Video Summarization Techniques: A Review

Hafiz Burhan Ul Haq, M. Asif, Maaz Bin Ahmad

Abstract: With the advancement of technology, the enormous amount of video data is generated on the daily basis. The processing of such huge chunky videos requires high storage, high computational processing power, and consumes a lot of time. Extraction of features from the video is a time-consuming task because the user has to watch the entire video. A large number of editing tools exist that require expertise that is highly expensive. The video summarization is used to overcome these issues that deal with lengthy videos and condense those, based on the various features. This paper presents a review and comparative analysis of video summarization techniques. Initially, these methods have been classified into different categories based on their characteristics. After that, the discussion is made related to the domain directions, applications, pros/cons, and challenges for existing video summarization approaches.

Index Terms: Analysis, Video Summarization, Trajectory, Sparse dictionary, Comparison.

1 INTRODUCTION

In this era, the digital camera plays a vital role in numerous fields including security surveillance, entertainment, event management, news, sports, education, advertisement, etc. Several kinds of security surveillance cameras (i.e. static and moveable) are installed at public places, homes, shops, airports, and banks, etc. that generate data 24/7 [1]. The analysis of video contents for retrieving useful or interesting information from the video is a complicated and time-consuming task. The reason is that users need to watch the whole video for extracting meaningful information. Similarly, the manipulation and management of such type of large data are far from the text of images due to the temporal nature of the video. Therefore, the recognition of anomalous or suspicious/ abnormal behavior from high dimensional data is also quite difficult and time-consuming.

Moreover, it demands high storage space and processing power along with human interaction and concentration [2]. To address this problem, the concept of video summarization is introduced. However, video summarization is defined as the generation of a summary of extensive video content by detecting and presenting the material to potential users which is most informative and contains interesting information [3]. In these days, video summarization is used for several purposes. The following are its main applications:

1.1 Summarization of personal videos

Video sharing is a common process in the family, where people share the video at different occasions like marriage, tourism, etc. but watching such type of lengthy videos are vague tasks. The automatic video summarization of personalized is very efficient for generating an interesting moment from a lengthy video in a short time [4].

1.2 Sport highlights

The sports highlights are produced through manual methods that require expertise to edit such type of video by using any type of video editing tool. On the other hand, video summarization automatically detects an interesting moment from a video based on knowledge given to it regarding specific sports [5].

1.3 The trailer of automated movies and serial

Video summarization is also useful to generate the trailers of movies and serials to convince a large number of the viewer to watch the movies or serials more interestingly [2].

1.4 Video search engines

The different search engines are used containing the themes of the video in the form of a short form of video clips for producing the user's interest [6].

2 CLASSIFICATION OF VIDEO SUMMARIZATION TECHNIQUES

Multiple techniques/approaches have been developed which have the main concern to precise the video contents and generate a video summary. These techniques are classified into four major categories based on their properties and characteristics. The classification of these techniques is given below in Fig.1.

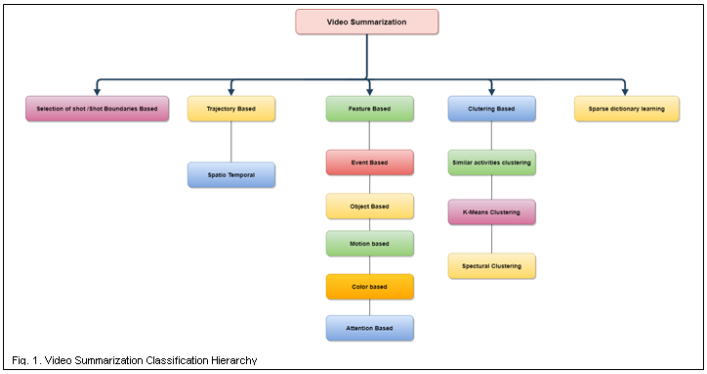

Fig. 1. Video Summarization Classification Hierarchy

- *Hafiz Burhan Ul Haq, Department of Computer Sciences, Lahore Garrison University, Lahore, 75500, Pakistan. E-mail: burhanhashmi64@lgu.edu.pk*
- *M.Asif, Department of Computer Sciences, Lahore Garrison University, Lahore, 75500, Pakistan. E-mail: drmuhammadasif@lgu.edu.pk*
- *Maaz Bin AhmadDepartment of Computer Sciences, PAF Karachi Institute of Economics and Technology, Karachi, 75500, Pakistan. E-mail :*

2.1 Video summarization based on Features

As video contains the information related to different features based on an object, event, color, and motion, etc. These features are very useful for the representation of the video and work efficiently. If the users focus on the video features, for instance, to recognize the objects from video by visualizing color features or by detecting activities/objects from the scenes, then user adopts various summarization techniques based on these desired features. Several feature-based video summarization techniques existed such as an event, object, color, motion, and attention-based techniques as mentioned in Fig. 2. The descriptions of these techniques are given below:

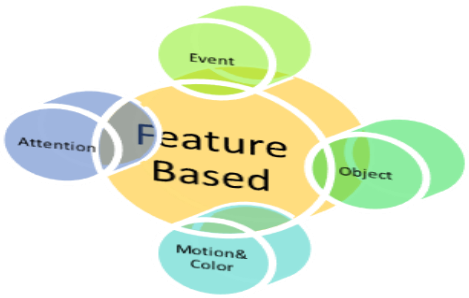

Fig. 2. Classification of Features based Approaches

2.1.1 Event-based video summarization:

Event-based approaches are useful for the identification of normal as well as abnormal events, existing in videos. For example, tracking and recognition of sudden changes appearing in the environment like robbery scenes, Mobile snatching, terrorism, etc. by using some detection models to observe suspicious/abnormal features. The video summarization algorithm is applied that consolidates all the frames having abnormal scenes to construct a video summary. P. Kalaivani & S. M. M. Roomi [7] described the event detection along-with the methods that are useful for event detection and constructing summaries of the video. Kumar et al [8] improved the selection of key frame accuracy by using the Bootstrap AGGregatING approach. Damnjanovic et al. [9] proposed an event-based video summarization technique. In this technique, initially, the energy of each frame is calculated by taking the sum of the absolute difference between the current frame and the reference frame pixel values. After that, all of the frames in which events existed are identified. Finally, the video summarization algorithm is used for these frames to create a video summary. Thomas et al. [10] described the perceptual video summarization. The researchers introduced the Human Visual System (HVS) that percepts the significant events and removes redundancy from the videos to generate summarized videos. Cricri et al. [11] described the method of event detection videos that detects an interesting event in sports by using a magnetometer. Therefore, their method takes clear camera shots, and then the joint analysis is performed to reduce computational complexity and false positive. Andaloussi et al [12] described the methods of précising video based on goal detection, score-box, and audio features. Cote et al. [13] discussed a method that detects the abnormal/anomalous/suspicious behavior during the exam by using the Hidden Markov model. The behavior is observed based on the student's head poses and considered abnormal by computing the probability with normal behavior. Kumar et al. [14] presented a method of multi-view event summarization, in which comparisons of video sequences based on DNA structures is performed. After that frames are divided into NE, SH, SE, SV that show the categories of event. Finally, less important events are removed with the help of object tracking and generates the summary. Events detection techniques perform better where the background is static. However, these approaches are limited in case of changing background.

2.1.2 Motion and Color Based video Summarization.

Motion-based video summarization is difficult, especially when the camera is involved. Initially, Wolf [15] introduced the extraction of key frame with the help of motion-based key frames. Later on, Bulut et al [16] also performed work on frame extraction based on motion. Furthermore, Li et al. [17] discussed the technique that works on constant speed motion and generating a summary with relative motion. He also described the emotional and spatial information for analysis. Ajmal et al. [18] described a method in which human motion can be tracked through the Kalman filter and the obtained trajectory. The color features are also a useful aspect for video, in which mostly color histogram is used for shots detection. Almeida et al. [19] presented the method of the video summarization with the help of a color histogram. However, in this research, the color histogram selects the most representative frame to get a unique video summary. Zhang et al. [20] also described the technique, where key frames selection is performed based on texture and color properties. Initially, the first frame in each shot is selected as key frame and the color histogram is used for the identification of further key frames that are combined to produce a video summary. Omar et al [21], describe the deep learning method for detection human action identification with the help of Histogram of the oriented gradient (HOG) and Temporal Difference Map (TDMap). However, in this method is extracted the sequences into the number of shots in order to predict similar shots and similarity among the frames. For experimental analysis, MHAD dataset has been used in this paper, that contains unusual action of humans. The motion-based approaches are beneficial for medium-level motion existing in the video, but these are limited to work for huge motion or no motion in the video. Similarly, color-based techniques are simple but having low accuracy.

2.1.3 Object-Based video summarization

The object-based techniques are helpful in the detection of an object from the video, like person, car, cat, etc. In this method, video summarization can be done by collecting all frames from the video having the desired object. However, if the video does not contain any type of the desired object, then this method will not perform efficiently. Feng and Chong-Wah [22] described a method of summarizing rushes videos based on objects and events in therefore Hierarchical Hidden Markov Model is used for generating video summary. Initially, HHMM identifies motion features and then semantically categorized the low-level features to a high level and that features of the object are identified to generate a video summary by selecting the most representative frame from the video. Lee et al. [23] discussed a method in which the appearance and disappearance of the object can be detected with varying video context by the regression model. M. Miniakhmetova &

M. Zymbler [24] described the method of summarizing personal videos based on objects. This method works in two stages. First, to structures the video by using various scenes detection techniques. Second, to generates a video summary in the form of a subset of video scenes. The object detection is performed on a subset of the video scene at the second stage. Finally, the Resultant video summary is generated that contains only the shots/scenes having objects. Rav-Acha et al [25] also presented the synopsis of the video, where all activities and events of the video are combined to produce a video summary. This work can be done by directly detecting the objects that are in moving state and then optimization is performed on video based on detected objects. K. Davila & R. Zanibbi [26] described the selection of frames in lecture videos, based on segmentation by diminishing the conflicts between removing objects, content regions and reconstruct each frame to produce a summarized video. Neeraj et al. [27] describe a query-conditioned method for object-based video summarization. In this method, a mathematical is also used that reduced redundancy depending on the loss function and summary variance and identifying the score in order to generate a video summary Object-based Query Image (OQI) dataset is used and achieved 57.06% F1-score. Object detection is applicable for both surveillance as well as non-surveillance video, but these approaches not much better performed in rushes videos and tend to miss more objects.

2.1.4 Attention-based video summarization

The attention-based models are used to identify the region of the user's interest. In this method, the system assigns the score to the key frames or shots that exist in an interesting region. These frames have high scores and are considered as important to produce a video summary. Ngo et al. [28] presented the approach for summarization that relies on both the perceptual quality and content balance of the video summary. Therefore a clustering method is applied to edit/cut the video, and a motion attention model is used to measure the shot's quality and clusters. The temporal graph is created with the help of both techniques that prescribes the cluster's importance. Finally, this temporal graph and is used for grouping the scenes from clusters. Its attention values are used to select the appropriate scenes to produce a video summary. Nahian et al. [29] described the prediction method for the importance of the shot by using CNN. Xu et al. [30] presented a method of optimization depends on RCNN, cosine similarity, and constraint sub-modular maximization to summarize the egocentric videos. Srinivas et al. [31] also discussed the method of video summarization because retrieving such type of extensive data from the internet is complicated. However, the video summarization can be performed by the computation of three-factor. First, scoring each frame based on various features such as quality, color, hue, attention. The second step is weighting each frame to get key frame. Finally, to eliminate the redundant frame to produce a video summary. 2.2 Clustering-based Video summarization. The clustering method is used to find similar activities or properties within a frame. However, it is also very useful to eliminate irregular frames. The clustering method summarized the video in a short duration as compared to other methods discussed earlier. Clustering-based Video summarization is classified into similar activities such as K-means, and spectral clustering.

2.2.1 Similar activities clustering

Pritch et al. [32] discussed that video can be précised by presenting the common or specific activities which existed in different scenes of the video. In their method activities are defined at the start that existed in multiple frames, that first extract the activity features by comparing tublets, where the tublets are the subpart of activity. The second, to determine the similarity and distance between the activities of the clusters. Finally, to generate a video summary of the desired cluster. The irregular activities can be easily detected with the help of this approach, but false predictions may lead to the false summary. Shruti and Mahmood [33], describe the deep learning method of video summarization with help of the clustering method by performing key frames extractions. The SumMe dataset is used for experimental analysis and also indicated that deep learning-based frame extraction performs better in dynamic videos. Otani et al. [34] also designed a deep neural network for video summarization. In this method extraction of deep features is performed from each segment and then used Clustering-Based summarization in order to summarize the video. Vrushali and Reena [35] also describe different video summarization-based features, key frame, and clustering methods that are useful in the IoT environment.

2.2.2 K-means clustering

Chen et al. [36] described a method that uses the k-means algorithm with a histogram. This mechanism firstly divides the input file into many segments, where the first frame of each segment is considered as a representative. However, these frames can be computed by the histogram. Finally, the histogram can be clustered by k-means algorithm. In this way, the required segments are collected in the form of a video summary. Mohan and Nair [37] describe the static method of video summarization, which is used for describing input videos with the help of interesting frames subsets. In this method HOG descriptors of Gabor are used for mapping the input video and finally, video summary can be generated with the help of candidate frames that are obtained by K means clustering. K-means clustering approach is not suitable where the scenes are static, because it causes redundancy.

2.2.3 Spectral Clustering

Peker, K., & Bashir, F. [38] described the human face detection along with spectral clustering. In these techniques, the human face can be detected by the computation of the number of faces, size, and location. The spectral method is used for clustering the desired frame. This technique is not suitable for detecting multiple faces. Chasanis et al. [39] described the algorithm in which the spectral clustering method and shot boundary detection are used for generating a video summary. Damnjanovic et al. [9] also used the Spectral Clustering algorithm to summarize the video based on detected events. The spectral clustering is not suitable for detecting multiple faces because this method only detects the human face.

2.3 Selection of shot /shot boundaries based video summarization.

This method extracts the key frames from the video, where extraction of the first image is considered as a shot key frame. However, this method is suitable and applicable just only for the variation of small content [40]. Generally, a shot is considered as the main component of the video, which is continuously used concerning different scenes and time. For

example, during the shooting of the movie and serial, when the camera moved from one actor to another, then it contains two shots. In this way multiple shots are collected, so that the selection of key-frames from the shot is a complicated task [41]. Uchihachi et al [42] described the packing algorithm to determine the superlative layout. Therefore, their algorithm packs the frame to have a best sequences containing that block. The frame packing algorithm is applied that organizes different shot and generate video in precise form. Lu, Z., & Grauman, K. [43] presented the method in which the importance of each shot can be estimated by sub shot of the original video sequence. Lu et al [44] also described the measurement of relative importance each shot of videos. Varghese, J., & Nair, K. R [38] described the method of video summarization based on detecting shot boundaries and the elimination of repetitive frames by Structural Similarity Index (SSI) and Stroboscopic effect. Ma et al. [45] used the collaborative representation of adjacent frame technique for detecting the abnormal and noisy data from the video. For removing abnormal data, minimum sparse reconstruction is used for the selection of key frames. Moreover, a greedy iterative algorithm is applied to optimize the model that controls the number of key frame by using the average percentage of reconstruction (APOR) and sparse boundary. Wang et al, [46] presented a modality correlation method for video summarization. This method performs three major tasks where first identify the correlation, and then fuse two-correlations to get each shot score. Finally, shots having a high score are considered for the video summary. Muhammad et al [47], proposed an efficient video summarization method for surveillance videos by Convolutional Neural Network (CNN). This method learns deep features using shot segmentation. Furthermore, this framework generates a video summary with the help of entropy and memorability. The shot, having a high score of memorability and entropy is taken as a keyframe.

2.4 Trajectory-based Video Summarization

This technique works efficiently when video data is used for analyzing the dynamic environment. It is used in surveillance applications for the detection of objects, where the fixed cameras are used along with a static background. The Trajectory method is also useful for tracking the object visited places/location. Therefore, it is widely used in surveillance videos to locate the objects and remove uncertainty [2]. A trajectory-based video summarization technique is discussed below:

2.4.1 Spatio-temporal /Spatial-temporal segments

The objects that are in moving state can be detected by three-dimension, where x and y dimensions are spatial and t is a time dimension. Therefore the video can be produced by the extraction and analysis of trajectories where the node is identified and considered as a critical point in the video. Lai et al. [49] described a method where clustering algorithm, optical flow, and background-subtracted are adopted for object detection. After object detection, a sliding window is used that combines the detected objects in consecutive frames to construct a Spatio-temporal trajectory. Finally, to combine all the Spatio-temporal trajectories to generate a video summary. These techniques are not much useful where the camera in motion.

2.5 Sparse dictionary learning:

This technique belongs to an unsupervised learning method that helps to create data logs and monitor the changes concerning time. However this technique is still time-consuming, but most probably in the future, it will be helpful for the creation of summaries depending on dynamic summarization. Li et al [50] presented a method that constructs a videography dictionary to show every video as a sequence of words for uncompelled videos. However, clustering and shot boundary techniques are used for dictionary formation. Mei et al [51], described a sparse representation-based video summarization that classifies the frame and features of the video. First, each frame of the video is divided into a number of patches with the help of features by using conventional sparse representation and then generated video summary after getting unique frames with the help of a simultaneous block version of block-based Orthogonal Matching Pursuit (SBOMP) algorithm. Ma et al [52], also used SBOMP for video optimization, in this method two datasets are used for experimental analysis such as VSumm and TVSum. Table 1 describes the brief overview of the video summarization techniques method.

TABLE 1

OVERVIEW OF VIDEO SUMMARIZATION LATEST TECHNIQUES

Authors	Approach/Methodology	Applications	Dataset	Remarks/Pros /Cons
Varghese and Nair (2015).	Summarized video by using SSI and Stroboscopic effect.	All kind of videos	Random Videos	55% reduce the volume of the original video
Miniakhmetova and Zymbler (2015).	User preference analysis for summarizing personalized visual data.	Personalized videos	Personal videos Database (PVD)	By tracing previous actions video summarizes are generated with the help of user- preferences estimations
Srinivas et al. (2016)	A Rank-based Approach is used for scoring, ranking and eliminating the frames	All kind of videos	Open Video Project	Frame Blocks Features Method improved by 1.8 %.
Almeida et al. (2010).	Video summarization by macroblocks and color histogram	Compressed Videos	Open Video Project	Poorly detects the object in lengthy video and the model is hardware dependent that having high computational power/speed
Wang and Feng (2019).	Two-level Hierarchical Hidden Markov Model (HHMM) Précising rushes video-based events, objects, and motions	All kind of rushes videos	TRECVID BCC video datasets	Ignoring events, that do not motion insensitive
Chasanis et al. (2008).	Video Summarization by shot boundary detection and spectral clustering	All kind of videos	TV-series/Sports	Over-segmentation occur in case of continuously changing scenes and

				shots.
Nahian et al. (2017)	Predicting the importance of frame shots based on CNN.	All kind of video-based application	TVSum50	Discussed method work efficiently based on existing feature methods in terms of F1 measure, absolute error variance and Mean Absolute-Error.

3. DOMAIN-BASED ANALYSIS FOR VIDEO SUMMARIZATION TECHNIQUES

In this research several types of techniques have been discussed, that belong to various types of domains. However, these domains utilize the approaches of video summarization based on their desired features. Therefore a brief analysis has been presented that classifies these domains into different categories. The classification of these domains based on their characteristics and methodology as shown in Table 2.

TABLE 2
CLASSIFICATION OF DOMAINS WITH DIFFERENT SUMMARIZATION TECHNIQUES

Domain	Useful/Desired Approaches of Summarization	Worked in literature
Movies/Serial Clips	Event, Motion, Objects, Attention, and Shot detection based Approaches.	[7],[8],[9],[10],[11],[12],[13],[14,[15],[16],[17],[20],[21],[22],[23],[24], [25],[26],[27],[28],[33],[34],[35],[36],[37], [38], [39].
Sports Video	Event, Motion, Color, Objects, Attention, and Clustering-based Approaches.	[7],[8],[9],[10],[11],[12],[13],[14],[15],[16],[17],[18],[19[20],[21],[22],[23],[24], [25],[26],[27],[28],[29],[30],[32], [33].
Personal Videos	Event, Color, Objects, and Attention-based Approaches.	[7],[8],[9],[10],[11],[12],[13],[18],[10],[20],[21],[22],[23],[24],[25],[26],[27],[28].
News Highlights	Event and Attention-based Approaches	[7],[8],[9,][10],[11],[12],[13],[25],[26],[27],[28].
Rushes Videos	Event, Motion, Objects, and Attention-based Approaches	[7],[8],[9],[10],[11],[12],[13],[14],[15],[16],[17],[20],[21],[22],[23],[24],[25], [26], [27], [28].
Surveillance Videos	Events, Motion, Color, Objects, Attention, Clustering, and Trajectory based Approaches.	[7],[8],[9],[10],[11],[12],[13],[14],[15],[16],[17],[18],[19],[20],[21],[22], [23],[24],[25],[26],[27],[28],[29],[30],[31],[32],[33],[40]

Different approaches have been mentioned that are useful for summarizing the videos. These techniques belong to different areas/domains for video summarization. The graphical representation of these approaches regarding specific domains and the number of papers that fall in a specific category are mentioned in Fig 3.

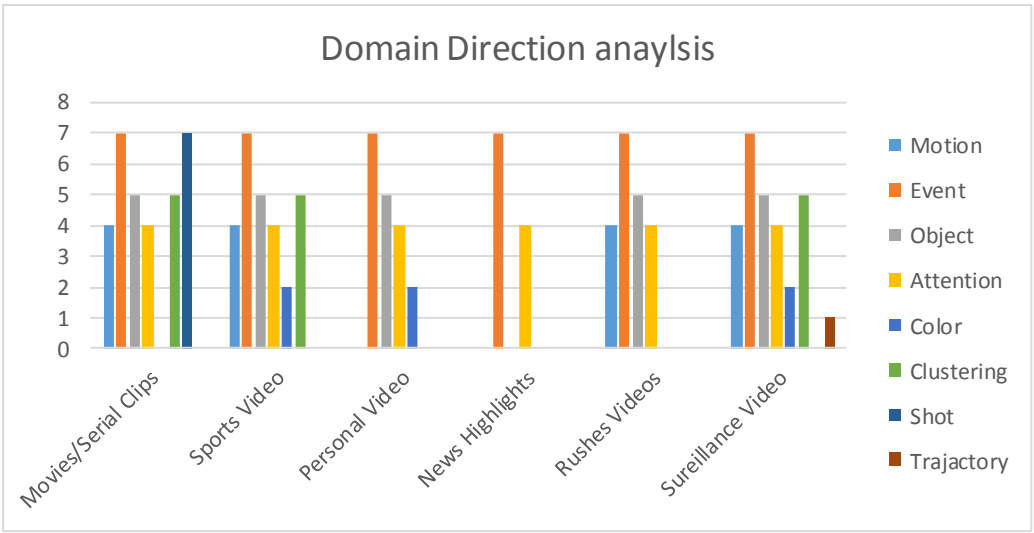

***Fig. 3**. Analysis based on domain direction*

In video summarization, the process of précising the original video content is important. During the summarization, there is the possibility of different problems faced by the user. These challenges/problems are highlighted in this research:

- Processing unbalance length of a dataset.
- In-expensiveness of training video.
- Subjectivity (Variation in the selection of shot)
- Redundancy (Minor difference of frames is considered as a key frame).
- Highly changes in content make clustering difficult.
- The complication in the development of an application.
- Integration of application and model.

5. RELATED WORK

Research related Video summarization is significantly advanced in the last few years. Therefore, various types of approaches have been developed that work on video summarization. Different surveys have been discussed here that cover the various type of video summarization approaches based on different features. Tinumol S & Jiby JP [53] described the survey of different methods that are useful for the extraction of key-frames for video summarization. Khan, S., & Pawar, S. [54] Presented the survey of summarization of soccer videos and the detection of events. Haq et al. [55] also discussed a variety of tools that are useful for video summarization like Python, Matlab, etc. Patil, S., & Phalke, D. [56] also described the survey of the video summarization method along with approaches performance. Workie et al. [57] also provided a review of video summarization approaches, techniques, and applications. Table 3 showing the comparative analysis of this research with the existing researches based on different factors. However including factors are video summarization techniques, domain direction approaches findings, Pros, Cons, Dataset, and challenges. The factor which is discussed in the particular research paper is represented by 1 and the factor which is not discussed in the particular research paper is represented by 0.Table 3 showing the comparative analysis of this research work with the existing researches based on different factors. These factors are: type of video summarization techniques, domain direction approaches findings, Pros, Cons, Dataset, and challenges. The factor which is discussed in the particular research paper is represented by 1 and the factor which is not discussed in the particular research paper is represented by 0. The total score is computed for each research work based on targeted factors. The comparative analysis shows this research work perceptively covers most of the factors, while the rest of the research work targeted only limited factors. Most of the researches have not covered the challenges, pros, and cons related to any specific technique. The information about video summarization applications and Datasets are not significantly discussed so far in the literature. Finally, domain direction, that specifies the user's attention regarding usage of any techniques for any domain is also not discussed in most of the researches.

TABLE 3

COMPARATIVE ANALYSIS OF VIDEO SUMMARIZATION METHODS BASED WITH EXISTING RESEARCHES

Factors Covered	Bora and Sharma (2018).	Murugan et al. (2018).	Khan and Pawar (2015).	Patil, and Phalke (2019).	Workie et al. (2020).	This Paper
Domain Direction	0	0	1	0	1	1
Motion Based	0	1	1	1	0	1
Event Based	1	1	1	1	1	1
Attention Based	1	0	0	1	1	1
Object Based	1	1	1	1	1	1
Color Based	0	1	1	1	0	1
Clustering Based Approaches	0	0	1	1	0	1
Shot Detection Based Approaches	0	1	1	1	1	1
Trajectory based Approaches	1	0	0	0	0	1
Sparse dictionary learning:	1	0	0	0	0	1
Applications	1	0	0	0	1	1
Datasets	1	1	0	0	0	1
Total Score	7	7	7	7	7	14

6. CONCLUSION

This research work conducted highlights a brief review of video summarization, classifications, and methods. Various techniques have been discussed and compared, to guide the user in the selection of the most suitable technique. The two analyses are performed in this paper. The first analysis is based on approaches, applications, datasets, and pros/cons/findings. Resultantly, some of the feature-based approaches are poorly performed for object detection in lengthy videos and require high hardware specifications. The clustering-based methods summarized the video more precisely as compared to other methods used. Moreover, a trajectory method is suitable for a dynamic environment such as surveillance videos where the camera is fixed. The second analysis is based on various factors as discussed in this paper. The analysis shows that other studies discussed only the applications or pros/cons rather than reporting all aspects that are related to précising video. This research also highlighted the domain's direction to describe the behavior/usefulness of these techniques regarding specific domain for video summarization. Current work will assist the users in the selection of specific techniques regarding the desired domain.

REFERENCES

[1] A. S. Murugan, K. S. Devi., A. Sivaranjani, P. Srinivasan, “A study on various methods used for video summarization and moving object detection for video surveillance applications”, Multimedia Tools and Applications, 77(18), pp. 23273-23290,2018.

[2] A. Bora, S. Sharma, “A Review on Video Summarization Approcahes: Recent Advances and Directions”. In 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), pp. 601-606,2018.

[3] R. Kansagara, D. Thakore, M. Joshi, “A study on video summarization techniques”. International journal of innovative research in computer and communication engineering, 2(2014).

[4] K. Darabi, G. Ghinea, “Personalized video summarization by highest quality frames”. In 2014 IEEE International Conference on Multimedia and Expo Workshops (ICMEW) (pp. 1-6), 2014.

[5] Z. Zhao, S. Jiang, Q. Huang, G. Zhu, “Highlight summarization in sports video based on replay detection”. In 2006 IEEE international conference on multimedia and expo, pp. 1613-1616, 2006.

[6] G. Irie, T. Satou, A. Kojima, T. Yamasaki, K. Aizawa, “Automatic trailer generation”. In Proceedings of the 18th ACM international conference on Multimedia, pp. 839-842, 2010.

[7] P. Kalaivani , S. M. M. Roomi, Towards Comprehensive Understanding of Event Detection and Video Summarization Approaches”. In Recent Trends and Challenges in ComputationalModels (ICRTCCM), 2017 Second International Conference, IEEE (2017, February): 61-66

[8] K. Kumar, D. D Shrimankar. & N. Singh. “Event BAGGING: A novelevent summarization approach in multiview surveillance videos”.In Innovations in Electronics, Signal Processing and Communication(IESC), 2017 International Conference. pp: 106-111, 2017.

[9] U. Damnjanovic, V. Fernandez, E. Izquierdo, “Event Detection and Clustering for Surveillance Video Summarization”. In: Proceedings of the Ninth International Workshop on Image Analysis for Multimedia Interactive Services. IEEE Computer Society, Washington, USA (2008).

[10] S. S. Thomas, S. Gupta, V. K. Subramanian, Perceptual video summarization—“A new framework for video summarization”. IEEE Transactions on Circuits and Systems for Video Technology, 27(8), pp. 1790-1802, 2016.

[11] F. Cricri, S. Mate, I. D. Curcio & M. Gabbouj. “Salient event detection in basketball mobile videos”. In Multimedia (ISM), 2014 IEEE International Symposium. IEEE pp. 63-70, 2014.

[12] S.J. Andaloussi, A. Mohamed, N. Madrane & A. Sekkaki. “Soccer video summarization using video content analysis and social media streams”. In Proceedings of the 2014 IEEE/ACM International Symposium on Big Data Computing, IEEE Computer Society,pp.1-7, 2014.

[13] M. Cote, F. Jean, A. B. Albu & D. Capson. “Video summarization for remote invigilation of online exams”. In Applications of Computer Vision (WACV), 2016 IEEE Winter Conference, pp. 1-9, 2016.

[14] K. Kumar & D. D. Shrimankar. “F-DES: Fast and Deep Event Summarization”. IEEE Transactions on Multimedia, 20(2), pp. 323- 334, 2018.

[15] W. Wolf, Key frame selection by motion analysis. In: ICASSP, vol. 2, (1996) pp. 1228– 1231.

[16] E. Bulut, T. Capin, “Key Frame Extraction from Motion Capture Data by Curve Saliency”. In: Proceedings of 20th Annual Conference on Computer Animation and Social Agents, Belgium (2007).

[17] C. Li, Y.T. Wu, S.S. Yu, T. Chen, “Motion-focusing key frame extraction and video summarization for lane surveillance system”. In: 16th IEEE International Conference on Image Processing (ICIP), pp. 7–10, 2009.

[18] M. Ajmal, M. Naseer, F. Ahmad, A. Saleem. “Human Motion Trajectory Analysis Based Video Summarization”. 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA) (2017):550-55510.1109/ICMLA.pp. 0-203, 2017.

[19] J. Almeida, R. D. S. Torres, N. J. Leite, “Rapid video summarization on compressed video”. In 2010 IEEE International Symposium on Multimedia , pp. 113-120,2010.

[20] H.J. Zhang, J. Wu, D. Zhong, S.W. Smoliar, “An integrated system for content based video retrieval and browsing”. Pattern Recognition 30, pp.643–658, 1997.

[21] O. Elharrouss, N Almaadeed, S. Al-Maadeed, A. Bouridane, and A. Beghdadi,. “A combined multiple action recognition and summarization for surveillance video sequences”. Applied Intelligence, pp. 1-23, 2020

[22] F. Wang, C.W. Ngo, “Summarizing rushes videos by motion, object and event understanding”. IEEE Transactions on Multimedia 14 (2012).

[23] Y.J. Lee, J. Ghosh, K. Grauman, “Discovering Important People and Objects for Egocentric Video Summarization”. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR (2012).

[24] M. Miniakhmetova & M. Zymbler, “An approach to personalized video summarization based on user preferences analysis”. In Application of Information and

Communication Technologies (AICT), 2015 9th International Conference, pp.153-155, 2015.
[25] A. A. Rav, Y. Pritch, S. Peleg, Making a long video short: Dynamic video synopsis. In 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), Vol. 1, pp. 435-441, 2006.
[26] K. Davila , R. Zanibbi. "Whiteboard Video Summarization via SpatioTemporal Conflict Minimization". In Document Analysis and Recognition (ICDAR), 2017 14th IAPR International Conference, vol 1, pp. 355-362, 2017.
[27] N. Baghel, S. C. Raikwar and C. Bhatnagar. 'Image Conditioned Keyframe-Based Video Summarization Using Object Detection". 2020 arXiv preprint arXiv:2009.05269.
[28] C. W. Ngo, Y. F. Ma, H. J. Zhang, "Video summarization and scene detection by graph modeling". IEEE Transactions on circuits and systems for video technology, 15(2) pp. 296-305, 2005.
[29] M. A. Nahian, A. S. M. Iftekhar , M. T. Islam, S. M. Rahman, D. Hatzinakos. "CNN-Based Prediction of Frame-Level Shot Importance for Video Summarization", arXiv preprint arXiv:1708.07023, 2017
[30] J. Xu, L. Mukherjee, Y. Li, J. Warner, J. M. Rehg ,V. Singh. "Gazeenabled egocentric video summarization via constrained submodular maximization". In Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference, pp. 2235-2244, 2015.
[31] M. Srinivas, M. M. Pai, R. M. Pai, "An Improved Algorithm for Video Summarization–A Rank Based Approach". Procedia Computer Science, 89, pp.812-819, 2016.
[32] Y. Pritch, S. Ratovitch, A. Hendel, S. Peleg, "Clustered synopsis of surveillance video". In: 6th IEEE Int Conf. on Advance Video and Signal Base Selection (AVSS 2009), Genoa, Italy, pp. 2–4, 2019.
[33] S. Jadon, and M. Jasim, "Video summarization using keyframe extraction and video skimming". 2019. arXiv preprint arXiv:1910.04792.
[34] M. Otani, Y. Nakashima, E. Rahtu, J. Heikkilä, and N. Yokoya. "Video summarization using deep semantic features. In Asian Conference on Computer Vision, pp. 361-377, 2016.
[35] V. Raut, and R. Gunjan, "Video summarization approaches in wireless capsule endoscopy: A review". In E3S Web of Conferences, Vol. 170, pp. 03005, 2020.
[36] F. Chen, M. Cooper, J. Adcock, "Video Summarization Preserving Dynamic Content". In: Proceedings of the International Workshop on TRECVID Video Summarization, 2007.
[37] J. Mohan, and M. S. Nair, "Domain independent static video summarization using sparse autoencoders and K-means clustering". Journal of Intelligent & Fuzzy Systems, 36(3), pp. 1945-1955, 2019.
[38] K.A. Peker, F.I. Bashir, "Content-Based Video Summarization using SpectralClustering". Mitsubishi Electric Research Laboratories Cambridge, MA. Universityof Illinois at Chicago, Chicago, IL (2009).
[39] V.T Chasanis, C.L Likas, N.P Galatsanos,"Scene detection in videos using shot clustering and sequence alignment". IEEE Trans. Multimed. 11(1), pp. 89–100, 2009.
[40] W. Sabbar, A. Chergui, A. Bekkhoucha, 'Video summarization using shot segmentation and local motion estimation", Innovative Computing Technology (INTECH), 2012 Second International Conference. pp.190- 193, 2012.
[41] G. Evangelopoulos, K. Rapantzikos, A. Potamianos, P. Maragos, A. Zlatintsi, Y. Avrithis, "Movie Summarization Based on Audio-Visual Valiency Detection". In: IEEE Intl Conf. Image Processing (ICIP), San Diego, CA,2008.
[42] S. Uchihachi, J. Foote, L. Wilcox, "Automatic Video Summarization Using a Measure of Shot Importance and a Frame Packing Method". United States Patent 6, pp.535-639, Mar.2003.
[43] Z. Lu, K. Grauman, "Story-Driven Summarization for Egocentric Video. 2013 IEEE Conference on Computer Vision and Pattern" Recognition, pp. 2714–2721, 2013. http://doi.org/10.1109/CVPR.2013.350.
[44] S. Lu, M. R. Lyu, I. King, H. K. Sar, "Semantic Video Summarization Using Mutual Reinforcement Principle and Shot Arrangement Patterns", 2005.
[45] J. Varghese, K. R. Nair, "An Algorithmic Approach for General Video Summarization". In 2015 Fifth International Conference on Advances in Computing and Communications (ICACC), pp. 7-11, 2015.
[46] X. Wang, X. Nie, X. Liu, B. Wang, and Y. Yin, "Modality correlation-based video summarization. Multimedia Tools and Applications", pp. 1-16, 2020.
[47] K. Muhammad, T. Hussain, and S.W. Baik, "Efficient CNN based summarization of surveillance videos for resource-constrained devices". Pattern Recognition Letters, 130, pp. 370-375, 2020.
[48] M. Ma, S. Mei, S. Wan, Z. Wang, D. D. Feng. "Robust video summarization using collaborative representation of adjacent frames". Multimedia Tools and Applications, 78(20), pp.28985-29005, 2019.
[49] P. K. Lai, M. Décombas, K. Moutet, R. Laganière, "Video summarization of surveillance cameras". In 2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 286-294, 2016.
[50] K. Li, S. Li, S. Oh & Y. Fu. "Videography-Based Unconstrained Video Analysis". IEEE Transactions on Image Processing, 26(5), pp. 2261-2273, 2017.
[51] S. Mei, M. Ma, S. Wan, J. Hou, Z. Wang, D.D. Feng, "Patch based Video Summarization with Block Sparse Representation". IEEE Transactions on Multimedia, 2020.
[52] M. Ma, S. Mei, S. Wan, J. Hou, Z. Wang, and D.D. Feng, "Video summarization via block sparse dictionary selection". Neurocomputing, pp. 378, 197-209, 2020.
[53] S. Tinumol, J.P. Jiby, "A survey on video summarization techniques". Int J Comput Appl 132(13) pp.31–33, 2015.
[54] S. Khan, S. Pawar. "Video summarization: survey on event detection and summarization in soccer videos". International Journal of Advanced Computer Science and Applications, 6(11), 2015.
[55] H. B. U. Haq, H. U. R. Kiyani, S. K. Toor, S. Zafar, I. Khalid. The Popular Tools Of Data Sciences: Benefits, Challenges and Applications. IJCSNS International Journal of Computer Science and Network Security, VOL.20 No.5, 2020. http://search.ijcsns.org/07_book/html/202005/202005008.html
[56] S. Patil, D. Phalke, International Journal for Research in Applied Science & Engineering Technology (IJRASET). Retrieved from https://www.ijraset.com/fileserve.php?FID=19635, 2019.
[57] A. Workie, R. Sharma, Y. K. Chung. "Digital Video Summarization Techniques: A Survey", 2020.