International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

A SURVEY REPORT ON VIDEO SEQUENCE AND SURVEILLANCE

BALAJI. GUGULOTH
Assistant Professor, Department of Computer Science and Engineering, Siddhartha Institute of Technology and Sciences, Narapally, Hyderabad, Telangana, India

ABSTRACT : In the present situation, the globe is totally under observation to furnish individuals with security and validation. The overhead with video recording is 24*7 checking, storage and recognizing crucial events in the recorded video. Video summarization is the way toward making and displaying the important dynamic video of the whole video in a brief timeframe. Video summarization techniques endeavor to enhance speedier perusing of recordings and effective video retrieval and access. It is the Short representation of an original video content and reduces the storage size. It provides indexed based accessing to the stored video stream. It helps in recovery of huge gathering of video information without losing essential perspectives. This article gives a complete review on different video summarization methods and its application to various continuous frameworks.

Keywords - Video Summarization, Surveillance, Indexing, Video Sequence

I. INTRODUCTION

Now a day because of the time limitation, people are not interested in watching full length news, television shows, sports etc. It would be helpful if the user is provided with a summary of such videos where the maximum information can be appended. The aim of video summarization is to eliminate the redundancies so that the user will get a reduced view of the video. It is a short outline of a unique video content which successfully and productively oversee and store a tremendous measure of information which gives easy to understand access to the put away information. [1] It is utilized as a part of classifying, ordering and recovering. [2] Retaining just the fundamental data in the video grouping enhances storage, transmission capacity and review time. Video summarization is carried out in two different ways.

i.Static Video Summarization

ii.Dynamic Video Summarization.

Static Summarization is also called as video summary. It is a collection of prominent images separated from the hidden video source. It considers only visual information and it does not handle audio and textual information. The advantage of static video summarization is - It is shown more easily since there are no scheduling or synchronization issues and there is no confinement over watching a gathering of key frames. Clients can ready to get a handle on the video content all the more rapidly. The Dynamic video summarization or video skimming is an accumulation of image successions and in addition the relating audio is removed from the original one. That

is it produces short video that represents original video. The advantage of dynamic video summarization is its expressiveness and entertaining due to audio and motion elements.

II. LITERATURE SURVEY

John, Alvina Anna, Binoy B. Nair, and P. N. Kumar [3] prepared a study on the application of clustering techniques for video summarization and evaluated the effectiveness of data mining techniques in video summarization. They assess the video summarization systems based on histogram and entropy features extracted from three different color spaces: RGB, HSV and YCBCR and clustered using K-Means, FCM, GMand SOM on fifty video datasets. The authors projected that clustering based video summarization technique can be effectively used for generating video summaries. Hamza, Rafik, et al [4] proposed their work on Wireless case endoscopy (WCE) since it is more valuable for remote IoT-helped human services administrations. Throughout the WCE technique, tremendous measure of excess video information is produced and the gastroenterologists have to examine and analyze this data. It is additionally troublesome and tedious for gastroenterologists to break down this enormous volume of gastrointestinal video information for desired content. To beat these issues, the authors proposed a secure video outline structure for outdoor patients experiencing WCE methodology. In the proposed framework, keyframes are extricated exploiting a light-weighted video synopsis conspire, making it more reasonable for WCE. Next, a cryptosystem is displayed for security of extricating keyframes in view of 2D Zaslavsky disorganized guide.

Jacob, Hugo, et al. [5] proposed another computational visual consideration model, enlivened on the human visual framework and in sight of computer vision techniques like face identification and movement estimation to assess the accumulation of key frames separated from the first recordings called static video abstracts. They developed a computational model of visual consideration regarding the automatic summarization of computerized recordings from the TV chronicles. The authors emphasize the normal application on video synopses to encourage browsing, indexing and recovery operations. Their approach gives a solution for the issue of automatic video summarization, delivering video outlines with comparative quality to the ground-truth physically made by a gathering of 50 users.

Sahoo, Anurag, et al. [6] addressed their review on searching and automatic summarization in visual information containing recordings, live streams and picture accumulations in an integrated manner. The authors proposed a system for multi-faceted summarization which extricates keyframes, video summaries and element outlines like summarization at the level of elements - things, scenes, people and faces in the video. The client can either see these as extractive outline, or query focused summarization. The authors first

pre-process the video once so that immensely vital visual components are extracted. Then they provide an interactive system for the user to summarize the video based on their option. The authors mainly focused on the clarification capability of various, differences, scope and demonstration models and their adaptability. They also summarized hours of video information in no time flat, and the framework proposed enables the client to produce summaries of different lengths and sorts intuitively on the fly.

Cirne, Marcos Vinicius Mussel, and Helio Pedrini [7] proposed and assessed a novel video summarization approach called VISCOM, which depends on color co-occurrence matrices to represent the video frames and create a summary with the most illustrative one. The authors worked on two data collections of different classes and show the viability of the proposed technique as far as quality.

Vasudevan, Arun Balajee, et al [8] also focused on the issue of automatic video summarization and making a video summary which additionally highlights components related to a search query. The authors tended to this issue by posing queries-relevant summarization which allows them to optimize for summaries which are illustrating the whole video, and related to a test query. In addition to that, they extended the model to hold query-independent properties, for example, quality of the frame. Furthermore, the authors introduced another dataset, interpreting on diversity and query-specific relevance labels. On this dataset, they tried the entire model for video summarization and demonstrate that it beats standard baselines like Maximal Marginal Relevance.The users sometimes may not get enough time or interest to watch the complete video or entire content of the video. In such cases, users may just want to go through the summary of the video instead of watching the entire video. Mishra, Deepak Kumar, and Navjot Singh. [9] Proposed an approach for event summarization in videos based on clustering method which provides a set of key frames as a summary of a video. The key frames which are closer to the cluster heads of the optimal clustering are combined to form the summarized video. The evaluation this model is done on a publicly available data set and compared with ten state-of-the-art models in terms of precision, recall and F-measure.

Li, Xuelong, Bin Zhao, and Xiaoqiang Lu. [10] Developed a common summarization system for both of altered video and basic video summarization. They separated the general work into three modules. In the primary module, four models are intended to catch the properties of video summaries, i.e., containing significant individuals and objects, representative to the video content, no comparable key-shots (differing qualities) and smoothness of the storyline (storyness). In particular, these models are applicable to both altered and raw recordings. In the second one, a comprehensive score function is developed with the weighted combination of the previously mentioned four models. In the

third module, the training set is developed with both altered recordings and raw recordings keeping in mind the end goal to make up the absence of preparing information. Especially, each training video is furnished with a couple of mixing coefficients which can diminish the structure mess in the training set created by the harsh mixture. The system is tested on three datasets, including altered recordings, short raw recordings and long raw recordings. These days, the Surveillance camcorders capture huge volumes of video streams each day. To dissect or explore any noteworthy events, it is a long and troublesome work to recognize these occasions from the immense video information on the off chance that it is done manually. Existing methodologies sometimes ignore key frames with critical visual substance and additionally select some irrelevant frames with low/no movement. To take care of this issue,Salehin, Md Musfequs, and Manoranjan Paul [11] proposed a video summarization system by consolidating three multimodal human visual sensitive components, for example, foreground objects which are the most imperative bits of a video as they contain more point by point data and assume a noteworthy part in critical occasions , the visual attention cue which essentially measure the cue of human attraction model for deciding key edges. As these components separately can't perform extremely well, they are joined to acquire better outcomes. For this reason, a versatile linear weighted combination scheme is proposed to join the components to rank video frames for summarization.

The research work done by Rabbouch, Hana, Foued Saâdaoui, and Rafaa Mraihi. [12] is a real time application of video summarization in transportation.. Since, the advent of computer vision technology, video-based surveillance of road vehicles has become a key component in developing autonomous intelligent transportation systems. In this perspective, the authors proposed a Pattern Recognition system which employs an unsupervised clustering algorithm with the objective of detecting, counting and recognizing a number of dynamic objects crossing a roadway. This strategy defines a virtual sensor to extract from the traffic video streaming a number of signals containing disordered information about the road traffic. Then, the set of signals is filtered with the aim of conserving only motion's significant patterns. The resulted data later processed by a statistical analysis technique so as to estimate and attempt to identify a number of clusters corresponding to vehicles. Finite mixture models fitted by the EM algorithm are used to assess such clusters, which provide important information about the traffic such as the instantaneous number of vehicles, their weights, velocities and intervehicular distances.

Ajmal, Muhammad, et al. [13] classified the video summarization methods based on the methodology used and presented a broad study on important methods in each group including their benefits and drawbacks. The paper also gives a complete description about the condition in which each system is most appropriate to apply.

Mundur, Padmavathi, Yong Rao, and Yelena Yesha [14] projected a system based on the Delaunay Triangulation algorithm which is completely automatic with no user specific parameters and is better suitable for batch processing to cluster the frames in videos. Using this, one can automatically gather the frames of interest in a video for purposes of summarization. The frame contents are represented as multi-dimensional point data and Delaunay Triangulation are used for clustering them. The authors also mentioned that, using Delaunay clusters produce good quality summaries with less frames and low redundancy when compared to other schemes.

Ma, Yu-Fei, et al. [15] developed a standard system of video summarization which depends on the modeling of the viewer's attention. The main advantage of using this framework is to understand the video content and eradicate the needs of difficult heuristic rules in video summarization. The authors also projected some audio-visual attention model features methods.

Ngo, Chong-Wah, Yu-Fei Ma, and Hong-Jiang Zhang [16] built up an approach for video summarization in view of the study on video structures and video highlights. Scene displaying and highlight identification are the main segments in this approach which are accomplished by normalized cut algorithm, temporal graph analysis and motion attention modeling. Here a video is mentioned as a complete undirected graph and the normalized cut algorithm optimally partition the graph into video groups. The subsequent groups then structures a directed temporal graph. The authors used Shortest path algorithm to identify the video scenes. The consideration values are then calculated and joined to the scenes, subshots, shots and clusters in a temporal graph. Therefore, the graph can illustrate the growth and perceptual significance of a video.

Cernekova, Zuzana, Ioannis Pitas, and Christophoros Nikou [17] introduced new strategies for identifying shot boundaries in video sequences and for getting the keyframes using the metrics based on information theory as it adventures the inter frame data in a more minimized manner. The strategy for shot boundary detection depends on the mutual information MI and the joint entropy (JE) between the frames and the technique identifies both and prompt cuts with high precision.

Otsuka, Isao, et al. [18] proposed a highlight scene detection function by means of only audio features and recognized a browsing function for the recorder that allows entire automatic recognition of sports highlights by identifying portions with "commentator's excited speech" using Gaussian mixture models (GMM's) trained using the MDL criterion.

Tseng, Belle L., Ching-Yung Lin, and John R. Smith. 19] Applied a video semantic summarization system, a video annotation tool, VideoAnn, to interpret semantic labels related with video shots and a video transmission system, Universal Tuner, for wireless video streaming. This system

transcodes MPEG-1/2 videos or live TV broadcasting videos to the BW or indexed color Palm OS devices. Videos are primarily divided into shots based on their visual-audio characteristics. And then they are played back using an interactive interface, which helps and speedup the annotation process. Users can interpret the video content with the units of temporal shots or spatial regions. The interpreted results are later stored in the MPEG-7 XML format.

Divakaran, Ajay, et al. [20] proposed video summarization and indexing techniques using MPEG-7 motion activity descriptor which are straightforward, flexible, and easy to extract and match. The authors described the video summarization techniques based on sampling in the cumulative motion activity space and then illustrate the combinations of the motion activity based techniques with generalized sound recognition that enable completely automatic generation of news and sports video summaries. The advantage is that the summarization allows quick creation of a summary outline which is of any preferred length

Zhang, Ke, et al.[21] proposed a novel subset selection technique to achieve automatic keyframe-based video summarization. The authors transfer the summary structures from the interpreted videos to hidden test videos. They generalized the method to subshot-based summarization which reduces the computational costs and provides more flexible ways of significant visual similarity across subshots of several frames.

Thomas, Sinnu, Sumana Gupta, and Venkatesh Subramanian [22] exhibited a structure for video summarization by presenting components of the human visual system (HVS) inside the summarization system itself to take into account the significance of perceptually imperative events while at the same time expelling the perceptual repetitions from the summaries.

Yin, Yifang, Roshan Thapliya, and Roger Zimmermann [23] proposed a novel technique of automatic video summary generation with individual adaptations. Inorder to diminish the semantic gap, the writers separate visual representations in view of a novel semantic tree (SeTree) which is a hierarchy that catches the conceptual associations between the visual scenes in a codebook. The SeTree is made by considering a standardized graph cut clustering algorithm by conjunctively abusing visual elements, textual data and social user image connections. To advance the intriguing parts of a video, the authors extracted a space-time saliency map and approximate the appearance of segments by kernel fitting and coordinating. A linear function is utilized to combine the two factors, based on which the playback rate of a video is adapted to generate the summary.

Zhang, Ke, et al. [24] proposed a new supervised learning method for summarizing recordings via automatically choosing keyframes or key subshots. The authors utilized Long Short-Term Memory (LSTM)

to demonstrate the variable-range temporal dependency among video outlines, to infer both representative and conservative video summaries. They introduced a procedure with address the requirement for a large amount of interpreted information for preparing complex learning ways to deal with summarization and created secondary annotated video summarization datasets.

Song, Xinhui, et al [25] proposed a novel approach of video summarization by large scale surveillance on the basis of event detection. First the authors obtained the trajectories of vehicles and pedestrians in a tracking-by-detection manner, and then detect the abnormal events using the trajectories. Finally, they designed a disjoint max-coverage algorithm to generate a summarized sequence with maximum coverage of interested events and minimum number of frames. The most important benefit of this method is, significant data can be efficiently extracted from the redundant contents. Abnormal events are successfully detected by combining the Random Forest classifier and the trajectory features. Third, the abnormal events are designed to display, and hence further reduces the compression ratio

Martins, Guilherme B., Joao P. Papa, and Jurandy Almeida [26] proposed a temporal- and spatial-driven approach that uses Optimum-Path Forest (OPF) clustering technique to automatically get the number of keyframes, as well as to extract them to create the final summary.

J.Kavitha, Dr.P.Arockia Jansi Rani [27] developed a video summarization system to extract significant features of interest from a given video by considering both the static features and wavelet features. This architecture combines both the static attention method and a method based on DWT for extracting most important key frames from a video. Here the input video is split into frames. Edges are identified in two consecutive frames using the sobel edge detection algorithm. Based on these edges, it compares the blocks of the video frames with one another. If the comparison exceeds a discrete threshold, it shows that the scene has been changed. The authors used Static attention method SAM to find the color opponency of each frame in each shot.

III. CONCLUSION

The world is mostly under video surveillance, but the issue is with video recording, to provide security by continuous supervisor monitoring which is really difficult. Video summarization is one good technique to minimize the total video into a few frames of an important event, which reduce the overall monitoring time and storage space. The article overview different newly developed video summarization techniques applicable to different applications and video format, which may help the new researchers in the field.

REFERENCES

1. BA TU Truong and Svetha Venkatesh, “Video Abstraction: A Systematic Review and Classification”, ACM Trans. on Multimedia Computing,

Communications and Applications, Vol.3, No.1, 2007.

2. John, Alvina Anna, Binoy B. Nair, and P. N. Kumar. "Application of Clustering Techniques for Video Summarization–An Empirical Study." Computer Science On-line Conference. Springer, Cham, 2017.
3. Hamza, Rafik, et al. "Secure video summarization framework for personalized wireless capsule endoscopy." Pervasive and Mobile Computing (2017).
4. Jacob, Hugo, et al. "A video summarization approach based on the emulation of bottom-up mechanisms of visual attention." Journal of Intelligent Information Systems (2017): 1-19.
5. Sahoo, Anurag, et al. "A Unified Multi-Faceted Video Summarization System." arXiv preprint arXiv:1704.01466 (2017).
6. Cirne, Marcos Vinicius Mussel, and Helio Pedrini. "VISCOM: A robust video summarization approach using color co-occurrence matrices." Multimedia Tools and Applications (2017): 1-19.
7. Vasudevan, Arun Balajee, et al. "Query-adaptive Video Summarization via Quality-aware Relevance Estimation." arXiv preprint arXiv:1705.00581 (2017).
8. Mishra, Deepak Kumar, and Navjot Singh. "Parameter Free Clustering Approach for Event Summarization in Videos." Proceedings of International Conference on Computer Vision and Image Processing. Springer, Singapore, 2017.
9. Li, Xuelong, Bin Zhao, and Xiaoqiang Lu. "A General Framework for Edited Video and Raw Video Summarization." IEEE Transactions on Image Processing (2017).
10. Salehin, Md Musfequs, and Manoranjan Paul. "Adaptive fusion of human visual sensitive features for surveillance video summarization." JOSA A 34.5 (2017): 814-826.
11. Rabbouch, Hana, Foued Saâdaoui, and Rafaa Mraihi. "Unsupervised Video Summarization using Cluster Analysis for Automatic Vehicles Counting and Recognizing." Neurocomputing (2017).
12. Ajmal, Muhammad, et al. "Video summarization: techniques and classification." Computer Vision and Graphics (2012): 1-13.
13. Mundur, Padmavathi, Yong Rao, and Yelena Yesha. "Keyframe-based video summarization using Delaunay clustering." International Journal on Digital Libraries 6.2 (2006): 219-232.

Ma, Yu-Fei, et al. "A user attention model for video summarization." Proceedings of the tenth ACM international conference on Multimedia. ACM, 2002.