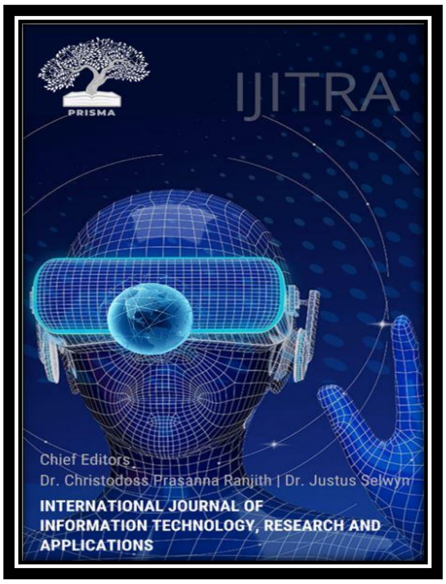

International Journal of Information Technology, Research and Applications (IJITRA)

Kumar K, Vimal Kumar Stephen (2022). A Novel Shot Segmentation Method to Synthesize Video Dataset. International Journal of Information Technology, Research and Applications, 1(3), 23-28.

ISSN: 2583 5343

DOI: 10.5281/zenodo.7386003

The online version of this article can be found at:
https://www.ijitra.com/index.php/ijitra/issue/archive

Published by:
PRISMA Publications

International Journal of Information Technology, Research and Applications (IJITRA) is a journal that publishes articles which contribute new theoretical results in all the areas of Computer Science, Communication Network, and Information Technology. Research paper and articles on Big Data, Machine Learning, IOT, Blockchain, Network Security, Optical Integrated Circuits, and Artificial Intelligence are in prime position.

PRISMA
PUBLICATIONS
https://www.prismapublications.com/

A Novel Shot Segmentation Method to Synthesize Video Dataset

Kumar K[1], Vimal Kumar Stephen[2]
[1,2] Lecturer, Department of Information Technology, University of Technology and Applied Sciences, IBRA, The Sultanate of Oman

Article Info

Article history:

From the proceedings of *2nd Symposium on Information Technology and Applied Mathematics* – SITAM 2019, 6th November 2019.

Host: University of Technology and Applied Sciences, Shinas, Sultanate of Oman.

Keywords:
Shot
Frame
Threshold

ABSTRACT

Many methodologies have been developed and variety of applications have been investigated including video data indexing or retrieval extracting representative features from raw video data before the mining process and integrating features obtained from multiple sources. There is a requirement for improved methods that will be used to extract knowledge while mining video datasets. We propose a suitable Shot Segmentation method to synthesize video dataset.

Corresponding Author:

Kumar K
Lecturer,
Ibra college of Technology,
Sultanate of Oman

1 CONCEPTS BEHIND MINING VIDEOS

Video mining can be defined as the unsupervised discovery of patterns in audio visual contents [1]. Video data motion is more complicated when the choice of mining images are compared. Further challenges are how to generate relationships between objects or segments within the video components, how to extract patterns in sound, how to categorize speech and music or how to track objects in video datasets. Present directions in mining video data include extracting data or metadata from the video databases, storing them in structured databases and apply tools to the structured video databases, integrating data-mining techniques with the information retrieval and to operate directly on the unstructured video databases.

We can state that video mining is a process to discover patterns of video structure, object activities, video events from vast amounts of video data with an assumption of their contents. The temporal (motion) and spatial (color, texture, shapes and text regions) features of the video can be used for mining [2]. The framework for real time video data mining for the raw videos is shown in Figure 1.

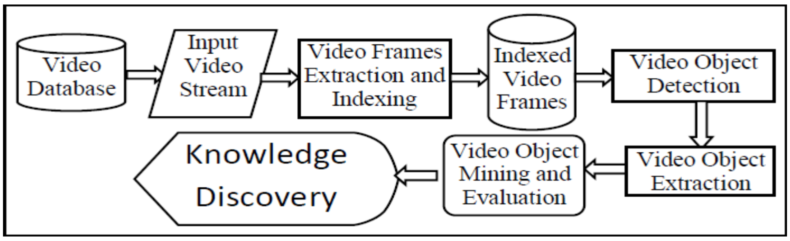

Figure 1 Stages in video mining

In the first stage, grouping of input frames takes place. In the second stage it extracts some of the features from each segment. In the third stage, the decomposed segments are clustered into similar groups. Following stages are actual mining of the raw video sequences and the video data compression for the storage of these raw videos [3]. The knowledge and patterns can discover and detect the object identification, modeling and detection of normal and abnormal events, video summarization, classification and retrieval. The fast proliferation of video data archives has increased the need for automatic video content analysis. Such automatic analysis greatly eases the authoring of content structure and increases data accessibility, which is critical for video database management and multimedia applications including broadcast video, video-on-demand, web search, etc. As an essential step to facilitate automatic video content analysis, the video clip is segmented into certain analysis units and their representative features are extracted. In order to effectively characterize the video documents, there are quite a number of research efforts devoted in this step to extract feature descriptors at low-level, mid-level and object-level, where a shot is generally adopted as the analysis unit. The second step is called the decision-making process that extracts the semantic index from the feature descriptors [4] [5]. Decision-making can be roughly grouped into knowledge-based approaches and statistical approaches. Knowledge-based approaches [15][16] typically combine the output of different media descriptors into rule-based classifiers, whereas the statistical approaches include the use of C4.5 decision trees, support vector machines, dynamic Bayesian Network[12], etc.

1.1 Video data mining versus pattern recognition

Both areas share the feature extraction steps but differ in pattern specificity. The aim of pattern recognition is to recognize or classify patterns, pattern generation and analysis. Pattern recognition [6][7] is indulging in a research on classifying special samples with an existing model while video mining is involved in discovering rules and patterns of samples with or without image processing. The objective of video data mining is to generate all significant patterns without prior knowledge of what they are.

1.2 Video information retrieval versus video data mining

Video mining involve three main tasks which are:

(1) Video preprocessing with high quality video objects such as blocks of pixels, key frames, segments, scenes, moving objects and description text.

(2) The extracting of the features and semantic information of video objects such as physical features, motion features, relation features and semantic descriptions of these features.

(3) Video patterns and knowledge discovery using video, audio and text features.

2 SHOT SEGMENTATION

Using histogram techniques we can extract the key frames [8] [9]. Video is a highly structured media and segmentation of this media into useful temporal units is normally the beginning process to retrieve the channel of data associated in it. We look into the possible techniques for extraction of key frames from the video stream. We focus on streamlining all the requirements of a user who are interested in viewing those scenes which he desires (frequent items) and also the scenes of a particular emotion. The technique considered herein is Shot Detection (SD).

2.1 Shot Detection (SD)

This technique is performed based on methods such as pixel per Inch or better known as "PPI-pixel intensity", histogram oriented approach, vector based edge and motion analysis. After comparing them, we understood that the most popular one is the "Histogram Difference". Method based on histogram for shot detection is constituted of the global percentages of various colours that a video shot can possibly contain. This technique based on histogram neglects space distribution and pixel. Histogram technique is based on separating each frame into smaller blocks and taking a "Histogram difference" [10] of frames in succession, and then specific weights are assigned to the blocks in consideration. Threshold [4] is formulated by the calculation of mean and standard deviation. If the difference between the frames is above than the threshold of a frame then that frame is the key frame.

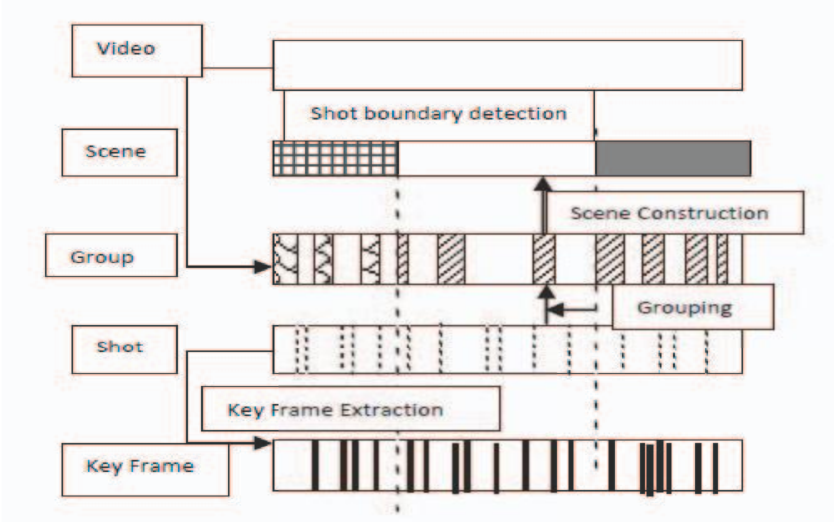

Figure 2 Stages in Key Frame Detection

Algorithm:

Step 1: A video sent as input is read, and then processed by the ShotDetection method.

Step 2: Using this method, the method divideSubFrames is called to convert frames into sub frames. These sub frames are a part of respective blocks.

Step 3 : The block difference for each sub frame is calculated by the divideSubFrames method.

Step 4 : Add up the block differences by the given Formula,

a. Block difference=BD
b. Histogram of 1st block= H1
c. Histogram of 2nd block=H2
d. No of Gray Levels=GL

$$BD= (H1-H2)*GL.$$

Step 5 : Once the block difference is obtained, calculation for the mean deviation (MD) and standard deviation (SD) is done. Threshold is calculated in following manner:

$$T = MD + (a* SD)$$

Step 6 : When block difference of a frame is above than the calculated threshold then that frame becomes the key frame.

Step 7 : Step1 to Step 6 are repeated till the complete video has been processed and every key frame has been determined and stored.

2.2 Key Frame Extraction Algorithm:

The algorithm discussed below focuses only on techniques that are considered to be different and from variable viewpoints, the fundamental dynamics of the video sequence. There exists an algorithm for key frame extraction which is defined as follows:

Step 1: Computation of dissimilarity between all reference and general frames is done using the above algorithm. (Max (i))

Step 2: Search the greatest difference within a Shot.

Step 3: Determine the type of shot depending on the relationship which exists between max (i) and mean deviation (MD): Static or Dynamic Shot.

Step 4: Determine the position of the key frame as follows:

If ShotType = 0, choose the frame in the middle as key frame. If there is even number of shot frames, then choose any one of the two middle frames as the key frame.

If ShotType=1, choose the frame with the greatest difference as key frame.

Figure 3 Sample Video Clip

Figure 4 Extracted Key Frames

2.3. Comparison among Popular Threshold methods

(a) Manual method

This looks into peaks of histogram of frequencies and then the valleys between them. Analysis is done independently for each image and the threshold is selected based on the entire image. It was carried out by Crowley, Grau and Arendt (2000).

(b) Isodata algorithm

This technique was developed by Ridler and Calvard (1978). Initially, a guess is made at a possible value for the threshold. The threshold is repositioned to lie exactly half way between the two means. Mean values are calculated again and a new threshold is obtained, and so on until the threshold stops changing a value. The mean pixel value (μ) is calculated by,

$$\mu = \sum_{i=0}^{t} i - h(i) / \sum_{i=0}^{t} h(i) \qquad \text{-------- (1)}$$

(c) Otsu's algorithm

This method (Otsu, 1979) is one of the most popular techniques of optimal thresholding (Nixon & Aguado, 2002). It is worked on the discriminant analysis and maximize the 'between-class variance' of the GL histogram to give the best separation of classes. Threshold is,

$$\sigma_B^2(t^*) = \max_{t \in GL} \sigma_B^2(t) \qquad \text{-------- (2)}$$

(d) Minimum error algorithm

It is developed by Kittler and Illingworth (1986), the GL histogram is viewed as an estimate of the probability density function of the mixture population comprising of the grey levels of the objects and background (j = 1,2). Each of the components is normally distributed with mean m(j) and standard deviation r(j) and a probability P(j). A criterion function J(t) is used as follows,

$$J(t^*) = \min_{t \in GL} J(t) \qquad \text{-------- (3)}$$

(e) K-means clustering algorithm: This approach (adapted from Hartigan,1975) considers the values in the two regions of the histograms (background and foreground pixels) as two clusters. The objective is to pick a threshold such that each pixel on each side of the threshold is closer in intensity to the mean of all pixels on that side of the threshold than the mean of all pixels on the other side of the threshold. The t value that holds the lowest partition error (PE) will be the threshold.

$$\mathrm{PE}(t) = \sum_{i=0}^{L-1} [h(i)(i - \mu(t))] \quad \text{-------- (4)}$$

(f) Entropic method (Pun method):

This algorithm use the entropy of GL histogram by applying information theory. Threshold is calculated by maximising the upper bound of the a posteriori entropy as,

$$P(t) = \sum_{i=0}^{t} p(i) \quad \text{-------- (5)}$$

(g) Moment-preserving algorithm:

In this method (Tsai, 1985), threshold is computed in such a way that the moments of an image to be thresholded are preserved in the binary image. The jth moment m(j) is calculated as,

$$m(j) = \frac{1}{N} \sum_{i=0}^{L-1} i^j h(i) \qquad j = 1, 2, 3. \quad \text{-------- (6)}$$

3 RESULT (based on suitable inputs for Human Interaction video dataset)

Table 1. Threshold Method Values

Manual Method	IsoData algorithm	Otsu's algorithm	Minimum error algorithm	K-means clustering algorithm	Entropic method	Moment-preserving algorithm	Shot Detection method
2.236	1.893	1.761	2.323	2.975	1.592	1.453	1.212

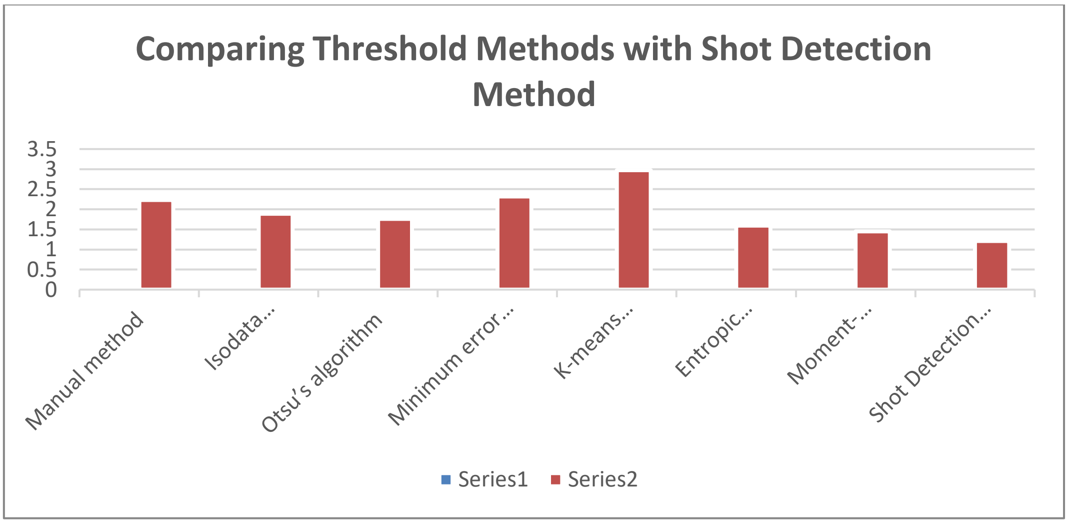

Figure 5 Comparing Threshold Methods with Shot Detection Method

4 CONCLUSION

The system is defined to generate temporal association rule for frequent and infrequent items. The real time application of this work is fetching highlights from a video or bookmarking certain portions of videos by the user to watch later. Video mining is an under research field and the voluminous nature of videos are much growing. In this situation we have taken steps to propose few effective methods to assist video mining tasks. The proposed methods and algorithms will perform with optimal memory usage and better execution time.

REFERENCES

[1] Shu C, Mei, Cheng, Jeff(2002), Information from Traffic Video Sequences, Journal of Intelligent Information Systems¸ 1, 61–77.
[2] Ding J and Yau(2009), An innovative data structure for mining association rules among infrequent items, Computers and Mathematics with Applications,2,290-301.
[3] Lin and Mei-Ling Shyu(2010), Weighted association rule mining for video semantic detection, International Journal of Multimedia Data Engineering and Management,1,37-54.
[4] Lin C and Hong T (2014), Maintenance of pre-large trees for data mining with updated records, Information Sciences, 3, 1-16.
[5] Mannila, H., Toivonen and A. Verkamo (1994), Efficient algorithm for discovering association rules, Knowledge Discovery and Data Mining, 3, Issue 4,181-192.
[6] Naqvi M, Hussain K, Asghar S and Fong S (2011), Mining temporal association rules with incremental standing for segment progressive filter, Communications in Computer and Information Science, 373-382.
[7] Vijayakumar,V and Nedunchezhian, R (2012), A novel method for mining video association rules using weighted temporal tree, ARPN Journal of Engineering and Applied Sciences, 7, 1335-1341.
[8] Wu X., C. Zhang and S. Zhang (2004), Efficient mining of both positive and negative association rules, ACM Transactions on Information Systems, 4, 381-406.
[9] Zhu X, Elmagarmid A, and Feng Z(2005), Video data mining: semantic indexing and event detection from the association perspective, IEEE Transactions on Knowledge Data Engineering, 5, 665-677.
[10] D.Saravanan, Dr.S.Srinivasan (2013), Matrix Based Indexing Technique for video data, Journal of computer science, 9(5), 2013, 534-542.
[11] D.saravanan, Dr.S.Srinivasan (2012), Video image retrieval using data mining Techniques, Journal of computer applications (JCA), Vol V,Issue 01, 2012. 39-42.
[12] D.Saravanan V.Soma sundaram(2014), "Matrix Based Sequential Indexing Technique For Video Data Mining, Journal of Theoretical and Applied Information Technology,Vol. 67, 3.
[13] D.Saravanan, Dr.S.Srinivasan(2013), "Video information retrieval using Chemelon Clustering" International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume-02,Issue 01, January – February 2013, Pages 166-170.
[14] D.Saravanan, A.Ramesh Kumar(2013),"ContentBased Image Retrieval using Color Histogram", International journal of computer science and information technology (IJCSIT), Volume 4(2), Pages 242-245, ISSN: 0975-9646.
[15] Sotiris Kotsiantis and Dimitris Kanellopoulos(2006), "Association Rules Mining: A Recent Overview", GESTS International Transactions on Computer Science and Engineering, Vol.32, No: 1, pp. 71-82.
[16] Anurag Choubey, Ravindra Patel,J. L.Rana(2011), "A Survey Of Efficient Algorithms And New Approach For Fast Discovery Of Frequent Item Set For Association Rule Mining", International Journal of Soft Computing and Engineering.