IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 6, June 2016

A Study On "VIDEO SUMMARIZATION"

Tanuja Subba[1], Bijoyeta Roy[2], Ashis Pradhan[3]

M.Tech, Computer Science & Engineering, Sikkim Manipal Institute of Tech[1]

Assistant Professor, Computer Science & Engineering, Sikkim Manipal Institute of Tech[2, 3]

Abstract: Nowadays, a huge amount of multimedia data is being processed, browsed, retrieved which makes its delivery slower and computation cost expensive. The technique video summarization is one of the ways of managing video and browsing and is extended in order to process entire video information in minimum amount of time. This technique makes user easier for quick browsing of large amount of data. It is a process of creating a summarized view or an abstract view of an entire video in minimum amount of time and also removing duplication or redundant features by extracting key frames and video skims. There are two ways to generate a sample video they are static and dynamic which are categorized into two main technique of video summarization i.e. key frame based and video skimming respectively. This paper focuses on the approach of static and dynamic technique.

Keywords: static video summarization, dynamic video summarization, key frame, video skim.

I. INTRODUCTION

The compact representation of sequence of video for user and letting the user browse and retrieve the large collection of video data easily and quickly is now becoming a large and important topics in video processing.

The popularity of the internet videos, Google videos has increased the availability of videos tremendously in the internet and for this reason automatic process of generating a concise representation of moving video or still video is necessary.

This refers to video summaries, which provides the user with information about the contents of the video in short period of time. The need for generation of video summary automatically is pushed for both the user and the production viewpoints.

Therefore, the goal of video summarization is to process video sequence with more interesting, valuable and useful to the user. 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 most informative and concise representation of the whole video[1].

There are two main category of video summarization that is static summary and dynamic skimming.

The process of extracting features from the original raw video and abstracting the video summary is shown in the figure 1. The extracted features may be sequence of stationary images (key frames) or moving images (video skims).

This paper focuses only on the type of video summarization technique that is static and dynamic that is used for feature extraction and abstraction in order to summarized video for the user to understand and retrieve the video easily irrespective of time.

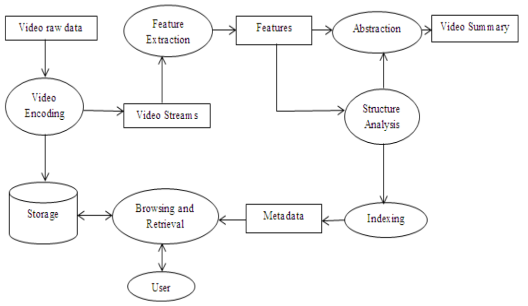

Figure 1: General application of abstraction and indexing of a video [1]

Figure 2 shows the general architecture of video summarization.

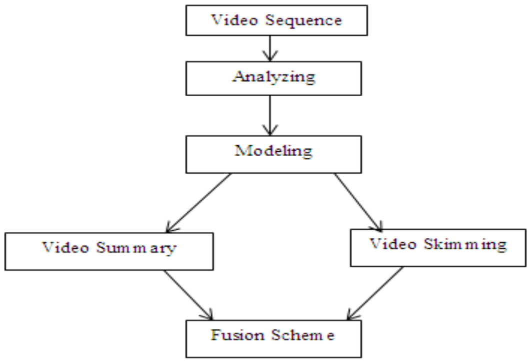

Figure 2: Block Diagram for video Summarization

II. RELATED WORKS

A number of studies and the application based on video summarization techniques have been proposed so far as because the use of the videos in the web has increased over the years. Some of them can be classified as follows.
1. R-sequence method: it is a static video summarization where fixed number of representative frames is extracted to summarize a given digital video[2]

2. Color based Selection: it quantize color space into n cells and compute histogram with number of pixels in each cell and compute the distance between the histogram[3]
3. Perceived motion energy detector: A triangle model of perceived motion energy is used to model motion patterns in video and a scheme to extract key frames based on this model[10]

III. STATIC VIDEO SUMMARIZATION

This is also called a key frame based video summarization techniques or still image abstract or storyboard. There are some criteria that come across for key frame based techniques, which are as follows:
1. Redundancy: frames with minor difference are selected as key frame.
2. When there are various changes in content it is difficult to make clustering.
The Figure 3 shows selection of key frame.

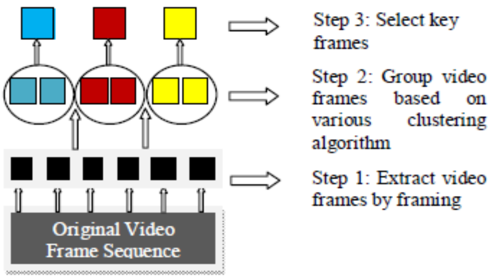
Step 3: Select key frames
Step 2: Group video frames based on various clustering algorithm
Step 1: Extract video frames by framing
Original Video Frame Sequence

Figure 3: key frame based video summaization[4]

The key frame based summarization can be classified in three different ways
1. Classification based on sampling
It selects key frame uniformly or randomly without considering the video content.
2. Classification based on scene segmentation
It extracts key frames using scene detection, it includes all semantic link in the video.
3. Classification based on shot segmentation
It extracts first image and last image as a shot key frame.
Example: Video summarization by clustering using euclidean distance.

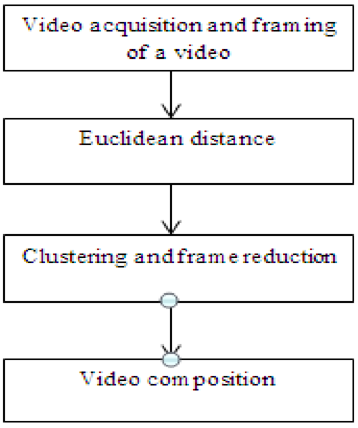
Video acquisition and framing of a video
Euclidean distance
Clustering and frame reduction
Video composition

Figure 4: block diagram of video summarization using Euclidean distance [5]

The detailed explanation for the flowchart, shown in figure 4 is as follows:

1. Video acquisition and framing of a video
Video is sampled a constant rate which is divided into set of frames. An ordered set of input digital video sequence V with cardinality N is defined as
$V= \{f_1, f_2, f_{3,\ldots\ldots,} f_N\}$

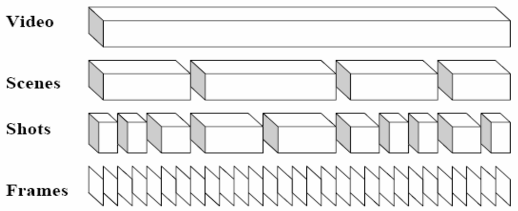

Figure 6: key frame extraction [6]

2. Euclidean distance calculation
The resulting frame is analyzed to obtain the feature frame matrix. Every image has RBG associated with each of its pixel. The first is taken as a reference frame.
The word count for each pixel value is taken in a vector form i.e. the vector represents a single frame and the whole video is represented in a set of vectors.

The distance between this frame is calculated using Euclidean distance

$$E=\sqrt{\Sigma(\boldsymbol{Xj}-\boldsymbol{Yj})^2} \quad \ldots\ldots\ldots.. \text{equ. (i)}$$

Where x and y represents 2 different frames of an images and j represents columns. The Euclidean distance between two consecutive images is calculated and a threshold value is given, when a distance exceeds a given threshold, a key frame is claimed and that frame serves as a new reference frame.

3. Clustering and frame reduction
The extracted features of the frame are clustered and are classified into different classes based on the distance calculated using Euclidean distance measure. Each individual class is classified under same frame name.

4. Video composition
The individual frame that has been extracted is considered as a key frame which when combine together will compose a summary of a video.

IV. DYNAMIC VIDEO SUMMARIZATION

The idea of dynamic summarization called as video skimming is a short video composed of informative scenes from original video presented to the user to receive in video format that is it condenses the original video into shorter form while preserving the important content of a video in short time. It also preserves the motion information.

DOI 10.17148/IJARCCE.2016.56164

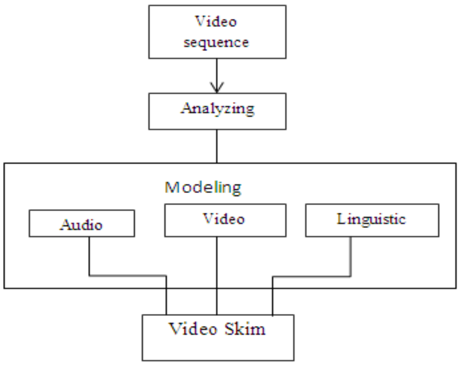

Figure 7: Dynamic video summarization [7]

Example: Video Skimming and Characterization through the Combination of Image and Language Understanding.
This technique summarize the video in which video and audio portion also consists significant audio or spoken words, instead of simply understanding the synchronized portion corresponding to the selected video frames.

1. Video Skim: A video is a collection of scene. The relative importance of each scene can be evaluated by 1) the objects that appear in it, 2) the associated words, and 3) the structure of the video scene. The significance of video is achieved through the integration of image and language understanding which is needed to realize the level of characterization and is essential to skim creation.

2. Video Characterization: Through this techniques scenes, segments and individual frames in video can be characterize. Understanding and identifying the most significant words in a given scene, and image understanding, it entails segmentation of video into scenes, detection of objects of importance (face and text) and identification of the structural motion of a scene.

3.

2.1 Audio and Language Characterization: Language analysis works on transcript to identify important audio regions known as “keywords”. In this approach a well-known technique of TF-IDF (Term Frequency Inverse Document Frequency) is used to measure relative importance of words for the video document [8].

$$TF\text{–}IDF = \frac{f_s}{f_c}$$ ……….. equ. (ii)

The TF-IDF of a word is its frequency in a given scene, fs, divided by the frequency, fc, of its appearance in a standard corpus.
Using individual keywords the audio skim may be fragmented and incomprehensible for some speakers. Therefore, to increase comprehension, this approach uses longer audio sequences, “keyphrases”, in the audio skim. A keyphrase may be obtained by starting with a keyword, and extending its boundaries to areas of silence or neighboring keywords. Another method for extracting significant audio is to segment actual phrases. To detect breaks between utterances a modification of Signal to Noise ratio (SNR) techniques is used which compute signal power. This algorithm computes the power of digitized speech samples where Si is a pre-emphasized sample of speech within a frame of 20 milliseconds.

$$Power = \log\left(\left(\frac{1}{n}\right)\cdot\sum\left(Si^2\right)\right)$$ ……….. equ. (iii)

Each keyphrase is isolated from the original audio track to form the audio skim.

2.2 Scene Segmentation: A color histogram difference technique is used to segment scene. By detecting significant changes is the weights of histogram of successive frames video sequence are separated into scenes. This technique is simple and robust and gives 91%of accuracy.

2.3 Camera Motion Analysis: Interpretation of camera motion is one of the important aspects in video characterization. There are two types of motion, object motion and actual camera motion and this are distinguishes by the global distribution of motion vectors. Object motion typically exhibits flow fields in specific regions of an image. Camera motion is characterized by flow throughout the entire image. This flow pattern is given by affine model.

$$u\left(x_i, y_i\right) = ax_i + by_i + c$$
$$v\left(x_i, y_i\right) = dx_i + ey_i + f$$ ……….. equ. (iv)

Affine parameters a, b, c, d, e, and f are calculated by minimizing the least squares error of the motion vectors. Average flow V and U is also calculated.

2.4 Object Detection: Face and Text: Identifying significant object in video is complex and important task for video skimming. This approach deals with two of the interesting objects in a field i.e. human faces and text. To reduce computation this fields is detected in every 15th frame. Text in the video provides significant information as to the content of a scene. A typical text region can be characterized as a horizontal rectangular structure of clustered sharp edges, because characters usually form regions of high contrast against the background. By detecting these properties we extract regions from video frames that contain textual information.

4. Technology Integration and Skim Creation: A video is characterized by scene breaks, camera motion, object appearance and detecting keyphrases. Skim creation requires all this appropriate characteristics to create video skim.

4.1 Audio Skim: Skim creation is done by creation of the reduced audio track, which is based on the keyphrases. Those phrases whose total TF-IDF values are higher than a fixed threshold are selected as keyphrases. Keyphrases with words that is repeated throughout the transcript may create skims with redundant audio. Therefore, keyphrases which repeat within a minimum number of frames (150 frames) and limit the repetition of each keyword in a keyphrase is discarded.

4.2 Video Skim Candidates: The contents of the audio and video are not necessarily synchronized. Therefore, for each keyphrase, analyzing the surrounding video frames and selecting a set of frames which may not align with the audio in time, but which are most appropriate for skimming is done.

4.3 Image Adjustments: With prioritized video frames from each scene, a suitable representation for combining the image and audio skims for the final skim is available. A set of higher order Meta-rule are used to complete skim creation. Figure 8 shows characterization of all the steps for video skimming and characterization.

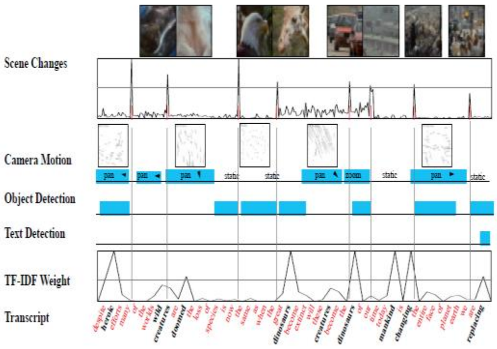

Figure 8: video characterization: keywords, scene breaks, camera motion, objects (faces and text)[9]

V. CONCLUSION

The emergence and the retrieval of high volume video libraries have clearly shown the need of summarization techniques for creation of abstract representation of video within a short period of time. Static video is used when key frame is to be extracted and dynamic approach of video skimming and characterization is described in order to give video skimming considering all the important aspects like audio, video, text and also camera motion. Both static and dynamic technique of summarization gives the summary of video which gives all the necessary details of entire video in short period of time.

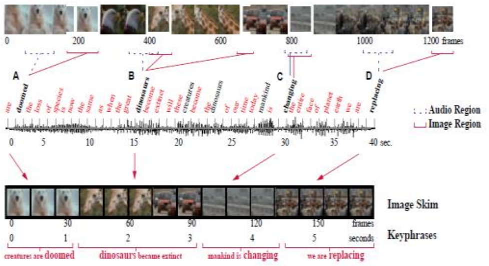

Figure 9: skim creation incorporating word relevance and significant object (faces and text)[9]

REFERENCES

[1] Sachan Priyavada Ranjendra, Dr. Keshaveni N(2014),"A Survey of Automatic Video Summarization Techniques", International Journal of Electronics, Electrical and computational System , Vol 3,Issue 1,pp.1-5.

[2] Xinding Sun,Mohan S. Kankanhalli," Video Summarization Using R-Sequences", Article in Real-Time Imaging 6(6):449-459, December 2000.

[3] Nishant Kumar, Amit Phadikar, "Video Summarization using color featurs and global thresholding", Digital Signal Processing, Vol 6, No 6 (2014).

[4] Sony, A.; Ajith, K.; Thomas, K.; et.al., "Video summarization by clustering using euclidean distance" , International Conference on Signal Processing, Communication, Computing and Networking Technologies on 21-22 July 2011, 2(8), pp.642-646,.

[5] Zaynab El khattabi, et.al., "Video Summarization: Techniques and Applications", World Academy of Science, Engineering and Technology International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol:9, No:4, 2015.

[6] Yu-Fei Ma, Lie Lu et.al."A User Attention Model for Video Summarization", Microsoft Research Asia, Beiing 100080,China.

[7] Mauldin, M. "Information Retrieval by Text Skimming," PhD Thesis, Carnegie Mellon University. Aug. 1989.

[8] Michael A. and Smith Takeo Kanade, "Video Skimming and Characterization through the Combination of Image and Language Understanding," , IEEE International Workshop on Content-based Access of Image and Video Databases (ICCV98 - Bombay, India).

[9] T. Liu, H. J. 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, 13(10), Oct 2003, pp 1006-1013.

DOI 10.17148/IJARCCE.2016.56164