Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 11, Number 6 (2015), pp. 4457-4467

http://www.ripublication.com

Various Assorted Cluster Performance Examination Using Video Data Mining Technique

D. Saravanan

Faculty of Operations & IT, IFHE University IBS Hyderabad 501203, Telangana.
Sa_roin@yahoo. com

Abstract

The amount of information produced every year is rapidly growing due to many factor among all media, video is a particular media embedding visual, motion, audio and textual information. Given this huge amount of information the researchers need effective clustering techniques for the video files. The resemblance is obtained from the values of the objects. The basic attributes are distance, pixel value, and other common factor if any.. Clustering can be termed here as a grouping of similar images. The purpose of clustering is to get similar result, fast retrieval, reduce the storage in various areas. The clustering methods are divided into: hierarchical, partitioning, density-based, model-based and grid-based methods. The goal of this survey is to provide a comprehensive review of different clustering techniques. Due to the importance of clustering a number of algorithms have been proposed but based on the image that is inputted the algorithm should be chosen to get the best results.

Key terms: Data mining, Clusters, Knowledge Extraction, Image Processing, Multimedia Data, Data Extraction.

1. Introduction

The availability of digital video contents over the web is growing at an exceptional speed due to the advances in networking and multimedia technologies and to the wide use of multimedia applications: videos can be downloaded and played out from almost everywhere using many different devices (e. g., cell phones, palms, laptops) and networking technologies. The large popularity is highlighted by the enormous success of web sites like Google-Video, YouTube and iTunes Video, where people can upload/download videos. In such a scenario, a tool for performing video browsing would be really appreciated. To handle the enormous quantity of video contents, many proposals have been presented for indexing, retrieving and categorizing digital

video contents. Considering the limited man-power, it is much expected to develop retrieval methods which use features automatically extracted from videos. However, since features only represent physical contents (e. g. color, edge, motion, etc.), retrieval methods require knowledge of how to use/integrate features for retrieving relevant videos to a query[1, 4, 5] To obtain such knowledge, this dissertation concentrates on *video data mining* where videos are analyzed using data mining techniques which extract previously unknown, interesting patterns in underlying data. Thereby, patterns for retrieving relevant videos are extracted as explicit knowledge[2].

2. Clustering Algorithms

The related terms data dredging, data fishing, and data snooping refer to the use of data mining methods to sample parts of a larger population data set that are (or may be) too small for reliable statistical inferences to be made about the validity of any patterns discovered. These methods can, however, be used in creating new hypotheses to test against the larger data populations.

The availability of digital video contents over the web is growing at an exceptional speed due to the advances in networking and multimedia technologies and to the wide use of multimedia applications[9, 10, 11]: videos can be downloaded and played out from almost everywhere using many different devices (e. g., cell phones, palms, laptops) and networking technologies. The large popularity is highlighted by the enormous success of web sites like Google-Video, YouTube and iTunes Video, where people can upload/download videos. In such a scenario, a tool for performing video browsing would be really appreciated. To handle the enormous quantity of video contents, many proposals have been presented for indexing, retrieving and categorizing digital video contents. Considering the limited man-power, it is much expected to develop retrieval methods which use features automatically extracted from videos. However, since features only represent physical contents (e. g. color, edge, motion, etc.), retrieval methods require knowledge of how to use/integrate features for retrieving relevant videos to a query[6]. To obtain such knowledge, this dissertation concentrates on *video data mining* where videos are analyzed using data mining techniques which extract previously unknown, interesting patterns in underlying data. Thereby, patterns for retrieving relevant videos are extracted as explicit knowledge.

Classification is a way to categorize or sassing class labels to a pattern set under the supervision. Decision boundaries are generated to discriminate between patterns belonging to different classes. The data set is initially partitioned into segments and the classifier is trained on the former. A framework to enable semantic video classification and indexing in a specific video domain was proposed. A method for classification of different kinds of videos that uses the output of a concise video summarization technique that forms a list of key frames was present.

3. Literature Survey:

As a result, our works motivated by strong application that needs and image/video indexing and retrieval, those are also active research topic in the multimedia research

community. Several projects, such as Columbia CuVid Search System [15, 2], Informedia Project [15, 16], IBM Marvel [15], and other activities based on TRECVID video benchmarks [15, 19, 17], have shown promising results in image and video indexing and retrieval. So, video database access is a very active research area. Many aspects of video databases must be investigated. The Color, Texture and Motion are very important attributes of video that are not fully used in indexing and retrieving.

4. Existing System

In existing system, a new approach is carried out for deep concept-based multimedia information retrieval, which focuses on high-level human knowledge, perception, incorporating subtle nuances and emotional impression on the multimedia resources.

- Existing system utilized color moments to classify video frames as narrative or text based (slide, web or whiteboard) using Decision Trees.

4. 1 Issues in Existing System

There is no separate tool for analyzing various clustering techniques.

- High complexity
- High time consumption for data retrieval

5. Proposed System

RGB feature used for eliminating redundant frames.

- Grid Based, Model Based, Density Based, Hierarchical clustering mechanisms are used for clustering the frames.

5. 1 Advantage of Proposed System

Our proposed framework can efficiently analyze the performance of various clustering mechanisms.

- Reduced Complexity
- Increased efficiency

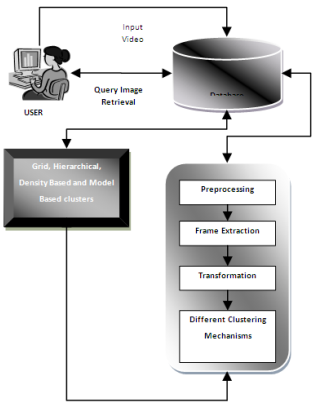

Fig 1. Proposed Architecture

5. Experimental setup

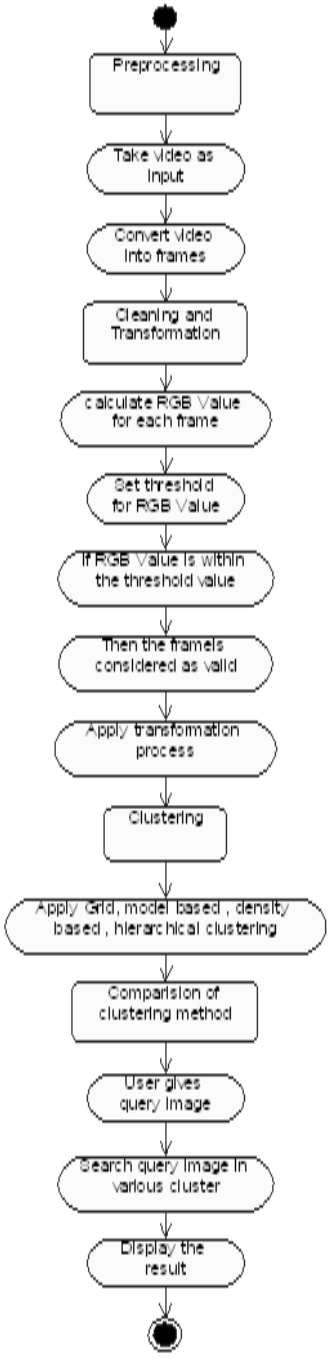

Fig 2. Flow diagram of Overall process of proposed system

6. 1 Video Preprocessing

Preprocessing is the preliminary process of our proposed framework. Initially, user gives input video for preprocessing. In preprocessing process, the video was converted into number of frames. The converted frames are stored into the database.

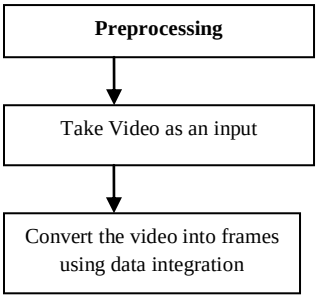

Fig 3. Flow of video preprocessing

Proposed new algorithm for training the video set:
Step 1: Segment the given video file.
Step2: Select the key frame among the segmented set.
Step3: Based on the avg frame value key frame is selected. (The average feature value of all the frames within the segment)
Step4: Store in the Data Base.
Step5 : Repeat step 2 and Step 3 for all type of video file.

6. 2 Cleaning and Transformation

This module describes the details of cleaning and transformation process. After preprocessing, calculate RGB value for each frame in the database. After calculate the RGB value set the threshold value for frame cleaning. If the RGB value is within the range then that frame is consider as valid. If the distance value is not within the range then that frame is consider as not valid. So eliminate these fames. Apply transformation process for re arranging the valid frames.

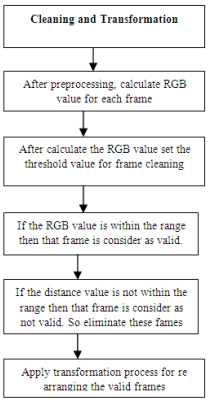

Fig 4. Flow of Cleaning and Transformation

Code for eliminating duplication:
Σ I=0 to image height

Σ j=0 to image width
Clrpixel= pixel (i, j)
Grey value=lngGrayScaleValue = (0. 299 * clrPixel. R) + (0. 587 * clrPixel. G) + (0. 1114 * clrPixel. B)
Grey = Σ Grey value
'Grey' gave the value of the grey value of the whole image.
Grey value= Image1. Greyvalue-Image2. Grey Value
If Grey value < threshold then
Duplicate image

6. 3 Clustering

Clustering is the process of grouping the data into classes or clusters so that objects within a cluster have high similarity but are very dissimilar to objects in other clusters. A good clustering algorithm should be able to identity clusters irrespective of their shapes[7, 13, 14]. In this module, model based clustering, grid based clustering, density based clustering, and hierarchical clustering mechanisms are utilized to cluster the segmented frames. Using various types of clustering mechanisms, the extracted key frames are clustered.

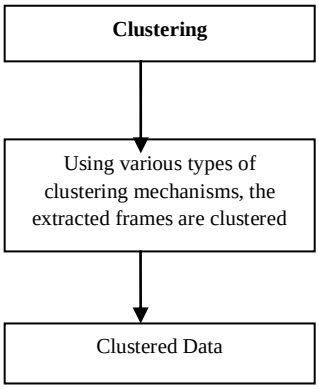

Fig 5. Flow diagram of Clustering Video file

6. 4 Comparison of Clustering Methods

This module describes the details of image retrieval process or comparison of clustering mechanisms. Initially, the user inserts a query image. The query image was compared with proposed grid based clustered data and other clustering mechanisms like hierarchical clustering, Density based clustering and model based clustering dataset. The matching results are then analyzed and the detection results are returned. Based on the results, better clustering mechanism could be identified.

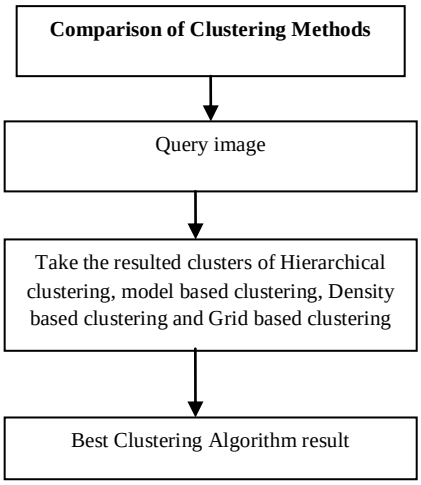

Fig 6 Flow of Cluster Comparison

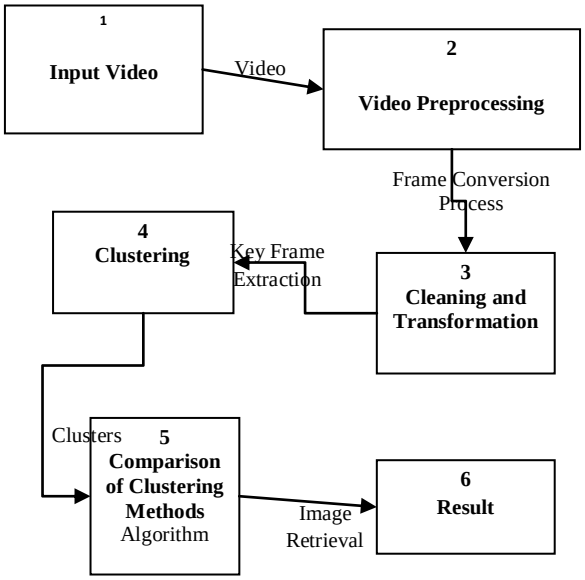

Fig 7. Overall Process of Cluster Comparison

7. Experimental outcome:

Fig 8. Browse video

Fig 9. Select the video

Fig 10. Extract the frame

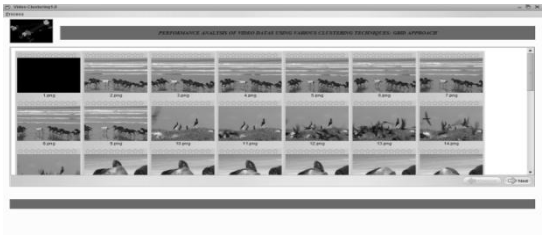

Fig 11. After extract the frame

Fig 12. Segment the frame

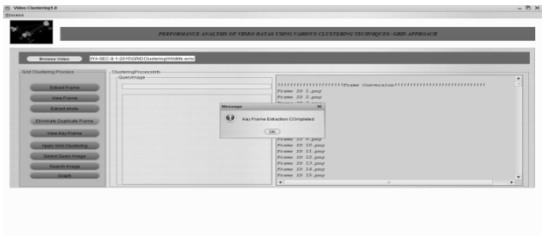

Fig 13. Eliminate the duplicate frame

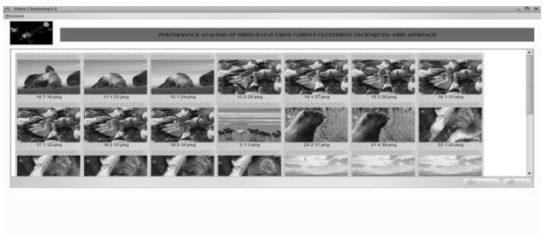

Fig 14. After eliminating the duplicate frame

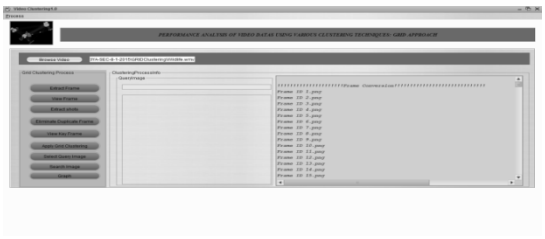

Fig 15. Apply clustring method

Fig 16. Select the query image

Fig 17. After searching the image

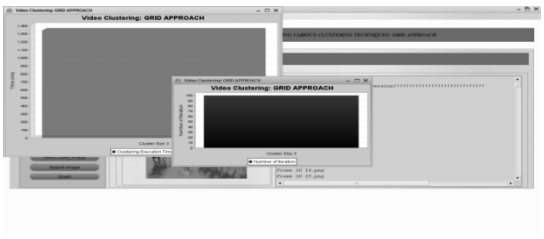

Fig 18. Graph

7. 1 Comparison of clustering method

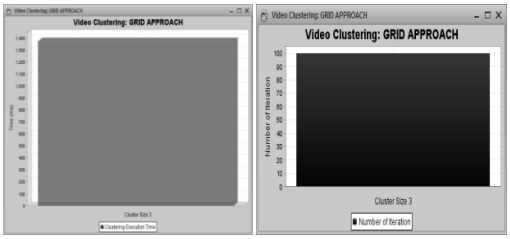

Fig 19. Grid based approach

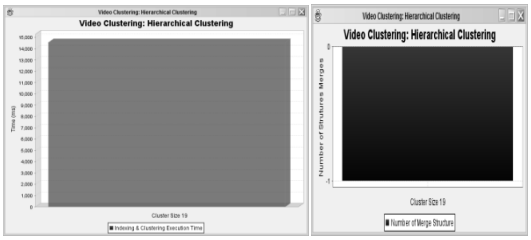

Fig 20. Hierarchical based approach

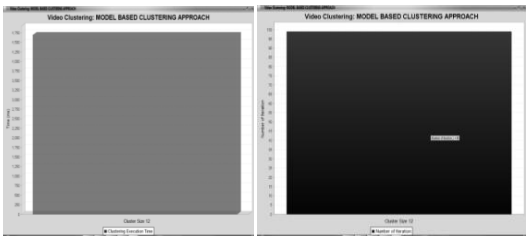

Fig 21. Model based approach

Fig 22. Density based approach.

8. Conclusion and Future Enhancement

In this research work presented a framework for analyze the performance of various clustering mechanisms like Hierarchical, Density based, Grid Based, Model Based, Partition clustering mechanisms. The proposed framework based on the analysis, we utilized RGB values to describe video frames. Initially, the number of frames is extracted from a video by using RGB value for each frame. Then, the extracted frames are clustered by using various clustering mechanism. If a user gives a query image, the corresponding image was searched in all the respected clusters. Based on the result, we can identify the better performance cluster

8. 1 Future Enhancement

The future enhancement of this project is to add additional features and to implement a new similarity calculation will produce additional results.

9. References

[1]. Tamizharasan. C, and Dr. S. Chandrakala, 2013, ' A Survey on Multimodal Content Based Video Retrieval', International Conference on Information Systems and Computing, pp. 66-69.

[2]. Akira Yanagawa, Dong-Qing Zhang, Eric Zavesky, Lexing Xie, Lyndon Kennedy, Shih-Fu Chang and Winston Hsu, (2005), 'Columbia University TRECVID-2005 video search and high-level feature extraction', in TRECVID Workshop, Waishington DC, pp 1-12.

[3] Yuen J et al., 2009, 'video building a video database with human annotations', In: Proceedings of international conference on computer vision, pp. 1451-1458.

[4]. D. Saravanan and Dr. S. Srinivasan, 2013, 'Matrix Based Indexing Technique for video data', *Journal of computer science,* 9(5), pp. 534-542.

[5] D. saravanan and Dr. S. Srinivasan, 2012, 'Video image retrieval using data mining Techniques', *Journal of computer applications (JCA),* Vol V, Issue 01, pp. 39-42.

[6] W. Chu et. al., 2010, ' An iterative image registration algorithm by optimizing similarity measurement', *Journal of Research of the National Institute of Standards and Technology* 115 (1), pp. 1-6.

[7] S. R. Pande et al., 2012, ' Data Clustering Using Data Mining Techniques', *International Journal of Advanced Research in Computer and Communication Engineering,* Volume 1, Issue 08, pp. 494-499.

[8] Guangnan Ye et al., 2012, ' Discovering Audio-Visual Bi-Modal Codeword's for Video Event Detection', *proceeding of the ACM*, pp. 1-12..

[9] D. Saravanan and A. Ramesh Kumar2013, 'ContentBased Image Retrieval using Color Histogram', *International journal of computer science and information technology (IJCSIT*), Volume 4(2), 2013, pp. 242-245.

[10] D. Saravanan and Dr. S. Srinivasan2013, 'Video information retrieval using :CHEMELEON Clustering', *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS*), Volume-02, Issue 01, pp. 166-170.

[11] D. Saravanan and Dr. S. Srinivasan 2011, ' A proposed new Algorithm for analysis for analysis of Hierarchical clustering in video Data mining', *International journal of Data mining and knowledge engineering*, vol 3, no 9. pp. 9-13.

[12] C. Schuld et. al., 2004, 'Recognizing human actions: a local SVM approach', *Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04),* pp. 1051-1054.

[13] D. Saravanan and Dr. S. Srinivasan2010, 'Indexing ad Accessing Video Frames by Histogram Approach' *In Proceedings of the International Conference on RSTSCC 2010,* pp. 196-199.

[14] D. Saravanan and V. Somasundaram 2014, 'Matrix Based Sequential Indexing Technique for Video Data Minin', *Journal of Theoretical and Applied Information.* Vol. 67 No. 3, pp 725-731.

[15] Chang. S. F, Zhang. H. J, Zhong. D(1996), "Clustering methods for video browsing and annotation", IS & T/SPIE Symposium on Storage and Retrieval for Image and Video Database, San Jose, pp 15-24.

[16]. Barber. R, Equitz. W, Faloutsos. C, Flickner. M, Glasman. E, Niblack. W, Petkovic. D and Yanker. P(1993), "The QBIC project: Querying images by content using color, textuand shape", volume SPIE Vol. 1908, pp 173-188.

[17]. Barber. R, Equitz. W, Faloutsos. C, Flickner. M, Niblack. W, Petkovic. D (1994), "Efficient and effective querying by image content', Journal of

Intelligent Information Systems", Special issue: advances in visual information management systems Volume 3, Issue 3-4, pp 231-262.

[18]. Abdelkader Belkoniene. U, Boryczka and Kilian Stoffel (2009), "Parallel k/h-Means Clustering for Large Data Sets", Finding groups in data: Cluster analysis with ants, Applied Soft Computing Journal, vol. 9, pp. 61-70.

[19]. Acharya. T and Mitra. S (2005), " Data mining: Multimedia, soft computing, and bioinformatics", John Wiley & Sons.