

EXPLORING VIDEO ABSTRACTION

Abstract

Video abstraction is the collection of representative key frames or the shortest version of the original video sequence from the given video sequence. This chapter discusses the basic technologies behind the video abstraction in order to understand this research work. This chapter concentrates on the general approaches of video abstraction and shot boundary detection methodologies. The research work on video abstraction has a long history and also has shown the specific surveys on the video abstraction. Some of the research teams gave domain specific projects for video abstraction which is also discussed in this chapter. The advantages and disadvantages of the existing methods are given.

Keywords: Video, abstraction, detect, research.

Author

Kavitha. J
Associate Professor,
Department of Computer Science and Design,
Vel Tech Rangarajan Dr.Sagunthala R&D
Institute of Science and Technology,
Avadi, Chennai.
vsskavitha@gmail.com

I. INTRODUCTION

This chapter discusses the basic technologies behind the video abstraction in order to understand this research work. This chapter concentrates on the general approaches of video abstraction and shot boundary detection methodologies. The research work on video abstraction has a long history and also has shown the specific surveys on the video abstraction. Some of the research teams gave domain specific projects for video abstraction which is also discussed in this chapter. The advantages and disadvantages of the existing methods are given following the proposed method.

Initially, the research work is concentrated on images retrieval process. That time, the video databases are less and retrieval of the videos is based on keywords annotated manually. Video is an essential one for visual media information. Video has the unique characteristics compared to images such as it has much richer content, enormous amount of raw data and less amount of prior structure. So the indexing and retrieval of dynamic video are relatively difficult when compared to images.

II. MATHEMATICAL EXPLANATION OF VIDEO ABSTRACTION TYPES

Two types of video abstracts are available such as static abstract and dynamic abstract.

- 1. Static Video Abstract:** The collections of representative key frames are selected from the given video sequence.

Hence the key frame set R is defined by Ba Tu Truong and SvethaVenkatesh [1].

$$R = A_{Keyframe}(V) = \{f_{r1}f_{r2} \dots \dots \dots f_{rk}\} \quad (1)$$

Where, $A_{Keyframe}$ denotes the key frame extraction procedure.

f_{ri} - i^{th} key frame of the video sequence

- **Merits**
 - The static summary is shown in a compact and faster way utilizing visual information and not considering audio and textual information.
 - It does not consider any timing and synchronization for displaying generated video abstracts.
 - Users easily understand the contents of the message which are arranged in a spatial order.
 - If users can easily take hard copy of the abstract whenever is needed.
 - It supports nonlinear browsing of the video content.
 - It has less complexity of generation of video abstract when compared to dynamic video abstract
- **Limitations**
 - The audio has much information for some type of videos such as education, movie, sports, etc. That time it fails to show the efficiency of the video.
 - Sometimes the abstracts are difficult to understand when the video is complex.
 - The motion information of the video cannot be described.

- **Size of the Key frames:** The size of the key frames or optimal key frame set is determined by various ways in an automatic key frame extraction process. It categorized into three types such as priori, posteriori and determined. The automatic key frame extraction techniques are mostly select any one of the type for the size of the key frame set.
 - **Priori:** the size of the key frames is fixed beforehand by the constraint of the algorithm.
 - **Posteriori:** Consider the level of visual changes in the video sequences. The high dynamic videos produce high number of key frames.
 - **Determined:** the size of the key frames is determined internally with in the abstraction process.
- 2. **Dynamic Video Abstract:** The shortest version of the original video sequence which contains the important segments selected from the original video.

$$R = A_{Skim}(V) = E_{i_1} \circ E_{i_2} \circ \dots \circ E_{i_k}, \quad E_i \subset V \quad (2)$$

Where, A_{Skim} denotes the video skim generation procedure

\circ it concatenated separate segments to create a video abstract by using video transitions such as cut, dissolve, wipe or fade.

E_i denotes the i^{th} segment of a video skim

- **Merits:**
 - It includes the corresponding audio and textual information of the original video into the abstract.
 - It gives more enjoyment and informative preview to the user over than a collection of key frames.
 - The video sequence's original time-evolving quality is preserved.
- **Limitations**
 - It takes time to display the video skims.
 - When the video highlights are accentuated, the reliability of the content can occasionally be compromised.
- **Size**

The length of the dynamic abstract is also specified by the two ways such as priori and posteriori.

 - **Priori:** The duration of the video sequence is specified as either a time period or a ratio.
 - **Posteriori:** The size of the video abstract is decided internally by the characteristics of the video sequence such as content entropy and interesting events.

III. METHODOLOGIES USED IN STATIC SUMMARY OR KEY FRAME EXTRACTION

1. Sampling Based Approach: The sampling based key frame extraction work is used in most of the earlier works of video summarization. The selection of keyframes is performed in random basis or uniformly sampling basis from the given video sequences at certain time intervals. Michael Mills et. al proposed Video Magnifier and generated the still-frame samples at regular intervals from the video sequence such as a sample for approximately every 8000 frames [2]. It is also applied in VBTool [3] and MiniVideo system [4]. The example of sampling based key frame extraction is given in figure 1.

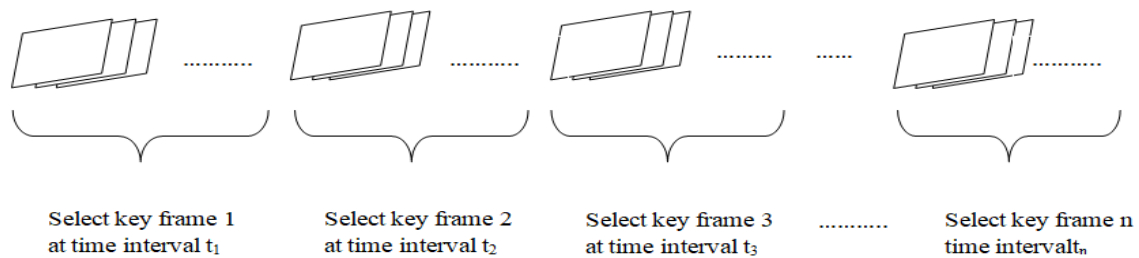


Figure 1: Sampling based key frame extraction

It is simplest and fastest way to select keyframes from the original video sequence. Sometimes, it does not produce the actual keyframes, if important event has no key frames.

2. Shot Based Approach: The shot in the levels of video structure is the most sophisticated level for key frame extraction. Collection Shot transitions like cut, dissolve, wipe, and fade-in and fade-out are employed for integrating the shots. Some of the techniques are chosen key frames as the first frame of each shot [5] and [6]. If the video content has dynamic visual content, then one key frame for each shot does not satisfy the whole content coverage of the original video sequence. So, the number of key frames per shot is determined by the underlying semantic content of the video. The shot boundary detection technique is performed by using the visual features or non-visual features of the video sequence. The visual cues for shot boundary detection can be defined as color-based method, motion-based method, texture-based method, and object-based method. Based on the frame attributes that are taken from the frames in the video sequence, it is determined how much two succeeding frames differ from one another. In general, the shot boundary is detected by using the threshold value. If the difference is more than a certain threshold value then the shot boundary is detected between the two frames. The pipeline of shot based key frame extraction process given in figure 2.

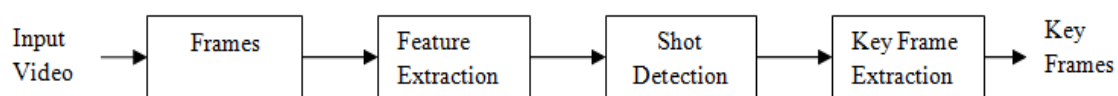


Figure 2: Shot based key frame extraction process

- 3. Segment Based Approach:** Researchers also concentrate on higher level video structure such as scene or event for generating video summary. Some type of video genre, it is helpful to create accurate summary of the video sequence. In that time, shot based key frame extraction is not favor for generating video summary from the video sequence. Chia-Ming Tsai et al [7] generated video summary of movie based on role community networks. This algorithm used analysis stage and summarization stage for generating scene based video summary. In the analysis stage, scene boundary is detected and human face clustering is used to detect and cluster the faces of roles. Then, the relationships between role communities are characterized and role community network is constructed. Based on the role community network and user perception the highlights of the video are detected.
- 4. Perceptual Features Approach:** Perceptual features or visual features of the frames in the video sequences are used in key frame extraction such as color, motion and object.
- **Color Feature Based Approach:** The color is the most effective human perception based feature for video analysis. The color features can easily extract from frames and have less sophisticated computing. Colour models like RGB, HSV, YCbCr, YUV, and others are used to extract color information. The color features are extracted from an entire frame or blocks of partitioned frame. The color based features are extracted by using color histogram, color moments, correlations between colors, a mixture of Gaussian models and etc. The procedure of creating a histogram involves choosing N colour bins to represent the colour space of a frame and counting the number of pixels in each bin. Zhang [8] quantized the color space into 64 super-cells. The normalized count of the number of pixels is allocated to each bin of the histogram. Using the following formula, the distance between two adjacent frames is determined.

$$D_{hist}(f_i, f_{i-1}) = \sum_{r=1}^{N-1} \sum_{g=1}^{N-1} \sum_{b=1}^{N-1} |f_i(r, g, b) - f_{i-1}(r, g, b)| \quad (3)$$

Where, $f_i(r, g, b)$ is the number of pixels of color (r, g, b) in frame f_i of N bins and $f_{i-1}(r, g, b)$ is the number of pixels of color (r, g, b) in frame f_{i-1} of N bins.

The distance between two consecutive frames is compared with the predefined threshold. The current frame is chosen as the key frame if the histogram distance exceeds the threshold value. A threshold value is used to limit the number of key frames. The high threshold value produces less number of key frames and the low threshold value gives more number of key frames. When texture and shape features are crucial in applications when two pictures have completely distinct objects but contain very similar colours, color-based features can be unsuccessful in some situations.

- **Motion Based Approach:** The motion is the important feature in video processing applications for detecting the movement between two consecutive frames. The common techniques for motion estimation are optical flow and block matching which are used by the most of the researchers. A motion metric calculates the magnitudes of the optical flow's components at each pixel, $M(t)$ for frame t. Then, the curve of $M(t)$

is generated. According to the N criterion, the algorithm determines local minima of motion between significant peaks. [9].

$$M(t) = \sum_i \sum_j |o_x(i, j, t)| + |o_y(i, j, t)| \quad (4)$$

Where, $o_x(i, j, t)$ and $o_y(i, j, t)$ are x and y components of optical flow at pixel i,j in frame t.

In the block matching process, the current frame is divided into a matrix of macro blocks. Then, the vector is created by comparing the block of the current frame with corresponding blocks of the current frame and its adjacent neighbor's blocks of the previous frame. The vector quantifies the movement of a macro block from the current frame to the previous frame. The motion is estimated by calculating the movement of a macro block from one location to another in the previous frame. On all four sides of the relevant macro block in the preceding frame, the search area for a successful macro block match is limited to p pixels. The cost functions are used for matching of one macro block with another such as mean absolute difference and mean squared error.

- Object Based Approach:** The first frame in the object-based key frame extraction approach is assigned as the key frame in the first phase. After that, it is calculated how many regions have changed between the key frame and the present frame. The present frame is viewed as a new key frame, meaning that a different event might have taken place, if the difference rises above a predetermined value. The city block distance measure, which gauges the spatial proximity of two feature vectors, is then used to calculate the distance between two seven-dimensional feature vectors. A frame is referred to as a key frame if its value exceeds a given threshold. The integrated scheme for object based video abstraction [10] is given in figure 3.

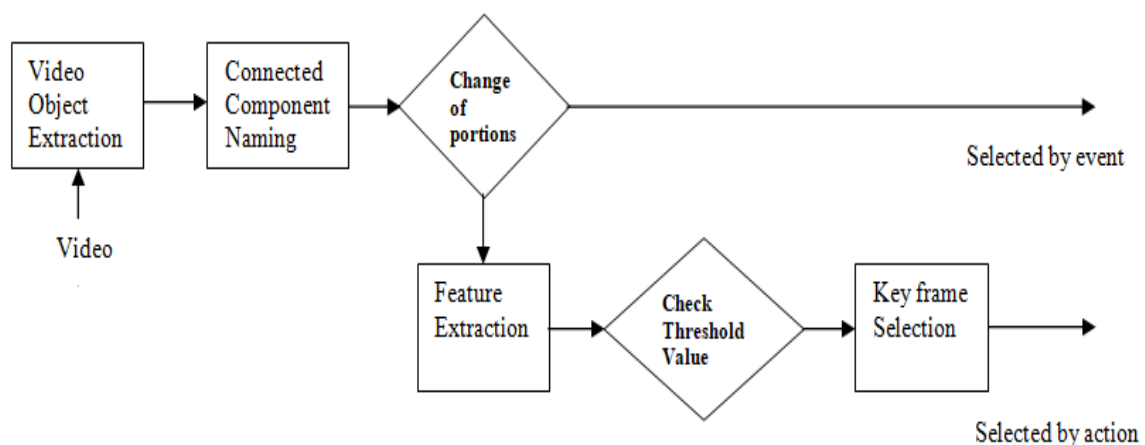


Figure 3: Block Diagram of Object based Video Abstraction

5. Feature Vector Based Approach: In the feature vector space-based key frame extraction method, multiple features are considered for characterizing the frames in the video sequence [11]. Curve splitting algorithm is used to represent the video as a trajectory curve in a high-dimensional feature space. [12]. The feature vectors' curve representation and curve splitting mechanism are displayed in figure 4(a) and figure 4(b) respectively.

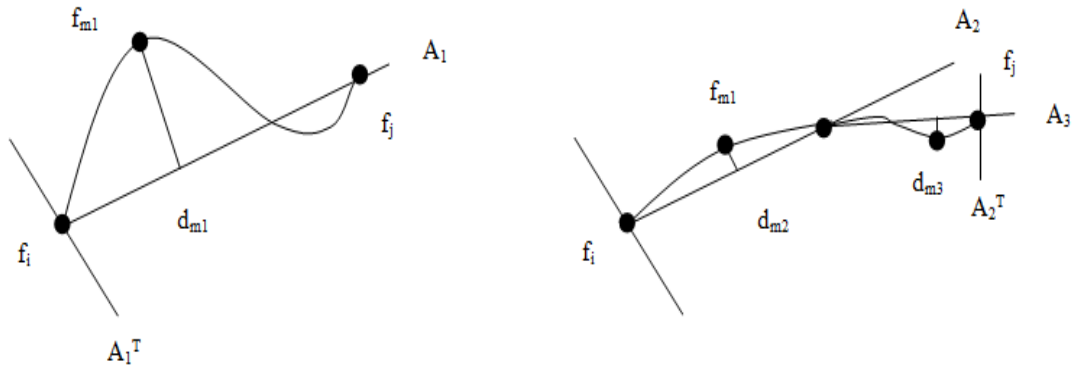


Figure 4(a): Curve Representation **Figure 4(b):** Curve Splitting Process

6. Cluster Based Approach: Based on the similarities between the frames, the video's number of frames (N) are divided into a certain number of clusters (M). The similarity value is measured between the centroid of the first cluster and the next frame f_2 . When the similarity value is greater than the certain threshold value, then the frame is added into first cluster. If not, the frame forms a new cluster.

The creation of new clusters or added to existing clusters is continued to all frames in the video sequence based on the above said process. Ngo et al. [13] used k-means clustering algorithm for extracting key frame from the video file. The general procedure of cluster based key frame extraction method is given in figure 5.

Where, N – Number of frames in the video sequence such as $\{f_1, f_2, f_3, \dots, f_N\}$

C - Number of clusters such as $\{C_1, C_2, C_3, \dots, C_M\}$

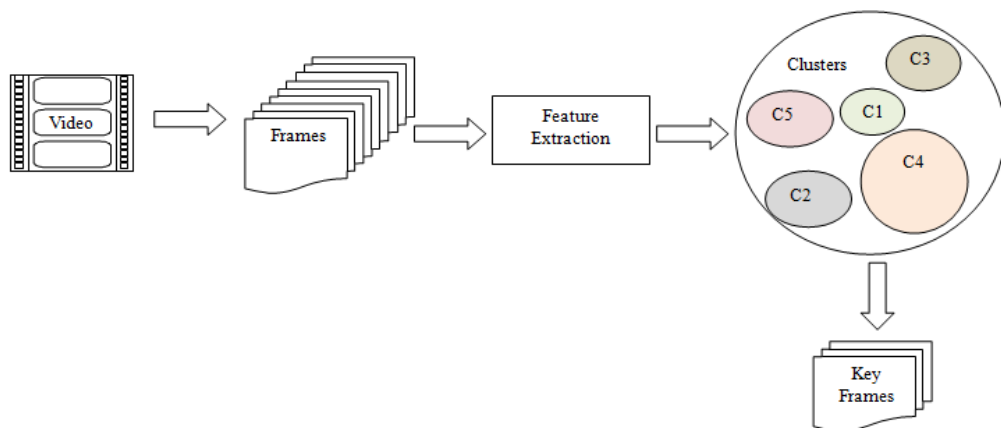


Figure 5: Process of cluster based key frame extraction

IV. METHODOLOGIES USED IN DYNAMIC SUMMARY

The two general categories of dynamic summary are highlight and summary sequence.

- 1. Highlight:** It presents the most exciting and interesting events in the original video, such as a movie trailer. The first dynamic abstract name as VAbstractis implemented with the following five properties of the good cinema trailer[14]. It explains that highlight of a movie should have important people and objects from an original movie, interesting events, dialog contains important information and hide the ending of a film. First, shot boundary is detected using the existing shot boundary detection techniques [15, 16, 17]. High contrast scenes are extracted for identifying important objects and people. The mood of the movie is identified by color perception of the movie. Finally, the five features are combined together in for generating highlights.
- 2. Summary Sequence:** Summary sequence extracts the impression of the content of an entire video. It delivers the highest level of semantic meaning of an original video. Time-compression based, model based and speech recognition based method are used for extracting summary sequences. Time compression based method is used to speed up the watching video. It contains two aspects likely audio compression and video compression. The selective sampling method is appliedfor audio compression. The model based method is applied to the fixed structures of the videos namely sports and news. In this method, the start of the play and end of the play are initially identified [18]. Then, the most exciting scenes are identified by the waveform of the audio. Finally, the summary sequence of the sports video is generated by combining them.

Summary sequence is generated by text and speech recognition based method which has four steps such as video segmentation, extract dominant text information, find corresponding shots and concatenate the corresponding shots [19, 20].This method is applied to various types of videos that haven't fixed structures. First, the video is segmented into shots or scenes based on its visual and audio contents. Using speech recognition software or collecting caption data, significant text information can be obtained from the video. The informative shots are identified which has dominant words or phrases. Finally, the corresponding shots are concatenated for generating summary sequence.

V. SHOT BOUNDARY DETECTION

In this section, the basic shot boundary detection techniques are discussed.The shot boundary detection refers the video is segmented into shots. Initially, the shot boundary detection techniques extract and measure features of the frames. The discontinuity between the frames is identified using the frame feature measures. Based on the discontinuity values, the shot transition between the frames are identified such as hard cuts or softcuts. Then, the shots boundaries are detected from the underlying video sequence using discontinuity values. The basic techniques are frame difference based (pixel based or block based), histogram based, motion based, edge based and etc.

- 1. Pixel Based Approach:** The simplest method for applying threshold to find the discontinuity between frames is pixel-based comparison. When two frames are compared side by side, the differences in the intensity or colour values of the respective pixels are assessed. A threshold value is used to compare the absolute sums of pixel differences. [21].

$$D(i, j) = \frac{\sum_{x=1}^m \sum_{y=1}^n |P_i(x, y) - P_j(x, y)|}{mn} \quad (5)$$

For gray level frames,

$$D(i, i + 1) = \frac{\sum_{x=1}^m \sum_{y=1}^n \sum_c |P_i(x, y, z) - P_j(x, y, z)|}{mn} \quad (6)$$

For color frames

Where, i, j – two successive frames with dimension $m \times n$

$P_i(x, y)$ - intensity value of the pixel at the coordinates (x, y) in frame i

C – index for the color components (in case of RGB color system, $c \in \{R, G, B\}$)

$P_i(x, y, c)$ - color component of the pixel at (x, y) in frame i.

The pair wise segmentation algorithm [22] computes the number of pixel changes by using the threshold t value from the frame to the next frame. The number of pixel changes are represented as a binary function $DP(i, j)$. Then, the segment boundary is identified if more than given percentage of the total number of pixels have changed using threshold T .

$$DP(i, j) = \begin{cases} 1 & \text{if } |P_i(x, y) - P_j(x, y)| > t \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

$$D(i, j) = \frac{\sum_{x=1}^M \sum_{y=1}^N DP(i, j, x, y)}{MN} \times 100 > T \quad (8)$$

- 2. Block Based Approach:** For video temporal segmentation, the block-based or global image feature comparison is used. Frame is divided into a number of blocks that are compared to the subsequent blocks' matching blocks. The difference between the two consecutive frames is measured by

$$D(i, j) = \sum_{b=1}^n c_b DP(i, j, b) \quad (9)$$

Where, i and j are two consecutive frames

b – block (there are n number of blocks in frame i and j)

c_b – predetermined coefficient for the block b

DP(i, j, b) - partial match value between b^{th} blocks in the frame i and j.

The blocks of the two successive frames are compared using likelihood ratio [23]. When the likelihood ratio is greater than the threshold, the shot boundary is detected between the frames.

$$\lambda_b = \frac{\left[\frac{\sigma_{b,i} + \sigma_{b,j}}{2} + \left(\frac{\mu_{b,j} - \mu_{b,i}}{2} \right)^2 \right]}{\sigma_{b,i} \cdot \sigma_{b,j}} \quad (10)$$

$$DP(i, j, b) = \begin{cases} 1 & \text{if } \lambda_b > T \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

- 3. Histogram Based Approach:** The fundamental tenet of shot boundary identification based on the colour histogram is that the colour content does not change quickly within but across shots. It is also easiest algorithm for computation. The shot boundary is detected as single peaks in the time series of the differences between color histograms of consecutive frames or of frames a certain distance k apart [24]. The color histogram differences (CHD_i) between two color frame P_i and P_j

$$CHD_i = \frac{1}{N} \sum_{r=0}^{2^B-1} \sum_{g=0}^{2^B-1} \sum_{b=0}^{2^B-1} |P_i(r, g, b) - P_j(r, g, b)| \quad (12)$$

Where,

$P_i(r, g, b)$ – number of pixels of color (r,g,b) in frame i of N pixels.

$2^B - 1$ – color component is discretized to 2^B different values

Local threshold or global threshold can be used in the color histogram difference value for detecting shot boundary.

- 4. Edge Based Approach:** The edge change ratio and edge contrast feature are the significant method to find the shot boundary between the frames. Edge change ratio is used to find shot boundary based on the edge pixels from the underlying video sequence. The basic idea is that the number of incoming and outgoing edge pixels does not change within in shots. The edges are computed by the canny edge detector and before calculating of the ECR global motion compensation based on the Hausdorff distance. The edge pixels in one frame which have edge pixels nearby in the other frame are not regarded as entering or existing edge pixels. The edge change ratio is calculated as follows

$$ECR_n = \max \left(\frac{X_n^{in}}{\sigma_n}, \frac{X_{n-1}^{out}}{\sigma_{n-1}} \right) \quad (13)$$

Where, σ_n and σ_{n-1} – Number of edge pixels in frame n and n-1

X_n^{in} – Number of entering edge pixels in frame n

X_{n-1}^{out} – Number of exiting edge pixels in frame n-1

The ECR value is from 0 to 1. According to the ECR values, the transition between the frames are recognized as hard cuts, fades, dissolves and wipes [25]. Hard cuts are identified as isolated peaks and fades are recognized in predominated edges in the ECR time series. In dissolve, the two images are displayed concurrently but one image fades out and the other fades in. The ECR time series values of frames and the types of recognized transitions are given in figure 6.

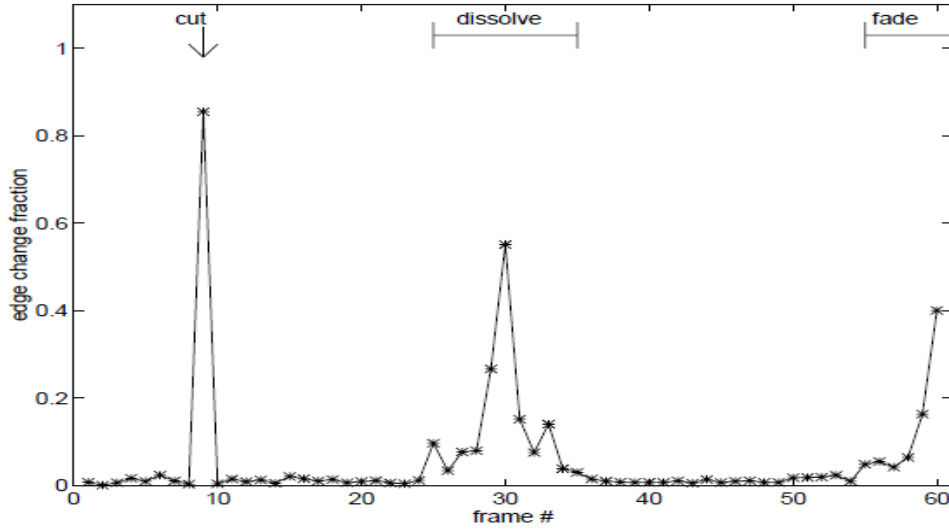


Figure 6: ECR time series values for frames and types of recognized transitions

The idea of the edge based contrast(EC) technique, the loss of contrast and sharpness are occurred in the frames during a dissolve transition that generally reaches its maximum in the middle of the dissolve [26]. So, edge based contrast technique is used to detect dissolve transition. This feature identifies the relation between stronger and weaker edges. The canny edge detector [27] is used to find the edges in the frames. Lower threshold value is used to detect weak edges and higher threshold value is used to detect strong edges in the frames.

$$w(K) = \sum_{x,y} W_K(x, y) \tag{14}$$

$$s(K) = \sum_{x,y} S_K(x, y) \tag{15}$$

With

$$W_K(x, y) = \begin{cases} K(x, y) & \text{if } \theta_w \leq K(x, y) < \theta_s \\ 0 & \text{else} \end{cases}$$

$$S_K(x, y) = \begin{cases} K(x, y) & \text{if } \theta_s \leq K(x, y) \\ 0 & \text{else} \end{cases}$$

$$EC(K) = 1 + \frac{s(K)-w(K)-1}{s(K)+w(K)+1}, EC(K) \in [0, 2] \tag{16}$$

It contains the following features,

- If EC is equal to 0, frame has not strong edges.
 - If EC values from 0 to 1, the frame has large number of weak edges then the strong edges.
 - If EC value is 1, the number of weak edges and number of strong edges are roughly equivalent.
 - If EC values between 1 and 2, the frame has large number of strong edges then the weak edges.
- If EC value is 2, the frame contains only strong edges.

- 5. Cluster Based Approach:** Shot boundary is detected using unsupervised clusterprocess. The clustering process treats a video segmentation as a two class problem such as ‘scene change’ or ‘no scene change’. The frames of interest are determined by applying k-means clustering algorithm on color histogram similarity measures between successive frames [28].

The clustering process applied on YUV color space of frames which has luminance and chrominance information for applying scene change detection. The color content dissimilarity of each frame pair is computed using frame comparison metric such as histogram difference and χ^2 test. The metric values are used to classify frames to two classes using k-means algorithm. The initial cluster means are selected in k-means algorithm using Euclidean distance value which is used as performance index. Each sample is assigned to the class of the nearest mean. The two cluster means are updated by sample mean of all samples in the cluster. The largest mean of the frame in the cluster is labeled as scene change. Again, the frames between two successive scene changes are labeled as scene changes. Then, the adjacent scene change frames are classified as edit effect frames such as ‘fade’, ‘dissolve’ and ‘wipe’. Finally, the shot boundaries are marked.

VI. ADVANTAGES AND DISADVANTAGES OF THE EXISTING TECHNIQUES

The pros and cons of the various techniques in the key frame extraction and shot boundary detection are reviewed and presented here.

- 1. Key Frame Extraction Approach:** Sampling based technique is the very simplest way to extract key frames from the video sequence. Sometimes, the significant parts have no representative frames and several frames have similar contents. So, it fails to identify the actual key frames.

If long videos, segment based method produces key frames in an efficient way which could be a scene, an event or entire sequence. But other works produce one or more key frames for each shot which does not scale up long videos. Since scrolling through hundreds of images are time-consuming, tedious and ineffective. It is a complicated work compared to the others.

In perceptual features based key frame extraction, the color, motion and object features are considered in general. The color based feature is an important feature and

widely used key frame extraction techniques. Also it is insensitive to camera and object movements. It is heavily dependent on threshold and cannot well capture the underlying dynamics when there is lots of camera and object motion. Motion based techniques are controlling the number of frames based on temporal dynamics in the scene.

The feature vector based method is an efficient method for extracting key frames. Clustering based technique is one of the popular methods in key frame extraction techniques and also it is widely used in data analysis. This method will extract a set of visually dissimilar key frames while their temporal order information is totally lost. The successful extraction of semantic meaningful clusters is very difficult. Because, it has to maintain large intra-class and low inter-class visual variance between clusters.

Shot based key frame extraction technique is the most sophisticated way to represent the key frames by adapting to the dynamic video content. But the segmentation of shot is a challenging work in this area.

- 2. Shot Boundary Detection Approach:** Pixel based shot boundary detection is very simple method. But it is very sensitive to camera and object motion or movements. The robustness to camera and object motion is increased by block-based approaches, which are also computationally challenging. It achieves better than pixel based method but is still sensitive to camera and object motion.

The histogram-based approach to recognizing hard cuts is straightforward and efficient, and it is more resistant to object and camera motions. It disregards the spatial data contained in the frames. Therefore, it is ineffective if the histograms of two distinct images are identical.

In the idea of edge based method is temporal visual discontinuity usually comes along with structural discontinuity. It performs well in fade, dissolve and wipe detection. It is robust against object motion and it does not outperform the above simple color histogram method in the hard cut detection. But it is computationally much more expensive. False positives could be caused by the quick changes in the frames, such as the overall shot brightness and very dark or very light frames.

Motion based methods are computationally expensive and performed well in hard cut transition detection. It cannot differentiate illumination changes and motion. The performance can be enhanced by using numerous characteristics at once when segmenting data using a cluster-based approach, which avoids the need for thresholding employed in earlier techniques. But it is not able to recognize the type of the gradual transitions.

REFERENCES

- [1] J.-C. Ren et al. Determination of shot boundary in MPEG videos for TRECVID 2007. In TREC Video Retrieval Evaluation Online Proceedings, 2007.
- [2] M. Mills, "A magnifier tool for video data", Proc. of ACM Human Computer Interface, pp. 93-98, May 1992.
- [3] K. Otsuji, Y. Tonomura and Y. Ohba, "Video browsing using brightness data", Proc. Of SPIE, vol. 1606, pp. 980-985, 1991.

- [4] Y. Taniguchi, "An intuitive and efficient access interface to real-time incoming video based on automatic indexing", Proc. of ACM Multimedia, pp. 25-33, Nov. 1995.
- [5] H. Ueda, T. Miyatake, S. Sumino and A. Nagasaka, "Automatic structure visualization for video editing", Proc. of INTERCHI'93, pp. 137-141, 1993.
- [6] S. W. Smoliar and H. J. Zhang, "Content-based video indexing and retrieval", IEEE Multimedia, pp. 62-72, 1994.
- [7] C. Tsai, L. Kang, C. Lin, and W. Lin, "Scene-based movie summarization via role-community networks," IEEE Transactions on Circuits and Systems for Video Technology, vol. 23, no. 11, pp. 1927–1940, 2013.
- [8] H. J. Zhang, J. Wu, D. Zhong and S. W. Smoliar, "An integrated system for content based video retrieval and browsing", pattern Recognition, vol. 30, no. 4, pp. 643-658, 1997.
- [9] W. Wolf, "Key frame selection by motion analysis", ICASSP'96, vol. 2, pp. 1228-1231, 1996.
- [10] Kim, C., & Hwang, J. (2001). An integrated scheme for object-based video abstraction. Proceedings of ACM Multimedia 2001, Los Angeles, CA, 303-309.
- [11] DeMenthon, D., Kobla, V., & Doermann, D. (1998). Video summarization by curve simplification. Proceedings of ACM Multimedia 1998, 211-218.
- [12] Ramer, U. (1972). An iterative procedure for the polygonal approximation of plane curves. Computer Graphics and Image Processing, 1, 244-256.
- [13] Ngo, C.W., Pong, T.C., & Zhang, H.J. (2001, Oct.). On clustering and retrieval of video shots. Proceedings of ACM Multimedia 2001, Ottawa, Canada, 51-60.
- [14] Pfeiffer, S., Lienhart, R., Fischer, S., & Effelsberg, W. (1996). Abstracting digital movies automatically. Journal of Visual Communication and Image Representation, 7(4), 345-353.
- [15] Kang, H. (2001). A hierarchical approach to scene segmentation. IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL 2001), 65-71.
- [16] Sundaram, H., & Chang, S. (2000). Video scene segmentation using video and audio Features. ICME2000, 1145-1148.
- [17] Wang, J., & Chua, T. (2002). A framework for video scene boundary detection. Proceedings of the 10th ACM international conference on Multimedia, Juan-les-Pins, France, 243-246.
- [18] Li, B., & Sezan, I. (2002). Event detection and summarization in American football broadcast video. Proceedings of SPIE, Storage and Retrieval for Media Databases, 202-213.
- [19] Agnihotri, L. (2001). Summarization of video programs based on closed captions. Proceedings of SPIE, Vol.4315, San Jose, CA, 599-607.
- [20] Alexander, G. (1997). Informedia: News-on-demand multimedia information acquisition and retrieval. In M. Maybury (Ed.), Intelligent Multimedia Information Retrieval, pp. 213-239. Menlo Park, CA: AAAI Press.
- [21] T. Kikukawa, S. Kawafuchi, Development of an automatic summary editing system for the audio-visual resources, Trans. Electron. Inform. J75-A (1992) 204}212.
- [22] H.J. Zhang, A. Kankanhalli, S.W. Smoliar, Automatic partitioning of full-motion video, Multimedia Systems 1 (1) (1993) 10}28.
- [23] R. Kasturi, R. Jain, Computer Vision: Principles, IEEE Computer Society Press, Washington DC, 1991, pp. 469}480.
- [24] B.-L. Yeo and B. Liu. Rapid Scene Analysis on Compressed Video. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 5, No. 6, December 1995.
- [25] R. Lienhart, "Comparison of automatic shot boundary detection algorithms," in Proc. IS&T/SPIE Storage and Retrieval for Image and Video Databases VII, vol. 3656, Jan. 1999, pp. 290–301.
- [26] J. Canny. A Computational Approach to Edge Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 8, No. 6, pp. 34-43, Nov. 1986.
- [27] T.N. Pappas, An adaptive clustering algorithm for image segmentation, IEEE Trans. Signal Process. 40 (1992).
- [28] Koprinska and S. Carrato, "Temporal Video Segmentation: A Survey," Signal Processing: Image Communication, vol. 16, no. 5, pp. 477–500, 2001.