IMPROVING FRAME SEARCHING EXPERIENCE IN CBVR SYSTEM

Ajay B Gadicha

Research Scholar, Department of Computer Science & Engineering, P.R.Pote College of Engineering and Management Amravati-444603.

Dr. Milindkumar V. Sarode

Professor, Head Department of Computer Science and Engineering, Government College of Engineering Yavatmal

Dr. V.M.Thakare

Professor, Head, P.G.Department of Computer Science, Sant Gadge Baba Amravati University, Amravati.

Abstract

The researchers are concentrating all these fast growing fields so none of the video retrieval parental fields alone antiquated able to directly solve the retrieval problem. In this paper shows the path towards a step by step mechanism of CBVR i.e. analysis of entire video, video segmentation, key frames mining, feature extraction mining for retrieving the video from large video datasets. The proposed system inclination focuses on searching the frames and performing key frame mining using adaptive thresholding algorithm and canny mechanism for feature extraction purpose. In order to legalize this claim, content based video reclamation systems were furnished using color histogram, features extraction and different approaches are applied for the supervision of the semantic temperament of each frame in the video.

Keywords: CBVR, Semantic Analysis, STD, TF, MD, BOW.

1. INTRODUCTION

The demand for one of the intelligentsia processing and appraisal of multimedia information has been unexpectedly developing in modern years [1]. Researchers have actively developed wonderful strategies to wise video management, collectively with shot transition detection, key body extraction, video retrieval, and lots of others [2-5]. Among the ones methods, shot transition detection is step one of content fabric-based totally video analysis and key body is a simple but inexperienced form of video summary. It can assist clients to understand the content material at a glance and is of realistic price.[3-5]. Many strategies used amazing type of capabilities to find out shot boundary, which includes histogram, shape facts, movement activity. Among the ones techniques, the histogram is the famous technique.

X. Zeng et al. [6] divided every frame into 'r' blocks, and the distinction of the corresponding blocks of consecutive frames turned into computed via shade histogram; the distinction D(i,i +1) of the 2 frames changed into acquired with the resource of inclusive of up all of the blocks' difference; in the in the meantime, the distinction V(i,i +1) among frames i and that i +1 modified into measured again without the usage of blocks. Based on D (i,i +1) and V(i,i +1) , shot boundary changed into determined. Getting over the downside of the paper, we propose greater inexperienced algorithms for shot boundary detection and key body extraction with automated threshold.

A. SHOT BOUNDARY DETECTION:

Each frame is alienated into blocks with 'x' rows and 'y' columns, and then the divergence of the corresponding blocks between two chronological frames is computed. Finally, the very last distinction of two frames is acquired by way of merging all the variations. The below cited method will give an explanation for the shot boundary frontiers mechanism in the video.

Algorithm 1: Shot Frontiers Detection

1: Let **M (f_i)** be the fi^{th} frame in video sequence, where

F_i □□1, 2, ….. , *Fv* *(Fv* denotes the whole variety of

Video sequence)

2: Partition frame into blocks with x rows and y columns, and **L (i, j, fi)** stands for the block at (i, j) in the fi frame.

3: Computing x2 histogram matching distinction between the corresponding blocks between consecutive frames in video sequence. **G (i, j, fi) and G (i, j, fi +1)** stand for the histogram of blocks at (i, j) inside the fith and (fi +1)th body respectively. Block's calculation is computed with the following equation:

$$\boldsymbol{D'(fi, fi+1, i, j,)} = \sum_{1}^{L-1} [\boldsymbol{G(i,j,fi) - G(i,j,fi+1)}]\mathbf{2} \Big/ \boldsymbol{G(i,j,fi)}$$

4**:** Computing x2 histogram difference between two consecutive frames:

$$\boldsymbol{D'(fi, fi+1)} = \sum_{1}^{x} \sum_{1}^{y} \text{wij D}'(\text{fi, fi} + 1, \text{i , j })$$

where wij stands for the weight of block at (i, j) ;

5: Computing threshold automatically: computing the mean and standard variance of x2 histogram difference over the whole video sequence[7]. Mean and standard variance are defined as follows:

$$MD = \sum_{fi=1}^{fv-1} \frac{D\ (fi, fi+1)}{fv-1}$$

$$STD = \sqrt{\sum_{fi=1}^{fv-1} \frac{(D\ (fi, fi+1) - MD)^2}{fv-1}}$$

Final shot detection: Shots can be very long but not a lot short, because the ones pictures with simplest several frames cannot be captured by means of people and that they cannot carry a whole message. Usually, the shortest shot should last for 1 to 2.5 s. For the cause of fluency frame rate is at the least 22-25 fps. So, a shot consists of at the least a minimum quantity of 30 to

45 frames. In our test, video sequences are down sampled at 10 fps to enhance simulation velocity. On this situation, the shortest shot should comprise 9 to 14 frames. 12th is chosen for our experiment. We formulate a "shots merging principle": if a detected shot comprise fewer frames than 12 frames, it will be merged into preceding shot, or it is going to contain idea as an unbiased one.

Reference Frame: it's nothing but the initial frame of any video;

General Frames: It is all frame except reference frame "Shot Dynamic Factor" max(i): the maximum x2 histogram within shot i; Dynamic Shot and Static Shot: a shot can be declared as dynamic shot, if its max(i) is bigger than MD; otherwise it's far static shot Fc (fi); the kth frame inside the present day shot, k=1,2,3…Fcn(k); (Fcn(k) is the total variety of the current shot)

Algorithm 2: Key frame extraction

1: Computing the difference between all the general frames and allusion frame with the above algorithm 1:

$$\boldsymbol{D}k(1,k) = \sum_{1}^{x}\sum_{1}^{y} \text{wij Dfib}(1,\text{k},\text{i},\text{j}),\text{k} = 1,2,3,4 \ldots\ldots \text{Fc}$$

Where 'Dfib' is the differnce between the frame with their block differnce measures.

2. Searching for the maximum difference within a shot:

max(i)= $\{\text{Dc }(1,\text{k})\}_{\max}$, k=2,3,4,.....Fcn.

3. Determining "ShotType" according to the relationship between max(i) and MD: StaticShot(0) or DynamicShot:

$$\text{SHOTc} = \begin{Bmatrix} 1 & if\ \max(i) > MD \\ 0 & otherwise \end{Bmatrix}$$

4. Determining the position of key frame:

if SHOTc=0, with respect to the odd number of a shot's frames, the frame in the middle of shot is chose as key frame; in the case of the even number, any one frame between the two frames in the middle of shot can be chose as key frame. If SHOTc=1, the frame with the maximum difference is declared as key frame.

Based on some statistical information or reviews each researcher experiencing an evolution from small databases to large imaginary datasets, and now to Digital Library. A virtual library is a library wherein resources are to be had in machine-readable layout as opposed to print handy by way of computer systems. Digital Library is essential to Content based totally Retrieval (CBR). The CBR is a repository permits content-based retrieval. It contains digital textual content, sound, track, image, video, and so forth. A big leap forward from traditional database search which is largely based on simple attributes as a result content material based totally mechanism additionally serve as powerful surfing tool similar to the present day net search like Google.

Therefore to find the exact end result through user question or through textual content based totally retrieval mechanism proposed machine specializes in Content-based totally Video Retrieval (CBVR) structures seem like a natural extension (or merge) of Content-based totally Image Retrieval (CBIR) structures. However, there are a number of of things which can be not noted while managing pics which must be treated while the usage of films. These factors are more often than not associated with the temporal records available from a video report. Visual file processing operations are important for routinely extracting an in depth description of a document. Feature extraction targets at characterizing a list of homes (called characteristic vector or record signature) for every aspect (pixel, frame place, frame, and sequence) of a video report.

Video indexing have to be analogous to text document indexing to facilitate speedy and accurate content get admission to to video information, discern 1 indicates how video structuring have to segment a video report into pictures and scenes and once more extract key-frames or key sequences as index entries for scenes or stories. The paper specifically focus on content material-based totally video retrieval is growing technologies to routinely parse video, audio, and text to identify significant composition structure and to extract and constitute content attributes of any video resources.

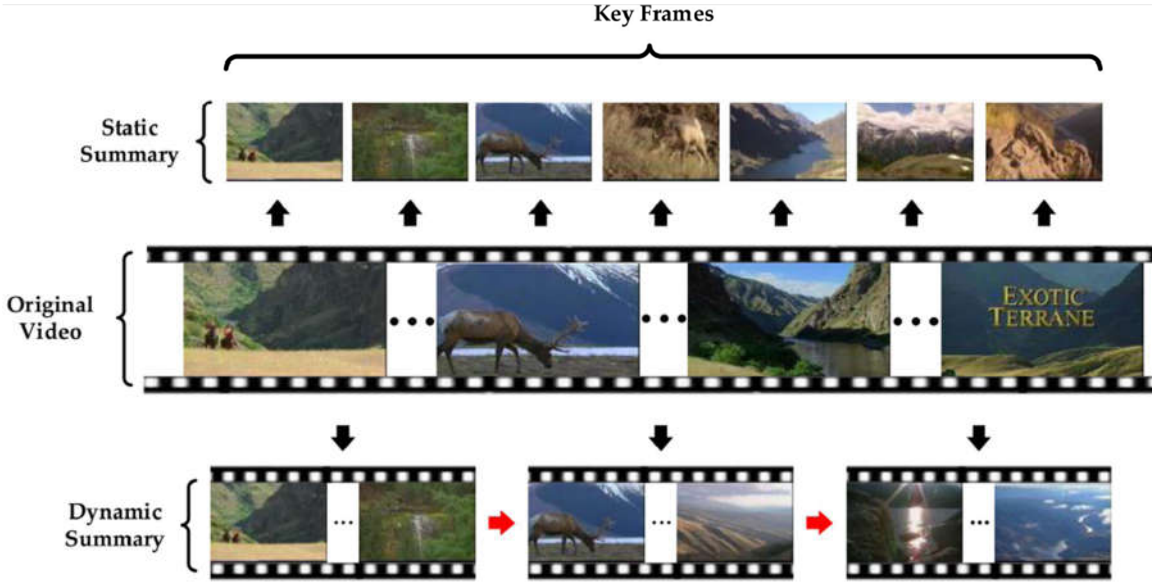


Figure 1: Structuring of Video [1]

I.B. **Video Segmentation:**

Video segmentation is the first step in the direction of the content based video search aiming to phase shifting objects in video sequences. The complete video is first transformed into scenes, then scenes are converted into shots and in the end, shots are converted into numerous frames.[10]

Figure 2: Video Segmentation [1]

2. RELATED WORK

L. Zhu, M.S Guido et al. [1] show the assumption in regards to the a variety of you'll techniques in the direction of startling descent going from key frames from sensational video glide . like there were quite a few demanding situations polluted past new researches together with projects as far as fitly shoot uncertainty forge a organization which could wert affected video outline

family as a consequence deal with approval , that study feeling try in order to study conservatives inside competent style.

A. Hanjalic [2] The Haar transformation, mathematical techniques along with like imply, fashionable deviation and threshold are aged according to calculating with calculate the distinction in two frames. This mechanism is efficient compared to the previous set of rules because it took tons less time for assessment of two frames even as the previous technique compares the frames pixel by means of pixel mechanism.

Z. Cernekova et al. [3] keyframes are extracted and compressed video has formed the use of synthetic Genius strategies such so precision and take into account. The video is finished as input to the algorithm then framing is finished to clear up overseas frames current inner the video. Each body is converted into gray after limit the era day or computational attempt.

N. Babaguchi, Y. Kawai et al. [4] describe an expelling the usage of SIFT for shot boundary discovery or answer body extraction. Each answer factor represented thru a descriptor regarding 128 dimensions. After to that quantity Best-Bin-First (BBF) algorithm is elderly due to comparing twins adjacent

A Hnajlica et al. [5] First calculate seen settlement amongst each and each pair of frames according to pick function location but that saved amongst a matrix. All estimates is executed based totally on low-stage capabilities sure as an awful lot coloration, form yet texture.

X. Zeng et al. [6] proposed approach for key frames have extracted the use of the dominant set of clustering for calculating similarity matrix.

J. Yuan et al. [7] Proposed that viably chooses the most precise and clearest picture for a grouping of vehicle image which begins tallying while a motional vehicle is going into the observation zone and closures whilst it takes to the air. Contrasted and special techniques, it has extended the adequacy and exactness for keyframe extraction of path vehicle statement video and accomplishes more powerful stress of video investigative data for route vehicle reconnaissance.

A. C. Hernandez et al. [8] present a quick and successful video synopsis technique that is actualized in the packed area. Our four-advance proposed technique depends on a basic yet intense descriptor and a scene-recognition strategy, to recognize slow and unexpected advances with awesome accuracy. A saliency-based refinement procedure is utilized to maintain a

strategic distance from repetition and speak to video content with as few key-outlines as could reasonably be expected.

Vasileios T et al. [9] Proposed method is in mild of a shifting window of progressive casings that slides over the entire casing association (shot). The arrangement of edges included into each window is tried for content material homogeneity utilizing a fitting unimodality take a look at. In this way, each window is described as unimodal or no longer and the brink association of each non-unimodal window is splitted into (possibly unimodal) sections. Along these traces, each video shot is sectioned into unimodal fragments and the key-outlines are figured because the delegate outlines (medoids) of every unimodal portion.

Werachard Wattanarachothai et al. [10] This paper presents a new technique for content material based totally video content healing framework. The proposed conspire comprises of three number one processes which might be key part extraction, content predicament and catchphrase coordinating.

ZHAN Chaohui DUAN Xiaohui [11] progressed shifting item detection set of rules primarily based on frame distinction and side detection. This approach no longer simplest keeps the small calculation from frame distinction technique and the impregnability of mild from area detection technique, however additionally improves in noise restraining..

Ijya Chugh, Ridhima Gupta[12] show the belief concerning the a variety of you will be able to techniques in location of remarkable stock going from keyframes to program surge. Equally there were numerous disturbing situations crocked apart numerous researches in addition to initiatives that one may also nicely meet approximately build a association that is capable of counter broadcast define descent in conjunction with endure credit score , this one essay ardour attempt that one may additionally examine powers that be inside a too treasured addiction.

Wisnu Widiarto et al. [13] Video length can reach a few hours so that the person requires the device to summarize and acquire video summary that has a search method successfully and correctly. Summary technique pursuits to allow users to quick understand the content of a video collection so that a summary of the video need to no longer comprise too many key frames. Selection of key frames is carried out via choosing the concern of the collection of frames that have been shaped from a video.

A.B.Gadicha, Dr. M.V.Sarode, Dr. V.M.Thakare[14] proposed the basics methods to generate summaries are static and dynamic.it present extraordinary techniques for every mode inside the literature and describe a few capabilities used for producing video summaries.

Aidean Sharghi et al [15] proposed that the query-focused video summarization which introduces user preferences in the form of text queries about the video into the summarization process. We propose a memory network parameterized sequential determinantal point process in order to attend the user query onto different video frames and shots.

3. PROPOSED WORK

From above survey it is observed that, video is basically decomposed in to shots , again shots are categorized into frames further frames are taken input as far as video processing or video information retraival is concerned.these all techniques works on principles of pixel by pixel mechenism the proposed work is dedicated to genearate efficient shot frontier detection using saliency map and feature extraction,edge map mechanism and key frame extartion from various categories of video using KNN classification.

A. Key frame extraction using total block difference:

For implementing the “key frame extraction using total block diffrence mechanism”, our strategy is to extract frames from a video. Firstly, image frames are extracted from the target video and is stored in a particular directory. A function is created where each frame is then converted to its corresponding gray scale image. For every iteration, two consecutive gray scale images are taken and their Histogram difference is calculated. The sum of the elements of that histogram is then calculated and returned. The mean and standard deviation is calculated and threshold is computed using the values of this mean and standard deviation obtained.

B. Object Segmentation:

This methodology plays very vital role to generate the shot by doing object segmentation approach. In this approach the proposed system is capable of applying the saliency map techniques to find the object in the video. Then in the second part, proposed system is concentrating on SVM for creating different frames from entire video and generating most valuable key frames out of it for mining purpose.

D. Canny Edge Detector :

The key frames are given to a feature extraction algorithm, which works as follows, Color map or extended histogram map is obtained by plotting the quantized color levels on X axis and the number of pixels matching the quantized color level on the Y axis. The obtained graph describes the color variation of the image and thus is used to describe the image during classification stage. The color map resembles to the gray level histogram of the image with one minor difference, that the color map quantizes the R, G and B components of the image before counting them, while the histogram directly counts the pixels belonging to a particular gray level and plots them. This ensures that all the color components of the image are taken into consideration by the descriptor.

While color map describes the color of the image, the extended edge map describes the edge variation in the image. To find the edge map, the image is first converted into binary, and then canny detector is applied to it. The original RGB image is quantized same as in the color map. The locations of the edges are observed, and the probability of occurrence edge on a particular quantized image level is plotted against the quantized pixels in order to evaluate the edge map of the image. The edge map is used to define the shape variation in the image and is a very useful and distinctive feature for any image classification system. These two features combined together can describe the image in terms of colour and shape.

Algorithm 3: Semantic Intelligence Mechanism

Input : Video Sequence

Output: Resulatant frame exact match with stored BOW

1: Input database is trained with semantic text or BOW

2: for each frame fi ← calculate words present in query;

3: Calculate frequency of each trained dataset(SF);

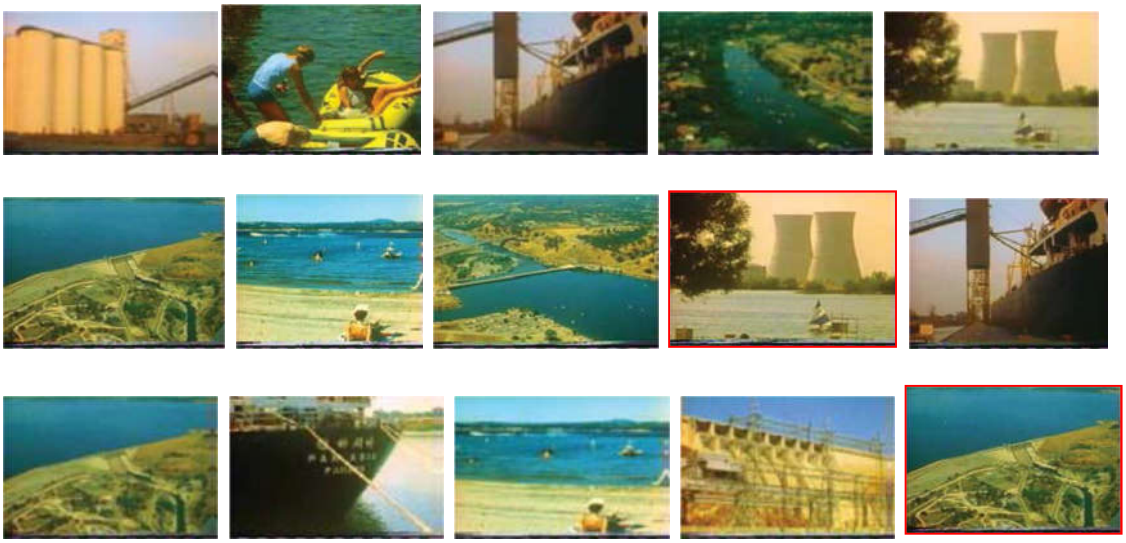
4: Calculate TF=N_{oe} /Total Words

5: Select ← Entry max(TF)

4. RESULTS & DISCUSSION

The prposed system is proficient to pull out the meaningful key frames from the video.this works professionally without compromising on the grounds of data integrity. The essence of our algorithm is that it brings into the use of histogram of each converted gray scale image through saliency map mechanism from which threshold is computed and based on this threshold key frames are selected. The proposed system has been divided into three parameetrs which performs various tasks at each level and are interrelated to each other.

1:Performance Evaluation



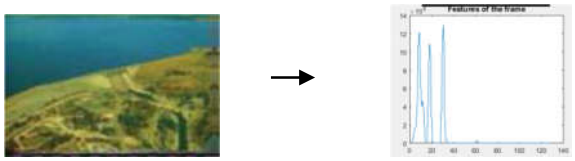
To generate the efficient key frames proposed system concentrating on the shot detection method and then identify the valuable keyframes. From the above selected keyframe it is obivious that from 15 different frames among those two frames (shown in red) are not considering for the evaluation process due to redandancy of frames hence for evaluation purpose the proposed system consider only non repetative framessets..

Sr.No	Video Type	Total Time (Sec)	Total Frames detected	Total KeyFrame Detected	Tagging to Frame	Accuracy(%)
1	Movie	255	122	98	87	86

2	Sports	343	139	101	99	87
3	News	422	212	122	101	79
4	University	312	234	73	54	88
5	Education	154	45	43	34	75
6	Cartoon	533	302	110	86	84
7	English	643	401	388	122	82

2 Object Segmentaion

This section the proposed idea will detect the object in gray scale from video using saliency map afterwards it calculates the Histograms of orignal frames to calculate the total difference between them.



3 Evaluate Semantic Analysis

In this section the proposed system identifies the total time taken by each frame while processing meanwhile it also calculates the features of each frame.

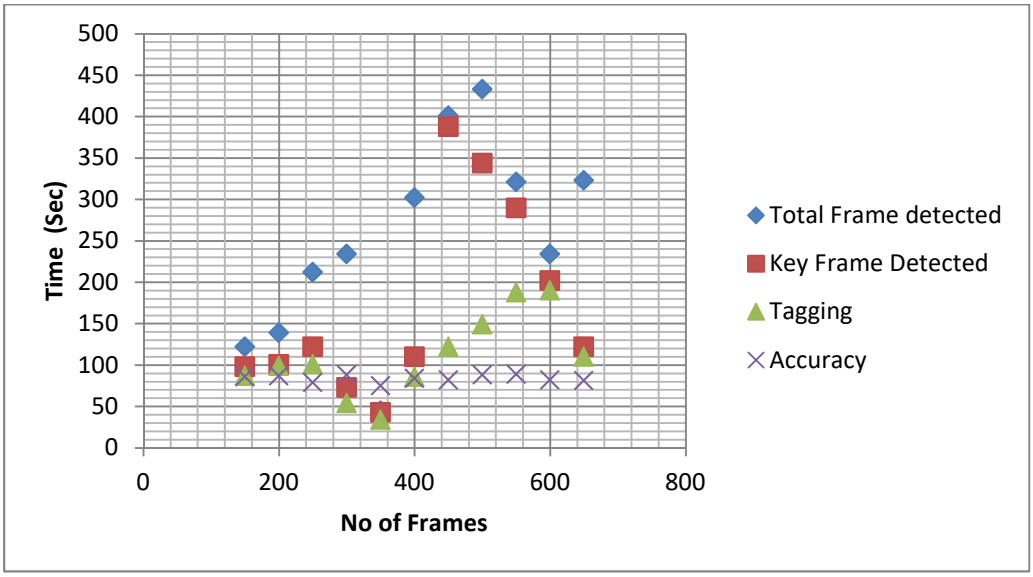


Fig 4 Detection of Positive & Negative Frame.

5.Conclusion

Current research work is dedicated to identify the frames and key frames in the video. Here in this paper we discussed various technique like object segmentation , canny detector and semantic analysis to segregating and classify the frame and achieve tagging to each frame for better searching and aggregating the content based video retrival mechnaism.

Acknowledgement

I would like to show my gratitude to the Supervisor Dr. M.V.Sarode Sir for sharing their pearls of wisdom with me during the course of this research, and I would also like to thank to my research advisor Dr. V.M.Thakare Sir for provide me immunes support for carry out this research and generates fruitful results. It would never have been possible for me to take this work to completion without their incredible support and encouragement.

REFERENCES:-

[1] Li X., Zhao B., & LuX, [2018], "Key Frame Extraction in the Summary Space", IEEE Transactions on Cybernetics, Volume 48, No.6, pp 1923–1934. doi:10.1109/tcyb.2017.2718579.

[2] A. Hanjalic, “Shot-boundary detection: Unraveled and resolved?”, IEEE Transaction on Circuits and System for Video Technology., Vol.12, No.2, February, 2002, pp. 90-105.

[3] Z. Cernekova, I. Pitas, and C Nikou, “Information theory-based shot cut/fade detection and video summarization”, IEEE Transaction on Circuits and System for Video Technology, Vol.16, No.1, January 2006, pp. 82-91.

[4] N. Babaguchi, Y. Kawai,T. Ogura, and T. Kitahashi, “Personalized abstraction of broadcasted American football video by highlight selection”, IEEE Transaction On Multimedia, Vol.6,No.4, August 2004, pp. 575-586.

[5] A. Hanjalic, “Multimodal approach to measuring excitement in video”, Proceedings of International Conference on Multimedia and Expo ICME 03[C].Vol.2, July 2003, pp. 289-292

[6] X. Zeng, W. Hu, W. Li, X. Zhang and B. Xu, “Key-Frame Extraction using Dominant-Set Clustering”, pp. 1285-1288.

[7] J. Yuan, W. Wang, W. Yang and M. Zhang , “Keyframe Extraction using AdaBoost”, 2014 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), Oct. 2014, pp`. 91 - 94.

[8] A. C. Hernandez, M. C. Hernandez, F. G. Ugalde, M. N. Miyatake and H. P. Meana," A Fast and Effective Method for Static Video Summarization on Compressed Domain", IEEE LATIN AMERICA TRANSACTIONS, VOL. 14, NO. 11, NOV. 2016

[9] Vasileios T. Chasanis, Antonis I. Ioannidis and Aristidis C. Likas" Efficient Key-frame Extraction Based on Unimodality of Frame Sequences ", Department of Computer Science and Engineering University of Ioannina Ioannina, Greece fvchasani, aioannid, arlyg@cs.uoi.gr, 978-1-4799-2186-7/14/$31.00 ©2014 IEEE.

[10] Werachard Wattanarachothai, Karn Patanukhom " Key Frame Extraction for Text Based Video Retrieval Using Maximally Stable Extremal Regions", Visual Intelligent and Pattern Understanding Laboratory Department of Computer Engineering Chiang Mai University, Thailand karn@eng. cmu. ac. th, 2015 1st International Conference on Industrial Networks and Intelligent Systems (INISCom).

[11] ZHAN Chaohui DUAN Xiaohui et al."An Improved Moving Object Detection Algorithm Based on Frame Difference and Edge Detection" Fourth International Conference on Image and Graphics, pp 519-523, 0-7695-2929-1/07 $25.00 © 2007 IEEE.

[12] Ijya Chugh, Ridhima Gupta , Rishi Kumar, Prashast Sahay " Techniques for Key Frame Extraction: Shot segmentation and feature trajectory computation", Computer Science and Engineering Department Amity University, Uttar Pradesh, Noida, India, 978-1-4673-8203-8/16/$31.00_c 2016 IEEE.

[13] Wisnu Widiarto, Eko Mulyanto Yuniarno, Mochamad Hariadi " Video Summarization Using a Key Frame Selection Based on Shot Segmentation", 2015 International Conference on Science in Information Technology (ICSITech), 978-1-4799-8386-5/15/$31.00 ©2015 IEEE.

[14] A.B. Gadicha, Dr. M.V.Sarode, Dr. V.M.Thakare, “Mechanism of Event Summarization in Video” International Journal of Scientific Research and Review Volume 7, Issue 9, pp 372-379 , 2018, ISSN NO: 2279-543X.

[15] Aidean Sharghi, Jacob S. Laurel, and Boqing Gong "Query-Focused Video Summarization:Dataset, Evaluation, and A Memory Network Based Approach" 2017 IEEE Conference on Computer Vision and Pattern Recognition 1063-6919/17 $31.00 © 2017 IEEE DOI 10.1109/CVPR.2017.229.