IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Survey On Content Based Video Retrieval

[1]Urvi Patel, [2]Dr. Narendra M. Patel

[1]M. Tech Student, [2]Professor

[1]Department of Computer Engineering,

[1]Birla Visvakarma Mahavidyalaya, Vallabh Vidyanagar, India

Abstract: Content based video retrieval has many intelligent applications in the field of education, security, medicine, agricul- ture, etc. Retrieval of a specific video story based on the textual query is the key idea behind research work. The video is first segmented in order to get shots representing small stories. To represent each shot efficiently, key frame extraction needs to be performed on each video. Various key frame extraction algorithm is also explored in this research work. This research work also involves natural language processing steps such as tokenization, punctuation and extra symbol removal, stemming of words to the root words, etc. for text processing of textual query. Later on, indexing algorithm needs to be implemented in order to obtain appropriate results based on the query. This research work explores the area of video retrieval based on textual query and image query.

Index Terms **- Key frame extraction, Optical Character Recognition(OCR), shot boundary detection, indexing, stemming, tok- enization**

I. INTRODUCTION

Today in the era of digital information, the amount of content available over the internet is increasing rapidly. Most of the content today over the internet is in the form of images or videos. Video data contains richer content as compared to image content. The video data can be of various categories such as sports, news, entertainment, multimedia messages, tutorials, lectures, e-learning videos, etc. All these videos consist of information such as text, objects, shapes, textures etc. Other information that are available in the video include text, description, audio, visual information, etc.

Content based video retrieval can be described as a process of retrieving desired videos from a large collection of video data. The Content based video retrieval (CBVR) system are capable of extracting the useful videos based on the extracted features from the video. CBVR has good application in the field of education, video surveillance, criminal data retrieval, biomedical information, entertainment, news, broadcasting, etc.

Content based video retrieval involves various tasks such as shot boundary detection, key frame extraction, feature extraction, indexing and similarity matching. The video is retrieved in the form of a query. The query can be an image query or a text query. Many video retrieval systems also take a shot, video clip or face related data as query to perform content based video retrieval. A video is comprised of many short stories or scenes. Each of these scenes are further divided into shots [1]. Shots in a video sequence can be termed as successive frames in a video without any significant changes in the visual content [14]. In order to achieve shot boundary detection various key frame extraction methods needs to be implemented. Key frames are still videos extracted from the video that best represent the video content [9]. In recent years many key frame extraction methods have been proposed such as Adaptive Threshold, Entropy based methods, Histogram difference methods, twin comparison algorithm, optical flow- based algorithm, etc. After obtaining the key frames, feature extraction needs to be performed on these frames. The features involved with the video are low level and high-level features. The low-level features include colour, shape and edge of the video. The high-level features include the text, description, audio, etc. Lastly indexing is performed on the video database. Video indexing can be defined as a process of tagging videos and effectively organizing them for efficient and fast access of videos and their retrieval [14]. For text query-based video retrieval the text from the video is extracted using OCR on key frames extracted. Later, based on the similarity measure between the extracted text and the query text, the appropriate video clip is retrieved. For image-query based video retrieval the various object images are passed as query and based on similarity measures, the appropriate videos are listed out.

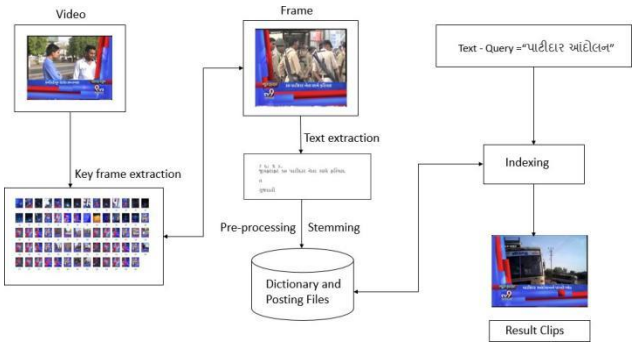

Figure 1: Flowchart of Content based video retrieval using textual query

Figure 1 provides a brief overview of overall process involved in content-based video retrieval using text as query. As the digital age is developing, the research work in this field is constantly increasing. This paper provides a brief overview of various research work carried out so far in this research area. The remaining section of this paper is divided into literature review of various works carried out in this research area, challenges and finally conclusion and future work.

II. LITERATURE REVIEW

As shown in the Figure 1 CBVR process contains various tasks involved in it such as shot boundary detection and key frame extraction, feature extraction, video indexing and matching.

The first task is to divide the video into segments of related frames. Shot boundary detection is one of the methods to detect key frames from the videos. These common frames, termed as shot are used to present the video content in abstract form. Shot boundary detection can be defined as aprocess to find the visual content discontinuity values between consecutive video frames.The mismatch between two frames is found when a shot changes. The shot changes are mostly represented using two forms: abrupt changes and gradual changes. Abrupt changes are very easy to detect but gradual changes are hard to detect [17]. Abrupt changes include hard cut and gradual changes include dissolve, fade in, fade out and wipe [16]. Various shot boundary detection algorithms have been proposed so far. O'Toole et al. [18] proposed using thresholding based method to identify shot changes. Patel et al. [9] in their paper proposed using entropy method to detect scene changes to obtain better result.

Liu et al. [15] proposed using Scale Invariant Feature Transform (SIFT) to perform shot boundary detection and key frame extraction. In this method the authors have extracted SIFT key points from each frame of videos in terms of their temporal sequence. They have applied Best-Bin-First [19] algorithm to match key points between two adjacent frames. Later, the ratios of match key points number to total number are used to detect shot boundaries and key frames [15].Baber et al. [20] have proposed a methodology to detect abrupt boundary changes accurately. The method proposed by the author also detects fade boundaries with high accuracy. Here, the authors obtain gradual shot boundary detection by entropy and then use Speeded up Robust Features algorithm to find the correspondence across a candidate shot boundary. It can be utilized to detect abrupt changes.

Shot similarities in a video can be identified using key frames. Keyframe can be described as a frame chosen to represent the content of the entire shot [1]. Dang et al. [21] proposed image epitome-based method to extract key frames from videos. In this method the authors have utilized image epitome as a feature vector and have applied an information divergence-based distance measure on the feature vector to measure dissimilarity between frames of the video. Sheena et al. [22] have proposed key frame extraction algorithm based on absolute difference of histograms of consecutiveimage frames. Dave et al. [2] have proposed the use of HSV color model to find the frame difference between successive frames of the video. Saric et al. [13] have proposed twin comparison method to perform shot boundary detection. Lu et al. [23] have proposed a joint learning framework over optical flow and color histogram for shot boundary detection.

After performing shot boundary detection and key frame extraction, the next task is to extract text from the framesusing Optical Character Recognition (OCR) for content based video retrieval using textual query. Later, indexing is performed for proper searching and retrieval of the video data. Video Indexing are of two types: 1) Syntactic Indexing and 2) Semantic Indexing. Syntactic features are used as a basis for matching in syntactic indexing and they employ either Query by-example or Query-through-dialog box to interface with the user [25]. Semantic indexing is confined to data modelling in specific domains in an unstructured manner. It is useful input for knowledge creation [17]. Wagh et al. [3] have used Hierarchical Clustering Tree algorithm for indexing. Gitre et al. [24] have used B+ Tree for indexing technique. The B+Tree automatically reorganizes itself with small local changes during insertion and deletion.

Many research works have been carried out in the field of content-based video retrieval using textual query and image query.In the work by Wagh et al. [3], the authors have proposed text-based video retrieval and video search system. In the first stage they have performed automatic video segmentation and keyframe detection for video content navigation. The textual metadata from the video were extracted using video OCR technique on key frames and Automatic Speech Recognition (ASR) on the lecture audio track by the authors. In the pro- posed methodology they have used Support Vector Machine (SVM) for classification purposes. Clustering is done with the k-means clustering [27] algorithm. Here, Indexing is done with the help of the Hierarchical Clustering Tree (HCT) algorithm. In the work by Dave et al. [1], the authors have proposed retrieval of specific news video based on gujarati textual query. Since the dataset for the Gujarati language videos was not available to work on, the author had created their dataset from three Gujarati language news video channels such as ETV Gujarati, DD 11, and Sandesh. Here in this research work, the authors have proposed the use of singular value decomposition and the rank of a matrix to perform keyframe extraction. Text is extracted from keyframes for further indexing data. With the proposed approach they have achieved 82.5 percent accuracy on the ETV news dataset.

In the work by Asha et al. [5], the authors have pro- posed a "Content-based video retrieval system using multiple features". The proposed method transfers each video of the database into scenes using a color histogram-based scene change detection algorithm and keyframes are extracted. For keyframes, multiple features are obtained using straightfor- ward rules. Subsequently, using the multiple feature vectors we compare the query and database videos by measuring Euclidean distance. The proposed system is compared with the CBVR systems with single features and it was observed that CBVR systems with multiple features provide better performance.

In the work by Yang et al. [4], the authors have proposed an approach for automated video indexing and video search in large lecture video archives. They have applied automatic video segmentation and key-frame detection to perform video content navigation. Later on, after performing the keyframe extraction, the authors extracted textual metadata by applying video OCR technology on key-frames of the video and ASR on the audio tracks. For key word extraction, the OCR and the ASR transcripts as well as the detected slide text line types are used. Using this both video and segment level keywords are obtained for video browsing and searching based on specific content. The effectiveness and performance of proposed indexing functionalities are proven by evaluation.

In the work by Chivadshetti et al. [28], the authors have proposed a personalized content based video retrieval system. The authors have proposed to extract both ASR and OCR features of the video to perform video retrieval. The keywords were extracted using the ASR technology. Later, OCR tech- nology was applied on the video frames to extract the HOG (Histogram of Oriented Gradients), OCR text and Gabor Filter from selected frames. It will also extract the Color, Texture and Edge detector on selected frames to perfrom appropriate content based video retrieval.

Along with text-based video retrieval, image query-based video retrieval is also performed. Araujo et al. [6] have performed content-based video retrieval based on image query. In this research work, the authors have proposed asymmetric comparison technique. The authors have also proposed a new kind of video descriptors that can be directly compared with the image descriptors. The authors constructed the Fisher Vectors for the video database using different aggregation techniques. They have created video descriptors using the combination Fisher embeddings with hashing techniques. The authors have used Bloom Filter based Indexing algorithm. They have claimed that this technique made the video retrieval process much faster as compared to frame based approach.

In another work by Fernandez-Beltran et al. [29], the authors have proposed a novel content based video retrieval system to cope with the semantic gap between low and high features of video. The authors here proposed a supervised topic model to transform the classical video retrieval problem into a class discovery problem. Following this, the authors identified a new probabilistic ranking method to deal with the semantic gap between the low and high level features of video. Finally, queries were generated to retrieve appropriate content based video.

Some of the authors have also proposed deep learning based approaches to perform content based video retrieval. In the work by Jyothi et al. [30], the authors have used Deep learning based approach to perform video retrieval. In this work the authors have performed content based video retrieval of flower database. They have trained their network in three forms, with keyframes, with segmented flowers and with gradient of flowers as input. The video related to the query is retrieved using Multiclass Support Vector Machine (MSVM).

In other work by Szűcs et al. [26] the authors have proposed "scene retrieval from video based on image query". Here, the author has proposed performing sampling of video and then the key points are extracted from all the images. After that based on clustering of key points an inverted index is constructed. Then, the extracted key points of the query image are compared to the inverted index and based on this the key points are filtered. Later, the most similar images are searched and appropriate video parts are retrieved and presented.

In order to retrieve videos based on the textual query and image query, in most of the works mentioned here they have performed feature extraction, video segmentation, and text feature extraction, image features matching, etc. in order to obtain the desired result.

III. CHALLENGES

Today the amount of video database is constantly increasing over the digital platform. This video data contains various information which proves to be important for researchers, students, entertainment, etc. purposes. These fields can bring advantages and become useful to people only if the video retrieval systems are efficient and fast enough to retrieve necessary video data from large-scale video databases. Conventional web search engine based approach to retrieve video are not very efficient these days. So, novel approaches have been introduced to retrieve video from the database based on content.

These retrieval systems still pose some challenges. Content based video retrieval systems perform key frame extraction, video indexing, etc. to retrieve video results. Some of the content based video retrieval system uses video clip and images as query. To retrieve video results by calculating the similarity between key frame features of the video database and that of the query requires a lot of computational time. These features include many low level and high level features, such as color histogram, edge and texture features, motion feature, temporal feature, etc. This large computational time results in long response time.

Along with image query and video clip query, content-based video retrieval is also performed using a textual query. In these systems, the textual content is extracted from the keyframes of video. The video database is later annotated. The resulting video is retrieved by measuring the similarity between the query text and the indexed features of the video. These video retrieval techniques require efficient OCR algorithms for accurately recognizing the text present in the video database. These systems also require performing text pre-processing for efficient video retrieval. Stemming, tokenization, etc. for the English language are easily available. In the work by Dave et al. [1] they have performed content based video retrieval from Gujarati news video using textual query. Systems like these face challenges of availability of efficient OCR algorithms, stemming and tokenization algorithms, etc. for local languages. Other than this efficient and robust indexing algorithms are also required to index, search and retrieve videos from large video database.

IV. Conclusion and Future Work

Many research works have been carried out in the field of context-based video retrieval. The number of works in this fields are constantly increasing day by day due to the interesting features of this research area. In order to perform this task, the researchers have performed shot boundary detection, key frame extraction, stemming, indexing, etc. Webelieve that the ongoing and future work on these particular points will benefit the specific task of context-based video retrieval. To further improve the performance of the retrieval system, Content based video retrieval from image part for this system can also be added and different feature extraction, classification and indexing can be explored. In future, both image and text-based methodology will be merged to increase the efficiency of the system.

V. Acknowledgment

I Urvi Patel express my sincere thanks and gratitude to my guides Dr. Narendra M. Patel and Dr. Mayur M. Vegad who guided me through the various stages of this research work. I am thankful for their constant support and guidance in identification of appropriate area of research work.

I am also thankful to all my colleagues who have given me support and assistance during this work. I am thankful to my parents and friends for their constant support and motivation. Last but not the least I am thankful to all the people who have directly or indirectly helped me realize this work.

References

[1] N Dave, M Holia, "Content based Video Retrieval", IJTE special issue of ICRISET 2017, PP 156-160.

[2] Namrata Dave, Dr. Mehfuza Holia "Shot Boundary Detection for Gujarati News Video", IJRASET Volume 6 Issue III, March 2018.

[3] Wagh K. D., Dr. M. U. Kharat "Content Based Video Retrieval", International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 1, January 2016

[4] Yang, Haojin, and Christoph Meinel. "Content based lecture video retrieval using speech and video text information." IEEE Transactions on Learning Technologies 7.2 (2014).

[5] D. Asha, Dr. Y. Madhavee Latha, Dr. V. S. K Reddy, "Content based video retrieval system using multiple features", International Journal of Pure and Applied Mathematics Volume 118 No. 14 2018.

[6] A. Araujo and B. Girod, "Large-Scale Video Retrieval Using Image Queries," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 28, no. 6, pp. 1406-1420, June 2018. doi:10.1109/TCSVT.2017.2667710.

[7] Liu, Ying, et al. "A survey of content-based image retrieval with high- level semantics." Pattern recognition 40.1 (2007): 262-282.

[8] Milan R. Shetake, Sanjay. B. Waikar, "CONTENT BASED IMAGE VIDEO RETRIEVAL" International Journal of Advances in Electronics and Computer Science, ISSN:2393-2835 Volume-2, Issue-9, Sept.-2015.

[9] B. V. Patel and B. B. Meshram, "CONTENT BASED VIDEO RETRIEVAL SYSTEMS" International Journal of UbiComp (IJU), Vol.3, No.2, April 2012.

[10] Weining Hu, senior member, IEEE, Nianhua Xie, LiLi, Xianglin Zeng, And Stephen Maybank, "A survey on Visual Content Based Video Indexing and Retrieval", IEEE Transactions, On System, Man, And Cybernetics-Part C: Applications and reviews PP1-23, January 2011.

[11] A. Hanjalic, "Shot-boundary detection: unravelled and resolved? "Cir- cuits and Systems for Video Technology, IEEE Transactions on, vol. 12, pp. 90- 105, 2002.

[12] C. V. Jawahar, Balakrishna Chennupati, Balamanohar Paluri and Nataraj Jammalamadaka, Video Retrieval Based on Textual Queries, Proceedings of the Thirteenth International Conference on Advanced Computing and Communications, Coimbatore, December 2005.

[13] M. Saric, H. Dujmic, and D. Baricevic, "Shot Boundary Detection in Soccer Video using Twin-comparison Algorithm and Dominant Color Region", JIOS, vol. 32, no. 1, Jun. 2008.

[14] Matheel E. Abdulmunem, Eman Hato, "Semantic Based Video Retrieval System: Survey", Iraqi Journal of Science, 2018, Vol. 59, No.2A, pp: 739-753 DOI:10.24996/ijs.2018.59.2A.12.

[15] G. Liu, X. Wen, W. Zheng and P. He, "Shot Boundary Detection and Keyframe Extraction Based on Scale Invariant Feature Trans- form," 2009 Eighth IEEE/ACIS International Conference on Computer and Information Science, Shanghai, 2009, pp. 1126-1130, doi: 10.1109/ICIS.2009.124.

[16] Pal G., Rudrapaul D., Acharjee S., Ray R., Chakraborty S., Dey N., "Video Shot Boundary Detection: A Review", In: Emerging ICT for Bridging the Future - Proceedings of the 49th Annual Convention of the Computer Society of India CSI Volume 2. Advances in Intelligent Systems and Computing, vol 338. Springer, Cham. https://doi.org/10.1007/978-3-319-13731-5 14.

[17] Hetaxi Kamli, Mahasweta Joshi, U. K. Jaliya," A Review Paper on Content based Video Retrieval", IJSRD International Journal for Scientific Research Development— Vol. 3, Issue 09, 2015.

[18] O'Toole, C., A. Smeaton, N. Murphy and S. Marlow, "Evaluation of automatic shot boundary detection on a large video suite". In: 2nd U.K. Conference Image Retrieval: The Challenge of Image Retrieval, Feb. 25-26, Newcastle, U.K., 1999.

[19] D. G. Lowe, "Distinctive image features from scale-invariant Key points," International Journal of Computer Vision, vol. 60, pp. 91- 110, 2004.

[20] Junaid Baber, Nitin Afzulpurkar, Matthew N. Dailey, And Maheen Bakhtyar – "Shot boundary detection from videos using entropy and local descriptor ", 2011 IEEE.

[21] C. T. Dang, M. Kumar and H. Radha, "Key frame extraction from consumer videos using epitome," 2012 19th IEEE International Conference on Image Processing, Orlando, FL, 2012, pp. 93-96.

[22] Sheena C V, N. K. Narayanana, "Key-frame extraction by analysis of histograms of video frames using statistical methods", Procedia Computer Science 70 (2015) 36 – 40.

[23] Z. Lu, L. Wu, M. Jian, S. Zhang, D. Wang and X. Wang, "Shot Boundary Detection with Key Motion Estimation and Appearance Differentiation," 2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP), Chongqing, China, 2019, pp. 1-7, doi: 10.1109/IC- SIDP47821.2019.9173023.

[24] Madhav Gitre, Harshal Bawaskar, Sourabh Sethi, Ajinkya Shinde – "Content Based Video Retrieval System", International Journal of Research in Engineering and Technology, vol 3, Issue 6, Jun-2014.

[25] P. Geetha, Vasumathi Naryanam – "A Survey of Content Based Video Retrieval", Journal of Computer Science 4 (6), 474-486,2008 Science Publications.

[26] G. Szu˝cs, "Scene Retrieval in Videos by Query Image", Journal of Engineering Science and Technology Review 8 (5) ,12-18, 2015.

[27] JAIN A K, DUBES R C, "Algorithms for clustering data",New Jersey:Prentice-Hall,1988.

[28] P. Chivadshetti, K. Sadafale and K. Thakare, "Content based video re- trieval using integrated feature extraction and personalization of results," 2015 International Conference on Information Processing (ICIP), Pune, India, 2015, pp. 170-175, doi: 10.1109/INFOP.2015.7489372.

[29] Ruben Fernandez-Beltran, Filiberto Pla,"Latent topics- based relevance feedback for video retrieval", Pattern Recog-nition, Volume 51, 2016, Pages 72-84, ISSN 0031-3203, https://doi.org/10.1016/j.patcog.2015.09.007.

[30] Jyothi V. K., Guru D. S., Sharath Kumar Y. H.,"Deep Learning for Retrieval of Natural Flower Videos", Procedia Computer Scienec, 132, 1533-1542, 2018.

[31] Aasif Ansari, Muzammil H Mohammed, "Content based Video Retrieval Systems - Methods, Techniques, Trends and Challenges", International Journal of Computer Applications (0975 – 8887), Volume 112 – No. 7, February 2015.