

Multiple Techniques Analysis on Video Synopsis

¹M.Shanmugapriya, ²K.Selvakumar

¹Anna University, Department of Computer Science, Chennai,

²Anna University, Department of Mathematics, Nagercoil,

Abstract

Video synopsis is the process of preserving key activities and eliminates the less important parts to create a short video summary of long original videos. These techniques are used for fast browsing, extracting big data, effective storing and indexing. The video synopsis techniques are broadly classified into two types: object based approaches and frame based approaches. But these approaches cannot handle the complexity of the dynamic videos. In object movement method focus on the movements of a single video object, and remove the redundancies present in the object movement, it helps to generate the more compact and efficient video synopsis. Video synopsis is the most popular research in computer graphics and computer vision area and several researches started on this area. In this review focused some of the video synopsis techniques are providing a detailed description of various representation methods and find out their positive and negative aspects.

Keywords: Frame, Object, Action synopsis, Object movement based method, Video synopsis.

1 INTRODUCTION

The Field of digital videos and multimedia technologies are increasing widely and rapidly. Large number of videos are available on the web. Length and content of these videos are too long. Among these large quantity of data we need browsing, retrieving and indexing. It is time consuming an tedious process, so most of the videos are not watched in full. For this the important parts are preserved and less important contents are removed from the original video and create a short length video. This process is called video synopsis. Video synopsis is a video summary technology that is confined to the original video information for the greatest extent by analyzing and processing a long video file and extracts the key information. The traditional video summaries only compress the video content in timeline. However, video synopsis compresses the video content not only in time but also in space. The length of video synopsis is shorter than that of the original video and the storage space of surveillance video can be greatly saved. Now that, the basic content of the original video is retained and the user can quickly browse and search the video content. The video synopsis technology has the broad application prospects and the important research significance. Video synopsis is mainly based on two approaches, first is frame based approaches and another is object based approaches. Frame base approaches having key frames are directly extracted and these frames are the basic unit of the synopsis, but other less important and redundant frames are to be ejected. Each object is considered and removes the redundancies between the objects in object based

approaches. Frame based and objects based approaches cannot handle the dynamic videos. The complexity of dynamic videos can be resolved based on the movements of a single video objects, and remove the redundancies present in the object movements. It helps to generate the more compact and efficient video synopsis. The dynamic videos are resolute based on the movement of the objects. Simple ideas for the moving object based approach works at the level of object part and eliminate the non moving object parts. Object movement synopsis method will effectively remove the redundancies during the movement of a video object. The algorithm of video synopsis is different from those methods mentioned above. After background modelling and detecting of moving object, the Region of Interest (ROI) of moving object is extracted and a non interactive matting algorithm of moving object based on GrabCut is used to obtain the matting moving object. These matting objects are placed without overlapping in a background image and a series of these frame images can form a single camera surveillance video synopsis. In this survey analysing of various Video synopsis methods and algorithm to helps and identify the opportunities of future developments in this video synopsis field.

2 FRAME BASED APPROACH

The important frames are the basic units of the video synopsis, which are mined from the original video and reject the less important frames. For video synopsis, many technologies are available. From them the some effective methods are taken and discussed.

2.1 Replay Boundary Detection

Sports videos replay scenes only highlights the important or interesting events in the game. These highlights are the spirit of the game and it precise the whole video without evading the soul of the game. Replay Boundary Detection (RBD) present the technique to detect the replay boundary of sports videos with the help of macro blocks (MBs) and micro vectors (MVs). The RBD process carried out in three steps. i.e identify, extract and detect. Identify the replay boundary directly from MPEG compression domains. The features are directly extracted from compressed MPEG video; so it will avoids the expensive in-verse DCT. The inverse DCT is used to transfer values from the compressed domain to the image domain. The rules of macro block and motion vector to detect the replay boundary. The MPEG compressed domain features are macro blocks, motion vector and similar measure function based on static analysis. RBD is defined [1] as a 4-tuple $V, F, S, \text{ and } B$, where V is a group

of video sequence, F is a group of video feature vector, S is similar measure function and B is a set of replay boundary frame. Similar measure function can find a set of replay boundary frame by the similar metric based on the vid-eo feature vectors. The structure of MPEG Group Of Pictures (GOP) is divided into a triplet, that is, three frames having the form IBB or PBB.

$$F_{ij}=3i-1 \quad B_{ij}=3i \quad b_{ij}=3i+1$$

where $j=0,1,2,\dots,n$, $i=0,1,2,\dots,n/3$ and n be the number of frames in the video sequence.

Rules of Macro Blocks

- Ratio of the number of intro coded MBs to the total number of MBs in P and B frame will larger than pre-defined threshold T intra.
- Ratio of the number of motion compensation coded MBs to the total number of MBs in B frame is greater than predefined threshold Tcompen.
- Ratio of number of skipped MBs to the total no. of MBs in B frame is less than T skip.

Rules of Motion Vector

- Ratio of the number of small magnitude MVs to the total number of MVs in P and B frames are greater than T mvabs.
- Ratio of the number of dominant direction MVs in P and B frame exceeds the T mvdirect.

2.1.1 Advantages and Disadvantages

Replay Boundary Detection based frame based approach is easily identify the important video clips and edited well. The performance of this method is faster than the real time decoding with the help of macro blocks and motion vectors. But this system is not suitable for large amount of data. It cannot recognize the fast moving objects and cannot distinguish the gradual changes.

2.2 Shot Reconstruction Degree

Shot reconstruction degree [2] is a key frame selection approach. Key frame based approach primarily faces a critical questions, i.e. what types of frames are to be chosen? For this query some researchers focus on the frames with minimal motion energy and it should be chosen because they think cameras focus only on important objects. Some of them study the frames with maximal motion dynamics as key frames be-cause the frames are not included, some important contents loss due to the vivid changes of the video. Some standards are to find the similarity between the selected key frames and the original video. This method is called fidelity. Fidelity is the maximum of the minimal distance between the key frame set and video shot.

- Similarity function: It specifies the similarity between the two video frames.
- Frame Interpolation Function: It specifies to calculate the distance between the key frames and the whole video shots.

2.2.1 Advantages and Disadvantages

Shot reconstruction degree will focus the global and local de-tails of the dynamic video. It is better to capture the detailed dynamic shots and performance is good in terms of fidelity and shot reconstruction degree. But this method

will lose some contents due to the fast activities and computational cost is too high.

3 OBJECT BASED APPROACHES

In object based approaches each moving objects considered as a 3D tube and remove the spatiotemporal redundancies between objects by shifting the tubes along time axis. From them some effective object based approaches are discussed.

3.1 Non chronological Video Synopsis and Indexing

This paper describes the object based video synopsis methods. The main application of non chronological video synopsis is surveillance cameras. This video synopsis is a temporally compressed representation of video that allows video browsing and collection. This approach eases the spatiotemporal redundancy in video. If two processes carried out simultaneously, then shift the events from original time intervals to other time intervals when no other works take place at the same location. This video synopsis approaches satisfies two properties: 1) Only videos are used for video synopsis, i.e dynamic videos are to be used. 2) Relative timing between the activities may be differ, i.e reduce the spatiotemporal redundancy. In object-based method for video synopsis, moving objects are first and foremost detected and divided into space-time "tubes." An energy function is defined on the promising time shifts of the events, which completely covers the accurate proper properties of the video synopsis.

$$EM = \sum b \in B E_a(b) + \sum b, \text{for all } b \in B (\alpha E_t(b, b') + \beta E_c(b, b'))$$

Where M is the temporal mapping, shifting objects b, t is time from its input video, E_a is the activity cost, E_t is the Temporal consistency cost, and E_c is the Collision cost. This energy function will helps to protect the original activity of the video, while evading collisions between different shifted actions. The temporal compression will attain through the stroboscopic effects. Stroboscopic effects of video synopsis are the presentation of multiple dynamic arrival of a single object.

3.1.1 Advantages and Disadvantages

These methods have efficient activities in the video sequences and effective indexing. It enables the use of object constraints and reduces the spatio temporal redundancy. This method is less applicable in dense activity applications. Editing and change the chronological ordering is too difficult, so complexities of the dynamic videos are less. This method cannot handle the video with redundancies in the moving objects.

3.2 Global Spatiotemporal Optimization

Basic idea of global spatiotemporal optimization is to globally shift the active objects in the spatiotemporal video volume, and then generate a compact background controlled by the optimized object courses to fit the shifted objects. Global spatiotemporal optimization structures will shifts active objects in the spatiotemporal space. The optimization process is collected with two cost terms: One is data term and the other is smoothness term.

Data term is used to reserve process as much as probable and avoid objects new path from conflicting far from their original ones, and the smoothness is used to prevent object collisions and retain the spatiotemporal stabilities and relations of objects as much as possible. Multilevel patch relocation (MPR) method to produce the compressed background, and visual artifacts are removed for increases the movement space of shifted objects based on the environmental context. Finally, fuse the shifted objects into the compact background to produce the final video synopsis. With the help of synthesized dense background and simple user interactions, global spatial temporal optimization can produce more reduced video scenes with crowded but non-colliding objects from input videos containing sparse moving objects, and it is widely used in application such as crowd animation design, video summarization, video games and interactive media production. This work [4] having two contributions:

- a. Spatiotemporal domain, it densely condenses the activity information and preventing visual collision artifacts between moving objects.
- b. Synthesized compact background will give a larger virtual motion space for shifted objects using a MPR method in Markov Random Field (MRF) networks.

This video synopsis structure consists of four stages. First, analyze the input video and extract the background and moving objects from the video. In second stage, shift active objects in the spatiotemporal video volume by using the global spatiotemporal optimization, which find the new positions in the synopsis for clustered objects. Third stage, using a MPR method, synthesizes a compact background with the scene-path context to fit the clustered objects. Finally, fuse objects into the synthesized compact background using a gradient-domain editing tool and produce the compact video.

3.2.1 Advantages and Disadvantages

Global Spatiotemporal Optimization method will reduce the spatio temporal redundancies of the input video as much as possible. It will preserve the chronological consistency of the important events; it will help to reduce the collision between the moving objects. This method is widely used for efficient browsing for the huge video data sets. But it will fail to produce the video synopsis in the crowded activity in both spatial and temporal domain. The spatiotemporal optimization stage has no interaction with the background. As a result, even if the spatiotemporal optimization works well, the background synthesis will fail to produce a consistent result.

4. ACTION SYNOPSIS

4.1 Pose Selection and Illustration

Action synopses will present the motions in still images [5]. In this method first selects important poses based on the animation sequences of the skeleton, to express complex motions in a single image or a small number of concise views. This method is to embed the high-dimensional motion curve in a low-dimensional Euclidean space, where the main characteristics of the skeletal action are kept. The lower complexity of the

embedded motion curve allows a simple iterative method which analyzes the curve and locates significant points, associated with the key poses of the original motion. Action synopsis techniques consist of the following stages:

1. Extracting motion aspects. Motion data will be generate from the captured videos, animated sequences and video clips. In a sequence of human action there having a sequence of skeletal poses. Small numbers of motion aspects are calculated among the skeleton poses. Motion aspect is an attribute of the motion, which defines the inter-pose distance.
2. Dimensionality reduction. Dimension of the motion curve is reduced by Affinity Matrices with a Replicated Multi Dimensional Scaling. Affinity matrices define the dissimilarity between the poses. That is the differences in the location of the most dissimilar poses. The technique defines a reduced dimensional space in which the salient features of the various motion aspects are kept.
3. Pose selection. Locate the local extreme points along the motion curve, which are associated with extreme poses of the motion. It generates a hierarchy of prioritized poses.
4. Synopsis view. This is the final stage of action synopsis. The Frames associated with selected poses com-posed into an image. The selected frames can be presented side by side or composed into a single image. Further some instances of the images are rendered semi-transparently; it leads to reduce the clustering of the resulting images and highlights the important poses.

4.1.1 Advantages and Disadvantages

Pose Selection and Illustration based action synopsis satisfies the object movement but the structure recovery is not possible and fast movements cannot correctly identified.

5. OBJECT MOVEMENT SYNOPSIS

Previous methods don't handle the dynamic video and which are more complex. They don't handle redundancies in the moving objects. Object movement synopsis handles the moving objects that are dynamic videos. In this method the moving parts are considered as important and non moving parts are less important parts. The moving parts are preserved and non moving parts are to be eliminated.

5.1 Part Assembling and Stitching

Object movement synopsis effectively removes redundancies during the movements of a single videos object. The basic idea is to work at the level of object part, and to remove the non moving parts. This method consists of three stages:

- Part Based Object Movements Partition
- Part Based Movement Assembling
- Part Based Movement Stitching

5.1.1 Part Based Object Movements Partition

First step of this process is object partition. First extract an object movement sequence from an input video. Then Partition each object into several semantic parts, which produces several part movement sequences. It consists of 3 steps

1. Object Partition

In object partition first segment and track the video objects with the help of hard segmentation and soft segmentation. The hard segmentation will need more user interactions to obtain the accurate results. This user interaction will be reduced with the combine the techniques rotoscoping [7] and matting. Used these techniques to partition object and track the object parts. This rotoscopic tracking techniques based on the key frames. Based on these techniques, first tracks object contours in video sequences with some user interaction. Then user draw control curves on key frames but not all the frames of the video. The control curves do not need to match the object contour exactly. With the help of the motion scoping curves, generate tri-maps of an object part and avoiding pixel-wise adjusting of the control curves. It will help to reduce the space between the objects. To extract the each object part with the help of global sampling methods for matting.

2. Partition Boundaries

Structure Completion Method is used to repair the holes between the images and the background. Belief propagation algorithm is used to minimize the energy with complexity $O(2LN^2)$. Robust Thinning Algorithm is used to partition boundary, in which each object having only one one-pixel width, perfectly connected, and well defined. It helps to reduce the data storage and transmission time.

3. Importance Estimation

To preserve the most important moving parts in the synopsis. For this compute an importance value for each moving parts. The faster moving parts are considered as important parts and other parts as redundant parts. The important parts are determined with the help of optical flow method.

5.1.2 Part Based Movement Assembling

Partition the object movements and obtained a sequence of movements for each part. Select the same number of part movements from each part sequence, an assembled them together frame by frame to produce synopsis. The non moving parts are eliminated. The selection Scheme satisfies the following properties and it is solved by MRF.

- The synopsis length is less than the actual input video.
- Important moving parts are preserved in the synopsis.
- Spatially assembled parts should be compatible with each other.
- Chronological order preserved for the same part, while order between different parts are disrupted.

5.1.3 Part Based Movement Stitching

This step is used to stitch the assembled parts [14] to eliminate the gaps between the images. The moving parts on the same frame of synopsis comes from different frames of the input video, they may not correctly stitch each other. Shift the parts in the spatial domain, and to stitch the moving parts together. Shifting vectors are computed via part movement stitching, whose objective function is the linear combination of 3 weighted energy terms.

$$E = \omega_d E_d + \omega_t E_t + \omega_s E_s$$

where, ω_d , ω_t , ω_s are weights E_d Prevents the moving parts from shifting, E_t Temporal coherence of moving part after shift, E_s To reduce gap between assembled moving parts. The result is after shifting the moving parts by part stitching optimization. The gaps between moving parts are almost eliminated. But the boundaries do not exactly match with each other, there are small gaps between moving parts. To remove the small gaps, move the boundary of a moving part to overlap with the other boundary, and deform the rest of the parts. Then the small seam disappears. The final synopsis result without gaps will be showcased.

6. ALGORITHM OF MOVING OBJECT MATTING

The moving object matting method uses the non-interactive GrabCut algorithm to extract the objects from the ROI image and provides extract more accurately the moving object

6.1. Extraction of the Roi Image

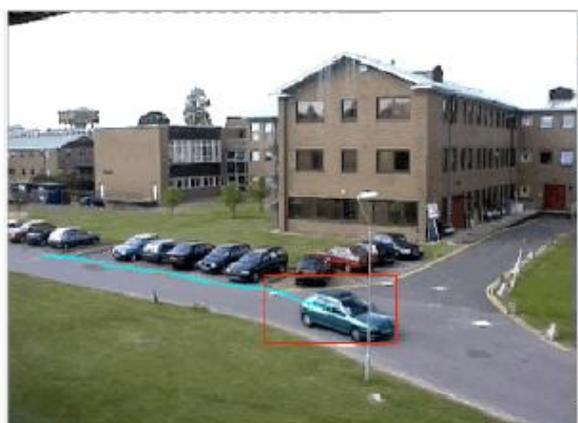
After the moving objects are detected by our method, we can obtain the binary foreground mask image. This image is processed in order to filter the smaller blobs and reserve the larger blobs. The position of centroid and the size of these blobs (moving objects) are computed and utilized to locate the moving object and determine the area of ROI. The center of rectangular ROI is the centroid of moving object and the width and height of rectangular ROI is 1.5 times width and height of the moving object respectively. The image area of rectangular ROI is extracted from the frame image.

6.2. Moving Object Matting

Moving object can be extracted by comparing the current frame with the background image, but the edge of the extracted moving object is not accurate and the holes in moving object are often produced. In order to extract more accurately the moving object, GrabCut algorithm can be used. But this algorithm is an interactive image segmentation method and the user must indicate some pixels of segmentation boundary between the foreground and background. In order to avoid the necessary manual interaction, this method is improved and a non-interactive matting algorithm of moving object is proposed based on GrabCut. After the moving object has been successfully detected and tracked, a rectangular region is drawn up around the moving object tracked. This rectangular region is called the ROI of moving object. The area outside ROI is directly considered as the background. The

centre part pixels of ROI are the foreground pixels and the edge position pixels of ROI are considered as the background. So some background and foreground pixels will be determined and GrabCut algorithm is used to segment the moving object in the continuous image of video. The edge of the matting image has a lot of burr, so edge smooth is used and the moving object boundary can be obtained accurately. This method avoids the common holes of moving object in the traditional object detection method.

Although the algorithm of moving object matting based on GrabCut can extract more accurately the moving object, the region of image with lower contrast at the transition from foreground to background may be lost. In Post processing steps we change the matting moving object image into a binary mask image and add it with the binary mask image obtained by our detection method in the same ROI position. The new binary mask image is used to extract the moving object from the original frame image. Fig.(1) shows matting example of moving object using our method in surveillance video. The moving object detected is wrapped in a rectangular frame and the centre pixels in the rectangular region are considered as the foreground, while the edge pixels of the rectangular region are considered as the background, and then the Grabcut algorithm is used to further segment. The accurate segmentation result can be obtained as below output, that this algorithm can accurately achieve the matting of moving object when the probable location of moving object has been obtained. Because the moving object tracking has been made and the moving object has been enclosed in a rectangular frame in the process of tracking, the manual reference point label of foreground and background needed in the Grabcut algorithm is avoided. So it is more convenient for moving object matting.



Output 1: Original image



Output (2a): Frame image



Output (2b):ROI



Output (2c):object

7.CONCLUSION:

This paper covers a survey on Video synopsis approaches and given a brief review of its advantages in terms of technology. In our survey we have seen we cannot handle the complexity of the dynamic videos to get the accuracy on the object detection. Tracking of moving object movement is done by object movement synopsis and the algorithm used is object movement matting algorithm. In this review and analysis of technology, the survey indicates the various approaches of object movement detection in video synopsis, including feature descriptors and object detection technique in video frames and various object detection methodologies. We expect that this survey on video synopsis will enrich theoretical details of the object detection methods and will give valuable contribution to research works on video synopsis and encourage new research.

REFERENCES

- [1]. J. L. J. Ouyang and Y. Zhang, "Replay boundary detection in mpeg compressed video," Proc. Int'l Conf. Machine Learning and Cybernetics, vol. 5, pp. 2800-2804, 2003.
- [2]. J. F. T. Liu, X. Zhang and K. Lo, "Shot reconstruction degree: A novel criterion for key frame selection," Pattern Recognition Letters, vol. 25, no. 12, pp. 1451-1457, 2004.
- [3]. A. R.-A. Y. Pritch and S. Peleg, "Nonchronological video synopsis and indexing," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 30, no. 11, pp. 1971-1984, Nov. 2008.
- [4]. H. S. Y. Nie, C. Xiao and P. Li, "Compact video synopsis via global spatiotemporal optimization," IEEE Trans. Visualization and Computer Graphics, vol. 19, no. 10, pp. 1664-1676, Oct. 2013.
- [5]. Y. C. J. Assa and D. Cohen-Or, "Action synopsis: Pose selection and illustration," ACM Trans. Graphics, vol. 24, no. 3, pp. 667- 676, 2005.
- [6]. P. L.-C. X. Yongwei Nie, Hanqiu Sun and K.-L. Ma, "Object movements synopsis via part assembling and stitching," IEEE Trans. Visualization and Computer Graphics, vol. 20, no. 9, Sept 2014.
- [7]. D. S. A. Agarwala, A. Hertzmann and S. Seitz, "Keyframe-based tracking for rotoscoping and animation," ACM Trans. Graphics, vol. 23, no. 3, pp. 584-591, 2004.
- [8]. C. R. X. T. K. He, C. Rhemann and J. Sun, "A global sampling method for alpha matting," Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR '11), pp. 2049-2056, 2011.

- [9]. J. J. J. Sun, L. Yuan and H.-Y. Shum, "Image completion with structure propagation," in *ACM Trans. Graphics*, vol. 24, no. 3, pp. 861-868, 2005.
- [10]. K. C. C. Shen, H. Fu and S. Hu, "Structure recovery by part as-sembly," *ACM Trans. Graphics*, vol. 31, no. 6, p. 180, 2012..
- [11]. T. M. S. Schaefer and J. Warren, "Image deformation using moving least squares," *ACM Trans. Graphics*, vol. 25, no. 3 pp. 533-540, 2006.
- [12]. C. Q. R.W. Zhou and G. Ng, "A novel single-pass thinning algo-rithm and an effective set of performance criteria," *Pattern Rec-ognition Letters*, vol. 16, no. 12, pp. 1267-1275, 1995.
- [13]. S. R. D. Sun and M. Black, "Secrets of optical flow estimation and their principles," *Proc. IEEE Conf. Computer Vision and Pat-tern Recognition (CVPR '10)*, pp. 2432-2439, 2010.
- [14]. T. M. S. Schaefer and J. Warren, "Image deformation using moving least squares," *ACM Trans. Graphics*, vol. 25, no. 3 pp. 533-540, 2006.
- [15]. Y. P. A. Rav-Acha and S. Peleg, "Making a long video short: Dynamic video synopsis," *Proc. IEEE CS Conf. Computer Vi-sion and Pattern Recognition*, vol. 1, pp. 435-441, 2006.
- [16]. B. Truong and S. Venkatesh, "Video abstraction: A syste-matic review and classification," *ACM Trans. Multimedia Computing, Comm., and Applications*, vol. 3, no. 1, p. 3, 2007.
- [17]. J. Wang and M. Cohen, "Image and video matting: A sur-vey," *Now Pub*, vol. 3, no. 2, 2008.
- [18]. R. W. Y. Nie, Q. Zhang and C. Xiao, "Video retargeting combining warping and summarizing optimization," *The Visual Computer*, vol. 29, pp. 785-794, 2013.
- [19]. Y. Pfitch, A. Rav-Acha, S. Peleg, "Nonchronological video synopsis and indexing," *IEEE Trans. PAMI*, vol. 30, no.11, pp.1971- 1984, 2008
- [20]. Y. Pritch, A. Rav-Acha, A. Gutman, and S. Peleg, "Webcam Synopsis: Peeking Around the World," in *11th International Conference on Computer Vision*, pp. 1-8, 2007.
- [21]. A. Rav-Acha, Y. Pritch, and S. Peleg, "Making a Long Video Short: Dynamic Video Synopsis," in *IEEE Conf. Computer Vision and Pattern Recognition*, pp. 435-441, 2006.