e-ISSN: 2349-9745
p-ISSN: 2393-8161
Scientific Journal Impact Factor (SJIF): 1.711
International Journal of Modern Trends in Engineering and Research
www.ijmter.com

A SURVEY ON VIDEO SEGMENTATION THE FUTURE ROADMAP

Abhishek Kundu[1], Prof. Nitin Janwe[2]
[1] *M. Tech student, Department of Computer Technology, Rajiv Gandhi College Of Engineering Research and Technology, Chandrapur, Maharastra, India.*
[2] *H.O. D , Department of Computer Technology, Rajiv Gandhi College Of Engineering Research and Technology, Chandrapur, Maharastra, India*

Abstract-Effective video segmentation is a challenging problem in digital storage media. This paper reviews the substantial research on the subject of video segmentation, expounding its basic terminology, techniques and algorithms. It further gives an overview of the state- of – art depicting some previous attempts to study video segmentation .Its practical and potential applications are also discussed, followed by the issues and challenges that will keep the field dynamic and lively for years to come.

I. INTRODUCTION

Video segmentation is an application of computer vision aimed at automating the extraction of an object from a series of video frames. However, it is a difficult problem, especially to compute at real-time, interactive rates. Although general application to video is difficult because of the high range of image scenarios, user interaction can help to reduce the problem space and speed up the computation.

The video segmentation is an imperative technique used for the improvement of video quality on the basis of segmentation [1]. The function of video segmentation is to segment the moving objects in video sequences [2]. Video segmentation is the first step towards automatic annotation of digital video sequences. Its goal is to divide the video stream into a set of meaningful and manageable segments (*shots*) that are used as basic elements for indexing. Each shot is then represented by selecting key frames and indexed by extracting spatial and temporal features. Shot segmentation is an essential first step to video segmentation. The color histogram-based shot boundary detection algorithm is one of the most reliable variants of histogram-based detection algorithms. It is not unreasonable to assume that the color content does not change rapidly within but across shots.

Replacing a long video by a small number of representative segments provides a synthetic description of the document, which can be exploited for numerous applications including both home video and professional usages. However, the construction of video summary remains an open problem at the source of active research activities. The main difficulty obviously relies on the detection of semantic events from low-level information.

II. LITERATURE SURVEY

Since last decade many different approaches for video segmentation have been proposed. The first attempts were aiming at automatically finding shot boundaries within videos. Shot boundaries can be defined as the physical boundaries where camera changes happen. Many shot boundary detection algorithms have been proposed [3]. The best ones achieve a high accuracy and thus this task is

regarded as essentially solved. The problem is that on the one hand even short videos can consist of a large number of shots and on the other hand, some kinds of videos consist of one single shot. Therefore, it is insufficient to index a video on the shot level.

This paper provides a starting point for researchers and developers how to tackle specific problems. Table 1 shows the brief survey of video segmentation

Table 1 : Brief Survey Of Video Segmentation

Author Name	**Type Dataset**	**Size Dataset**	**Similarity Matching**	**Discussion**
Kender and Yeo [7]	Sitcoms	1 video/18 scenes	color	Probable scene boundaries are measured by calculating a short-term shot-to-shot coherence
Rui et al. [8]	Movies	8 movies/-	color, activity, temporal constraint	This method select the first and the last frame from all shots of a video as key-frames. Scenes are identified by merging overlapping groups of shots.
Kwon et al. [9]	Movies	1 video/11 scenes	color Motion shot duration	It improve algorithm by using motion features and by applying an improved overlapping links method, which needs fewer shot comparisons to identify overlaps.
Adams et al. [10]	Movies	4 videos/	motion, shot duration	It extract expressive elements from motion pictures. In particular, they take advantage of the fact that film makers often use a different tempo for adjacent scenes.
Chen et al. [11]	MPEG-7 Test Videos, Movies	6 videos/511 scenes	color, texture	It introduces rule-based algorithm For each shot of a video several key-frames are extracted.
Tavanapong And Zhou [12][13]	Movies	2 videos/120 scenes	color, shot activity	They tries to identify common objects in the background, but only the four corner regions and the background region at the top are used to compare key-frames based on color and motion features.
Zhu and Liu [14]	MPEG Test Video, Movies	6 videos/202 scenes	color, texture, temporal constraints	It present an approach for the segmentation of continuously recorded TV broadcasts. A

				spatio-temporal clustering algorithm
Rasheed and Shah [15]	Movies, Sitcoms	5 movies/-	color, motion, temporal constraints	It introduce a shot similarity graph. Edges indicate the likelihood that connected shots belong to one scene.
Sakarya and Telatar [16]	Movies	3 movies/70 scenes	color temporal constraints	This paper propose a tree-based peeling strategy using dominant sets for scene segmentation. The idea of this approach is to perform a step-wise partitioning of the shot graph. Two sets are created in each iteration of this algorithm: (1) the dominant set and (2) the remaining shots.
Xie [17]	Soccer Videos	4 videos/	color, motion	Here ,uses Hidden Markov Models (HMM) for the segmentation of soccer videos. The dominant color ratio and the motion intensity are extracted from the compressed domain of MPEG videos. A manually labeled training set is used to train the HMM.
Ariki et al. [18]	Sports Videos	4 videos/	color, audio, text	A real-time highlight extraction algorithm for live baseball videos is introduced It is a partial segmentation approach. Pitcher scenes are extracted using visual features and a speech recognition system is used to extract text from the audio stream.
Hauptmann and Witbrock[19]	News Videos	13 videos/749 scenes	color, face detection, audio,text	Introduce an approach for news video segmentation. MPEG optical flow, the presence of black frames, face detection, and color similarity are used to identify visual scene breaks.

III. VIDEO SEGMENTATION

A. Basic Structure of Video Segmentation

The video segmentation is a technique of detecting changing frames in video and one of the important techniques required for efficient management of video data [1] several video segmentation algorithms have been proposed. They can be classified into types shape based video segmentation, edge information based video segmentation, image based video segmentation, texture based video segmentation and color based segmentation [2]. Basically, the approaches presented in video segmentation are categorized based on three classes of low-level features: visual features, audio features and textual features. Using these three basic classes and all possible combinations of them, this survey groups the presented approaches into seven categories as: (1) visual-based, (2) audio-based, (3) text-based, (4) audio-visual, (5) visual-
textual, (6) audio-textual, and (7) hybrid segmentation approaches. Figure 1 shows the classification of video segmentation methods.

During a video segmentation process, information about different hierarchical levels of a video can be extracted. For example, a video can be segmented into key frames, shots, scenes or even groups of scenes, but also other logical hierarchies are possible. In addition to the segmentation hierarchy, we distinguish between full and partial segmentation. A full segmentation partitions an entire video into disjoint units, in such a way that the original video can be reproduced again by composing all units together [3]. In partial segmentation only certain parts are extracted from a video and the rest is disregarded. The original video cannot be reproduced. This is common for surveillance scenarios or for highlight extraction in sports videos, where parts of the video where nothing happens are left out.

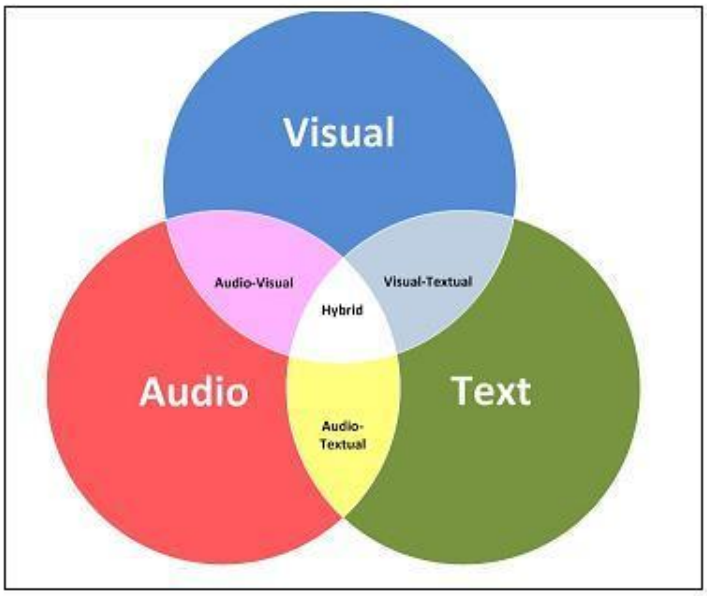
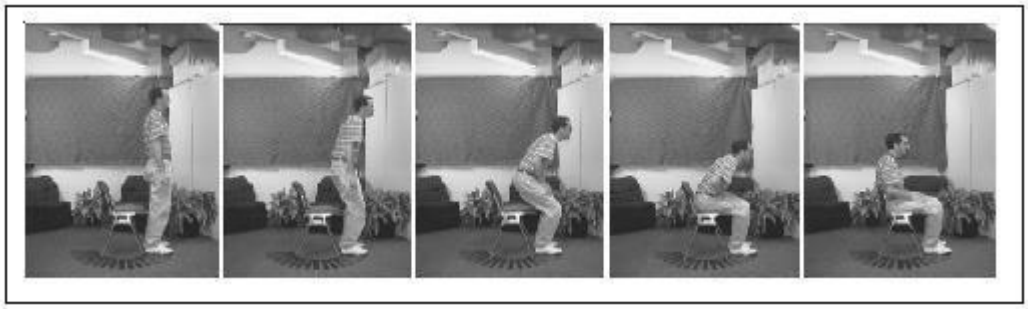


Figure 1: Classification of Video Segmentation Methods

Image difference is one of the simplest and most used techniques for detecting changes. It consists in thresholding the intensity difference of two frames pixel by pixel. The result is a coarse map of the temporal changes. An example of an image sequence and the image difference result is shown in figure 1. Despite its simplicity, this technique cannot be used in its basic version because it is really sensitive to noise. Moreover, when the camera is moving the whole image is changing and, if the frame rate is not high enough, the result would not provide any useful information.

Statistic theory is widely used in the motion segmentation field. In fact, motion segmentation can be seen as a classification problem where each pixel has to be classified as background or foreground. Statistical approaches can be further divided depending on the framework used. Common frameworks are Maximum A posteriori Probability (MAP), Particle Filter (PF) and Expectation Maximization (EM). Statistical approaches provide a general tool that can be used in a very different way depending on the specific technique [15].



B. The main attributes of a Video segmentation

The main attributes of a Video segmentation algorithm can be summarized as follows.

1. Feature-based or Dense-based: In feature-based methods, the objects are represented by a limited number of points like corners or salient points, whereas dense methods compute a pixel-wise motion [4].

2. Occlusions: it is the ability to deal with occlusions.

3. Multiple objects: it is the ability to deal with more than one object in the scene.

4. Spatial continuity: it is the ability to exploit spatial continuity.

5. Temporary stopping: it is the ability to deal with temporary stop of the objects.

6. Robustness: it is the ability to deal with noisy images (in case of feature based methods it is the position of the point to be affected by noise but not the data association).

7. Sequentially: it is the ability to work incrementally; this means for example that the algorithm is able to exploit information that was not present at the beginning of the sequence.

8. Missing data: it is the ability to deal with missing data.

9. Non-rigid object: it is the ability to deal with non-rigid objects.

Furthermore, if the aim is to develop a generic algorithm able to deal in many unpredictable situations there are some algorithm features that may be considered as a drawback,
1. specifically: Prior knowledge: any form of prior knowledge that may be required.
2. Training: some algorithms require a training step.

C. Evaluation of Video Segmentation Approaches

An overview of the datasets used for the evaluation
1 Datasets:
- The MPEG-7 dataset: Provides different types of videos including a ground truth for scenes. It is used for the evaluation of six approaches presented in this survey.

- The TRECVID : Initiative also provides large video collections and uniform scoring procedures for various tasks related to video retrieval issues.
- NIST also provides a database consisting of eight videos to enable the scientific comparison of solutions of digital video search, retrieval and display[76].
- Most of the approaches were evaluated with movies, TV shows or sitcoms and news videos. Only the remaining few focus on special domains like sports videos, documentaries or home videos.

2. Evaluations Methods: The presented approaches do not only differ in the used datasets, but also in the applied ground truths and evaluation measures. These Evaluations methods are based three types of ground truth namely 1) Manually obtained, 2) Provided and 3) DVD Chapters. The fact that only a minority of the approaches relies on common ground truths makes it hard to compare different approaches.

Different evaluation measures are used to quantify the video segmentation accuracy. In most cases recall and precision are used. In many information retrieval fields these measures are used to express the accuracy of retrieval results. This seems to be the reason, why they are also predominantly used for expressing the video detection accuracy.

Video segmentation is a ways of dividing a movie into meaningful segments. In the context of lecture capture, segmentation is best applied to captured screen presentation, that the presenter goes through slide after slide. As a result, video segmentation returns the exact time points of slide changes on the timeline, which allows for sophisticated ways for the learner to browse the lecture content, as shown in the slides section of the Matterhorn Media Player[].

IV. FUTURE CHALLENGES IN VIDEO SEGMENTATION

There are numerous applications where digital video is acquired, processed and used, such as surveillance, general identity verification, traffic, criminal justice systems, civilian or military video processing et al.

- To illustrate the complexity of video segmentation, we identify three major challenges namely temporal coherence, automatic processing and Scalability issue.

- There is more potential in enhancing the results by incorporating human knowledge in video segmentation approaches. Interactive segmentation of videos by combining automatically retrieved scene candidates with an interactive segmentation tool may be a possible solution.

- High-quality layer separation from a video is a very challenging problem because tightly coupled color, depth, and motion give rise to a large number of variables and significant inexactness in computation.

- It is not worth to put many efforts on improving and tuning automatic algorithms if only minor, hardly recognizable improvements can be achieved. There is more potential in enhancing the results by incorporating human knowledge in video segmentation approaches.

V. CONCLUSION

In the discipline of computer vision, image processing is a quickly moving field. Its growth has been fueled by technological advances in digital imaging, computer processors and mass storage devices. In this paper, we have discussed the basic concepts of video segmentation, which includes brief review of segmentation methods. These may suggest that the usage of this information may help to improve method performances, especially in terms of robustness and ability to deal with occlusions. Future work in expanding existing techniques to handle more general writings and crossing domains is an exciting opportunity for both academia and businesses.

ACKNOWLEDGMENT

We express our thanks to publishers, researchers for making their resource available & teachers for their guidance. We also thank the college authority for providing the required infrastructure and support. Last but not the least we would like to extend a heartfelt gratitude to friends and family members for their support.

REFERENCES

[1] Irena Koprinska!, Sergio Carrato Temporal video segmentation: A survey 2001 Elsevier Signal Processing: *Image Communication* 16 (2001) 477}500

[2] Ram Kumar Yadav, Sanjeev Sharma and Jitendra Singh Verma,"Deformation and Improvement of Video Segmentation Based on morphology Using SSD Technique" IJCTA, Vol. 2, no. 5, pp.1322- 1327,2011.

[3] Shiping Zhu Xi Xia Qingrong Zhang " A Novel Spatio-Temporal Video Object Segmentation Algorithm" IEEE.,2008.

[4] Boreczky, J. and Rowe, L., "Comparison of Video Shot Boundary Detection Techniques", Proc.SPIE Conference on Storage and Retrieval for Still Image and Video Databases IV, San Jose, CA, February, 1996, pp. 170-179.

[5] Manfred del Fabro and Laszlo B•osz•ormenyi State-of-the-Art and Future Challenges in Video Scene Detection: a Survey DelFabro_MMSJ2013.

[6] Weiming Hu, Senior Member, IEEE, Nianhua Xie, Li Li, Xianglin Zeng, and Stephen Maybank. IEEE Transactions On Systems, "A Survey on Visual Content-Based Video Indexing and Retrieval" Man, And Cybernetics—Part C: Applications And Reviews, Vol. 41, No. 6, November 2011.

[7] J. Kender and B.-L. Yeo. Video scene segmentation via continuous video coherence. In Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on, pages 367 {373, jun 1998.

[8] Y. Rui, T. S. Huang, and S. Mehrotra. Constructing table-of-content for videos. Multimedia Systems, 7(5):359{368, September 1999.

[9] Y.-M. Kwon, C.-J. Song, and I.-J. Kim. A new approach for high level video structuring. In Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on, 2000.

[10] B. Adams, C. Dorai, and S. Venkatesh. Toward automatic extraction of expressive elements from motion pictures: tempo. IEEE Transactions on Multimedia, 4(4):472{481, December 2002.

[11] L.-H. Chen, Y.-C. Lai, and H.-Y. Mark Liao. Movie scene segmentation using background information. Pattern Recognition,41:1056{1065, March 2008.

[12] W. Tavanapong and J. Zhou. Shot Clustering Techniques for Story Browsing. IEEE Transactions on Multimedia, 6(4):517{527, August 2004.

[13] J. Zhou and W. Tavanapong. Shot Weave: A Shot Clustering Technique for Story Browsing for Large Video Databases. In A. Chaudhri, R. Unland, C. Djeraba, and W. Lindner, editors, XML-Based Data Management and Multimedia Engineering EDBT 2002 Workshops, volume 2490 of Lecture Notes in Computer Science, chapter 17, pages 529{533. Springer Berlin / Heidelberg, Berlin, Heidelberg, November 2002.

[14] S. Zhu and Y. Liu. Video scene segmentation and semantic representation using a novel scheme. Multimedia Tools and Applications, 42(2):183{205, Apr. 2009.

[15] Z. Rasheed and M. Shah. Detection and representation of scenes in videos. IEEETransactions on Multimedia, 7(6):1097{1105, December 2005.

[16] U. Sakarya and Z. Telatar. Video scene detection using dominant sets. In Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on, pages 73 {76, 2008.}

[17] L. Xie. Structure analysis of soccer video with domain knowledge and hidden markov models. Pattern Recognition Letters, 25(7):767{775, May 2004.
[18] Y. Ariki, M. Kumano, and K. Tsukada. Highlight scene extraction in real time from baseball live video. In Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval, MIR '03, pages 209{214, New York, NY, USA, 2003. ACM.
[19] A. Hauptmann and M. Witbrock. Story segmentation and detection of commercials in broadcast news video. In Research and Technology Advances in Digital Libraries,

