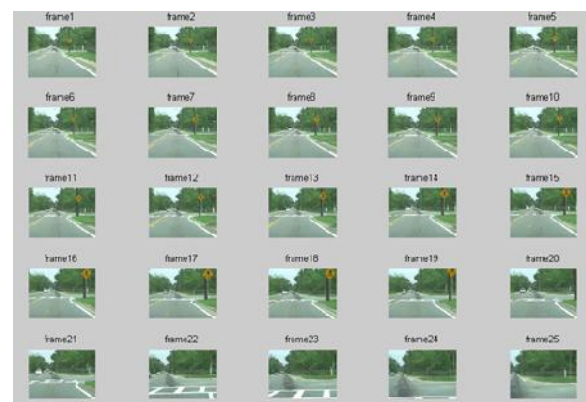
Block motion compensation divides up the current frame into non-overlapping blocks, and the motion compensation vector tells where those blocks come from (a common misconception is that the previous frame is divided up into non-overlapping blocks, and the motion compensation vectors tell where those blocks move to). The source blocks typically overlap in the source frame. Some video compression algorithms assemble the current frame out of pieces of several different previously-transmitted frames. Frames can also be predicted from future frames. The future frames then need to be encoded before the predicted frames and thus, the encoding order does not necessarily match the real frame order. Such frames are usually predicted from two directions, i.e. from the I- or P-frames that immediately precede or follow the predicted frame [10]. These bidirectional predicted frames are called B-frames. A coding scheme could, for instance, be IBBPBBPBBPBB [1].

- Key frame extraction

To start the extraction process, the first frame is declared as a key frame. Then the frame difference is computed between the current frame and the last extracted key frame. If the frame difference satisfies a certain threshold condition, then the current frame is selected as key frame. This process is repeated for all frames in the video.

4. RESULTS

Here the experiment is performed on the video having 25fps and 320×240 pixels to improve simulation speed. The performance is better than many current algorithms. A segment of video is selected to do the experiment. Firstly, let the first frame as a key frame, and the ratios are calculated according to equation. The frame where a ratio peak occurs is extracted as a key frame. If there are no transitions in a shot, the frames in the shot have high similarity, and there is no significant change among the characteristic curves. Then the first frame can be extracted as a key frame, and finally the key frames of the video can be obtained. Initially frame difference is calculated as shown in the fig 4 and the key frames extracted as shown in fig 5. The experimental result is as follows:



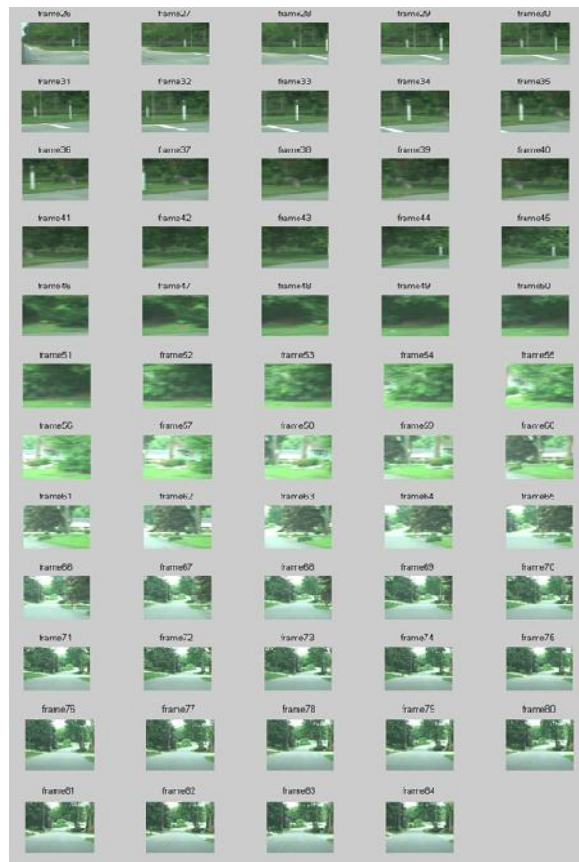


Fig 5: The result of key frame extraction from video stream

The video consists of 84 frames from that we have extracted the frame number 1,25,57,83. Thus using key frame extraction compression ratio is 98%.

5. CONCLUSION

In this experiment extracted frames from the video summarization by using histogram difference. The extracted key frames can satisfactorily represent the content of video. Our proposed system is able to extract the key frames from most of the videos. The method used is computationally simple and dynamically determines the number of key frames. The experimental results shows that the frame difference features using edge detection has high accuracy rate and low error rate.

REFERENCES

- [1] Guozhu Liu and Junming Zhao "Key Frame Extraction from MPEG Video Stream" Third International Symposium on Information Processing, 2011
- [2] Keyframe Based Video Summarization Using Automatic Threshold & Edge Matching Rate. Mr. Sandip T. Dhagdi, Dr. P.R. Deshmukh, 2012. International Journal of Scientific and Research Publications, Volume 2, Issue 7, July 2012 ISSN 2250-3153

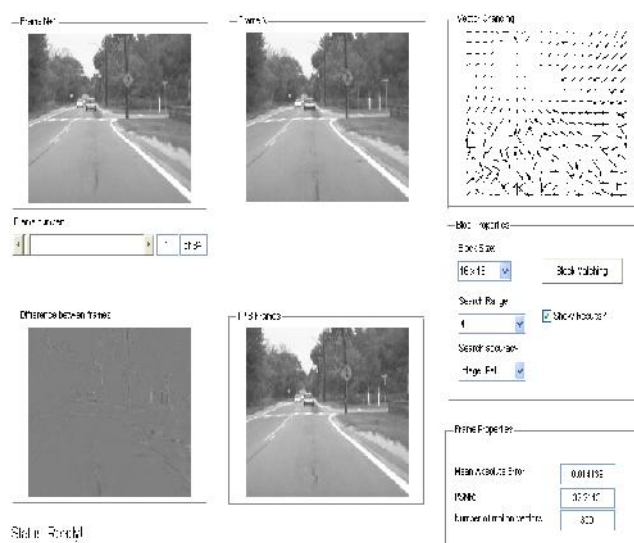


Fig 4: Calculation of I,P,B frames

- [3] HaiyanXie,” Key Frame Segmentation in Video Sequences”, 2008
- [4] T. Liu, H. Zhang, and F. Qi, “A novel video key-frame-extraction algorithm based on perceived motion energy model,” IEEE Transactions On Circuits And Systems For Video Technology, vol. 13, no. 10, pp. 1006- 1013, 2003.
- [5] “Efficient key-frame extraction and video analysis” Janko calic and ebroul izquierdo Multimedia and vision research lab, queen mary, university of london,
- [6] “Keyframe Extraction for Video Tagging & Summarization” Damian Borth, Adrian Ulges, Christian Schulze, Thomas M. Breuel, German Research Center for Artificial Intelligence (DFKI) and University of Kaiserslautern
- [7] ”Content based video retrieval systems” ,B v patell and b b meshram shah & anchor kutchhi polytechnic, chembur, mumbai, India
- [8] “Key Frame Extraction from Motion Capture Data by Curve Saliency” Eyuphan Bulut & Tolga Capin , Turkey
- [9] Liu, F., Zhuang, Y., Wu, F., Pan, Y.: 3d motion retrieval with motion index tree. Comput. Zx Vis. Image Understand. 92(2), 265–284 (2003).
- [10] Gong, Y., Liu, X: “Video summarization using singular value decomposition”. In Computer Vision and Pattern Recognition, pp 174-180 (2000).