International Conference on Information Engineering, Management and Security
2015 [ICIEMS 2015]

ISBN	978-81-929742-7-9	**VOL**	01
Website	www.iciems.in	**eMail**	iciems@asdf.res.in
Received	10 - July - 2015	**Accepted**	31- July - 2015
Article ID	ICIEMS038	**eAID**	ICIEMS.2015.038

Video Depiction of KeyFrames- A Review

Deepika Bajaj[1], Shanu Sharma[2]
[1,2] CSE Department, Amity University, Noida, Uttar Pradesh, India

Abstract: *Nowadays, there are numerous, unstructured and voluminous videos which leads to high collection of data on web. Searching and navigating through these videos for meaningful information is a time consuming task, whereas a good 'summarized video' can provides a user determined information about particular video sequence in definite time limits. So, there is great need of extraction of semantic and useful information from videos for various multimedia applications. The video summarization is the novel and promising method of detecting relevant and informative data from videos and also aims to provide effective and efficient storage of relevant information. This technique of summarization leads to abstraction of most representative and relevant scenes from videos and concatenates to display as one successive and uninterrupted video and thus has been powered up the rapidly progressing research domain. In this paper, all latest and enhanced approaches of video shot detection have been discussed and summarized.*

Keywords- key frame extraction, face detection, shot boundary detection, feature extraction.

I.INTRODUCTION

In recent times, the quantity of videos has been increased day by day. Videos are ample, unorganized and redundant data streams which contain images, graphics and textual information such as label, keywords, etc. Almost, every field such as entertainment, news channels or advertisements, etc. [1] involves wide use of videos. However, people use to spend their large amount of time to download huge videos in order to evaluate that these videos are relevant or irrelevant. Browsing through large amount of videos is a time consuming, tough and very tiring job for human beings. Therefore, it is hard and painful work to extract the meaningful content or desired scenes from videos. Video summarization is the best and efficient solution to transform huge and amorphous videos into organized, structured and systemized manner. Summarization of videos is process of creating concise, clear, succinct and meaningful information. Generally, it contains following three complexities-

☐ The summarized video should involve the important and desired parts or scenes from the video which means it should be as concise and brief as possible. For example, video summarization approach has the ability to provide semantic portion to user as per his requirements from the huge videos.

☐ The summarized video should maintain the semantic meaning which means it should represents the good continuous association between scenes.

☐ The summary of video should not contain any redundancy/duplicity among scenes.

Cite this article as: Deepika Bajaj, Shanu Sharma. "Video Depiction of KeyFrames- A Review." *International Conference on Information Engineering, Management and Security (2015)*: 224-230. Print.

In order to generate optimal video summaries, the two different ways and means for creating efficient and effective retrieval of video are:- summarization and highlights [2]. The prior aims to deliver a summarized and concise storyline demonstration of a video while the later aims to extract the affective data from the video. Video summarization is a technique that is important and suitable in context of searching and retrieving desired part of video. Summarizing the large videos into small videos is very helpful to people who can get the main concept or idea about the movie without watching the full video. This aims to provide the new view mode to audience. For example, by watching movie trailer, the people get to know that movie is romantic, comedy or action movie. Therefore, this technique is very significant and useful in classifying the videos and movies as well and explains the main concept and idea of video.
The summarized videos have wide variety of advantages in various fields for various purposes. Summarizing videos has wide variety of applications such as video indexing, estimating the rating of movie, etc. [3]. The parsing of videos is shown below-

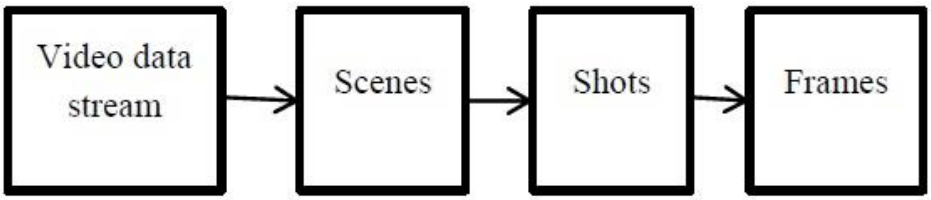

Fig1: Flowchart of parsing of video

A. Shot detection and scene detection in videos

Initially, input is taken as "video" and it consists of distinct scenes, then from those scenes further different shots are analyzed and extraction of frames is performed. The frame extraction leads to generate the most representative frames (keyframe extraction) from already extracted frames. A scene can be defined as "a section/ division of movie or video in which the set is permanent, time is uninterrupted and the action is established in one place". The scene can also be described as a collection of video shots that satisfy certain similarity along with semantic and meaningful information [4]. A scene must abide three rule i.e. unitary space, time and action.

According to Bordwell et al. [5], three types of scenes exists in videos i.e. action scene, conversational scene and suspense scene. In order to detect scene, there are many approaches and procedures such as graph based segmentation method, color histogram based method etc. A scene can contain various transitions in videos. Transitions are the variations from one scene to another. During the changing of shot or scene, discontinuities are found. There are four types of transition [6]:

1) Hard cut- This is the most simple and basic kind of transition. When one shot replaces the other instantly, then hard cut occurs. In normal feature film, there are thousands number of hard cuts. Cuts are useful for the continuous and enhanced movement of the movie.

2) Fade- Fade is further divided into two parts-

☐ *Fade in* – It is basically used in the starting of the movie. Fade-in occurs when scene changes gradually from black /solid color to picture/image.

☐ *Fade out*- It is basically used in the ending of the movie. Fade-out occurs when scene changes gradually from current image to solid paint.

Cite this article as: Deepika Bajaj, Shanu Sharma. "Video Depiction of KeyFrames- A Review." *International Conference on Information Engineering, Management and Security (2015)*: 224-230. Print.

Fig2: Hard cut effect and fade effect

3) Dissolve- It is also known as overlapping. Dissolve occurs when there is replacement of one shot with the next gradually. One shot disappears as the next shot appears. For some time both shots overlap each other and used to represent the passage of time.

Fig3: Dissolve effect

4) Wipe- Wipes are dynamic kind of transition. When one shot pushes the other off frame, then wipe takes place.

Fig-4 Wipe effect

Scenes are the combination of extracted shots. For each detected shots, variable count of keyframes can be determined. The keyframe which is extracted first is always lies near a shot boundary. These frames can be further combined to form semantic and most representative scenes.

B. Key Frame Extraction

Key frames extraction is the most significant step in abstraction of videos. Extraction of key frames leads to set of meaningful images from the sequences of videos. Many researches are still working for the better and improved automatic system of key frame extraction. This step is quite effective and efficient in providing the summarization of video.

Below some definitions to extract keyframes are listed [7]:

☐ Reference Frame- refers to first frame in every shot.
☐ General Frames- refers to rest of the frames other than reference frames;
☐ "Dynamic shot factor" max (i)- The max x2 histogram within shot i;
☐ Static shot and dynamic shot- if max(i) is larger than mean calculated then it is considered to be dynamic shot otherwise it is static shot.
☐ $Fm(k)$: The kth frame within the current shot, k=1,2,3… $Fm(k)$ ($Fm(k)$ is the total number of the current shot).

The following steps are used generally in extraction of key frames:

☐ Find the difference between reference frame and general frame.
☐ Look which shot has the maximum difference.
☐ Determine the shot type i.e. static or dynamic shot.
☐ Then, see the position of key frame.

Cite this article as: Deepika Bajaj, Shanu Sharma. "Video Depiction of KeyFrames- A Review." *International Conference on Information Engineering, Management and Security (2015)*: 224-230. Print.

II. Video Shot Boundary Detection (SBD) Algorithms

Todate, many scholars and researchers are doing work to develop more reliable and accurate algorithms that can results into more precise shot boundaries. Earlier technology was more focused on cut detection whereas latest approaches are more concentrated on gradual transitions detection. Broadly categorizing or classifying the shot boundary detection methods are listed below:
(a) Basic approach
(b) Feature based approach
(c) Segmentation based approach
(d) Texture based approach

☐ *Pixel based difference*- This is the basic approach towards shot boundary detection. In this method, intensity of pixels is evaluated by taking two consecutive frames and comparing pixel by pixel. When the intensity of pixels is more than threshold, then it is referred to scene change [8]. A number of algorithms have been implemented for calculation of pixel difference. 3 X 3 averaging filter is used in [9] [10]. The limitations in this method is setting threshold manually. Hence, this method is slow and setting threshold manually is not acceptable concept.

☐ *Statistical based difference*- It is modified and overcome the limitations of pixel based approach. It is sensitive to noise. In this method, each frame is divided into blocks and further each block"s mean and standard deviation is evaluated. Each block some characteristics or features of each pixel in that particular block between consecutive frames are equated [6]. The limitations of this approach are it is sensitive to noise and quite complex as it includes statistical parameters and calculations.

☐ *Transform based difference*- It includes various transformation methods like Discrete Cosine Transformation (DCT) coefficients. Using various transformation approaches, it computes compression difference [6].

☐ *Histogram based difference*- Histograms are the most significant and important technique to find the shot boundaries in videos. In this method, find the difference between frame n and frame n+1 that results to change of color content within shots. Histograms are based on certain concept such as bin to bin, distance formula and intersection. In bin to bin, calculate the difference between color component (R, G, B) of two consecutive frames. If the difference is greater than threshold, then a shot is declared. Distance based approach include Chi- square distance, Manhattan distance, Swain and Ballard distance. Color based histogram is also one of the advanced version of histogram based detection algorithms.

☐ *Edge change ratio*- It is a feature based approach. First of all, edge operator is selected in order to find out the edges. After applying canny edge detector [11], calculate no of pixels in each edge. Edges of successive frames are identified. Then, in order to find new edges are appeared in image or old edges are vanished, edge pixels are combined with neighboring pixels in other image.

☐ *Graph theory based*- It is a latest and feature based approach to evaluate the shot boundary. In this procedure, "color" feature is defined and HSV color model is used to extract color content. Then, difference between frames is computed and apply graph based algorithm on frames of videos which are further divided into numerous different sets. Cut and gradual shots are detected through this algorithm. Cut and gradual changes [12] refers to when two consecutive frames appropriate to different set have different characters on those frames, then it is said to be cut sand gradual shot.

☐ *Information theory based*- It is a texture based approach and wavelet transformation is used to extract the texture and then define the difference based on MI (mutual information) and co-occurrence MI of texture feature. To analyze the image in different scales i.e. information of high frequency defines basically the texture feature and information of low frequency defines the color feature. So, to measure the dissimilarity, the MI and Information entropy is used. The wavelet coefficient is evaluated by discrete wavelet transform along with some advantages of less complexity and orthogonality.
Performance of video shot boundary detection methods can be evaluated by following measures [13]:
(a) Recall- This measure is also known as function of true positive or sensitivity. It is the ratio of detections of correct experimental to the detection of correct and missed.

Recall= correct
correct+missed
(b) Precision- It is the ratio of detections of correct experimental to the detection of correct and false.

Precision=correct
correct+false
(c) F-measure- It is defined as:

F-measure=2*Recall+Precision

Cite this article as: Deepika Bajaj, Shanu Sharma. "Video Depiction of KeyFrames- A Review." *International Conference on Information Engineering, Management and Security (2015)*: 224-230. Print.

Recall+Precision

III. Extraction Of Features

Feature extraction is the most important and essential step in summarization of videos. Depending on various parameters, features are extracted and are classified as low level features and high level features. Low level features are extracted on the basis of texture, color and shape. Every domain have distinct technology for evaluating the graphical and pictorial information and moreover, it has own pros and cons. Sometimes, the information present in video and the understanding of an individual differs due to understanding of graphic information which is basically depending on low level features. This difference of understanding is known as semantic gap. The problem of semantic gap is the biggest complexity that the technical society faces. In order to overcome, feedback of users is collected and updates the required information but then again this information may not be attainable every time and it is somehow a supervised approach. Therefore, now high level features comes into picture in order to deal with the shortcomings of semantic gap.
High level features acts as the bridge and filling the gap between user understanding and information exists. Human plays a vital role in understanding the graphic content of Human determined research. To recognize the face and upper portion data of human is very useful and supportive in understanding gender and age of person. The high level features are listed below:-
(a) Processing of face
(b) Motion magnitude
(c) Duration of shot
(d) Identification of gender

(a) *Processing of face*- Many studies have been done in face processing domain from last many years. Processing of face is very important and basic element in field of security, networking and surveillance videos. The facial expressions and motions of human face help us to recognize the emotions, fitness quality and social interaction information. Processing of face contains mainly three types of steps-

(i) Detection of face
(ii) Clustering of face
(iii) Recognition of face

(i) Detection of face- It is very basic and first step in face processing [14]. A number of algorithms have been introduced for the purpose of detecting faces in videos. The importance and significance of person in video depend on the number of occurrences of particular face in particular video. In [15] Li et al. provides promising results for faces along with different measures in video. Using this method, extraction of face features provides the size and position of all face detected and the total of hits [16]. The technique used for face detection is [17] local successive mean quantization transform (SMQT) and Sparse network of Winnows classifier (SNoW) is experimentally verified to be optimal choice for face detection process. Face detection is a challenging task due to following factors such as postures of faces, intensity, different backgrounds and invariant level of contrast.

The biggest challenge in detecting faces from videos is that characters do not face the camera in full front. So, this problem gives rise to the concept of head turning and head rotation [18]. This concept can easily deal with above mentioned limitation of face direction. Everingham et al. [19] proposed an algorithm in which characters in videos and films are label automatically. Foucher et al. [18] used spectral clustering to detect faces of actors.
(ii) Clustering of face- It is performed by considering the degree of interaction between characters and can be computed by Mutual information (MI). MI is a significant and beneficial process to find out the similarities between characters.
(b) Motion magnitude- The information of magnitude and orientation is contained in vector. Motion magnitude is quite useful for indicating highlights. The fast motion frames as well as slow motion frames both are involved in highlights. In [20], there is a scene in movie in which a actress‟s was crying ,so therefore a scene was slow moving scene and can be considered as emotional scene. In order to identify fast and slow motion frames, we need to perform two steps- first, normalize the
(iii) Recognition of face- The process has two pros which have to be deal in face recognition. The first one is face detection process also involve frontal face along with profile between characters. And, second is the huge number of faces detected for recognition process is a heavy task to be considered. The faces are tagged with face number, shot number and frame number when the faces are obtained after face detection process. The recognition of face is implemented with eigenfaces technique [13] is highly appropriate because our foremost concern is effectiveness and speediness along with easiness. In [16], multiple classes face recognition procedure is implemented which leads to reduce the computing time. To improve the speed of process, interaction score computing is calculated and along with it interaction graph and phenograph between characters in videos are represented.
(b) Motion magnitude- The information of magnitude and orientation is contained in vector. Motion magnitude is quite useful for indicating highlights. The fast motion frames as well as slow motion frames both are involved in highlights. In [20], there is a scene in movie in which a actress‟s was crying ,so therefore a scene was slow moving scene and can be considered as emotional scene. In order to identify fast and slow motion frames, we need to perform two steps- first, normalize the average motion „a‟ of every block within

Cite this article as: Deepika Bajaj, Shanu Sharma. “Video Depiction of KeyFrames- A Review.” *International Conference on Information Engineering, Management and Security (2015)*: 224-230. Print.

the i-th frame. After normalizing the average motion, in second step, imply both fast frames as well as slow motion frames and suppress average motion frames to calculate the highlight score.

(c)Duration of shot- It is most important parameter in summarization of video and also specifies and defines the time limit of shot appeared in video. To attract viewer"s attention, directors use long duration shots [20] and highlights extracted from the videos also involve long duration shots. The duration of shot and motion magnitude and highlights have relationship and association between them. Low motion magnitude and long shot duration refers to highlight love scenes and emotional scenes and high motion magnitude and short shot duration are used to represent an action scene.

(d)Identification of gender- It is very important to identify the gender i.e. male or female in videos and parameters such as eyes, nose, and chin all this helps in identification of gender. The two basic methods that are defined for identification of gender are- (a) related to facial features and (b) observe the relationship between facial features [3]. A lot of researchers and scholars have been developed numerous algorithms and techniques for detecting gender. The principal component analysis (PCA) and neural network have provided very fine results. In order to detect frontal features from face, a selection of genetic features subset was used.

IV. DISCUSSION AND CONCLUSION

Video summarization has many applications in various domains such as scene tagging, retrieving scenes, indexing of videos, highlighting useful and semantic scenes, etc. This paper summarizes all the recent methods and techniques used for video summarization. Many researchers and scholars have done lots of work in current research area and still work is going on high peak. These technologies can be advanced and refined in order to detect shot boundary more accurately. The extraction of keyframes needs to be more accurate and represents the semantic and meaningful portion from videos. It is an essential step so that no important information should be missed. Furthermore, research and work can be done in this area to develop more efficient and fast video shot detection algorithms. This application provides the user with most representative and needful information from video as per to his desired requirement and delivers the new mode of viewing the video. Moreover, summarization of videos saves time, energy and provides ease to an individual.

REFERENCES

[1] *Yihong Gong and Xin Liu,* "Generating Optimal Video Summaries", Multimedia and Expo, 2000. ICME 2000, IEEE International Conference 2000 (Volume:3), pp. 1559 – 1562.

[2] Chen, Oscal T.-C. ,Jhen Jhan Gu, Chih-Chang Chen and Ping-Tsung Lu, "Automatic Highlights Extraction for Drama Video Using Music Emotion and Human Face Features", 4th International Conference on Awareness Science and Technology (iCAST) 2012, pp.104 – 108.

[3] Muhammad Nabeel Asghar et.al, "A Framework for Feature Based Dynamic Intravideo Indexing", Proceedings of the 19th International Conference on Automation & Computing, Brunel University, London, UK, 13-14 September 2013.

[4] Ruxandra Tapu, Bogdan Mocanu and Ermina Tapu ; "Automatic Scene / DVD Chapter Extraction in Hollywoodian Movies" in IEEE, 2013.

[5] D. Bordwell and K. Thompson "Film Art: An Introduction, 5th edition". *New York: McGraw-Hill*, 1997.

[6] Nikita Sao and Ravi Mishra**, "**A survey based on Video Shot Boundary Detection techniques", International Journal of Advanced Research in Computer and Communication Engineering, (Volume: 3, Issue: 4), April 2014.

[7] Ganesh. I. Rathod and Dipali. A. Nikam ," An Algorithm for Shot Boundary Detection and Key Frame Extraction Using Histogram Difference", International Journal of Emerging Technology and Advanced Engineering,(Volume: 3, Issue: 8),August. 2013, pp.155-163.

[8] Mohini Deokar and Ruhi Kabra, " Video Shot Detection Techniques Brief Overview", International Journal of Engineering Research and General Science (Volume:2, Issue:6), October-November, 2014, pp-817-820.

[9] Ravi Mishra ,S.K.Singhai,M. Sharma , "Comparative study of block matching algorithm and dual tree complex wavelet transform for shot detection in videos", Electronic system, signal processing and computing technologies(ICESC), 2014 International Conference, Jan 2014.

[10] Zhe Ming Lu and Yong Shi —Fast Video Shot Boundary Detection Based on SVD and Pattern Matching-‖Image processing IEEE Transactions (Volume:22 , Issue: 12), Dec. 2013.

[11] Rainer Lienhart, "Comparison of Automatic Shot Boundary Detection Algorithms", [Online]. Available: http://www.vis.uky.edu/~cheung/courses/ee639_fall04/readings/spie99.pdf

[12] Ravi Mishra ,S.K.Singhai,M. Sharma, "A Comparative based study of Different Video-Shot Boundary Detection algorithms", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) (Volume :2, Issue: 1), Jan. 2013.

[13] Zuzana Cˇernekova´et.al., "Information Theory-Based Shot Cut/Fade Detection and Video Summarization", Circuits and Systems for Video Technology, IEEE Transactions on (Volume:16, Issue: 1),jan 2006, pp-82-91.

[14] *Songhao Zhu and Yuncai Liu,* "A Novel Scheme for Video Scenes Segmentation and Semantic Representation", IEEE, 2008.

[15] L.Chaisorn, T. S. Chua and C. H. Lee, "The segmentation of news video into story units", IEEE proceeding on International Conference on Multimedia and Expo, pp. 73–76, 2002.

[16] *Yi-Chong Zeng,* "Automatic Extraction of Useful Scenario Information for Dramatic Videos", ICICS, 2013.

Cite this article as: Deepika Bajaj, Shanu Sharma. "Video Depiction of KeyFrames- A Review." *International Conference on Information Engineering, Management and Security (2015)*: 224-230. Print.

[17] Hari R, Roopesh C. P. and Wilscy M., "Human Face Based Approach For Video Summarization", in IEEE Recent Advances in Intelligent Computational Systems (RAICS), 2013, pp. 245-250.
[18] S. Foucher, and L. Gagnon, "Automatic detection and clustering of actor faces based spectral clustering techniques," *Canadian Conf. on Computer and Robot Vision*, pp.113-122, May 2007.
[19] M. Everingham, J. Sivic, and A. Zisserman, "Hello! My name is …Buffy – automatic naming of characters in TB video," *in Proceedings of BMVC*, pp. 889–908, 2006.
[20] L. D. Giannetti, "*Understanding Movies*," 10th edition, Prentice Hall, 2004.

Cite this article as: Deepika Bajaj, Shanu Sharma. "Video Depiction of KeyFrames- A Review." *International Conference on Information Engineering, Management and Security (2015)*: 224-230. Print.