AP
ijpam.eu

A SURVEY ON SURVEILLANCE VIDEO SYNOPSIS TECHNIQUES

POOVIZHI. S.[1], DR.GANGATHARAN.N.[2], ELAKHYA N.[3]
[1],[2]DEPARTMENT OF ELECTRONICS AND COMMUNICATION
ENGINEERING, R.M.K COLLEGE OF ENGINEERING AND
TECHNOLOGY, CHENNAI, TAMIL NADU, INDIA
[3]DEPARTMENT OF ELECTRONICS AND COMMUNICATION
ENGINEERING, S.S.N COLLEGE OF ENGINEERING, KALAVAKKAM,
TAMIL NADU, INDIA

ABSTRACT. The rapid growth of digital videos and multimedia technologies is leading to enormous volume of video data being recorded. The main challenge involves processing of the huge amounts of video data within a short period of time. Video synopsis is a smart solution that automatically retrieves and indexes the lengthy surveillance video. It gives a quick abstract of the lengthy video by eliminating redundant information. This paper provides a thorough understanding of traditional video synopsis methods and various video synopsis techniques are reviewed with detailed representation by highlighting the merits and demerits in the surveillance domain. It also offers an insight over new synopsis methods that can be used to meet the ever growing demands in surveillance applications.

Key Words and Phrases: Video Surveillance, Video Synopsis, Video Skimming, Frame, Part Movement, Object Synopsis.

1. INTRODUCTION

At present, the usage of video surveillance systems has increased almost in all places for security purposes. The video camera operates 24 hours a day. Thus the amount of captured video data grows at an incredible speed which makes video browsing tedious and time consuming, so most of the videos are not watched. In most cases, surveillance video contains too much redundant information and the interesting activities are very limited. Video synopsis is a method in which the huge video data is condensed into a smaller

1

volume and the activities are time shifted. This technology detects the objects in motion without losing any key activities. Video Synopsis can be applied for enhancing public safety, security in defense and public sectors. The traditional video summarization methods condense the video data in timeline whereas the conventional video synopsis methods condense both in time and space domain. The main aspects of video summarization are static and dynamic. In summarization method the important frames are considered for reduction.
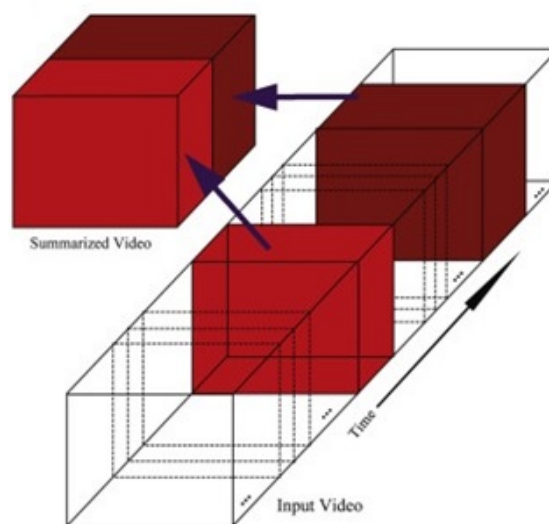

Figure 1: Description of video summarization

In static video summarization, essential key frames are selected and extracted from the surveillance video, whereas the lesser activities are skipped[1]. Dynamic video summarization is a method in which moving objects or a group of frames are segmented and retrieved from the video. This method is marked by the presence of unimportant events in every segment[2]. The earlier works of dynamic abstraction include video fast forward[3] and video skimming[4]. The summary obtained from video skimming is better than that of key frame summary method.Video synopsis is a video abstraction technique that can reduce the spatio-temporal redundancy optimally. The actual time of each moving object is displayed effectively.The

2

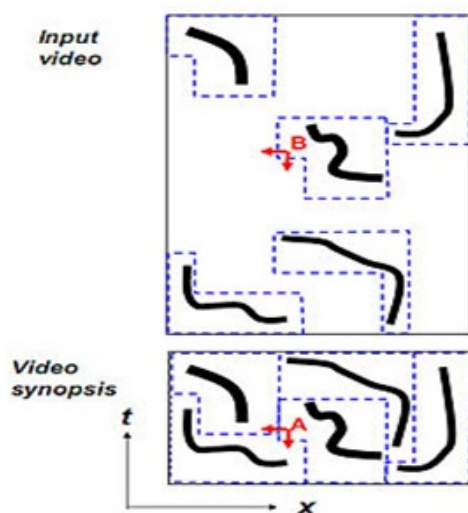object of interest can be found easily from the original video.

Figure 2: Description of video synopsis

Video synopsis method greatly reduces the video length and the space occupied by it. This paper is structured with Introduction in the first section, explanation of various video synopsis techniques in the second section and a comparison of merits and demerits are detailed in the third section. Finally, a conclusion is presented in the fourth section

2. SURVEILLANCE VIDEO SYNOPSIS TECHNIQUES

In general there are some quantitative standards in video synopsis field i) The output video should preserve maximum activities with reduced spatio temporal redundancy. ii) Collisions need to be avoided in the output video. iii) The chronological and spatial consistency of objects should be maintained well. Many approaches have been proposed for video synopsis. They are Frame Based Approach, Object Based Video Synopsis, Object Movement Video Synopsis and Potential Collision Graph methods.

Frame Based Approach

3

In frame based approach, frames are considered as the main building blocks. The two methods are key frames and video skimming. The key frame based abstraction methods[5],[6],[7] in which key frames are extracted from the input video. Unfortunately it cannot preserve the fast moving activities. McMillan[8] proposed a time lapse video summarization to match the users desired duration. Adaptive fast forward approaches[3] were developed to alleviate the loss of fast activities when dropping frames. Instead of selecting one frame, video skimming method[9],[10]] selects some video clips from the original video by skipping the less important activities. To generate the video clip, high level content analysis and uniform sub sampling are used to extract the fixed duration at fixed intervals. Frame based methods are simple and efficient. The performance of this method is faster than the real time decoding with the help of macro blocks and motion vectors. In the frame based approach the selected frames can represent the content of the original video effectively. However some activity information is lost and the jumping phenomenon between frames in the condensed video may result in unrealistic artifacts and the computation cost is too high.

Object Based Video Synopsis

The object based video synopsis methods extract moving objects reducing the spatio- temporal redundancies. A novel concept was introduced by[12],[13] and[14] which assembled the important activities from different times rather than selecting all the video frames. Space time video portions were extracted by Kang et al[15]] that were assembled using first fit and graph cut optimization algorithms. Undesirable seams appear between the portions that do not match within the optimal boundary. Rav-Acha et al[11] identified and extracted moving objects that can be shifted along time axis to create a compact synopsis. The chronological order of the events is disturbed in order to get a compact video like[13]. Pritch et al. expanded the work[16] to deal with the endless video streams captured by webcams. The local chronological order is preserved while moving the tubes temporally. Videos with backgrounds illuminated variably can be handled. This method is susceptible to collision artifacts. In object video synopsis due to high density of objects in the condensed video, collisions are usually unavoidable.

4

Nei[23] proposed a method in which moving objects are shifted both in space and time. The framework contains four main stages. The first stage involves extraction of the distinct objects in the video. In the second stage the active objects are shifted such that video size is reduced. By using MPR method the compact background are synthesized for stitching the objects. At the end the active objects are fused into the background .The above optimization does not have any interaction with the background.

Object Movement Based Synopsis

In this method the active objects are taken into account while the non-moving parts are considered as less important parts. The above object based synopsis does not handle redundancies well.The steps involved in movement synopsis are
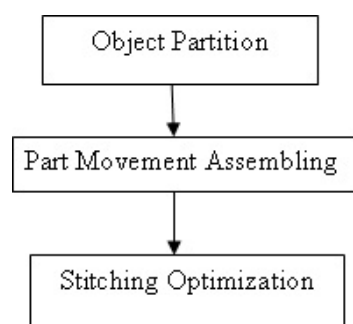
Nie et al[24] suggested a novel approach in which the objects are first decomposed into several parts and each part corresponds to a movement sequence. Then the subsequence changing is slightly discarded and the rest are assembled and stitched together to get a compact video.

In object partition the video objects are segmented and tracked with the help of hard segmentation and soft segmentation. The segmentation described in[18],[19]requires higher degree of user interaction for obtaining suitable results. The user interactions are greatly reduced by combining rotoscoping and matting techniques[20]. Rotoscoping is a technique which contours and tracks object contours in video. Then user draws control curves on key frames.

5

With the help of rotoscoping trimaps, object parts are generated and avoid pixel wise adjusting of the control curves. It helps to reduce the space between the objects. After object partition, the structure completion method is used to repair the holes between the images and background[24]. Finally, robust thinning algorithm is used to partition boundary in which each object having one pixel width, perfectly connected and well defined. It greatly reduces data storage. A synopsis is obtained by choosing the same number of movements from each sequence corresponding to the portioned sections and assembling them together. The selection of movements is based on constraints such as ensuring that the output synopsis video length is shorter than the input video, and that the synoptic video contains the most important movements. Also, compatibility is to be observed between the parts that are spatially combined and within the same part the chronological sequence must be maintained.

In order to enable stitching together of moving parts coming from different frames into the same frame of the synoptic video, the moving parts are shifted by vectors whose objective function is the linear combination of three weighted energy terms. This enables the reduction of gaps in the final video.

Potential Collision Graph Method

This technique as proposed by Yi He[29] which is used for carrying out tube rearrangement in the video synopsis process, is known to have decreased computational complexity as opposed to conventional techniques. An advance determination of the potential collision status between tubes can be made which can be further utilized in an online framework. Collisions can occur when two tubes act on the same area at a given time instance. This is avoided when the tubes pass the area at different times. The two distinct collision relationships that are identified enable the probability of collision to be determined. Tube rearrangement is then carried out deterministically. Other methods that can produce compact synoptic videos rely upon spatio-temporal shifting of tubes[25],[27],[28] and utilization of abnormal tubes for synopsis. A solution that is accelerated by a multicore technique is given by Zhu et al.[12] for tube

6

rearrangement modeled as a stepwise optimization problem.

3. COMPARISON OF VARIOUS SYNOPSIS TECHNIQUES

Table 1

Methods	Advantages	Disadvantages
Frame Based	Easier to identify the important video clips	It cannot recognize the fast moving objects and cannot distinguish the gradual changes.
	The performance is fast	This method lose some contents due to fast activities.
	Simple and Efficient	Computational cost is high.
Object Based	Reduces spatio temporal redundancy	Chronological ordering is very difficult.
	Efficient browsing for the huge video data sets .	The background synthesis fails to produce consistent result
Synthesis Based	Time and memory intensive.	This method cannot handle long and dynamic videos.
Object Movement Based	The parts in motion are conserved and other redundancies are removed.	The overlapping between the parts increases collision in the synoptic video.
Potential Collision Method	It reduces computational complexity with less collision artifacts	Difficult for dynamic background videos.

4. CONCLUSION

Through the survey, different approaches and methods for efficient compact video synopsis have been demonstrated. The comprehensive survey along with a comparison of various synopsis techniques were illustrated. The challenges of the different techniques have been brought to the forefront. On the basis of this survey, researchers can be quickly acquainted with the various approaches in

7

video synopsis and make an appropriate choice depending on their requirements and constraints..

References

[1] B.Truong., Svetha Venkatesh, *Video abstraction: a systematic review and classification*, ACM Trans. Multimedia Computer Communication. (2007).

[2] Nam J and Tewtik, *Video abstract of video*,IEEE Third Workshop Multimedia Signal Processing. (1999).

[3] Petrovic, N., Jojic, N., Huang, S.T, *Adaptive video fast forward*,Multimedia Tools Appl. (2005).

[4] Smith, M., Kanade, T, *Video skimming and characterization through the combination of image and language understanding techniques*,Proc. Conf. Computer Vision and Pattern Recognition. (1997).

[5] Pritch Y., *Clustered synopsis of surveillance video*. Proc. IEEE Int. Conf. Advanced Video and Signal Based Surveillance (2009).

[6] T. Liu, X. Zhang, J. Feng, and K. Lo, *Shot Reconstruction Degree,A Novel Criterion for Key Frame Selection*,Pattern Recognition Letters (2004).

[7] C. Gianluigi and S. Raimondo, *An Innovative Algorithm for Key Frame Extraction in Video Summarization*,Real-Time Image Processing (2006).

[8] E. P. Bennett and L. McMillan, *Computational time-lapse video*,ACM Trans. Graph (2007).

[9] M. A. Smith and T. Kanade, *Video skimming and characterization through the combination of image and language understanding*,Pattern Recognition Letters (2004).

[10] Z. Longfei, C. Yuanda, D. Gangyi, and W. Yong, *A computable visual attention model for video skimming*,Proc. 10th IEEE Int. Symp. Multimedia (2008).

[11] Rav-Acha A., Pritch Y., Peleg S, *Making a long video short: dynamic video synopsis*,Proc. IEEE Conf. on Computer Vision and Pattern Recognition (2006).

[12] C. Pal and N. Jojic,*Interactive montages of sprites for indexing and summarizing security video*,Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition (2005).

[13] H. Kang, Y. Matsushita, X. Tang, and X. Chen ,*Space-time video montage*,Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognitio (2006).

[14] C. Rother, L. Bordeaux, Y. Hamadi, and A. Blake, *Autocollage*,ACM Trans. Graph (2006).

[15] H. Kang, X. Chen, Y. Matsushita, and X. Tang, *Space-time video montage*,IEEE Computer Society Conference (2006).

[16] Y. Pritch, A. Rav-Acha, A. Gutman, and S. Peleg, *Webcam synopsis: Peeking around the world*,IEEE Transactions on Pattern Analysis and Machine Intelligence (2008).

[17] Y. Pritch, A. Rav-Acha, and S. Peleg, *Nonchronological video synopsis and indexing*,IEEE Transactions on Pattern Analysis and Machine Intelligence (2008).

8

[18] J. Wang, P. Bhat, R. Colburn, M. Agrawala, and M. Cohen,*Interactive Video Cutout*,ACM Trans. Graph (2009).

[19] X. Bai, J. Wang, D. Simons, and G. Sapiron, *Video Snapcut: Robust Video Object Cutout Using Localized Classifiers*,ACM Trans. Graph (2007).

[20] A. Agarwala, A. Hertzmann, D. Salesin, and S. Seitz,*Keyframe- Based Tracking for Rotoscoping and Animation*,ACM Trans. Graph (2004).

[21] C. Barnes, E. Shechtman, A. Finkelstein, and D. Goldman,*Patchmatch: A randomized correspondence algorithm for structural image editing*,ACM Transactions on Graphics (2009)

[22] C. Xiao, M. Liu, Y. Nie, and Z. Dong, *Fast exact nearest patchmatch for patch-based image editing and processing*, Visualization and Computer Graphics, IEEE Transactions (2011)

[23] Y. Nie, C. Xiao, H. Sun, and P. Li, *Compact Video Synopsis via Global Spatiotemporal Optimization*,IEEE Trans. Visualization and Computer Graphics (2013).

[24] Y. Nie, H. Sun, P. Li, C. Xiao, and K.-L. Ma, *Object movements synopsis via part assembling and stitching*,IEEE Trans. Vis. Computer. Graphics (2014).

[25] R. Zhong, R. Hu, Z. Wang, and S. Wang, *Fast synopsis for moving objects using compressed video*,Signal Processing Letters (2014).

[26] J. Zhu, S. Feng, D. Yi, S. Liao, Z. Lei, and S. Z. Li, *High-performance video condensation system*,Circuits and Systems for Video Technology (2015).

[27] S. Chakraborty, O. Tickoo, and R. Iyer,*Adaptive keyframe selection for video summarization*,Circuits and Systems for Video Technology (2015).

[28] K. Li, B. Yan, W. Wang, and H. Gharavi,*An effective video synopsis approach with seam carving*,Signal Processing Letters, IEEE (2016).

[29] Yi He, Zhiguo Qu, Changxin Gao, Nongsangi,*Fast Online Video Synopsis based on Potential Collision Graph*, IEEE Signal Processing Letters(2017)

E-mail address: POOVS1@GMAIL.COM

9