A Survey on Key Frame Based Video Summarization Techniques

[1]Himani Parekh, [2]Pratik Nayak
[1] Assistant Professor, CGPIT, Bardoli,[2] Teaching Assistant, SRIMCA, Bardoli

Abstract - The large amount of videos usage increase the volume of data, more time to access and more man power is required. Video summarization is the solution for this problem. Summarized video can be used to review the important aspect of particular video, indexing and faster browsing. Video summarization techniques are classified into key frame based classification and skim based classification. This paper focuses on process and techniques of key frame based classification.

Keywords — Key frame, Shot segmentation based classification, Inter frame difference, Clustering.

1. INTRODUCTION

As technology is increasing rapidly, it is very common to record the videos for even small events. The large amounts of videos for different events such as news, movies, advertisements, sports are available. Videos are also used for surveillance system such as traffic monitoring, crime investigation and to provide automation [1]. It is hard job to maintain and processed large amount of volume in real time. When user wants to know important context from video, it increases the man power requirement, the storage requirement and time to process the data. In such scenario, video summarization plays an important role to reduce the resources. Summarization can be used to generate highlights of news, movie trailer, event summary, highlights of sports [1].

II. OVERVIEW

Video summarization abstracts the important scenes, meaningful occurrence, and particular object in order to provide the core information of whole video. Summarization removes duplicate scenes by extracting key frames from video. Video Summarization is a process of creating & presenting a meaningful abstract view of entire video within a short period of time. [2]
The major task in video summarization is to segment the original video into shots and extract those video frames from the original video that would be the most informative and concessive presentation of the whole video. Such frames are referred as key frames. Basically, there are two different kind of Video Summarization

techniques. One is static Video summarization and the other is dynamic Video Summarization. Static video storyboard summary involves a set of key frames from original video and there is no restriction with time and sequence issue where as dynamic video selects the most significant, small, dynamic portions of audio and video in order to generate the video summary [2].

III. TECHNIQUES

Video summarization techniques are broadly classified into two categories: key frame based classification and video skim based classification. Key frame based classification contains the methods in which the distinct frames are identified and use them in preparing summary. Video skim based classification involves the methods in which the semantic contents such as colors, textures, motions are considered. And based on that short summary is created. This paper mainly focuses on key frame based classification. The following figure shows the classification of video summarization technique.
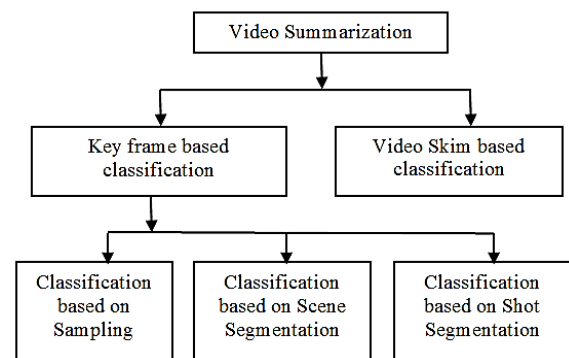
Fig: 1 Video summarization technique classification

IV. KEY FRAME BASED SUMMARIZATION

Key frame is the representative frame of shot. Key frames are extracted uniformly or randomly from the original video sequence. Key frame extraction is fundamental process in video content management. It involves selecting one or multiple frames that will represent the content of the video and used for generating video summaries [3]. The fundamentals steps for this type of classification has been shown in fig 2.

The key frame based classification can be classified into basic three categories: i) classification based on sampling, ii) classification based on scene segmentation, iii) classification based on shot segmentation. Details for each has been given below.
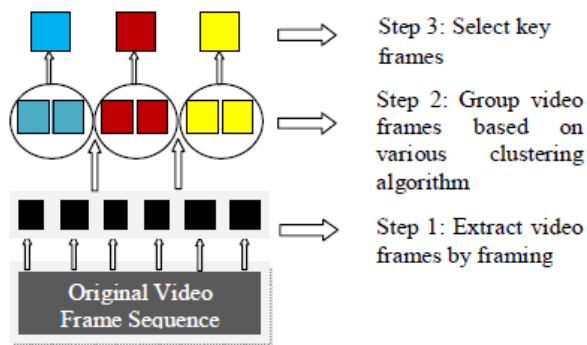

Fig: 2 Process of key frame based classification

A. Classification based on sampling:
The uniform or random sampling is done without considering the video content. As a result, the chance for appropriate summary is less. This type of classification can be used to summarize the video in which number of same type of frames are large [2].

B. Classification based on scene segmentation:
The boundary of the scene is detected and based on that different scenes are identified from video sequence. The scene can be identified based on same background, semantic link or from time period. After that, key frames are selected from each scene to represent that scene. The collection of key frames generates the summary video. But the video generated using scene segmentation does not hold temporal information [2].

C. Classification based on shot segmentation:
A scene is the collection of different shots. Each shot represents the small event. In this classification, shot detection is the first step and then from shot the key frames are identified. It gives more accurate summary video then the scene segmentation [2].

V. CLASSIFICATION BASED ON SHOT SEGMENTATION

This type of classification requires the shot to be identified. Shot is sequence of scene representing one whole event. Two or more scenes are connected semantically. So identifying such shot is difficult. The steps required to perform this classification shown in fig 3.
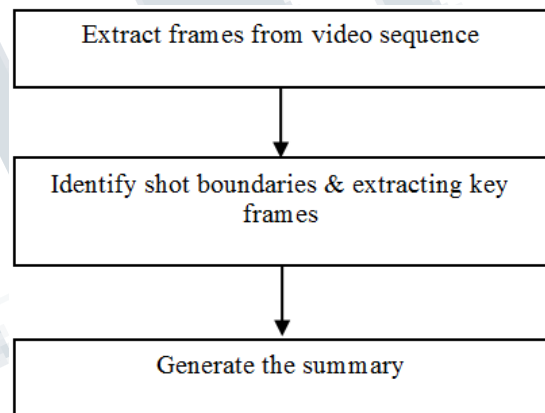

Fig: 3 Process of classification based on shot segmentation

A. Extract frames from video sequence:
Frames can be extracted from video based on number of frames required. Number of frames are predefined and based on that, frame rates are used. For example, for 60 seconds video, if 100 frames are required then frame rate will be 1.67 frames/sec.

B. Identify the shot boundaries & extracting the key frames from each shot:
Key frame extraction approaches can be classified into two categories: based on inter frame difference and based

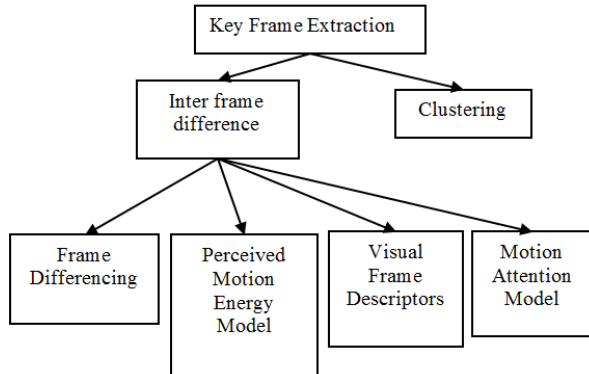on clustering. Fig 4 shows the classification of inter frame difference.

Fig: 4 Classification of key frame extraction

A. Inter frame difference:

In the approaches based on inter frame difference, a new key frame is extracted if the difference between two inter frame is greater than some threshold value. The difference between frames can be find by different distance equations, histograms, motion patterns, frame descriptors and frame visual features.

Frame Differencing [4]: In frame differencing method, no need to identify the shot boundary. Whenever drastic change is identified, the next frame is selected as key frame. The identification of change is depended on threshold value. If threshold value is large, then only drastic change can identified and if it is less, then minor change can identified. The more the threshold value, the less the number of key frames, the less the threshold value, the more the number f key frames. So the number of key frames can be controlled by the threshold value.

Perceived Motion Energy Model (PME): video sequence is segmented into shots using twin comparison method. The key frames are selected based on the motion patterns within the shots. For shots having motion pattern the triangle model is used to select the key frame, whereas for shots with no motion pattern, the first frame is chosen as a key frame [5].

Visual Frame Descriptors: It uses different features which have visual effects such as color histogram, wavelet statistics and edge direction histogram for selection of key frames. Similarity measures are computed for each

descriptor and combined to form a frame difference measure [5].

Motion Attention Model: In this method, shots are detected using color distribution and edge covering ratio. Key frames are extracted from each shot by using the motion attention model. Here the first and last frame of every shots are considered as key frame and the others are extracted by adopting motion attention model [5].

B. Clustering:

Clustering-based approaches group frames with similar low-level features and select the frame closest to each cluster centre as a key-frame. Cluster of similar frames are identified using different clustering algorithm. The cluster is known as a shot. The number of clusters can be taken as a choice of author. It can be controlled by choosing the appropriate clustering method. Most common frame from that cluster is taken as a key frame. The selection of key frame from each shot can be either randomly or any algorithm for identifying most relevant frame can be applied to find the key frame. These approaches may not grasp the interesting events and objects for viewers or they cannot find visually salient key-frames [6].

C. Generate the summary:

The final summary video is generated by combining all key frames identified from each shot. The key frame sequence should be same as shot sequence to maintain uniformity.

CONCLUSION

Video summarization is the recent research area and gains the attention of many researchers. This paper contains the review of video summarization classification techniques. One of the classification, key frame based classification are broadly classified in three categories and among them, shot segmentation based classification has further classified into various techniques. The basic steps are discussed which are the minimum required steps to generate summary.

REFERENCES

[1]   Muhammad Ajmal, Muhammad Husnain Ashraf, Muhammad Shakir, Yasir Abbas, and Faiz Ali Shah, "Video Summarization: Techniques and Classification", International Conference on Computer Vision and Graphics, pp. 1–13, 2012.

[2]   Dharmesh Tank, "A Survey on sport video summarization", International Journal for Science and Advance Research in Technology - Volume 2 Issue 10, October 2016.

[3]   Zaynab El khattabi, Youness Tabii, Abdelhamid Benkaddour, "Video Summarization: Techniques and Applications", World Academy of Science, Engineering and Technology, Vol:9, No:4, 2015.

[4]   Smt.M.Tirupathamma, "Key frame based video summarization using frame difference", International Journal of Innovative Computer Science & Engineering, Volume 4 Issue 3; May-June-2017; Page No. 160-165.

[5]   Sujatha C, Uma Mudenagudi, "A study on keyframe extraction methods for video summary", International Conference on Computational Intelligence and Communication Systems, 2011.

[6]   Yasmin S. Khan, Soudamini Pawar, "Video Summarization: Survey on Event Detection and Summarization in Soccer Videos", International Journal of Advanced Computer Science and Applications, Vol. 6, No. 11, 2015

[7]   Sabbar, W.; Chergui, A.; Bekkhoucha, A., "Video summarization using shot segmentation and local motion estimation," Innovative Computing Technology (INTECH), 2012 Second International Conference on, vol., no., pp.190, 193, 18-20 Sept. 2012.

[8]   A. Chitra, Dhawale, and S. Jain, "A novel approach towards key frame selection for video summarization," Asian Journal of Information Technology, vol. 7, no. 4, pp. 133–137, 2008.