

An Approach for Analyzing keyframes based on Self Adaptive Threshold and Scene Descriptors

Suruthi.K¹, Babu Kannan.K², Kumaresan.A³, Velu.S⁴

^{1,2,3,4}Computer Science and Engineering, ¹SKP Engineering College,
^{1,2,3,4}Tiruvannamalai, ^{1,2,3,4}Tamil Nadu, ^{1,2,3,4}India,

Abstract : In this paper, we propose a CBVR (content based video retrieval) method for retrieving a desired object from the abstract video dataset. Recording and storing enormous surveillance video in a dataset for retrieving the main contents of the video is one of the complicated task in terms of time and space. Even though, methods are available for retrieving the main content of a video based on ROI as well as threshold values for retrieving background information key frames, determining the threshold values manually is a complex scenario. So, we propose a method, where we use self-adaptive threshold for determining the background information along with the use of several descriptors to increase the efficiency of determining the contents of the key frames. We can also use CBVR to retrieve the information of a desired object from our abstract dataset.

Keywords : Self-adaptive threshold, Keyframes, Descriptors, CBVR

I. INTRODUCTION

The process of providing security plays a major role in all organizations these days. This security can be provided in many ways considering the criticality of the information being secured. These security methodologies include providing manual guards around the perimeter or providing electric fence around the infrastructure or any other available effective means of technology available. In spite of the availability of these methodologies, an effective and 24x7 security could be provided with the help of installation of cameras at the crucial areas of an organization which should be out of reach for the humans. The optimal number of cameras to be installed in an environment could be calculated with respect to [1]. Since these cameras are recording videos with a time scale of 24 hours, the recorded videos are to be stored and analyzed where storing these videos require an enormous database and analyzing these videos require humans to play through the entire video in order to analyze the incidents occurred where the biggest de-merit is that we cannot skip the videos being played since we would miss the important actions when we skip. So, we are in need of a method for

extracting the essential events been occurred from the prolonged surveillance videos and storing these events alone in a separate database which would minimize the memory space being utilized for data storage along with minimization of human work to look through the entire videos. We know that the first step in observing videos is to convert it into individual frames or images since the broadcasting of moving visual images form a video. This can be termed as image retrieval.

Image retrieval is the process of retrieving images from an enormous database based on the metadata added to the image which could be said as the annotations. But this annotations have some demerits. Annotating images manual is a time consuming work to be done and if images are annotated ambiguously, the user would never get the required results no matter the number of times he search the image database. Several methods for automatic image annotations have been under research due to the advancement in the field of semantic web and social web applications. In spite of the advancements, there is an effective methodology termed CBIR (content based image retrieval), in which feature extraction is basis. These features represent text based features representing keywords as well as annotations whereas visual features correspond to color, texture and faces along with shapes [2]. Since, features plays a major role here, when user inputs an input image, the pixel value of these images are compared with all the images prevailing in the database and the results given to the user would contain all the images containing a part of the queried image which is an effective way of avoiding annotations to avoid ambiguity. Since we are dealing with videos here, we need an advanced approach from CBIR.

II. RELATED WORK

Since we are dealing with extracting information from frames, we are in need of looking into the previous

researches done in this area in order to obtain the state of the art technique available. Siva Kumar Avula et al have worked on frame based video retrieval using video signatures [3]. The authors have designed a video similarity measure algorithm in which the representation of the color features in the video is done by a static size representation termed as the video signature. YCbCr Histogram, along with its mean values are used for the computation of these signatures. The computed video signatures are employed for the purpose of figuring out the videos having similar frames. Though this implementation is done for avi video files alone, this could be implemented for any video file irrespective of its format or compression methodology. The authors have hinted the future work to be focused on implementing this scenario at LAN level considering only the matrix form of the video files which can be removed from the system after processing where the essential extracted information could be stored in logs or any other file for reference. B.V. Patel have researched on CBVR systems and their survey have revealed that this era is gaining importance on systems working on content based retrieval rather than systems based on text based retrieval [4]. The features to be extracted are important as the output data solely depends on the features that we take into account. The authors have also considered the fact that the time and space complexities also plays an important role in considering the features to be extracted. The authors have also considered the common issues in the prevailing research in this area. The concept of video browsing systems based on compressed domain feature extraction have been researched by Ajay Divakaran et al [5]. A video browsing system have been designed by the authors using which the essential features and metadata could be extracted dynamically from the compressed video. The proposed methodology for feature extraction is fast and it also makes use of motion and color descriptors for efficient extraction of features from compressed video. Experimental results of the authors have proved that the combination of color and motion descriptors is simple which can provides accurate results. Changick Kim et al have worked on object based video abstraction for video surveillance systems [6]. The authors have proposed an video object segmentation (VOS) system which is used for video abstraction based on objects. Since keyframes are the vital part of an image as it represents the content of an image in an abstracted manner, the concept of 'keyframe' is been replaced by video object plane (VOP) in the proposed methodology. Video objects which are semantically meaningful are to be directly accessed from the video sequences in order to obtain an online video

surveillance system based on objects. For the extraction of key VOPs , we are in need of suitable descriptors which can extract key VOPs considering the shape and context of the image. So the authors have employed the region based descriptors. These key VOPs are sequentially extracted and the experimental results have proved that the proposed framework works well in summarizing the surveillance videos. Changsheng Xu et al have worked on semantic event detection in broadcast sports videos [7]. The prevailing video event detection scenarios faces a serious problem such as extracting huge semantic information from the video based on computer vision. In order to overcome this, the authors have proposed a method for sports video content extraction based on aligning the broadcast video with webcast text. The authors first get the webcast text from the internet and then analyze it and detect text events using probabilistic latent semantic analysis (pLSA). The clustering of video event with the text event is done by the application of conditional random field model (CRFM).

III.FAST CLUSTERING METHOD BASED ON ROI

Since users find easy to access online videos easily these days, we are in need of finding an effective way to store and maintain enormous amount of video files facilitating easy and quick access for multiple users. In order to support research in this area, Guang-Hua-Song et al have proposed the fast clustering based on the region of interest (ROI). The authors have employed the average histogram algorithm for the purpose of extracting key frames from each shot. A shot could be defined as the depiction of a particular scene or action. A single shot refers to the action covered by a camera between the start and stop of the recording time which would be normally in the same angle. The extracted key frames are used for the generation of edge maps which contribute the next step in the video abstraction scenario. Based on the above methodologies, the authors have determined the key points. Calculation of threshold values from the respective key frames would be the next step which is done for the purpose of expanding and identifying the area surrounding the key points [8]. The authors have proposed the observation of main content in each of the key frame based on the threshold values defined and the concept of key points. As the final step of their proposed method, they have utilized the ROIs of the key frames and have performed the fast clustering method on them. The different methodologies involve before implementing the fast clustering method along with the implementation of

fast clustering methodology is explained in the following sections.

A. Keyframe Extraction

The representation of video sequence would be in the form of a hierarchical structure considering the scene, shot and frame contributing different levels on the hierarchy [9]. Different researches on video sequences requires the researches to deal with the different levels of the video sequence hierarchy with respect to the information needed for their research. Shot is to be considered first for the purpose of key frame extraction. The shot level is chosen at the hierarchy among the other available levels due to certain reasons. The sequence of video frames captured continuously by a camera contributed a shot which also would include the moving objects, panning and zooming in terms of the recording camera. We also have a greatest merit with the shot as the two adjacent shot does not have the same content which would obviously eliminate redundancy. The authors have employed the use of algorithm proposed in [10] for the purpose of extracting key frames. The key frame extraction process also involves the average histogram method. A shot $S = \{ \}$ of length n is assumed. The k th frame in the assumed shot is represented as F_k . Considering F_k to be the gray level histogram containing L bins could be generated from frame, whereas the calculation of the average histogram H is done based on the following formula:

$$H(i) = \frac{1}{n} \sum_{k=1}^n H_k(i) \quad (i = 0, 1, \dots, L - 1)$$

Where $H_k(i)$ represents the value of the i th frame of frame k . After the extraction of key frame, ROIs are generated by adopting a series of key frame analysis this process is followed by saliency map generation and edge map generation.

B. Edge Map Detection

It is a general concept that we would focus on objects which has a whole shape in the video. So there would be edges within these components. We are in need of determining the key points which would be available inside the objects and so determining edges would make our tracking process easier. The authors have used the canny edge detection scenario with respect to [11]. This process is followed by the location of key points and generation of ROI.

C. Fast Clustering

In a video sequence, though each shot would be having a different content to portray, some of the shots may look similar to one another in camera angle or facial expression of the people involved or in any other means. Sometimes, a shot would be manually segmented into many shots and used at different places in a video sequence. The approach of the authors is to make the video sequence compact and thus they have clustered the key frames in order to avoid the redundant frames.

Normally, clustering before the entire process of extracting the key frames is done would be of no use since the new frames could not be taken into account. In order to overcome this traditional approach, the authors have used fast clustering in which clustering process starts once the key frame extraction and identifying ROI are done. Even though this approach was good enough to an extent, the authors have not used more effective descriptors to extract more features from the frames for better observation. In addition to this manually setting the threshold to obtain the background information would not be so effective.

IV. APPLICATION OF SELF ADAPTIVE THRESHOLD AND DESCRIPTORS

Though the use of assigning the threshold manually works in a better way, setting the threshold manually is a difficult task. So we are in need of an alternate way for setting the threshold which is the adaptive threshold methodology. We propose the use of adaptive threshold in our video abstraction method for the purpose of gaining more knowledge about the objects in the background. In addition to this, we have also made use of several descriptors such as *FCTH (Fuzzy Color and Texture Histogram)* and *SCD (Scalable Color Descriptor)*. A descriptor is generally used for extracting different kinds of features from an image based on the functionality of a descriptor. Features refers to the different kinds of information that could be extracted from an image which may refer to the color, intensity, pixels, etc. The methodology follow for displaying the image similar to the one given as the input is shown in figure.1 as a block diagram.

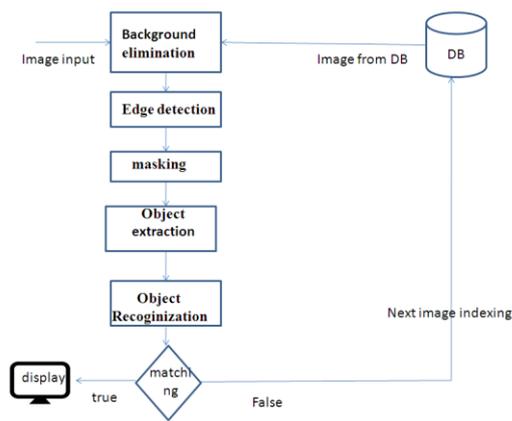


Fig.1 Block diagram for displaying the matched image.

The functionality of FCTH and SCD are discussed as follows:

A. FCTH (Fuzzy Color and Text Histogram)

In this type of descriptor, fuzzy is used for gathering information about colors which lie between the pure black and pure white. Here, fuzzy is made used of since the general concept of fuzzy is to deal with all possible scenarios (partial true / partial false) which lies between the True (1) and False (0) values.

B. SCD (Scalable Color Descriptor)

SCD is used here for the purpose of extracting information about the colors which are scalable. This scalable colors represent colors which are extended to the nearby boundaries and would be available in a different form within that boundary.

C. Algorithm: Distance Vector

We are using *Distance Vector* algorithm in this video abstraction process for the purpose of observing the distance travelled by an object in two subsequent frames in order to determine the motion of the object in a more likely scenario which involves the following steps:

- Detecting and identifying the boundaries of the moving objects.
- Extracting ROI (region of interest) of the object within the frame.
- Searching for the same object in the next subsequent frame.

- Detecting boundaries and location of the object.
- Comparing the location of the object and finding its distance moved from the previous frame to the current frame.
- Repeating the above steps for all the video frames would enable us to find the moving object distance covered for each frame.
- Updating the distance vector matrix.

This scenario is applied for minimizing the memory complexity in terms of storing and retrieving enormous 24x7 surveillance videos where recording and storing of the entire video would increase the demand of memory as well as looking through the entire video to verify a crime scene would be a more complex scenario. In order to overcome this complexity, our method extract the key frames from the entire video and store it in a desired database where only the distinct images would be available minimizing the work of the user to look through a full length video. In addition to that, saving images would have a memory demand much lesser than the demand of the videos. Since we are using descriptors, more detailed information could be extracted from the images. Self-adaptive threshold enables the user to get more details above the objects available in the background which is an added advantage of this methodology. Any sort of frame can be given as a query into the system and the user would get the relevant video containing the respective key frame. If the frame is not available in any of the dataset, user would be shown with an error prompt. This process is termed as CBVR. CBVR is similar to CBIR but differs in a way that user would be given a frame (image) as a result in case of CBIR whereas result would be the entire video in case of CBVR. But in both the cases, data is compared and retrieved based on the contents available in the frames.

V. EXPERIMENTAL RESULTS

We have conducted our experiment with videos available in the MATLAB dataset. First step would be the extraction of key frames based on self-adaptive threshold value which is shown in Figure 1.

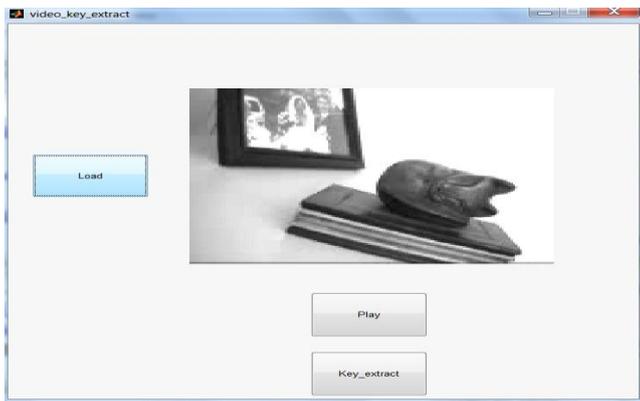


Fig 1. Window for Key frame Extraction

Key frames are extracted and stored in a destined folder as shown in the Figure 2.

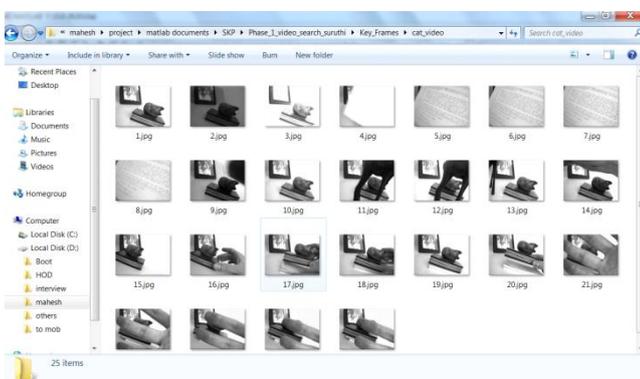


Fig 2. Key framesStored in the Destined Folder

After the key frame extraction, the user can input a key frame of their choice and the contents of all the available videos in the dataset are compared and the respective video containing the requested key frame would be found based on CBVR and retrieved as shown in Figure 3a. The user can click on the play button available at the bottom right to play the entire video containing the requested key frame. If the requested frame is to found, the user would be prompted with an error message as shown in Fig 3b.

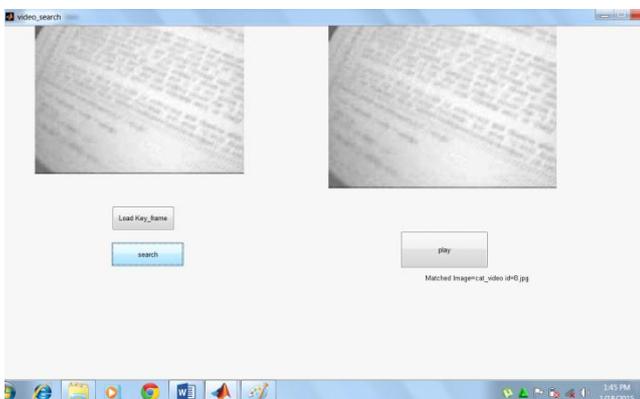


Fig 3a. Video is retrieved on the queried key frame using CBVR based

Our experiment have showed a compromising result with more than 80% accuracy. As explained above, this methodology can decrease the memory space demands and the time of the user to spend in looking through the entire videos.

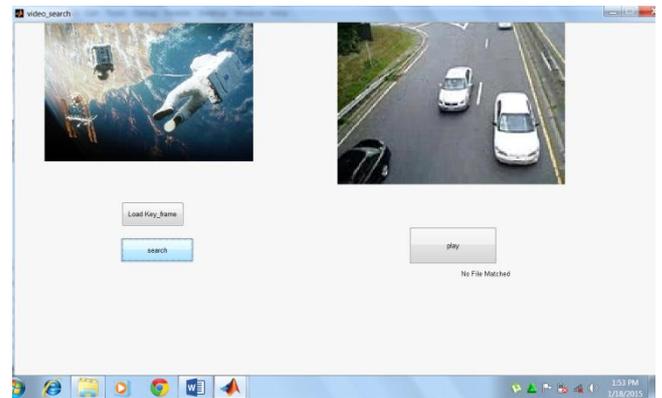


Fig 3b. User id prompted with an error message since the requested frame is not found

VI. CONCLUSION

In this paper, we have proposed a methodology for video abstraction based on several descriptors and self-adaptive threshold. This methodology facilitates user to minimize the memory demands and time demands for looking through the videos. Our methodology also makes use of CBVR for retrieving a video based on the contents with respect to the user requested key frame. The only problem that our methodology faces is the time taken for comparison if the key frame to be searched is available in the final video available in the dataset. Our future work is to concentrate on limiting the time space for comparison in a large video dataset.

VIII. REFERENCE

- [1]Junbin Liu, *Member, IEEE*, Sridha Sridharan, *Senior Member, IEEE*, Clinton Fookes, *Member, IEEE*, and Tim Wark, *Senior Member, IEEE*, "Optimal Camera Planning Under Versatile User Constraints in Multi-Camera Image Processing Systems", *IEEE TRANSACTIONS ON IMAGE PROCESSING*, VOL. 23, NO. 1, JANUARY 2014
- [2]Yong Rui and Thomas S. Huang, "Image Retrieval: Current Techniques, Promising Directions, and Open Issues" in *Journal of Visual Communication and Image Representation* **10**, 39–62 (1999)

[3]Siva Kumar Avula and Shubhangi C Deshmukh, "Frame based Video Retrieval using Video Signatures", International Journal of Computer Applications (0975 – 8887) Volume 59– No.10, December 2012

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

[5] Ajay Divakaran, Anthony Vetro, Kohatro Asal and Hirofumi Nishikawa, "Video browsing system based on compressed domain feature extraction", IEEE Transactions on Consumer Electronics, Vol. 46, No. 3, AUGUST 2000.

[6] Changick kim, member, IEEE and Jeng-Neng Hwang, fellow, IEEE, "OBJECT-BASED VIDEO ABSTRACTION FOR VIDEO SURVEILLANCE SYSTEM", IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL.12, NO.12, DECEMBER 2002.

[7] Changshang Xu, senior member, IEEE, Yi-Fan Zhang, Guangyu Zhu, Yong Rui, SENIOR MEMBER, IEEE, HANQING LU, SENIOR MEMBER, IEEE AND QINGMING HUANG, SENIOR MEMBER, IEEE, "USING WEBCAST TEXT FOR SEMANTIC EVENT DETECTION IN BROADCAST SPORTS VIDEO", IEEE TRANSACTION ON MULTIMEDIA, VOL.10, NO.7, NOVEMBER 2008.

[8] Guang-hua song, Qing-ge ji, Zhe-ming lu, Zhi-dan fang, Zhen-hua xie, senior member, IEEE, Yi-Fan Zhang, Guangyu Zhu, Yong Rui, "A NOVEL VIDEO ABSTRACTION METHOD BASED ON FAST CLUSTERING OF THE REGIONS OF INTEREST IN KEY FRAMES", INT.J.ELECTRON.COMMUN.(AEU)68(2014) 783-794.

[9] Jameel T, Gilani SAM, Mumtaz A, "CONTENT BASED VIDEO RETRIEVAL FRAMEWORK USING DUAL TREE COMPLEX WAVELET TRANSFORM", INNOV ADV TECHN COMP INF SCI ENG 2007:449-54 I.

[10] Bescos J, Cisneros G, Martinez JM, Menendez JM, Cabrera J. "A unified model for techniques on video shot transition detection". IEEE Trans Multimedia 2005;7(2):293–307.

[11] Canny JF. "A computational approach to edge detection". IEEE Trans Pattern Anal Mach Intell 1986;8(6):679–98.



Kumaresan.A was born in Pondicherry, in 1981. He received the Bachelor Engineering from VRS College of Engineering, Tamilnadu, India and Post Graduate Degree in Computer Science and Engineering from SKP Engineering College, Tamilnadu, India. He is having more than 9 years of Teaching and Research Experience.



Velu.S was born in Cuddalore, in 1992. He received the Bachelor Engineering from Krishnaswamy College of Engineering, Tamilnadu, India and Post Graduate Degree in Computer Science and Engineering from SKP Engineering College, Tamilnadu, India.



India.

Suruthi.K was born in Thiruvannamalai, in 1991. She received the Bachelor Engineering from The New Royal College of Engineering and Technology, Tamilnadu, India and Post Graduate Degree in Computer Science and Engineering from SKP Engineering College, Tamilnadu,



Tamilnadu, India. He is having more than 9 years of Teaching and Research Experience.

Babu Kannan.K was born in Tindivanam, in 1982. He received the Bachelor Engineering from Mailam Engineering College, Tamilnadu, India and Post Graduate Degree in Computer Science and Engineering from Arulmigu Meenakshi Amman College of Engineering,