MAT JOURNALS

Journal of Image Processing and Artificial Intelligence
e-ISSN: 2581-3803
Volume 5 Issue 1

Informative Content Extraction through Key Frames

Tejashree Mahajan[1], Prajakta Mahale[1*], Ashwini Ladekar[2], Akansha Matkar[1]
[1]UG Student, Department of Information Technology, Pimpri Chinchwad College of Enginering, Pune, Maharashtra, India
[2]Assistant Professor, Department of Information Technology, Pimpri Chinchwad College of Engineering, Pune, Maharashtra, India
****Email:**prajakta.mahale17@gmail.com*
***DOI:** http://doi.org/10.5281/zenodo.2567748*

Abstract

Nowadays, a huge amount of multimedia data is available on internet. As there is much redundancy, we can't go through all this data. For various purposes this data is browsed, retrieved and processed. But when time and speed constraints are taken into consideration, accessing the data is very inefficient. Video Summarization is one of the way which makes it possible to give non-redundant, effective, feature based abstract view of the video which is covering entire contents in terms of the selected key frames. Feature based selected frames will lead to the summarized video. There are two main technique of video summarization i.e. key frame based and video skimming. This paper focuses on key frame extraction using video abstraction, visual descriptors and bag of visual words approach.

Keywords: *Video summarization, Key-frame extraction, Abstraction, Visual descriptor, Bag of words*

INTRODUCTION

Recently, capturing and creating videos as being a craze among people has led to huge amount of video data containing redundancy. Today, time and space are the most important constraints. So, there is a growing need to create summarized video data. Video summarization is provides an efficient feature to search and go through this data quickly with maximum information gain. Salient features are focused so that data won't be lost. Video summarization is a mechanism provides the way to create a short, abstract, informative, feature based video which may contain stationary images or moving images(skim). Based on this video summarization is divided into video skimming and key frame extraction [1, 2].

Key-frames are also called as characteristic frames, representative frames and these contain the salient features form the original source video. Hence, the key-frame set K is defined as follows:

K = Key Frame (V) = {f1; f2,…….. fn}

There are various ways to extract these characteristic frames like histogram difference between neighbor or non-neighbor frames, shot difference calculation, calculating standard mean and deviation etc. Then the threshold is provided. Frames with difference greater than threshold are selected as key frames and further used to create summarized video. The quantity of recordings has been expanded and the capacity of people to catch or/and make advanced video has been developed in the meantime. So there is a developing requirement for video synopsis [3, 4].

RELATED WORK

Video Abstraction

Video abstraction, as the name infers, is a short synopsis of the substance of a more drawn out video report. Generally not over 5 minutes. In particular, a video unique is a succession of still or moving pictures speaking to the substance of a video. A collection of images that best represent the important content is the main focus of

video abstraction. Based on the way the key-frames are constructed abstract video will generate (Figure 1).

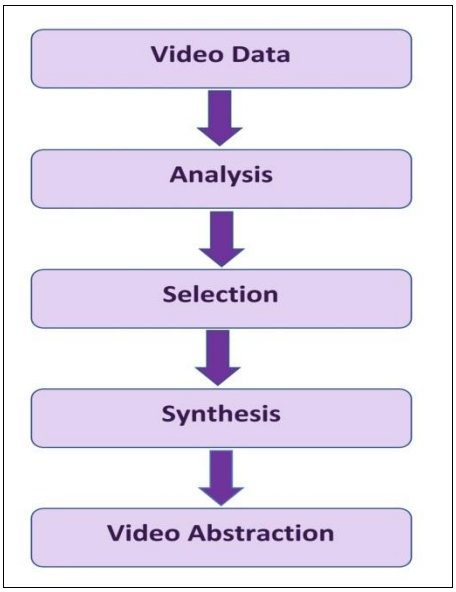

Figure1:*Video Abstraction.*

The video abstraction process usually has three phases: Video information analysis, meaningful clip selection and output synthesis. Video data is provided to as input after this first step is Analysis. Analysis means to find the different components. Then next step is to select meaningful frame selection. This needs to choose the appropriate video and audio sets which are relevant to each other. Then last step is Synthesis it means that whichever frames are selected they need to combine with each other. This finally gives the abstracted video[1]. Video abstraction mainly focuses on to give a sequence of stationary or moving images to quickly go through the entire content covering all the information in the video [5].

Visual Descriptors

Visual descriptors image descriptors are portrayals of the visual highlights of the substance in pictures, recordings. They portray rudimentary qualities, for example, the shape, the shading, the surface or the movement, among others. These descriptors have a decent information of the items and occasions found in a video, picture or sound and they permit the fast and effective hunt of the broad media contet. The institutionalization framework that bargains with broad media descriptors is the MPEG-7 (Motion Picture Expert Group - 7).

Visual Descriptors are dividedin Two Main Groups:

General Information Descriptors (Global): These are low level descriptors which give information about color, shape, regions, textures and motion.

Specific Domain Information Descriptors (Local): they give information about objects and events in the scene. These descriptors are used in summarization of visual content. Based on the features of the content summary is generated. Instead of considering complete video at a time, video shot is referred. Following dig shows the extraction process step by step (Figure 2) [6].

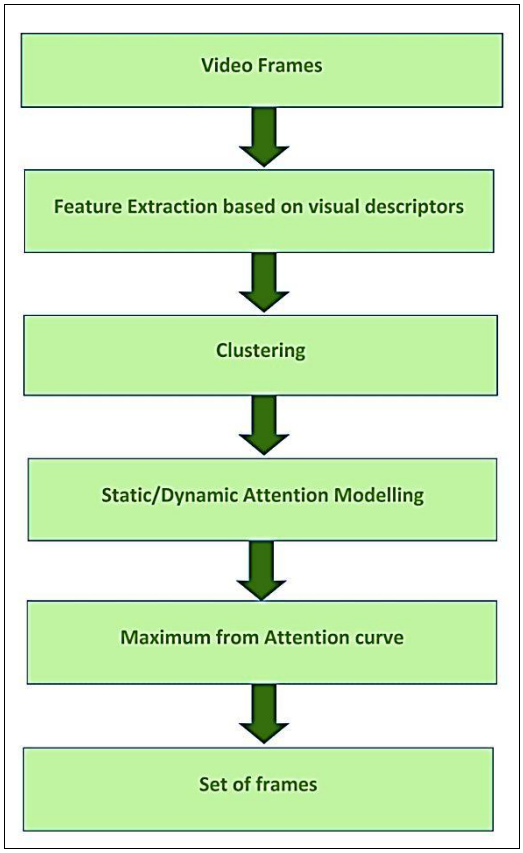

Figure 2:*Visual Descriptors.*

Identification of minute changes in frames is important so that the chances of content loss will be reduced. For this purpose instead of using single descriptor, combination is used.

Bag of Visual Words

Words of visual entities found in an image that describe object and represent semantic entities. Key frame extraction is done using local features of the image. Bag of words count how many number of times a particular image occurs in a video. Bag of words an approach extract feature of an image using SIFTS descriptors and produces a video summary. Bag of old model consists of:

Feature representation

Feature presentation consists of feature detection and description. In feature extraction local features are divided as detecting a local region known detector and describing the detected region known as descriptor. The most popular technique for extracting local features is using SIFT descriptor. SIFT describes local features of an image by storing waited at orientation histogram of salient corners of an image. SIFT algorithm detect salient image regions and extracts distinct descriptors of the appearances. Through Feature extraction and description setup feature vectors is generated [7].

Codebook generation

Testing is performed on set of feature vectors. Through clustering visual vocabulary is created. Code word is group of feature vectors that has similar properties. These code words are visual words and collection of this code words is code book and also called as visual word vocabulary.

Histogram generation

Histogram of visual voice is generated. Histogram count occurrences of visual words. Histogram of visual world is created for each frame.These histograms a clustered and the frame that represents the centroid of the cluster are considered as key frame. There is elimination of redundant keyframes. These key frames are ordered and a video summary is generated (Figure 3).

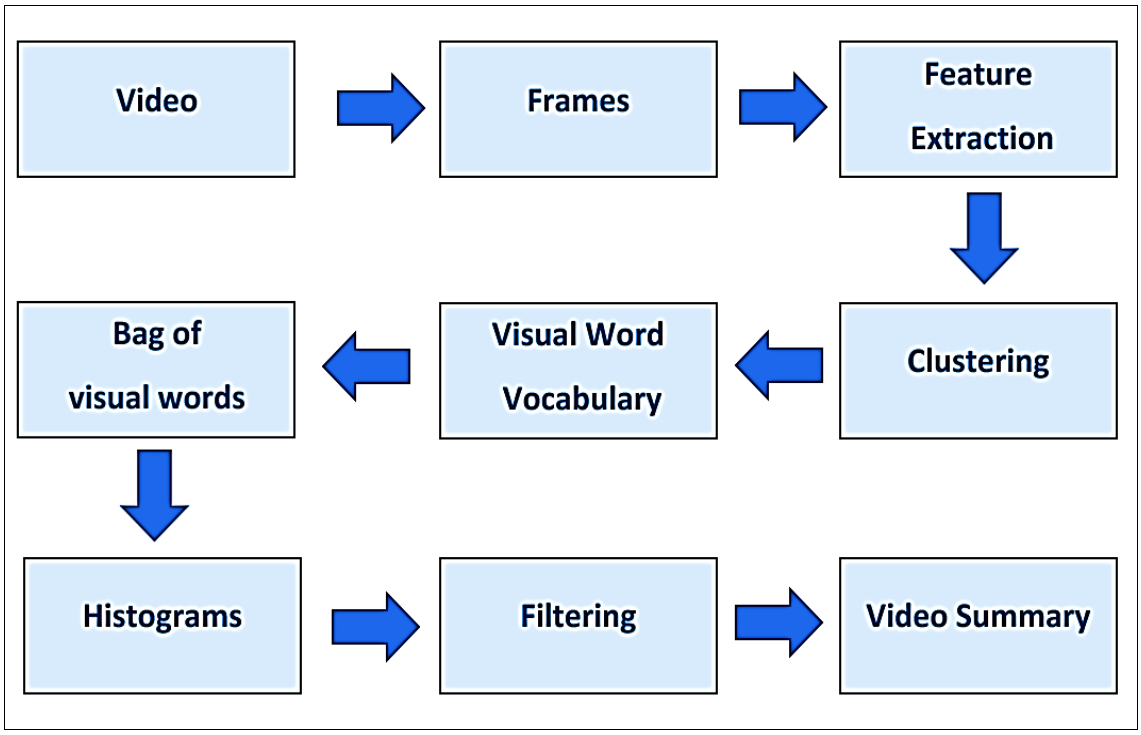

Figure 3:*Bag of Words.*

PROPOSED METHOD

This section explores the detailed step by step process of video summarization. Video segmentation is the very first step. The featured frames extracted using various algorithms, are used to find the cumulative difference .They may be neighbor or non-neighbor frames. Based on this difference, a frame which covers all the content without redundancy are selected as key frames to give abstract video (Figure 4).

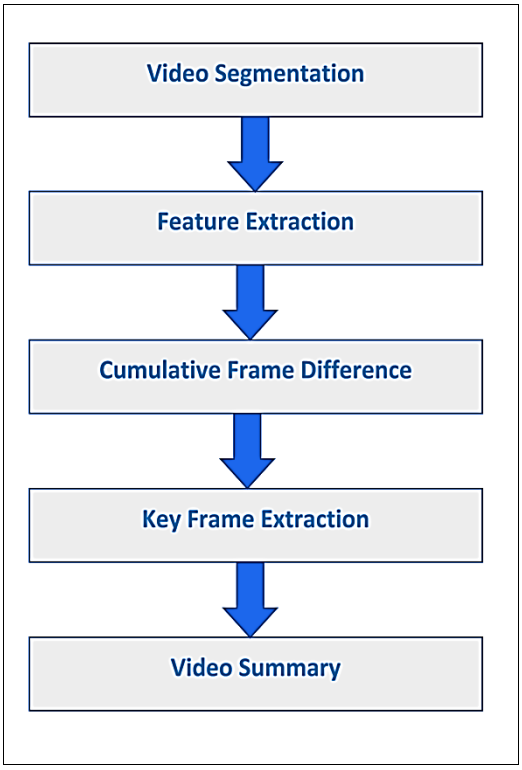

Figure 4:*Proposed Method.*

1. Read the input video and get the properties of the video.
2. Convert the input video into frames.
3. Extract features from video
4. Calculate the difference between the adjacent frames.
5. Quantize the difference for better result.
6. Extract key-frame using the difference.
7. Key frames extracted are combined to form a new video which is the summarized video of the given input video.
8. The summarized video is played in video viewer app using a command in MATLAB.

OBSERVATION

Among the various techniques of key frame extraction, Video abstraction mainly focuses on to give a sequence of stationary or moving images to quickly go through the entire content covering all the information in the video. Rather visual descriptors form the cluster of frames based on various features like color, shape, motion. Instead of applying single descriptor, combination can be used to have maximum details from the video by identifying minute differences. But it will lead to more complexity in terms of time.

But bag of words is more efficient technique to extract the frames.Selected features are represented in vector form and clustered based on similarities to give code words. The centroid of the cluster represents the key frame which is highly feature based and non redundant. It is efficient in all aspects.

RESULT AND DISCUSSION

To resolve the issue of redundancy and minimize the time, space constraints summarization plays important role.

The generated summary is more efficient in terms of detail abstract view with more accuracy. Due to feature extraction every detail is captured.

CONCLUSION

There are various approaches for summarizing thevideo. By focusing on video abstraction, visual descriptors and bag of words techniques used for summarizing the video, bag of words approach is more efficient and produces good results. Video can be summarized by using BOW and video temporal segmentation. Features of the objects are obtained by using bag of visual words and are trained using SIFT descriptor so that the feature matching points and the

performance of the identification of the objects is improved. In addition, video summaries created with our semantic analysis will be most similar to the user's summaries.

REFERENCES

1. Video Summarization: Techniques and Applications Zaynab El khattabi, Youness Tabii, Abdelhamid Benkaddour World Academy of Science.*Engineering and Technology International Journal of Computer and Information Engineering.*2015; 9(4).
2. Key frame extraction based on visual attention model Jie-Ling Lai,Yang Yi Computer Science Department, School of Information Science & Technology, Sun Yat-sen University, *Panyu Distric Guangzhou, China.*
3. Bag of Words Based Surveillance System Using Support VectorMachines,Nadhir Ben Halima, Osama Hosam. *International Journal ofSecurity and Its Applications.* 2016; 10(4).
4. A Study on "VIDEO SUMMARIZATION" Tanuja Subba, Bijoyeta Roy,Ashis Pradhan. M.Tech, Computer Science & Engineering, Sikkim Manipal Institute of Tech1Assistant Professor.*Computer Science & Engineering, Sikkim Manipal Institute of Tech International Journal of Advanced Research in Computer and Communication Engineering*June 2016; 5(6).
5. Truong B.T.& Venkatesh S.Videoabstraction:asystematic review and classification.*ACM Transactions on Multimedia Computing, Communications and Applications*.2007; 3(1): pp.1–37.
6. A Novel Key-Frame Extraction Approach for Both Video Summary and Video Index ShaoshuaiLei,Gang Xie,&Gaowei Yang Hindawi Publishing Corporation. *e Scientific World Journal*. Volume 2014, Article ID 695168, 9 pages http://dx.doi.org/10.1155/2014/695168.
7. Lai J.-L. &Yi Y.Key frame extraction based on visual attention model.*Journal of Visual Communication and Image Representation*. 2012; 23(1): pp.114–125.

Cite this article as: *Tejashree Mahajan, Prajakta Mahale, Ashwini Ladekar, & Akansha Matkar. (2019). Informative Content Extraction through Key Frames. Journal of Image Processing and Artificial Intelligence, 5(1), 25–29. http://doi.org/10.5281/zenodo.2567748*