Video Summarization: Techniques and Classification

Muhammad Ajmal, Muhammad Husnain Ashraf, Muhammad Shakir, Yasir Abbas, and Faiz Ali Shah

COMSATS Institute of Information Technology
M.A. Jinnah Building Defence Road, Off Raiwind Road, Lahore, Pakistan
{rajmal,fashah}@ciitlahore.edu.pk

Abstract. A large number of cameras record video around the clock, producing huge volumes. Processing these huge chunks of videos demands plenty of resources like time, man power, and hardware storage etc. Video summarization plays an important role in this context. It helps in efficient storage, quick browsing, and retrieval of large collection of video data without losing important aspects. In this paper, we categorize video summariztion methods on the basis of methodology used, provide detailed description of leading methods in each category, and discuss their advantages and disadvantages. Moreover, we discuss the situation in which each method is most suitable to use. The advantage of this research is that one can quickly learn different video summarization techniques, and select the method that is the most suitable according to one's requirements.

1 Introduction

Recently, digital video technology is growing at a rapid rate. Due to advancement in technology, it becomes very easy to record huge volume of videos. A huge bulk of digital contents such as news, movies, sports, and documentaries etc. is available. Moreover, the need for surveillance has increased significantly due to increase in the demand of security especially after 9/11. Thousands of video cameras can be found at public places, public transport, banks, airports, etc. resulting in large amount of information which is difficult to process in real time. Furthermore, storage of huge amount of video data is not that easy. It is very important to quickly retrieve and browse huge volume of data efficiently because end user want to get all important aspects of data. To solve this problem, numerous solutions are provided in literature [1,2,3,4,5,6]. Video summarization plays an important part in this regard, as it helps the user to navigate and retrieve through a large sequence of videos. In recent years, video summarization has become an emerging field of research. But pratical implementation is far behind due to complexity of mehtods. The purpose of our work is to provide a brief and categorized overview of video summarization techniques according to advantages, drawbacks, and methodology used. One can easily grasp the idea of different techniques and can choose the technique of his/her choice.

L. Bolc et al. (Eds.): ICCVG 2012, LNCS 7594, pp. 1–13, 2012.

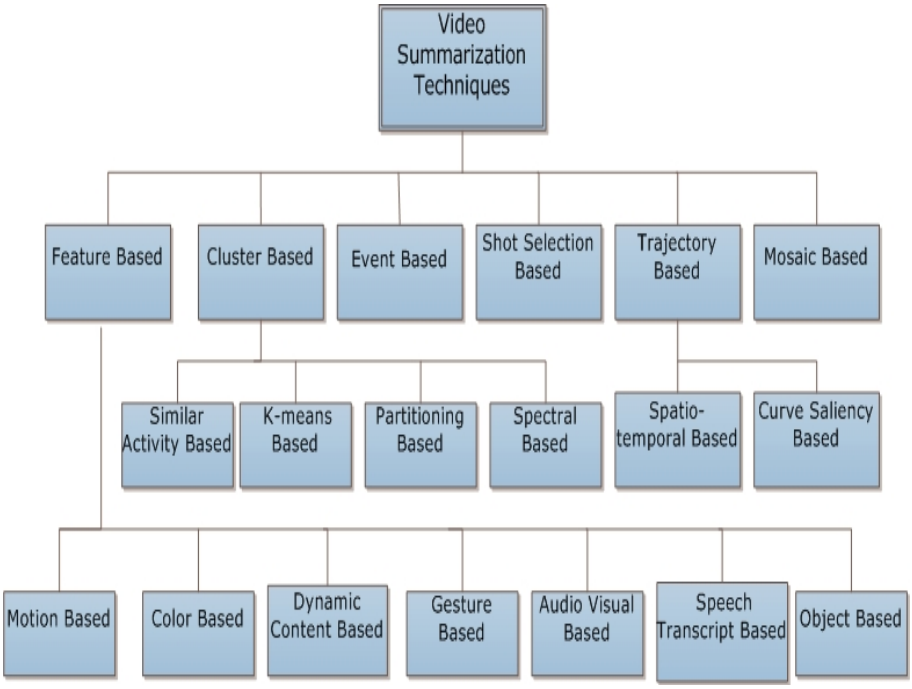

Fig. 1. Hirarchical structure of video summarization techniques

Surveys on video summarization techniques are already conducted that provides an overview and classification in different ways as video skims, highlights [7], still image and moving image abstracts [8,9], and video summarization techniques for mobile applications [10]. Techniques are based on features [11,12,13], clustering [14,15,16,17], event [12,15,18], shot selection [19,20], mosaic [21,22,23,24], and trajectory analysis [25,26,27]. Figure 1 represents a hierarchical classification of these techniques. As can be seen from the figure, we categorize video summarization techniques in 6 major categories based on mechanism and overall process. A detailed overview of these techniques is given in the following sections.

2 Feature Based Video Summarization

Digital video contains many features like color, motion, and voice etc. These techniques work well if user wants to focus on features of video. For example, if user wants to see color features then its good to pick color based video summarization techniques. As can be seen from the Figure 1, feature based video summarization techniques are classified on the basis of motion, color, dynamic contents, gesture, audio-visual, speech transcript, object. These techniques are described below.

2.1 Motion Based Video Summarization

It is always difficult to summarize video based on motion only . The task becomes more complex when camera motion is also involved. The concept of motion for key frame extraction was first used in [11]. After that, [1,2,28] also used motion for key frame extraction. A novel technique involving camera motion as well as object motion is presented in [29]. The changes in dominant image motion lead to temporal segmentation of video. The residual contents (i.e. after subtracting the camera motion from global image motion) of the video are identified. The extraction of meaningful dynamic events is done by statistical motion model. The technique described in [30] focuses on constant speed motion and generates summary by using the concept of relative motion. It also preserves the spatial and motional information for further analysis. Motion based approach are good when there is medium-level motion in a video. In contrast, it fails to do well for videos containing huge motion or no motion at all. Surveillance videos are good candidate for motion based approach.

2.2 Color Based Key Frame Extraction

Color is considered an important aspect of video. Thats why it has been used quite often for video summarization. A nice color based technique is proposed in [3]. Color histogram is used to store color information of a shot. By using probability model, different shots having color change patterns are identified. Utterance of a shot is also identified. Finally, integration of these utterances leads to the meaningful summary of video. Zhang also proposed color based techniques in [4,13], in which key frames are selected based on color and texture properties. The first frame within a shot is considered as key frame. Color histogram difference is used to identify next key frame. At the end, we have a collection of key frames. Color based summarization techniques are very simple and easy to use. However, their accuracy is not reliable, as color based techniques may consider noise as part of summary.

2.3 Dealing with Dynamic Contents

Such kind of techniques use motion, color, and audio features for video summarization [12,13]. In particular, a novel approach towards preserving dynamic contents was proposed in [31]. Firstly, it does color and motion based segmentation and then classifies dynamic and static segments. The next step is clustering of dynamic and static segments and then the selection of summary segments which results in a meaningful summary. This technique is very useful for videos containing rich visual, graphic, and audio contents as well as for moving camera, but summary is not generated as camera is zooming or panning.

2.4 Gestures Based Key Frame Extraction

Gestures are very important in real life. Recently, some studies have focused on gestures considering it a very important aspect of video. A compact framework

for summarizing lecture video using hand gestures recognition is presented in [32]. Similarly, the method proposed in [33] summarizes video on the basis of gestures like hand, head, and legs movement etc. The gesture energy is calculated using Zernike movement and then local maxima and minima are monitored for key frames. It is not a threshold dependent technique so the number of key frames extracted may vary. This technique is very useful for videos of sign language communication, human computer interaction and even for robot guidance. Accurate classification and segmentation of different gestures is a difficult task, due to which it may fail to produce a meaningful summary.

2.5 Audio-Visual Based Approach

Nowadays, music companies put their music online for business. The customers who want to buy music would like to see highlights first. Several methods for music video summarization has been proposed [34,35]. Most of which are for MIDI data. MIDI data summarization cannot be applied to real music videos. Music video structure consists of audio data that depicts what we see on the screen (visual data). Lectures and sports videos are also rich with audio-visual features. Moreover, many surveillance applications also use audio-visual features for summarization. The synchronization between audio and video is really crucial in this regard, so the summarization must be audio centric and image centric. To ensure synchronization, summary must be generated by choosing video segments that correspond to some audio segments. Such kind of techniques use audio-visual features alone or in combination with some other features like motion, color, and depth etc. A nice approach is used in [36] for summarizing movies to produce clips or trailers. A video summarization technique, which is specific to music videos, is proposed in [37]. The technique is based on the methods of video highlights. Using audio-visual features for key frames extraction is efficient. But this technique lacks efficiency when audio from a video is missing.

2.6 Speech Transcript Based Approach

In [38,39] frequency of word is used for creating summary of long video. This type of videos do not contain the close-caption text. First of all whole video is segmented into small coherent segments and audio pause boundary detection is done through temporal analysis of pauses between words. Segment boundary is declared between these words. But before doing this, speech transcript is generated using speech recognition systems such as IBM Via voice [40], and a shot table is generated. Ranking of each segment is done by detecting dominant word pairs using an adaptation method [41]. The rank of those segments increases that contain the first 30 of dominant word pairs. Video skims are generated by selecting those segments that maximize the cumulating score of summary and the duration of summary is provided by user. This technique is useful for those videos that contain gradual pauses between speech of the speaker, while the audience is silent. It fails in noisy videos and also for those in which there is continuous talking.

2.7 Object Based Approach

These techniques are used when user concerns more about objects in the video. Content change in video can be detected by appearance and disappearance of such objects [12,42,43,44]. Generally, such techniques are based on detecting, tracking, and keeping history of objects lifetime. However, different studies use different approaches to deal with each of the step described earlier. For example, technique of temporal segmentation to extract key frames is proposed in [45]. First of all, the temporal sub regions are defined and analysis is done on each sub region. Then motion signature for each sub region is defined and indexing is done. After indexing, key frames are extracted, that represents important stages in each object's life time. Object based approach is mostly used in surveillance applications. However, some non-surveillance applications also use this approach, but it depends on the application domain and interest of end users. One needs to be very careful using object based approach at rush places, as false objects may lead to a false summary.

3 Video Summarization Using Clustering

Clustering is most frequently used technique when we encounter similar characteristics or activities within a frame. It also helps to eliminate those frames which have irregular trends. As discussed earlier, other methods for video summarization enable more efficient way of browsing video, but create summary either too long or confusing. Video summarization based on clustering is classified into similar ctivities, K-means, partitioned clustering, and spectral clustering (Figure- 1).

3.1 Similar Activities Based Clustering

Video synopsis created by displaying similar activities which can be performed in different time periods is an effective way to present video data in summarized form [46]. First of all activity is defined; activity is a dynamic object appearing in multiple frames and defined by sequence of object masks in those frames. Activities are divided into subparts called tublets. A comparison is done on tublets and activity features are extracted. Next step is to determine similarity between activities and distance between activities is used for clustering. Then play time is assigned to objects within each cluster and play time is assigned to each cluster. At last, desired clusters are selected for summarization. Clustered summaries are very clear, easy to access. Moreover, irregular activities are easy to detect and presents a structured browsing of objects. Wrong perception about activity may lead to false summary.

3.2 K-means Clustering

K-means clustering uses k-means algorithm for clustering of key frames [47,48]. A method, using concept of histogram computation in combination with k-means

algorithm is proposed in [31]. It splits the input file into k segments and takes first frame of each segment, as a representative of this segment. It then computes histograms from these frames and cluster the histogram using k-means algorithm. Desired segments are selected and inserted into a list. Finally, lists are joined to generate desired summary. This technique is not very good for videos, in which scene remain static for long time as it selects repeated segments.

3.3 Partitioned Clustering

This technique works by removing the visual content redundancy that exists among the video frames [16,17]. First of all, the whole video is grouped into clusters such that each cluster contains frames of similar visual content. This clustering is done using partition clustering technique [49]. Cluster validity analysis is done in order to find the number of clusters that is optimal for the given sequence. Finally, key frames are concatenated to produce the video summary which is dependent on the number of shots that are contained in the video. It fails to generate good video summary for those videos where change in scene is rapid and every frame of video presents a different view.

3.4 Spectral Clustering

Spectral clustering uses the spectrum of similarity matrix of data in order to reduce dimensions for clustering. Spectral clustering technique for video summarization is proposed in [15]. There is also a novel technique based on human face detection and spectral clustering [50]. Human face detection is done to compute the location, size, and number of faces. Spectral clustering is used to cluster the desired key frames. This technique is useful in videos where our concern is only humans only. But in video which contains multiple faces in each frame this algorithm does not work well. Moreover, hidden faces may mislead to false summary.

4 Event Based Key Frame Selection

The most difficult task in such kind of techniques is the description and detection of interesting events [12,18]. Event detection consists of two steps; in first step, difference between pixel intensities of current and reference frame is calculated and absolute value of this difference is taken. Then for each difference frame, energy is calculated. In second step, finding of those frames is done that are showing some events. Then with respect to this frame, reference frame will be refreshed. Then these frames are used as input for spectral clustering algorithm.

Spectral clustering algorithm is used to create short video skims for dynamic summary and key frames are used for static browsing. Each cluster is used to create summary for some duration and more than one segment - that representing the events - from one cluster includes in the summary [15]. This technique is most suitable for those videos that are captured from static camera. The reason

is that in static camera environment, the background remains same, so reliability of event detection increases. This technique is not good for those videos where background is changing, since this change may be considered as event.

5 Shot Selection Based Approach

A shot is important component of a video, that is contiguous in appearance and time space. There are many techniques to segment video into shots [6,38]. Key frame selection from a shot [20] is also very challenging task. First important shots are measured. For example in film shooting, camera shifted from one actor to other, it consists of two shots. Hierarchical clustering methods exist in which each frame in video is considered as unique cluster and by merging two closest clusters, number of clusters is reduced. Many techniques are available for measuring distance between frames such as the color histogram distance [14], and transformation co-efficient distance [51]. Hierarchical structuring will result in a tree form, such as individual frames on the leaves of tree. Important clustered are determined by assigning weight to each cluster. Greedy algorithm or dynamic programming can be used to find optimal layout. But [19] proposed a frame packing algorithm to find best layout and this algorithm is better than dynamic programmig and greedy approach. This algorithm packs best sequence of frame in a block. Best sequence is found by finding all possible ways for a block. Iteratively it packs the blocks until all blocks are packed. So, by applying frame packing algorithm, it organizes the shots and make summary of long videos. This technique is suitable, when end user's focus is to select important shots of video. It is used only with moving camera, because it makes selection on shot boundary detection.

6 Video Summarization Using Trajectory Analysis

When it comes to analyzing dynamic environment in a video, this technique is equally efficient. In most of the surveillance applications, camera is fixed and background does not change. In this type of environment, the important element for video generalization is to detect the behavior of moving objects with respect to time.

6.1 Spatio-temporal Based Approach

Moving objects can be traced in three dimensional space (x,y,t). Here (x,y) is spatial dimension and (t) is time dimension. Summary is being generated by extracting and analyzing the trajectory. This technique is based on the identification of nodes in these trajectories and marks it as critical points in the video. [26,27] presented a decent approach towards trajectory analysis. The time corresponding to a node provide respective frame. Nodes are assigned with respect to spatio-temporal behavior, if spatio-temporal breakpoint occurs then more nodes

are assigned and if motion is smooth then fewer nodes are assigned [52]. Self organization map is used for node placement. Number of nodes is selected for controlling the generalization of video. Generalization of nodes and trajectories of objects make a hierarchical tree data structure which is used to describe the movement of objects in a scene. Spatio-temporal approach is highly useful in surveillance applications. However, it fails when camera motion is also involved.

6.2 Curve Simplification Approach

Curve simplification algorithms [5,28] are best choice to analyze motion, because they use the spatial data of moving objects. In first step, important frames are determined by curve saliency. In second step, finalized key frames are selected by clustering. Mesh saliency algorithms which takes important part of mesh can also be used for curve simplification [25].

Curve saliency is done by computing Gaussian weighted average. Reduction process is done if excessive number of frames has values greater than average. The key frames are selected by each angle of curve. In final step, those key frames are deleted having least importance among nearest key frames and the selected key frames make the summary of long video. This method deletes the redundant key frames and represents the motion capture sequence by only small percentage of all captured frames.

7 Mosaic Based Approach

Mosaic based approach is used to generate a panoramic image from a large number of consecuetive frames having some important content. This technique is proposed in [21,22,23,24,53,54] and known by different names as salient stills, video sprits, and video layers. Two fundamental steps that are used in these studies are: In first step, a global motion model is fitted between two successive motion frames and then a panoramic image is generated using motion model [55]. The drawback of this technique that it represents only static background information, but does not contain any information about moving objects. In order to deal with this drawback, [56] proposed a technique in which two types of mosaics are defined, static background mosaic and synopsis mosaic. In static background mosaic, the focus is on background scenes, whereas in synopsis mosaic, visual summary is generated by analyzing the trajectory of moving objects. After this, combination of these two mosaics is used to generate single panoramic image. This technique is ideal for situations in which multiple cameras are used and background is static.But it fails when background/foreground is rapidly changing.

8 Analysis

This study presents numerous techniques of video summarization and also classifies them into different categories based on methodology and characteristics.

Table 1. Suitable video summarization techniques in different domains

Types of videos	Summarization techniques
Sports videos	Motion based approach Color based approach Clustering Event based approach Object based approach
Music videos	Audio-visual based approach Clustering
Traffic videos	Motion based approach Object based approach Event based approach Clustering
Surveillance videos	Color based approah Motion based approach Event based approach Clustering Trajectory analysis
Movie highlights	Audio-visual based approach Similar activity based approach Motion based approach Shot selection based Mosaic based approach
Phone calls	Audio-visual based approach
Sign language communication	Gesture based approach
Rushy videos	Motion based approach Clustering Audio-visual based approach Shot selection based approach
Documentary	Motion based approach Shot selection based approach Mosaic based approach Clustering
News videos	Audio-visual based approach Mosaic based approach Gesture based approach
Rushy videos	Gesture based approach Audio-visual based approach

In Table 2, techniques are classified according to static summary, dynamic summary, fixed camera, moving camera, having knowledge of significant contents, and without knowledge of significant contents. Some techniques work well when camera is fixed and in contrast, others are good with moving camera. For example, spatio-temporal based approach is used with static camera and shot selection based approach is used with a moving camera. Similarly, mosaic based approach is a good choice, when the focus is on detailed and quick overview of entire scene. However, it is difficult to extract only few key frames representing whole scene.

Table 2. Quick analysis of video summarization techniques

Techniques	Static summary	Dynamic summary	Fixed camera	Moving camera	Significant contents knowledge	Without significant contents knowledge
Motion based	No	Yes	Yes	Yes	No	Yes
Color based	Yes	No	Yes	Yes	No	Yes
Dynamic contents based	No	Yes	Yes	Yes	Yes	Yes
Gesture based	Yes	No	Yes	Yes	Yes	No
Audio-Visual based	No	Yes	Yes	Yes	Yes	No
Speech transcript based	No	Yes	Yes	Yes	Yes	No
clustering based	Yes	Yes	Yes	Yes	Yes	Yes
Event based	Yes	No	Yes	No	Yes	No
Shot selection based	Yes	No	No	Yes	No	Yes
Trajectory based	Yes	No	Yes	No	Yes	No
Mosaic based	Yes	No	No	Yes	No	Yes

Different domains require different approaches towards video summarization. Some techniques do fit in more than one situations while others are specific to a single situation. Table 1 presents techniques with respect to application area. Each technique contains specific prons and cons regarding a specific situation. For example trajectory analysis technique outperforms in surveillance applications but it does not produce good results for non-surveillance applications as camera motion is very high. However, the presented techniques tend to save time and cost as well as reduce efforts needed for browsing of long sequence of videos.

9 Conclusion

The fast evolution of video technology has brought huge volume of video data. There is need to browse, retrieve and store this video data efficiently. Nowadays, people have no time to spend on watching whole videos. People want to see only important content of videos. So there is need of efficient video summarization techniques to facilitate users of all categories. Table 2 represents suitable techniques for specific domain. It will save time and cost as one can quickly select the most suitable technique according to one's need. Although, a lot of effort is done to efficiently store, retrieve and browse large video streams, but still there are many deficiencies in video summarization techniques that need to be addressed.

References

1. Divakaran, A., Peker, K.A., Sun, H.: Video Summarization Using Motion Descriptors. In: Conf. on Storage and Retrieval from Multimedia Databases (2001)
2. Ju, S.X., Black, M.J., Minneman, S., Kimber, D.: Summarization of Video-Taped Presentations: Automatic Analysis of Motion and Gestures. IEEE Transactions on CSVT (1998)

3. Fujimur, K., Honda, K., Uehara, K.: Automatic Video Summarization by Using Color and Utterance Information. In: Proceedings 2002 IEEE International (2002)
4. Zhang, H.J., Low, C.Y., Smoliar, S.W.: Video parsing and browsing using compressed data. Multimedia Tools and Applications 1, 89–111 (1995)
5. DeManthon, D., Kobla, V., Doermann, D.: Video Summarization by Curve Simplification. In: Proceedings of the Sixth ACM International Conference on Multimedia (1998)
6. Koskela, M., Sjberg, M., Laaksonen, J., Viitaniemi, V., Muurinen, H.: Rushes Summarization with Self-Organizing Maps. In: Proceedings of the International Workshop on TRECVID Video Summarization (2007)
7. Truong, B.T., Venkatesh, S.: Video Abstraction: a Systematic Review and Classification. ACM Transactions on Multimedia Computing, Communications, and Applications 3(1) (2007)
8. Li, Y., Zhang, T., Tretter, D.: An Overview of Video Aabstraction Techniques. Technical Report HPL (2001)
9. Barbieri, M., Agnihotri, L., Dimitrova, N.: Video summarization: methods and landscape. In: Proceedings of SPIE, vol. 5242, p. 1 (2003)
10. Adami, N., Benini, S., Leonardi, R.: An Overview of Video Shot Clustering and Summarization Techniques for Mobile Applications. In: Proceedings of the 2nd International Conference on Mobile Multimedia Communications (2006)
11. Wolf, W.: Key frame selection by motion analysis. In: ICASSP, vol. 2, pp. 1228–1231 (1996)
12. Wang, F., Ngo, C.W.: Summarizing rushes videos by motion, object and event understanding. IEEE Transactions on Multimedia 14 (2012)
13. Zhang, H.J., Wu, J., Zhong, D., Smoliar, S.W.: An integrated system for content based video retrieval and browsing. Pattern Recognition 30, 643–658 (1997)
14. Chheng, T.: Video Summarization Using Clustering. Department of Computer Science University of California, Irvine (2007)
15. Damnjanovic, U., Fernandez, V., Izquierdo, E.: Event Detection and Clustering for Surveillance Video Summarization. In: Proceedings of the Ninth International Workshop on Image Analysis for Multimedia Interactive Services. IEEE Computer Society, Washington, USA (2008)
16. Hanjalic, A., Zhang, H.: An integrated scheme for automated video abstraction based on unsupervised cluster-validity analysis. IEEE Transactionson Circuits and Systems for Video Technology 9, 1280–1289 (1999)
17. Vctor Valdes, J.M.M.: On-Line Video Skimming Based on Histogram Similarity. In: Proceedings of the International Workshop on TRECVID Video Summarization (2007)
18. Li, B., Sezan, M.I.: Event Detection and Summarization in Sports Video. In: Content-Based Access of Image and Video Libraries, CBAIVL IEEE Workshop (2001)
19. Uchihachi, S., Foote, J., Wilcox, L.: Automatic Video Summarization Using a Measure of Shot Importance and a Frame Packing Method. United States Patent 6, 535,639, March 18 (2003)
20. Evangelopoulos, G., Rapantzikos, K., Potamianos, A., Maragos, P., Zlatintsi, A., Avrithis, Y.: Movie Summarization Based on Audio-Visual Valiency Detection. In: IEEE Intl Conf. Image Processing (ICIP), San Diego, CA (2008)
21. Wang, J., Adelson, E.: Representing moving images with layers. IEEE Transactions on Image Processing 3 (1994)
22. Pope, A., Kumar, R., Sawhney, H., Wan, C.: Video abstraction: Summarizing video content for retrieval and visualization (1998)

23. Aner, A., Kender, J.R.: Video Summaries through Mosaic-Based Shot and Scene Clustering. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part IV. LNCS, vol. 2353, pp. 388–402. Springer, Heidelberg (2002)
24. Sawhney, H., Ayer, S.: Compact representation of video through dominent and multiple motion estimation. IEEE Trans. on Pattern. Analysis and Machine Intelligence 18 (1996)
25. Lee, C.H., Varshney, A., Jacob, D.W.: Mesh saliency. ACM Transaction on Graphics, 659–666 (2005)
26. Ngo, C.W., Ma, Y.F., Zhang, H.J.: Automatic Video Summarization by Graph Modeling. In: Proceedings of the 9th IEEE International Conference on Computer Vision (2003)
27. Qiu, X., Jiang, S., Liu, H., Huang, Q., Cao, L.: Spatial temporal attention analysis for home video. In: IEEE International Multimedia and Expo, vol. 23 (2008)
28. Bulut, E., Capin, T.: Key Frame Extraction from Motion Capture Data by Curve Saliency. In: Proceedings of 20th Annual Conference on Computer Animation and Social Agents, Belgium (2007)
29. Peyrard, N., Bouthemy, P.: Motion-Based Selection of Relevant Video Segments for Video Summarization 26(3) (2005)
30. Li, C., Wu, Y.T., Yu, S.S., Chen, T.: Motion-focusing key frame extraction and video summarization for lane surveillance system. In: 16th IEEE International Conference on Image Processing (ICIP), pp. 7–10 (2009)
31. Chen, F., Cooper, M., Adcock, J.: Video Summarization Preserving Dynamic Content. In: Proceedings of the International Workshop on TRECVID Video Summarization (2007)
32. Adnan, H., Mufti, M.: Video Summarization Based Handout Generation from Video Lectures: A Gesture Recognition Framework. In: 5th WSEAS International Conference on Signal Processing, Computational Geometry and Artificial Vision (2005)
33. Kosmopoulos, D.I., Doulamis, A., Doulamis, N.: Gesture-based video summarization. In: ICIP IEEE International Image Processing, pp. 11–14 (2005)
34. Furini, M., Ghini, V.: An Audio-Video Summarization Scheme Based on Audio and Video Analysis. In: IEEE CCNC (2006)
35. Divakaran, A., Peker, K., Radhakrishnan, R., Xiong, Z., Cabasson, R.: Video summarization using mpeg7 motion activity and audio descriptors. In: Video Mining, vol. 91 (2003)
36. Evangelopoulos, G., Rapantzikos, K., Potamianos, A., Maragos, P., Zlatintsi, A., Avrithis, Y.: Movie Summarization Based on Audiovisual Saliency Detection. In: ICIP (2008)
37. Shao, X., Xu, C., Maddage, N.C., Kankanhalli, M.S., Jin, J.S., Tian, Q.: Automatic summarization of music videos. ACM Transactions on Multimedia Computing,Communications and Applications (TOMCCAP) 2 (2006)
38. Taskiran, C.M., Amir, A., Ponceleon, D., Delp, E.J.: Auto-mated video summarization using speech transcripts. In: Proceedings of SPIE Conference on Storage and Retrieval for Media Databases volume, San Jose, CA, pp. 20–25 (2002)
39. Taskiran, C.M., Pizlo, Z., Amir, A., Ponceleon, D., Delp, E.J.: Automated video program summarization using speech transcripts. IEEE Transactions on Multimedia (2006)

40. Bahl, L.R., Aiyer, S.B., Bellegarda, J.R., Franz, M., Gopalakrisnan, P.S., Nahamoo, D., Novak, M., Padmanabhan, M., Picheny, M.A., Roukos, S.: Performance of the IBM Large Vocabulary Continuous Speech Recognition System on the ARPA Wall Street Journal Task. In: Proceedings of IEEE International Conference on Acoustic, Speech and Signal Processing, Detroit, MI (1995)
41. Dunning, T.E.: Accurate methods for the statistics of surprise and coincidence. Computational Linguistics 19(1), 61–74 (1993)
42. Liu, D., Chen, T., Hua, G.: A hierarchical visual model for video object summarization. IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (2010)
43. Kim, C., Hwang, J.N.: An Integrated Scheme for Object-Based Video Abstraction. In: Proceedings of the 8th ACM International Conference on Multimedia (2000)
44. Lee, Y.J., Ghosh, J., Grauman, K.: Discovering Important People and Objects for Egocentric Video Summarization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR (2012)
45. Ferman, A.M., Gunsel, B., Tekalp, A.M.: Object-Based Indexing of MPEG-4 Compressed Video. In: Proceedings of IS&T/SPIE Symp. on Electronic Imaging (1997)
46. Pritch, Y., Ratovitch, S., Hendel, A., Peleg, S.: Clustered synopsis of surveillance video. In: 6th IEEE Int Conf. on Advance Video and Signal Base Selection (AVSS 2009), Genoa, Italy, pp. 2–4 (2009)
47. Ali Amiri, M.F.: Hierarchical key frame-based video summarization using qr-decomposition and modified k-means clustering. EURASIP Journal on Advaces in Signal Processing (February 2010)
48. Farin, D., Effelsberg, W., Peter, H.N.: Robust Clustering Based Video Summarization with Integration of Domain Knowledge. In: Proceedings 2002 IEEE International Conference (2002)
49. Jain, A.K., Dubes, R.C.: Algorithms for Clustering Data. Prentice-Hall, Englewood Cliffs (1988)
50. Peker, K.A., Bashir, F.I.: Content-Based Video Summarization using Spectral Clustering. Mitsubishi Electric Research Laboratories Cambridge, MA. University of Illinois at Chicago, Chicago, IL (2009)
51. Girgensohn, A., Foote, J.: Video Frame Classification Using Transform Coefficients. In: ICASSP 1999 (1999)
52. Stefanidis, A., Partsinevelos, P., Peggy Agouris, P.D.: Summarizing Video Datasets in the Spatiotemporal Domain (2000)
53. Massey, M., Bender, W.: Salient stills: Process and practice. IBM Systems Journal 35 (1996)
54. Lee, M., Chen, W., Lin, C., Gu, C., Markoc, T., Zabinsky, S., Szeliski, R.: A layered video object coding system using sprite and affine motion model. IEEE Transactions on Circuits and Systems for Video Technology (1997)
55. Vasconcelos, N., Lippman, A.: A Spatio Temporal Motion Model for Video Summarization. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (1998)
56. Iran, M., Anandan, P.: Video indexing based on mosaic representation. IEEE Computer Society (1998)