© 2014 IJIRT | Volume 1 Issue 7 | ISSN: 2349-6002

Video Summarization using clustering – A Survey

Darshil J. Shah[1], Mr. Narendra Limbad[2]
[1]*Computer Engineering Department, L.J.I.E.T, Ahmedabad, India*
[2]*Computer Engineering Department [P.G Department], L.J.I.E.T, Ahmadabad, India*

***Abstract*— Video Summarization have a wide area of promising application. This paper offers a different technique for video summarization, focusing on simple method or technique for video summary including feature extraction, clustering and frame extraction for static video. Moreover here we are discussing which method suitable for which type of video.**

***Index Terms*— Video summarization techniques, feature extraction, Clustering methods**

I. Introduction

A video summary is defined as a sequence of still pictures that represent the content of a video in such a way that the respective output group is rapidly provided with concise information about the content, while the essential message of the original video is preserved [5].

Rapid development of computation, communications, and storage infrastructures, are contributing to an enormous and steadily growing availability of video content. Despite the enormous investments in digital video technologies, the capabilities of an average user to manipulate, interact with and manage videos are still far behind what average users can achieve with other types of media such as text or images. This is mainly due to the temporal and multi-model nature of video and the size of the associated medium [4].

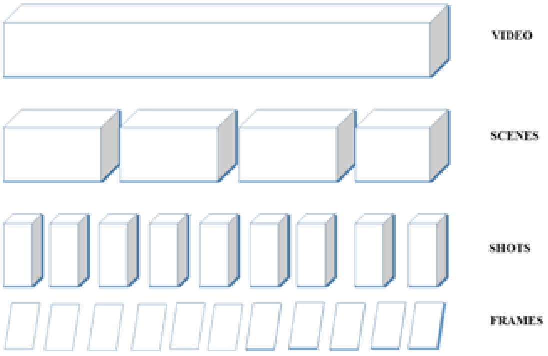

Figure 1 hierarchy structure of video [7]

A video is nothing but a synchronous sequence of a number of frames, each frame being a 2-D image. So the basic unit in a video is a frame. The video can also be thought of as a collection of many scenes, where a scene is a collection of shots that have the same context. This is shown in figure-1. A frame is single still image from video. It defines as fps (frame per second). Shot is the sequence of frame recorded by single camera operation. And scene is collection of shots. So frame is basic unit for all video. And mostly for video summary is frame is used.

Video summarization is used for give summary of original video so there so many application regarding this. It is used to summarize video in specific situation like Previews of movies, TV episodes, etc, Summaries of documentaries, home videos, etc, Highlights of football games, cricket match, etc, Interesting events in surveillance videos (major commercial application). Video summarization techniques also used for video retrieval for video browsing.

Basically video summarization method divided in two parts (1) Static video summarization and (2) Dynamic video summarization [8], [9], [10], [18]. Static video summaries consist of a set of frames (key frames) extracted from the original video, while dynamic video summaries are a video clip consists of a collection of video segments (and corresponding audio) extracted from the original video [4].

A. Static video summarization

The simplest static visualization method is to present one frame from each video segment, which may or may not correspond to an actual shot, in a storyboard fashion. The problems with this method are that all shots appear equally important to the user and the representation becomes impractically large for long videos.

Although video abstracts are compact, since they do not preserve the time-evolving nature of video programs, they present fundamental drawbacks. They are somewhat unnatural and hard to grasp for no experts, especially if the video is complex. Most techniques just present keyframes to the user without any additional metadata, like keywords, which can make the meaning of keyframes ambiguous. Finally, static summaries are not suitable for instructional, and presentation videos, as well as teleconferences, where most shots contain a talking head, and most of the relevant information is found in the audio stream. These deficiencies are addressed by dynamic visualization methods [8].

B. Dynamic video summarization

In these methods the segments with the highest scores are selected from the source video and concatenated to generate a video skim. While selecting portions of the source video to be included in the video skim, care must be exercised to edit the video on long audio silences, which generally correspond to spoken sentence boundaries. This is due to the experimentally verified fact that users find it very annoying when audio segments in the video skim begin in mid-sentence [8].

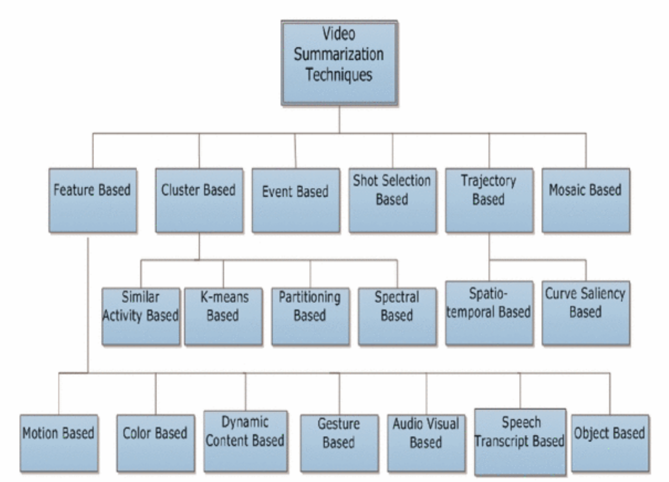

Figure 2 Hierarchical Structure of video summarization Techniques [6]

Techniques are based on feature, cluster, event, selection, trajectory and mosaic. This is shown in figure-2. Feature based have different method like motion, color, texture, gesture, object and etc. cluster based techniques have different methods like k-means, density based, etc.

There are some other techniques for summarization like video summarized on user preferences [3]. And some automated methods are available for video summarization. In [2], video summarization is done using simple action patterns. In that, they give summary on basis of finding action driven by the different object. They also define simple flow for pattern detection. First they detect moving object using movement analysis then using actor analysis actor is detected and actor tracking is done. Then periodicity extraction is done using HoG on the interested blob which get from last tracking method. Lastly summarization is done on basis of output from periodicity extraction steps. In periodicity extraction are extracted key frames from an action's most specific point.

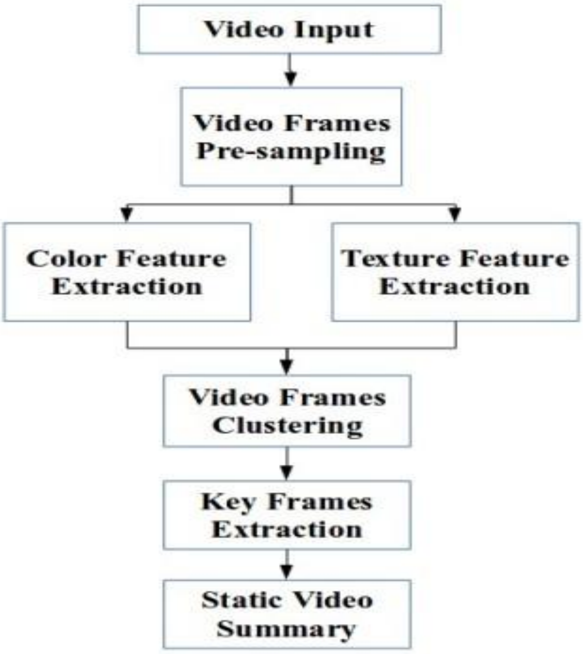

Figure 3 Generic framework of video summarization technique [5]

In figure 3, generic framework of video summarization is given. In this video is taken as input. In first step pre-sampling step is done. Then color features are extracted from video using different methods. Then texture features are extracted from video. Then video frame is clustering by cluster methods. And finally result of last steps is arranged in its original order to generate summary of video.

Rest of paper is organized as follows. Section II is Pre-sampling step. Section III is feature extraction in which different techniques is shown. Section IV is about clustering methods. Section V is description of key frame extraction. Section VI is analysis And finally we offer conclusion in last section.

II. Pre-Sampling

The first step towards video summarization is pre-sampling the original video which aims to reduce the number of frames to be processed. Choosing a proper sampling rate is very important. A low sampling rate leads to poor video summaries; while a large sampling rate shortens the video summarization time. In [5] approach, the sampling rate used is selected to be one frame per second. So, for a video sample of duration one minute, and a frame rate of 30 fps (i.e., 1800 frames); the number of extracted frames is 60 frames.

III. FEATURE EXTRACTION

Feature extraction is used for extract specific features from video. There are mainly two types for extraction (1) Local feature extraction (2) Global feature extraction [1].

Feature Extraction
Global Feature
Local Feature

Figure 4 Basic types of Feature Extraction [1]

Local features such as scale-invariant feature transform (SIFT) or (SURF) descriptor have played important role in many application [1], [19]. It is one of best method for feature extraction using local feature extraction. Global feature such as color, texture, shape is used in most video summarization techniques. There different methods for color, texture feature extraction [15].

Table 1, Contrast of global and local feature extraction [14]

Global features adopted	Local features adopted
Color histogram in HSV space	Color moments, Gabor texture, shape and size
Gabor wavelets texture and edge orientation histogram	SURF
Color histogram in HSV space	Sobel shape
Shape index	SIFT

A. Color feature extraction

Color is important feature of video. In many methods like VSUMM [16], STIMO [17] color feature is used for feature extraction. In video summarization systems, the color space selected for histogram extraction should reflect the way in which humans perceive color [5]. Different color method is shown in table 2 with its advantage and disadvantage.

Table-2, Contrast of different color descriptors [14]

Method	Pros.	Cons.
Histogram	Simple to compute, intuitive	High dimension, no spatial info, sensitive to noise
CM	Compact, robust	Not enough to describe all colors, no spatial info
CCV	Spatial info	High dimension, high computation cost
CSD	Spatial info	Sensitive to noise, rotation and scale
Correlogram	Spatial info	Very high computation cost, sensitive to noise, rotation and scale
DCD	Compact, robust, perceptual meaning	Need post-processing for spatial info

Some other methods are also used for color feature extraction such as color movements. Color Histogram is generally used for color feature extraction. One good choice is the HSV color space for color histogram. A different color based technique is proposed in [13]. Color histogram is used to store color information of a shot.

B. Texture Feature Extraction

Texture is also globally used feature to extract image for video summary. Texture is a important low-level feature for representing images. Texture can be defined as an attribute representing the spatial arrangement of the pixels in a region or image [5]. There are different methods of texture is as bellow:-

- DWT(Discrete Wavelet Transformation)
- Haar Wavelet Transformation [20]
- Gabor Transformation
- Gaussian Pyramid
- Ranklet Transform
- Discrete Fourier Transform
- Discrete cosine transform
- GLC

DWT is common method for texture feature extraction. Gabor is one of best method for texture feature extraction. DWT is used to extract texture feature by transforming it from spatial domain into frequency domain. Wavelet transforms extract information from signal at different scales by passing the signal through low pass and high pass filters.

Table-3, Contrast of different texture descriptors [14]

Method	Pros.	Cons.
Spatial texture	Meaningful, easy to understand, can be extracted	Sensitive to noise and distortions

Method	Pros.	Cons.
	from any shape without losing info.	
Spectral texture	Robust, need less computation	No semantic meaning, need square image regions with sufficient size

Gray co-occurrence matrix (GLC) is one of most effective and important methods for texture feature extraction and description. Its original idea is first proposed in Julesz (1975). Julesz found through his famous experiments on human visual perception of texture, that for a large class of textures no texture pair can be discriminated if they agree in their second-order statistics [12].

IV. Video Frame Clustering

In Video frame clustering process, there are different clustering methods used. All method make cluster using different concept. Clustering methods are as follows.

- DBSCAN
- K –means
- Fuzzy C-means [4]
- GDSCAN [22]

Clustering is the process of grouping a set of objects into classes or clusters so that objects within a cluster have similarity in comparison to one another, but are dissimilar to objects in other clusters [21].

DBSCAN is most likely used algorithm for clustering video frame. DBSCAN tens for density based clustering. In DBSCAN algorithm for clustering first they select any arbitrary point then using that point they checked others point it on boundary or inside the cluster or outside of cluster.

DBSCAN clustering method is enhanced in [5] and they used that method for clustering. In [22] , GDSCAN algorithm is introduced which is generalized version of DBSCAN algorithm and is also used for clustering a video frames.

V. Key Frame Extraction

In video summarization, key frame extraction is done on basis of which methods used in above steps. All methods have different output for this step so basis of given input type key frame extraction is done. In general, for key frame extraction they select frames with highest degree if we use fuzzy c-means clustering [4] or with middle core frame from all clusters if we use DBSCAN concept [5].

For key frame extraction we choose best frame from given input for video summarization and merge them in order to get summary of the given video

VI. Analysis

Different techniques for video summarization is studied so which techniques or methods uses for which relative types of video is analyzed.

Table-4, analysis of Video summarization Techniques [6]

Techniques	Static summary	Dynamic summary	Fixed camera	Moving camera
Motion Based	No	Yes	Yes	Yes
Color Based	Yes	No	Yes	Yes
Gesture Based	Yes	No	Yes	Yes
Dynamic contents based	No	Yes	Yes	Yes
Audio-visual Based	No	Yes	Yes	Yes
Speech transcript Based	No	Yes	Yes	Yes
Clustering Based	Yes	Yes	Yes	Yes
Event Based	Yes	No	Yes	No
Shot Detection Based	Yes	No	No	Yes
Trajectory Based	Yes	No	Yes	No
Mosaic Based	Yes	No	No	Yes

VII. Conclusion

The rapid evolution of video technology has brought large volume of video data. There is need store video efficiently. Now days, People have no time for watching full length video. People want small video which content important parts only. So for that video summarization techniques are required for efficient video summary.

References

[1] Genliang Guan, Zhiyong Wang, Kaimin Yu, Shaohui Mei, Mingyi He, and Dagan Feng, “Video Summarization with Global and Local Features,” 2012 IEEE International Conference on Multimedia and Expo Workshops

[2] M. Said Aydemir, Ugur Ergul, Adem Guclu, M. Elif Karsligil, “Video Summarization Using Simple Action Patterns,” 21st International Conference on Pattern

Recognition (ICPR 2012) November 11-15, 2012. Tsukuba, Japan

[3] Kannan, R., Ghinea, G., Swaminathan, S., & Kannaiyan, S. (2013, December). Improving video summarization based on user preferences. In *Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), 2013 Fourth National Conference on* (pp. 1-4). IEEE.

[4] Ebrahim Asadi*, Nasrolla Moghadam Charkari," Video Summarization Using Fuzzy C-Means Clustering", 20th Iranian Conference on Electrical Engineering, (ICEE2012), May 15-17, Tehran, Iran

[5] Mahmoud, K. M., Ismail, M. A., & Ghanem, N. M. (2013). VSCAN: An Enhanced Video Summarization Using Density-Based Spatial Clustering. In*Image Analysis and Processing–ICIAP 2013* (pp. 733-742). Springer Berlin Heidelberg.

[6] Ajmal, M., Ashraf, M. H., Shakir, M., Abbas, Y., & Shah, F. A. (2012). Video summarization: techniques and classification. In *Computer Vision and Graphics*(pp. 1-13). Springer Berlin Heidelberg.

[7] Rajendra, S. P., & Keshaveni, N. A Survey of Automatic Video Summarization Techniques.

[8] Taskiran, C., & Delp, E. (2005). Video summarization. *Digital Image Sequence Processing, Compression, and Analysis*, 215-231.

[9] Truong, B. T., & Venkatesh, S. (2007). Video abstraction: A systematic review and classification. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, *3*(1), 3.

[10] Ravi Kansagara , Darshak Thakore, Mahasweta Joshi. A study on video Summarization Techniques. *A International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization)Vol. 2, Issue 2, February 2014.*

[11] Hu, W., Xie, N., Li, L., Zeng, X., & Maybank, S. (2011). A survey on visual content-based video indexing and retrieval. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, *41*(6), 797-819.

[12] Patel, B. V., & Meshram, B. B. (2012). Content based video retrieval systems.*arXiv preprint arXiv:1205.1641*.

[13] Chary, R., Lakshmi, D. R., & Sunitha, K. V. N. (2012). Feature extraction methods for color image similarity. *arXiv preprint arXiv:1204.2336*.

[14] ping Tian, D. (2013). A Review on Image Feature Extraction and Representation Techniques. *International Journal of Multimedia and Ubiquitous Engineering*.

[15] Singha, M., & Hemachandran, K. (2012). Content based image retrieval using color and texture. *Signal & Image Processing: An International Journal (SIPIJ)*,*3*(1), 39-57.

[16] de Avila, S. E. F., & Lopes, A. P. B. (2011). VSUMM: A mechanism designed to produce static video summaries and a novel evaluation method. *Pattern Recognition Letters*, *32*(1), 56-68.

[17] Furini, M., Geraci, F., Montangero, M., & Pellegrini, M. (2010). STIMO: STIll and MOving video storyboard for the web scenario. *Multimedia Tools and Applications*, *46*(1), 47-69.

[18] Li, Y., Merialdo, B., Rouvier, M., & Linares, G. (2011, November). Static and dynamic video summaries. In *Proceedings of the 19th ACM international conference on Multimedia* (pp.1573-1576)ACM

[19] Iparraguirre, J., & Delrieux, C. (2013, December). Speeded-Up Video Summarization Based on Local Features. In *Multimedia (ISM), 2013 IEEE International Symposium on* (pp. 370-373). IEEE.

[20] Stanković, R. S., & Falkowski, B. J. (2003). The Haar wavelet transform: its status and achievements. *Computers & Electrical Engineering*, *29*(1), 25-44.

[21] Yang, X., & Cui, W. (2008, December). A novel spatial clustering algorithm based on delaunay triangulation. In *International Conference on Earth Observation Data Processing and Analysis* (pp. 728530-728530). International Society for Optics and Photonics.

[22] Sander, J., Ester, M., Kriegel, H. P., & Xu, X. (1998). Density-based clustering in spatial databases: The algorithm gdbscan and its applications. *Data mining and knowledge discovery*, *2*(2), 169-194.