International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-6 Issue-5, June 2017

Video Summarization: A Review on Local Binary Pattern and Classification Process

Aiswarya. N. R, Smitha. P. S

Abstract*— Video summarization system can yield good results if the high level features also called the semantic concepts in video frame are modeled accurately by considering the temporal aspects of the frames. The existing system is context aware surveillance video summarization which is a Domain dependent System. It works only on low level features and correlation between them is extracted and updated using dictionary algorithm in an online fashion. Thus dictionary size increases. In contrast to the existing method, the proposed system is a domain adaptive video summarization framework based on high level features in such a way that the summarized video can capture the key contents by assuring minimum number of frames. One of the high level features extracted is Local binary pattern (LBP).Key frames can be extracted after finding the Euclidean distance between the LBP descriptor in different methods. The key frames are classified using k-means clustering algorithm. The result is compared with several datasets thus showing the effectiveness of the proposed system. The entire work can be simulated using matlab.

***Keywords* — Euclidean distance; feature extraction; LBP; video summarization**

I. INTRODUCTION

There is a huge growth in video data that calls for an urgent need to develop tools that summarize events occurring in these videos. Large parts of most videos are often redundant or not informative. So manually watching for hours only to figure out the informative events is very time consuming. Furthermore, it is difficult for people to focus on watching videos for hours and not miss important events in the video. So, it is very important to develop tools that automatically select the most informative parts of a video sequence. This summarization technique is called Video Summarization.

Video summarization is divided into two:

a) Static Video Summarization

It is also called static video storyboard, that contains a set of key frames which is extracted. from the original video

b) Dynamic Video Summarization

It is also called dynamic video skimming that computes the similarity or relationship of each shot by collecting a set of shots.

In this paper, a simple approach for video summarization is proposed in a domain adaptive framework which is based on extraction of high level features from video frames and are classified using any learning techniques.

Manuscript published on 30 June 2017.
***** Correspondence Author (s)
Aiswarya. N.R, Department of Electronics and Communication Engineering, Sree Chitra Thirunnal College of Engineering, Trivandrum, India. E-mail: aisu.classmate@gmail.com)
Smitha.P.S , Department, of Electronics and Communication Engineering , Sree Chitra Thirunnal College Of Engineering, Trivandrum, India.

In addition, a new methodology for evaluating video summarization by comparing with VSUMM dataset and user summaries is considered for comparison. Thus evaluation of VSUMM is performed on different videos e.g.: cartoons, news, sports, tv-shows) etc.

II. LITERATURE SURVEY

Shu Zhang, Yingying Zhu, and Amit K. Roy-Chowdhury [1], proposed a method called Context-Aware Surveillance Video Summarization. There are two main algorithms used in this method. One is sparse group lasso optimization algorithm and other one is Online updates the dictionary of correlation algorithm. This method is mainly focused on summarizing surveillance videos. Features and correlation that exist among features of individual video frames are also considered

Zhuang et al. (1998) [2] proposed a method using unsupervised clustering for key frame extraction. Here, the video is segmented into shots and then a color histogram is calculated for every frame. The clustering algorithm uses a threshold that controls the clustering density. Before a new frame is classified, the similarity between the node and the centroid of the cluster is computed first. If this value is less than threshold, then this node is not close enough to be added into the cluster. The key frame selection is employed only to the clusters which are considered as key clusters. In that case, a representative frame is extracted from this cluster as the key frame. The key frame is selected as the frame which is closest, to the key cluster centroid for each key cluster. This proposed technique is efficient and no comparative evaluation is performed for validating such assertions

Hanjalic and Zhang (1999) [3] proposed a method for producing a summary of an arbitrary video sequence which is based on cluster-validity analysis and is designed to work without any human supervision. This entire video material is first grouped into clusters. Then each frame is represented by color histograms in the YUV color space. Now, a partitional clustering is applied n times to all frames .Then the prespecified number of clusters starts at one and is increased by one each time the clustering is applied. Thus, the system automatically calculates the optimal combination of clusters by applying the cluster-validity analysis. After this optimal number of clusters is found, each cluster is represented by one characteristic frame, which becomes a new key frame. Hanjalic and Zhang (1999) concentrated on the evaluation of the proposed procedure for cluster- validity analysis, rather than on evaluating the produced summaries.

Gong and Liu (2000) [4] proposed a technique for video summarization based on Singular Value Decomposition (SVD).

Firstly, a set of frames in the input video is selected. Then, color histograms in the RGB color space are used to represent video frames. Each frame is divided into 3 * 3 blocks, and a 3D-histogram is created for each of the blocks to incorporate spatial information. Then, these nine histograms are concatenated together to form a feature vector. A feature-frame matrix A (usually sparse) is created for the video sequence using this feature vector extracted from the frames, Then, SVD is performed on A to obtain the matrix V, in which each column vector represents one frame in the feature space. Then, the cluster closest to the origin of the feature space is found, and then the content value of this cluster is computed. This value is used as the threshold for clustering the remaining frames. Now from each cluster, the system selects the frame that is closest to the cluster center as key frame.

III. PROPOSED METHOD

Fig 1: shows the steps of our method that produces a domain adaptive static video summary. Firstly, the original video is split into frames and then high level features are extracted, i.e. here Local binary pattern (LBP) features are calculated and are classified using an unsupervised method called k-means clustering. In most of the existing methods, domain dependent system is used for summarization (for e.g. only: surveillance videos, sports videos etc.).

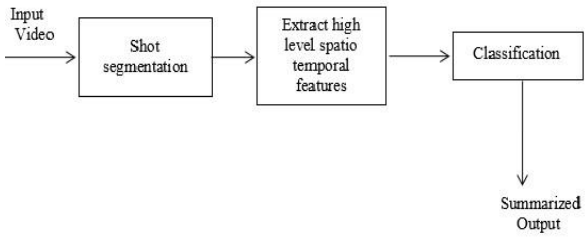

Fig 1: Block diagram for proposed method

VSUMM approach doesn't take all frames. So frame rate is calculated and corresponding frames are taken for feature extraction. Then these frames are classified to extract key-frame. The meaningless frames are removed from video sample. Then, by using k-means clustering the frames are grouped. By calculating Euclidean distance between frames of each cluster and within each cluster, one frame per cluster is selected which the selected frame is called Key frame. Now similar key frames are eliminated to refine static video summary. Finally, remaining key frames are arranged in temporal order.

Steps involved are:

3.1. Shot Segmentation

Here, the video stream is split into shots or frames of images. Frame rate is calculated for every video. This method is also called pre-sampling approach where sampling rate is fixed on one frame per second. For e.g.: normal frame rates are 24fps, 30fps, 60fps etc.

3.2. Feature Extraction

In the existing methods, low level spatio temporal features are extracted to detect motion regions and to detect multiple events. For e.g.: features like spatio-temporal interest point (STIP) detector, histogram of oriented gradients (HOG) and histogram of optical flow (HOF) features are extracted to detect motion regions. But this will not give an accurate result after extracting features. So a high level spatio temporal feature is extracted here.

Local binary pattern (LBP), which is a visual descriptor used for classification in computer vision. It's mainly used for texture classification. When LBP is combined with (HOG) descriptor, it improves the detection performance on datasets. LBP feature vector is created as:

- Firstly, divide the examined window into cells (e.g. 16x16 pixels for each cell).
- Now for each pixel in a cell, compare the pixel to each of its 8 neighbours. Follow the pixels along a circle, i.e. clockwise or counter-clockwise.
- When the center pixel's value is greater than the neighbor's value, set binary array to "0" else "1". This gives an 8-digit binary number (which is usually converted to decimal).
- Optionally normalize the histogram.
- Now concatenate (normalized) histograms of all cells which gives a feature vector for the entire window.

The feature vector can now be processed using the Support vector machine or some other machine-learning algorithm to classify images. Such classifiers can be used for face recognition or texture analysis. Here it is processed using k means clustering algorithm.

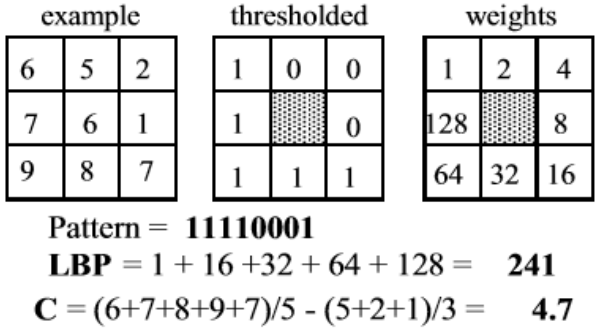

Fig 2: Calculation of LBP values

3.3. Clustering Technique

Clustering is a method of grouping similar frames within a cluster or in between clusters. In the existing methods, a High density peak search (HDPS) clustering algorithm and a Video representation based high density peak search (VRHDPS) clustering algorithm was used for integrating some important properties of video.

In this paper, the most efficient clustering method when compared to existing method is K-means clustering algorithm. This is one of the simplest methods of unsupervised learning algorithm. In this work, k-means clustering is applied to frames extracted using LBP feature descriptor. Now, Euclidean distances between LBP features are calculated and then classified using k-means clustering algorithm. For finding the centroid of each cluster, Euclidean distance between the clusters and within the clusters is calculated. Thus for each key cluster, the closest frame to the centroid cluster which is measured by Euclidean distance is selected as key frame.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

The value of k is user defined. When the value of k increases, redundancy will occur. If it decreases, there will be loss of key frames. So threshold value (i.e. k value) is set according to this. But different videos have different k value. This is the main drawback of k-means clustering.

IV. EXPERIMENTAL RESULT

The experiments were performed on VSUMM dataset which includes VSUMM summaries and user summaries of videos to improve the effectiveness of video summarization. VSUMM dataset contains several videos with different events for e.g.: cartoons, news, sports, TV-shows, home videos etc. Surveillance videos are also considered for summarization since they have lot of redundancy. A comparison with VSUMM dataset and user summary is performed on extracted frames after clustering. If these frames exist in the summary, then that frame is a key frame else redundant frame. Thus redundant or useless frames are removed. Experimental set up is done for more than 50 videos each with a duration varying from 1 to 10 min.

Fig 4: shows VSUMM summary and one of the user summaries of cartoon video. The extracted frames are compared. But result shows key frames are missing and no: of redundant frames increases when k value increases. Say for k=10, there are missing frames as well as useless frames. So accuracy is affected. Thus future scope depends on getting accurate result with minimum redundancy and maximum key frames.

Thus an alternate method to summarize the key frames is using LBP extraction and setting a threshold value i.e. mean or standard deviation instead of using any learning algorithms like k-means clustering for classification process. This results in better accuracy and performance when compared to k-means clustering process.

Fig 3: Result analysis of k means clustering for k=10

(a) VSUMM

(b) User #5

Fig 4: Dataset of VSUMM summary and one user summary of a cartoon video

V. CONCLUSION

Video Summarization has attracted a fast growing attention from researchers and thus several algorithms and techniques have been proposed. In this work, a review on domain adaptive framework using LBP and k-means clustering is carried out. An alternate method using threshold value calculation is also done for comparison and to produce better result. VSUMM dataset is used to produce static video summaries. The evaluation process includes comparison between VSUMM dataset, user summary and extracted key frames of video. Thus, this technique produces video summaries of high visual quality and also can be used for summarization of different types of compressed videos. Video summarization produces more informative summary if semantic features are combined with visual descriptors. Future work is based on overcoming the drawback of k-means clustering technique and this can be extended using any other semantic feature as well as different clustering algorithm.

REFERENCE

1. Shu Zhang, Yingying Zhu, and Amit K. Roy-Chowdhury. ContextAware Surveillance Video Summarization. IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 25, NO. 11, NOVEMBER 2016 .
2. Zhuang, Y., Rui, Y., Huang, T.S., Mehrotra, S., 1998. Adaptive key frame extraction using unsupervised clustering. In: Proc. IEEE Internat. Conf. on Image Processing (ICIP), vol. 1, pp. 866–870.
3. Hanjalic, A., Zhang, H., 1999. An integrated scheme for automated video abstraction based on unsupervised cluster-validity analysis. IEEE Trans. Circuits Systems Video Technology 9 (8), 1280–1289
4. Gong, Y., Liu, X., 2000. Video summarization using singular value decomposition. In: Proc. IEEE Internat. Conf. on Computer Vision and Pattern Recognition (CVPR).

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)