ISSN 2347-4289

Review Of Video Synopsis Approaches

Jeeshna P. V., Kuttimalu V. K.

M. Tech Student, Dept. of CSE, SCTCE, Pappanamcode, Trivandrum, India;
Assistant Professor, Dept. of CSE, SCTCE, Pappanamcode, Trivandrum, India.
Email: jeeshnapv@gmail.com

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, ectracting 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. Naturally this is not a complete review of the entire video synopsis techniques. In this review focuses some of the video synopsis techniques.

Keywords: Video synopsis; Frame; Object; Action synopsis; Object Movement Synopsis

1 INTRODUCTION

DIGITAL videos and multimedia technologies are developed widely and rapidly. Today hunddereds and millions of videos such as cartoons, films, camera surveillance, sports are existing in social networking sites, TV, youtube etc. These videos have too long and large filesize. In the busy life watching these long videos are the time consuming and uninteresting process. So these videos are converted into short videos with the help of video synopsis. Video synopsis is the process of eliminating the less essential data and preserves the key parts of the video. To exclude the redundant parts and protect the important activities are to produce a short video summary from the long videos. Video synopsis mainly used for fast browsing of the videos, recovering, indesing, storing and handle the big data. Video synopsis is one of the most popular exciting and interesting research area in computer graphics and computer vision. Video synopses are mainly based on two approaches, first is frame based approaches and another is object based approaches. Frame based approaches having key frames that 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 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.

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 macroblocks (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 inverse DCT. The inverse DCT is used to transfer values from the compressed domain to the image domain. The rules of macroblock and motion vector to detect the replay boundary. The MPEG compressed domain features are macroblocks, motion vector and similar measure function based on static analysis. RBD is defined [1] as a 4-tuple *V, F, S, 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 video feature vectors. The structure of MPEG Group Of Pictures (G0P) is divided into a triplet, that is, three frames having the form IBB or PBB.

$$F_i^{j=3}{}_i^{-1}\ B_i^{j=3}{}_i\ b_i^{j=3}{}_i^{+1}$$

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

Rules of Macro Blocks

- Ratio of the number of introcoded MBs to the total number of MBs in P and B frame will larger than predefined 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 T_{compen}.
- 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 macroblocks and motion vectors. But this

ISSN 2347-4289

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 because 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.

$$d_{sh}(S,KF) = \max_n \{d(t+n\Delta t)\},\ n=1,2,3\ldots..N$$

where S is the semi-Hausdorff distance and KF is the key frame. It doesn't focus local details of the frame. Shot Reconstruction Degree [2] is the latest method. In this method, key frame selection approach is based on the movement of the dynamic videos. It will concentrate on the local and global details of a video shot. For this method Inertia based frame interpolation algorithm is used. This algorithm is used for finding the distance between the key frames and the complete video shots.

$$\theta_0 = \arg_\theta \{SRD(\theta)\}$$

Where,

$$SRD(\theta) = \sum_{n=0}^{N-1} Sim(F(t+n\Delta t), F^*(t+n\Delta t, \theta))$$

Two items required to make SRD computable.

- 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 details of the dynamicvideo. 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 Nonchronological Video Synopsis and Indexing

This paper [3] describes the object based video synopsis methods. The main application of non chronological video synosis is survilance 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 proporties: 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 [3] on the promising time shifts of the events, which completely covers the accurate proper properties of the video synopsis.

$$E_M = \Sigma_{b\varepsilon B} E_a(b) + \Sigma_{b,\mathrm{forall}\varepsilon B} (\alpha E_t(\hat{b}, \hat{b}'') + \beta E_c(\hat{b}, \hat{b}''))$$

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 is 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 reduce 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 [4] 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 noncolliding 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:

1. Spatiotemporal domain, it densely condenses the activity information and preventing visual collision artifacts between moving objects.
2. Synthesized compact background will give a larger

ISSN 2347-4289

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 spati temporal 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 captuared 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 motin curve, which are associated with extreme poses of the motion. It generates a hierarchy of prioritized poses.
4. Synosis view. This is the final stage of action synopsis. The Frames associated with selected poses composed 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. [6] 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 video sobject. 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. Firat extract an object movement sequence from an input video. Then Partition each object into several semantic parts, which produces several part movement sequences. It consist of 3 steps

1. Object Partition

In object partion 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 keyframes. Based on these techniques, first tracks object contours in video sequences with some user interaction. Then user draw control curves on keyframes 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 rotoscoping curves, generate trimaps 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 gloabal sampling methods [8] for matting.

2. Partition Boundaries

Structure Completion Method [9] [10] [11] is used to repair the holes between the images and the background. Belief propagation algorithm is used to minimize the energy with complexity $O(2LN2)$. Robust Thinning Algorithm [12] 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.

ISSN 2347-4289

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 iportant parts and other parts as redundant parts. The important parts are determined with the help of optical flow [13] 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, ans 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 legth 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 synopsised result without gaps will be gets.

6 Conclusion

Users prefer briefer and compressed important contents without eliminating the essence. This achieved through the video synopsis. The synopsised videos retain only important objects, so it can effectively reduce the memory storage for large video sets. In this review the main techniques of video synopsis is discussed. There are two type of video synopsis: frame based and object based. In frame based approaches the less important frames are removed. This method has two basic forms, keyframe and videoskim. In object based approaches eliminate the spatiotemporal redundancies between the objects by shifting the events. These approaches cann't handle the redundancies in the moving objects. This complexity will be reduced using object movement based synopsis. The partbased redundancies are widely exist during the movement process of video object. It focuses on the movements of a single object and eliminates the less important parts present in the object movements and produce compact synopsis of video. First partition the video object into several semantic parts. Then with the help of part assembling optimization, select the same number of most important movements and assembled to form synopsized object movements. Finally apply the part stitching optimization to stitch assembled moving parts. This will effectively remove part-based redundancies and successfully generate more compact video synopsis. This review reveals the importance of the video synopsis in computer graphics and computer vision communities and to identify the applications of video synopsis in the day to day busy life. Also this review helps to identify the oppurtunities of future developments in this video synopsis field.

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 assembly," 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 algorithm and an effective set of performance criteria," Pattern Recognition Letters, vol. 16, no. 12, pp. 1267-1275, 1995.

ISSN 2347-4289

[13] S. R. D. Sun and M. Black, “Secrets of optical flow estimation and their principles,” Proc. IEEE Conf. Computer Vision and Pattern 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 Vision and Pattern Recognition, vol. 1, pp. 435-441, 2006.

[16] B. Truong and S. Venkatesh, “Video abstraction: A systematic 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 survey,” 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.