A STUDY ON AN APPROACH FOR SURVEILLANCE VIDEO SUMMARY BASED ON OBJECTS

Ajay Gopichand Barsagade\textsuperscript{1} and Dr. A.S. Alvi\textsuperscript{2}

\textsuperscript{1}M.E (Pursuing), Computer Science and Engineering PRMIT, Badnera, Amravati, India
\textsuperscript{2}Professor, Computer Science and Engineering PRMIT, Badnera, Amravati, India

Abstract - Video summary is an effective technique to provide a compress representation of the original video by neglecting spatiotemporal redundant activities and by keeping the only important activities. Most recent technics for video summary will cause collisions among active objects, especially when the video is condensed much. In this project, we present an approach for video summary to reduce the clashing between objects. Our approach first shifts active and important objects along the time axis to compress the original video. Then, the actual size of the objects are minimize when clashing occurs. In between, the geometric centroids of the objects will be remains unchanged to secure the spotting information. Our contributions will be in three phase. First, an approach is proposed to minimize the clashing in the summary video through reducing the sizes of the objects. Second, some suitable framework is developed to indicate the optimal time position and the appropriate trimming of coefficient for each object. Finally, some mathematical metrics are proposed, and some work will carry out to evaluate the proposed approach.

Keywords - Video summary, surveillance, reduce collision, reduce size, optimization.

I. INTRODUCTION

NOWADAYS, with the development of our society, the number of surveillance cameras is increasing continuously and most of these cameras work 24 hours a day. Thus, the amount of video data grows at an incredible speed which makes video browsing laborious and time consuming. In most cases, surveillance videos contain too much redundant information, and the interesting activities are very limited. Therefore, many surveillance videos are never reviewed after they were produced. Video summary or synopsis, a shorter condensed representation with important events preserved, is more preferable. The video synopsis can be an efficient index of the original video by calculating and displaying the original time of each object. According to the time label, the interesting objects can be found easily from the original video. In addition, video summary or synopsis techniques can also reduce memory storage for surveillance video data sets when it is not necessary to store the entire raw video. So far there is no unified quantitative measurement for video synopsis to determine whether the synopsis is good or bad, as the quality of the synopsis video highly depends on the content of the original video and the requirement of the application. Besides, different users may have different views, which make the criteria more intuitive. But in general, there are some qualitative standards in video synopsis field. They are as follows:

(1) The output video should preserve the activities presented in the original video as much as possible while reducing most spatiotemporal redundancies.
(2) Collisions among active objects should be avoided to the greatest extent in the output video.
(3) The chronological and spatial consistency of objects should be kept as far as possible.

Fig: A general Framework of Visual Surveillance System

The first step in high level video processing is to identify the all objects available in a scene. The next step is to see how these track down objects shift with respect to each other. The above two
steps combined, can be pronounced as “Object Tracking”. An important function of object tracing is video surveillance. Airports, Railway stations, malls, religious places, courts and public sector buildings are only some examples of locations where video surveillance has an extremely high priority. In addition to this, astronomy, navigation systems, military, robotics, augmented reality, medical imaging and road traffic regulation are some of the other major applications of object tracking. There are mainly two origins of information in video that can be used to identify objects: basic visual features (such as texture, color and shape) and movement information. A typical procedure is to segment a frame into regions based on shape, color and texture details first, and then merges regions based on common motion subject to certain conditions such as adjacency. Derivation of these two types of information can be done either in the pixel format or in the compressed format. Tracking in videos especially from big databases, requires an enormous amount of processing since the videos are generally saved in a compressed form and must be always decompress each time any processing is perform on the video frames for object tracing. In order to reduce processing time and save processing resources, it is essential that the tracing takes place in compressed domain itself. Surveillance videos are widely used in domains such as access control in traffic monitoring, airports, home door surveillance and human verification. Generally in this application surveillance camera capture detail information over long period of time, resulting in a large amount of data. Such a big amount of data need big amount of memory for storing and large bandwidth requires for their transmission. Therefore an effective compression mechanism is required for the efficient storage and transmission of such big amount of data. Most of the current video compression mechanisms are intended for general purpose video compression i.e. when no belief about camera motion is made.

II. LITERATURE SURVEY

Surveillance systems are meant to monitor, screen and track activities in public places such as airports, in order to ensure security. Many varied aspects like screening objects and people, maintaining the database of potential threats, biometric identification and video surveillance, act in tandem to monitor activity. The spectrum of surveillance applications include access control in special areas, human identification, event detection/recognition, traffic monitoring, patient monitoring at hospitals and activity monitoring at public places like shopping malls, industries, banks and government establishments. A major aspect of any surveillance system is to efficiently compress the huge volumes of recorded video, to facilitate subsequent processing. However, this aspect has not been sufficiently explored as compared to the other modules in the system, among the reported works. The importance of video compression for any surveillance system is manifold. Apart from archival of surveillance video, compression plays a crucial role in networked surveillance systems.

A. Using Scaling Down Objects

Xuelong Li, Zhigang Wang, and Xiaojiang Lu [1] showed that the Video synopsis is an effective technique to provide a compact representation of the original video by removing spatiotemporal redundancies and by preserving the important activities. Most current technics for video synopsis will cause collisions among objects, especially when the video is condensed much. In this approach, they present an approach for video synopsis to reduce the collisions. Their approach first shifts active objects along the time axis to compact the original video. Their contributions are as: Object Detection and Segmentation, in object-based approaches, object detection and segmentation are essential. Aggregated Channel Features (ACF) detection is also employed to extract the moving objects. The ACF detector can produce a precise bounding box for each object. Meanwhile, the bounding box and geometric centroid of each observation are also recorded. Although they try to generate the precise masks of objects, background pixels are included inevitably. Thus, Poisson Image Editing is utilized later to stitch objects into the background. The Optimization Framework, they will introduce the optimization framework in detail which penalizes the loss of activity information, the collision artifacts, the size reduction of an object and the violation of relationships between moving objects. Stitching Objects into the Background, after the optimization process,
objects are stitched into the background to create the synopsis video. For stitching, a background video with user-defined length is generated before adding moving objects into the synopsis.

Figure shows the flow chart of this approach. First, the original video is captured by a surveillance video. Then, a corresponding background is calculated using a temporal median approach over several video clips. After reconstructing the background, moving objects are detected using background subtraction approach and ACF detection. Continuous observations of the same object are generated by a motion-based multiple object tracking approach. Then, objects are shifted and zoomed out according to the result of the proposed optimization framework. Finally, stitch objects into the background using Poisson Image Editing approach.

B. Using Foreground Motion Compensation

R. Venkatesh Babu and Anamitra Makur [2] presented a fast and robust method for moving object tracking directly in the compressed domain using features obtained in MPEG videos. DCT domain background subtraction in Y plane is used to locate candidate objects in subsequent I-frames after a user has marked an object of interest in the given frame. DCT domain histogram matching using Cb and Cr planes and motion vectors are used to select the target object from the set of candidate objects. The target object position is finally interpolated in the predicted frames to obtain a smooth tracking across GOPs.

![Flow chart of the synopsis video creation process](image)

Fig. 2. The overview of the compression technique

This technique presents an object-based compression technique for surveillance videos. Very high compaction rate is obtained by calculating the moving objects with motion compensation. The performance of the proposed compression technique is compared with that of the standard MPEG-1 compression technique. Since the object segmentation algorithm is based on the edge information alone, and the motion compensation is done only for the object region, the proposed algorithm is suitable for real-time video coding.

C. Motion Estimation by Object-Matching

Aishy, Amer Mitiche and Eric Dubois[2] showed that Motion estimation plays a key role in many video applications, such as frame-rate video surveillance, video retrieval, video conversion, and video compression. The key issue in these applications is to define appropriate representations
that can efficiently support motion estimation with the required accuracy. In this, a low-complexity object motion estimation technique is proposed that is designed to fit the needs of high-level video representation such as in video surveillance or retrieval applications. This is a method to estimate object motion based on the displacements of the object MBB (minimum bounding box). MBB-based motion estimation is methods use the displacement of the centroid of the MBB. This is sensitive to noise and other image artifacts such as occlusion. Besides, many MBB motion estimators thought translational motion when motion type can be important data as in other object-based video retrieval, for instance. The author contribution is in the detection of the type of object composition, motion translation, scaling, and the subsequent estimation of one or more motion values per object depending on the detected motion type. Special consideration is given to at image margin and object motion in interlaced video. Also, proposed analysis of displacements of the MBB-sides allows the estimation of complex motion as when objects move towards or away from the camera. Object Segmentation and matching: Segmentation is realized in four steps [4,5]: binarization of the input gray-level images, morphological edge detection, contour analysis, and object labeling. Contour analysis transforms edges into contours and uses data from previous frames to adaptively eliminate noisy and small contours. Small contours remain completely inside a big contour are merged with that big contour according to a spatial similarity criterion [1]. The elimination of small contours is spatially adapted to the homogeneity criterion of an object and temporally to corresponding objects in previous images.

Figure 1: The motion estimation method.

**Matching**: Object matching is achieved by matching single object features and then combining the matches based on a voting scheme. Such feature-based solutions need to answer some questions concerning feature selection, monitoring, correction, integration, and filtering [1]. Feature selection schemes define good features to match. Feature monitoring aims at detecting errors and at adapting the matching process to these errors. Feature correction aims at compensating for segmentation errors during matching, especially during occlusion. Feature integration defines ways to efficiently and effectively combine features. Feature filtering is concerned with ways to monitor and eventually filter noisy features during matching over time.

**D. Using Background Subtraction and Motion Estimation**

Ashwani Aggarwal, Susmit Biswas, Sandeep Singh, Shamik Sural, and A.K. Majumdar [6] presented a fast and robust method for moving object tracing directly in the reduced domain using features available in MPEG videos. DCT domain background subtraction in Y plane is used to find out candidate objects in the I-frames after a user has notified an object of interest in the present frame. DCT domain histogram matching using Cb and Cr planes and motion vectors are used to mark the target object from the set of available candidate objects. The target object position is finally included in the concluded frames to obtain a smooth tracking across GOPs.

They proposed a novel tracking method that effectively tracks objects in a compressed video by combining background subtraction and motion estimation. The system consists of four main components: background subtraction, candidate object identification, target object
selection and object interpolation. Although they work strictly in the compressed domain, they still get high levels of accuracy with minimal processing time and computational cost. Several issues may further be addressed. This includes handling of full occlusions, fast camera motion, multiple object tracking and unsupervised tracking of objects.

E. Path Estimation and Motion Detection of Moving Object

Ritika and Gianetan Singh Sekhon [7] presented an efficient and effective approach for identifying and tracking of moving object from a video. A video is captured by static camera. Moving object tracing and identified from video sequences has applications in many areas such as video indexing, automatic video surveillance, human-computer interaction, motion-based recognition, vehicle navigation and traffic monitoring. In this work, she presents a computer vision-based approach for object tracking and detection. A method is proposed to identify and track moving object through video even if background motion is changed at any instant and capable of plotting a 3D graph mesh based on the moving object in between any number of frames per second. They use consecutive frame analysis technique to detect background changing criteria and use morphological filtering for image enhancement. Finally, they get the co-ordinates for the moving object and these co-ordinates are imported to any other 3D software’s like MAYA etc to analyses or edit the results calculated by the algorithm.

In their system, video is captured with a stationary camera. Their algorithm does not want any manual input for starting object tracking. They plot object tracking in real time and logging the data into memory frame via frame. According to their algorithm first frame remains constant at the end of the video as it is used to calculate the per pixel distance (in mm). They calculate the change in background with changing object and compare each frame with the previous one. But the computational costs of these methods are very high and have problem in stability [8]. The state of each object is calculated by morphology filter. Morphology is calculated in terms of set theory. Sets represent objects in an image; for instance, the set of all white pixels in a binary image is a complete morphological description of an image. They use morphological filtering for image enhancement. Enhancing the contrast among adjacent regions or features; clarifying the image via selective smoothing or approximation of features at certain scales and retaining only features at certain desirable scales [9]. They have proposed tracing of moving object from a video algorithm using successive frame evaluation technique. Video is captured with a stationary camera. An important component in such systems is the ability for a computer to track and identify moving objects. Hence their algorithm is capable to plot the movements of object through videos if the background is changed at any instant. Their objective is to provide software that can be used on a computer for performing object tracing along with video improvement. The algorithm improves the detection and tracking of moving objects in the real time video.

F. Object Based Fast Motion Estimation and Compensation Algorithm

Gopal Thapa, Kalpana Sharma and M. K. Ghose[10] showed that in surveillance systems, the storage requirements for video compression are a major factor because of recording of videos continuously for very long periods of time, resulting in big amounts of data. Therefore, it is essential to apply effective compression mechanism for compressing surveillance video. The mechanism used for the general video compression may not be the efficient and effective technique for the compression of surveillance video because of the use of stand-alone camera as compared to moveable camera in general purpose videos. Generally surveillance video consist of multiple objects, smaller in size as compared to the background and they have frequent occlusion with each other. In this a new object based motion estimation and compensation technique for surveillance video compression is proposed. Background differencing and summing technique (BDST) is generally used for the segmenting of the moving objects. This technique not only identifies moving object but also the maximum distance moved by the object in given group of frames. Their experimental results show that the approach achieves high compression ratios compared to MPEG-2 compression.
The motion activities for particular sub images at different resolution may be different but are highly correlated. The hierarchical search for the motion activities result in fast approximation of motion vectors with minimum computational complexity [12]. Therefore, in this paper, the hierarchical multi resolution motion estimation technique used in [11] has been used for the estimation of motion vector and compensation. The inverse wavelet transform is calculated in the reverse manner, i.e., starting from the lowest resolution sub image, the higher resolution images are calculated recursively. Once the object is segmented from the frames and bonding box is created. Three level discrete wavelet transform is applied to the current frame and the reference frame objects for the motion estimation and compensation. After the motion compensation, the error image is obtained which is then encoded along with the motion vectors using Huffman coding. The block diagram of the proposed algorithm is shown in Figure 1.

III. PROPOSED WORK

Many approaches have been proposed for video synopsis. Some approaches are based on frames, in which an entire video frame is treated as the fundamental building blocks which cannot be decomposed any more. In frame-based approaches, video frames are compressed along the time axis which allows the viewers to watch the summary of the video in a much shorter time. A common frame-based approach, called fast-forward [14], is often used in which several video frames are skipped. Unfortunately, it is prone to lose fast moving objects when skipping video frames. To alleviate this problem, some approaches have been proposed in [15]–[16], where either key frames are extracted according to certain principles or some video clips without activities are skipped. These approaches tried to preserve the most useful information and skip frames of low interest. However, frame based approaches tend to lose the dynamic aspect of the original video and retain large empty spaces of the background in the synopsis [13]. To reduce most empty spaces of the original video, a number of different approaches have been proposed to extract informative video portions from the video volume and montage these portions together [17]–[19]. In the final synopsis video, a frame may contain objects coming from different times. In this case, the stitching seams among different image patches are always very apparent. To address the problem, [13], [20] proposed an object based approach which has become the mainstream for surveillance video synopsis. In object-based approaches, objects are first extracted from the original video volume, and then they are transformed in the temporal domain. Although object-based approaches can reduce spatiotemporal redundancies to a large degree, it will cause undesirable collisions among different objects in the synopsis video. To reduce collisions and further improve the compression ratio, some spatiotemporal analysis approaches were proposed. Nie et al. [20] indicated that there are still some empty spaces in the
Synopsis video where no moving objects spread over, and then shifted the moving objects in the spatiotemporal domain by exploiting both spatial and temporal information. Although these approaches are capable to avoid most collisions, both temporal and location information of objects are violated.

To reduce collisions and keep the original positions of objects unchanged, an object-based video summary approach is proposed.

A. System Block Diagram

B. System Description:

Above figure shows the block diagram of working of surveillance video using motion object estimation. Any video we take as a input video then pre-processing procedure we do on that input video. Video made by number of frames so first we extract the frames in pre-processing procedure. if there is any noise in that video, we removed it. After removal of noise feature extraction process will be start. In feature extraction, we are going to estimation reliable motion of the frames using previous and current frames. Here we are going to find the respective change or difference in between previous and current frame. if the change is more than a certain limit. then it will not reliable. So on the basis of previous and current frame. we will decide threshold value because a small change of the object is also a change and we don’t want to consider this change. if sudden change will occur then will define this change on threshold value. we calculate reliable motion from threshold value. After that post processing procedure will apply on that frame. Using post processing we are going to track the object. Once the object get tracked or detect then anomalous detection procedure will start. In anomaly detection procedure two condition are there, one is abnormal and another is normal. If reliable motion will exceed than a certain level of action then it will detect that respective object or frame. That respective frame or object will notify and display the face of that object and at the same time on reliable motion. It detect that the object’s behavior. In most cases, surveillance videos contain too much redundant information, and the interesting activities are very limited. Video summary, a shorter condensed representation with important events preserved.

IV. CONCLUSION

Video summary is an effective technique to provide a compress representation of the original video by removing spatiotemporal redundancies and by preserving the active activities. The experiments have demonstrated that the summary or synopsis video produced by our approach has much fewer collisions while the compression ratio is high. This paper gives the basic idea about various techniques used for summary of surveillance video. Video surveillance system is very useful in the field of security. Most current approaches for video summary will cause collisions among active objects, especially when the video is compress too much. In this paper, we present an approach for video summary to reduce the collisions. Our approach first shifts active objects along the time axis to compact the original video. Then, the sizes of the objects are reduced when collisions occur. Meanwhile, the geometric centroids of the objects will be kept unchanged to preserve the location information. Our contributions are fourfold. First, study about video processing analysis and its application. Second, develop an algorithm of an approach surveillance video summary based on object. Third, verify and simulate system for different videos. Finally, analysis the system performance parameters evaluate the proposed approach. The main feature of all of above is to ability to compress the surveillance video as small as possible with preserving the active activities only.
REFERENCES


