



Video Segmentation And Summarization Using Machine Learning - Detailed Survey

Dr. H S Prasantha

Professor

Department of Computer Science and Engineering
K.S. Institute of Technology, Bangalore, India

Abstract: Many deep learning research works have shown successful results, primarily focusing on three data types: images, speech, and text. In addition, deep learning has also been successfully applied to communication signals/packets. Widely used applications of these kind of data are image classification, speech recognition, regression problem, pattern recognition, and text sentiment classification. The most fascinating of all is video data. Video data is also interesting for research from the perspective of its big size and dimension. Millions of video data are uploaded every day on YouTube; thus, it becomes a rich repository and empowered artificial intelligence (AI) research

Keywords- Video, machine learning, Segmentation, Summarization

1. Introduction

Millions of video data are uploaded every day on YouTube; thus, it becomes a rich repository and empowered artificial intelligence (AI) research. However, video data is challenging to analyse and process because of its large file sizes and complexity despite having rich data. Research on video processing using AI gained popularity after many AI algorithms were developed for Image processing for various applications. Video data is one of the most popular choices of users of different platforms like Twitter, YouTube, Facebook, etc. also the fastest-growing data type nowadays. However, most video representation and summarization approaches that have appeared in the literature relies on static arrangements of key frames. A set of key frames is selected from each video shot and arranged in a variety of pictorial summary forms within the video frame. Such compact representations of video provide viewers with a global picture of the entire video content on a single screen. This is often seen in our google browser search results therefore giving us an idea of what the video might contain

2. LITERATURE SURVEY

Author	Title	Methodology used	Remarks
Xue Yang, Zhicheng Wei	Video segmentation and summarization based on Genetic Algorithm” , 2011 4th International Congress on Image and Signal Processing	Binary Genetic Algorithm (BGA)	In this paper author compared BGA with DGA(Digital GA) and he obtained a result that is, BGA outperforms the DGA in convergence speed and possibility to get global optimum.Furthermore, he compared BGA with uniform approach and he concluded that BGA method can capture optimal result.In future research he will analyse the effect of parameter and fitness function
A. Sasithradevi, S. Mohamed MansoorRoomi , M. Mareeswari	Video object segmentation: A survey” , 2016 International Conference on Communication and Electronics Systems (ICCES)	Issue tackling mode, Complexity reduction mode, Inference mode	Author did a detailed survey on recent researches in object segment domain. This paper categorises video object segmentation algorithm into three parts Issue tackling model, reduction mode, Inference mode. By considering merits and demerits, future Research path are provided.
Hyuncheol Kim, Inhye Yoon, TaeYong Kim, Joonki Paik	Video summarization using feature dissimilarity“ 2016 International Conference on Electronics, Information, and Communications.	Iterative Dissimilarity Shrinkage	With this paper author shows that the proposed method generates the appropriate adaptation to the motion energy and produces video segmentation which have more motion energy and represent a significant event.
Hui Chen, Cuihua Li	“A practical method for video scene segmentation” , 2010 3rd International Conference on Computer Science and	Twin-threshold method, HLS colour distribution property	In this paper author have proposed a general framework of a practical method for scene segmentation. He followed three steps

	Information Technology		<p>detect raw shots , remove redundant shots ,calculate each shot's top 25 domain colours and correlation between different shots.</p> <p>In future author well focus on how to combine other information in the video that is sound for sound is a key factor despite the sharp difference between shots' colour information. Finally he would find the semantic correlation between different shot to get more satisfying results.</p>
MayuOtani, Yuta Nakashima, EsaRahtu, JanneHeikkil'a, and NaokazuYokoya	"Video Summarization using Deep Semantic Features"	Deep Neural Networks (DNN), Convolutional Neural networks (CNNs)	<p>This paper presents a video summarization technique for an Internet video to provide a quick way to overview its content. They have evaluated the video summaries using the SumMe dataset as well as baseline approaches. The results demonstrated the advantages of incorporating our deep semantic features in a video summarization technique.</p>
Hafiz Burhan UIHaq, M. Asif, Maaz Bin Ahmad.	"Video Summarization Techniques: A Review" , International Journal of Scientific Technology, November 2020	Spectral Clustering, K-means clustering.	<p>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.</p>

			The first analysis is based on approaches, applications, datasets, and pros/cons/findings.
Tomohito Shiraiwa , Hajime Nobuhara	“Efficient Video Summarization Based on Semantic Segmentation Model” 2021 IEEE 10th Global Conference on Consumer Electronics (GCCE)	FCSN and vsLSTM. Segmentation models such as U-Net and SegNe	This paper investigated the use of semantic segmentation models for video summarization. The three-layer U-Net and SegNet Basic were explored and their results outperformed existing methods in terms of the F-measure, or accuracy. In addition to further improving the accuracy, we plan to improve the model and algorithms to reduce the computation time and to perform video summarization reflecting the users preferences.
Amir H. Meghdadi, Pourang Irani.	“Interactive Exploration of Surveillance Video through Action Shot Summarization and Trajectory Visualization” 2013 IEEE Transactions on Visualization and Computer Graphics	Video visual analytics system sViSIT algorithm.	System that can help users search a video and find targets in a fast and efficient way. Showed that except for very short videos using sViSIT is considerably faster than waiting for the target event to happen in a 5x fast-forward video. This is expected as the power of our system is due to its ability to summarize movements individually and apply spatiotemporal filters to limit the search results
Bhattacharya, Koustav and Chaudhury, Santanu&Basak, Jayanta	“Video Summarization: A Machine Learning Based Approach” ICVGIP 2004, Proceedings of the Fourth Indian Conference on Computer Vision, Graphics & Image Processing, Kolkata, India, December 16-18,	Genetic algorithm and linkage algorithm	To sum up they have worked on the use of learning in Content Based Retrieval in the domain of videos. For exploring video analysis tasks like video filtering or shot summarization, the video is first divided into shots. Then described a novel

	2004		approach to the design of a content based filter which was based on learning the ICs of critical windows in a shot.
V. Sharma, M. Gupta, A. Kumar and D. Mishra	"Video Processing Using Deep Learning Techniques: A Systematic Literature Review", IEEE Access .	CNN (convolution neural network) DNN (deep neural network) RNN (recurrent neural network) Hybrid approach	Deep learning techniques can be used robustly for video understanding, video classification, video analysis, action recognition, and pose recognition. Significant work has also been done on video processing using deep learning for human action recognition, behaviour analysis, and crowd anomaly detection. This paper contributes a systematic literature review to investigate the up-to-date research in video processing using deep learning techniques.

3. MACHINE LEARNING ALGORITHMS

Genetic Algorithm

- A genetic algorithm (GA) is a heuristic search algorithm used to solve search and optimization problems. This algorithm is a subset of evolutionary algorithms, which are used in computation. Genetic algorithms employ the concept of genetics and natural selection to provide solutions to problems.
- These algorithms have better intelligence than random search algorithms because they use historical data to take the search to the best performing region within the solution space.

Iterative Threshold Algorithm

- Thresholding is the most widely used image segmentation method. Thresholding algorithm has histogram bimodal method(also known as the mode method), Otsu method and the iterative threshold method, etc.
- Histogram bimodal method is used to some simple images, which appear two separate peaks in histograms, and then the troughs which correspond gray value between two peaks was selected as threshold value.

Twin threshold approach

- This approach allows to reduce false detections caused by steep intensity fluctuations (due to noise, motion, visual effects, etc.), as well as to retrieve dissolves caught up in other visual effects or scene movements.
- Additionally, to overcome the restraint visual continuity of the animated movies, fading-out and fading-in pixels are selected at intensity level from a reduced time window of only several frames

Spectral clustering

- Communities of nodes (i.e. data points) Clustering is one of the most widely used techniques for exploratory data analysis, with applications ranging from statistics, computer science, biology to social sciences or psychology.
- Spectral clustering is an EDA technique that reduces complex multidimensional datasets into clusters of similar data in rarer dimensions.
- The main outline is to cluster the all spectrum of unorganized data points into multiple groups based upon their uniqueness “Spectral clustering is one of the most popular forms of multivariate statistical analysis” ‘Spectral Clustering uses the connectivity approach to clustering’, wherein that are connected or immediately next to each other are identified in a graph.
- The nodes are then mapped to a low-dimensional space that can be easily segregated to form clusters. Spectral Clustering uses information from the eigenvalues (spectrum) of special matrices (i.e. Affinity Matrix, Degree Matrix and Laplacian Matrix) derived from the graph or the data set.

K-Means clustering algorithm

K-Means clustering algorithm is defined as an unsupervised learning method having an iterative process in which the dataset are grouped into k number of predefined non-overlapping clusters or subgroups, making the inner points of the cluster as similar as possible while trying to keep the clusters at distinct space it allocates the data points to a cluster so that the sum of the squared distance between the clusters centroid and the data point is at a minimum, at this position the centroid of the cluster is the arithmetic mean of the data points that are in the clusters.

Linkage Algorithm

It determines the distance between sets of observations as a function of the pairwise distance between observations.

- In **Single Linkage**, the distance between two clusters is the minimum distance between members of the two clusters.
- In **Complete Linkage**, the distance between two clusters is the maximum distance between members of the two clusters.
- In **Average Linkage**, the distance between two clusters is the average of all distances between members of the two clusters.
- In **Centroid Linkage**, the distance between two clusters is is the distance between their centroids.

4. TOOLS

- VsLSTM which can perform video summarization with higher accuracy than a normal bidirectional LSTM.
- A three-layer U-Net, a model used for semantic segmentation and SegNet Basic.
- GoogLeNet to extract features for each frame of the video image.
- SumMe and TvSum which are data sets for video summarization.
- FSCN to compare the means and standard deviations of the F-scores generated .
- ViSIT(Video Summarization and Interaction Tool), a novel video visual analytics system for interactive and analytic exploration of video data.

5. APPLICATIONS

- I. A video visual analytics system (sViSIT) that can help users search a video and find targets in a fast and efficient way.
- II. Consumer video applications, Image-Video databases management and surveillance.
- III. In organizing and indexing large volumes of video data to facilitate efficient and effective use of these resources for internal use.
- IV. A model for finding low-level surprise at every location in video streams.
- V. Deals with problem of categorizing a given videos sequence into one or predefined video genre.

6. REFERENCES

1. Dr. H S Prasantha, "NOVEL APPROACH FOR IMAGE COMPRESSION USING MODIFIED SVD", International Journal of Creative Research Thoughts (IJCRT), Volume 8, Issue 8, Page 2234-2243, Aug 2020
2. Dr. H S Prasantha, "IMPLEMENTATION OF IMAGE COMPRESSION USING FAST COMPUTATION OF SVD ON DM642", International Journal of Creative Research Thoughts (IJCRT), Volume 8, Issue 8, Page 2364-2368, Aug 2020
3. Prasantha, H, H Shashidhara, K N B Murthy, and M Venkatesh. "Performance Evaluation of H.264 Decoder on Different Processors." International Journal on Computer Science & Engineering. 1.5 (2010): 1768. Web. 7 Apr. 2013.
4. H. S. Prasantha, H. L. Shashidhara, and K. N. Balasubramanya Murthy. Image compression using SVD. In Proceedings of the International Conference on Computational Intelligence and Multimedia Applications, pages 143–145. IEEE Computer Society, 2007.
5. Gunasheela K S, H S Prasantha, "Compressive sensing for image compression: survey of algorithms", Proceedings of Emerging Research in Computing, Information, Communication and Applications, ERCICA, Springer publication, Bengaluru, 2018
6. K N Shruthi, B M Shashank, Y. SaiKrishna Saketh, H.S Prasantha and S. Sandya, "Comparison Analysis Of A Biomedical Image For Compression Using Various Transform Coding Techniques", IEEE, pp. 297-303, 2016