Journal of Theoretical and Applied Information Technology
30th September 2014. Vol. 67 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

MATRIX BASED SEQUENTIAL INDEXING TECHNIQUE FOR VIDEO DATA MINING

[1]D.SARAVANAN [2]V.SOMASUNDARAM
Assistant Professor, Faculty of Computing, Sathyabama University
Chennai 600 119, Tamil Nadu, India
Email[1] sa_roin@yahoo.com [2] vssbss@gmail.com

ABSTRACT

Due to the technology development digital media play very vital role in information technology Availability of increasing volumes of non-text information such as audio, image and video, on web it increases the accuracy of the search process. Video data mining plays an important role in efficient video data management for information retrieval. Gaining and storage of video data is an easy task but retrieval of information from video data is a challenging task. Due to this rapid development and increasing demand of users, indexing process must be automated, from the collection of information. Indexing associated with this type of files differ from normal indexing techniques. With the help of that user can find and retrieve the relevant information effectively. Many successful data mining indexing techniques have been developed through academic research and industry .This paper describes the sequential indexing technique and the experimental result verifies that the proposed technique is better than the existing technique.

Keywords: *Data Mining, Information Retrieval, Knowledge Extraction, Video data mining, Indexing, Clustering, Sequential Indexing, Image Data*

.

1. INTRODUCTION

The usage of the images has been increasing day by day as they are not only being used for expressing and explanation but also for identification and analysing. They are being used in every field today. Large amounts of image data is being stored in the databases everyday. All the data can only be used efficiently only when the correct data is retrieved successfullyHalf a terabyte or 9,000 hours of motion pictures are produced around the world every year. Furthermore, 3,000 television stations broadcasting for twenty-four hours a day produce eight million hours per year, amounting to 24,000 terabytes of data. Although some of the data is labeled at the time of production, an enormous portion remains un indexed. For practical access to such huge amounts of data, there is a great need to develop efficient tools for browsing and retrieving content of interest, so that producers and end users can quickly locate specific video sequences.

Due to the technology improvement image is the essential tool to communicating the information in the form of displaying the detailed text message it is the alternative tool. A video sequence is a set of image frames ordered in time. Generally video indexing refers to indexing of individual video frames based on their contents and the associated camera operations involved in the imaging process. We note that the image indexing techniques can be applied individually to index each frame based on their content. Hence, for computational efficiency, the video sequence is segmented and index is created for fast retrieval.

2. RELEATED WORK

Indexing is the oldest technique for identifying the content and to assist in their retrieval. Indexing was created with reference to the original text. Another decision is made by indexing the item is creating the index for the document. [1 book]. Xuchecn, Alfred o.Hero, Silvio Savanvese[1] they propose a Multimodal video indexing and retrieval using directed information , in this paper they propose SODA (shrinkage optimized directed information assessment)method for video indexing and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

reterival.In this method they calculate empirical probability distribution of both audio and visual feature over successive frames, to calculate the joint probability density functions of audio and visual feature in order to fuse feature from different modalities. The proposed method is effective for joint probability function.D.Saravanan, Dr.S.Srinvasan [2] they propose a new clustering algorithm for grouping of video frames. In this technique video frames are extracted and frame compared with each other to define relationship. Hierarchal clustering technique is adopted to make group, to detect particular event as per user demand. Bahjat safadi, Georges Quenot [3], in their paper have proposed video retrieval using Re-Ranking technique, by re-evaluating the scores of shots by the homogeneity and the nature of the video they belong to .This method improves the system performance over 18% in avg. In case of non-homogeneity content system performance over 11-13%. Essam A.El-kwae, Mansur .R.Kabuka [4] have they proposed a two signature multilevel signature file, where (2SMLSF) is introduced to access a large image information.2SMLSF encode the image information into binary signature and create a tree structure that can effectively search to satisfy the user query. This method reduces storage space, created indexed structure answer more number of users queries by creating index structure in large image file the storage space is 78% and image retrieval performance improved by 98%.

3. VIDEO INDEXING AND RETRIEVAL

3.1 Image Mining

Image mining is a complex problem because of a variety of reasons. When we try to mine image information we need some type of domain knowledge. Second, feature extraction used to identify objects. Images can be indexed by two methods 1. Text based index 2. Image based indexing. According to text based indexing the user specifying the text base query to retrieve image or image content in the database. In Image based indexing, input is given in a visual form, images are retrieved by image characteristics such as color, texture and pixel values stored in the frame and shape are used. Among the different types of image contents, the shape feature usually plays an important role due to its relatively unique property. Both methods have their own limitations. The aim of image indexing is to retrieve the similar image which is stored in the database. Stored images have their own feature image indexing and can be implemented by comparing their features, which are extracted from the images. The Essential operation performed on image mining is

1. It should be unsupervised.
2. It should be flexible.
3. It should be computationally simple.
4. It should discover interesting patterns.
5. Generationof Semantic Labels.

Based on requirement 1 while using unsupervised techniques, knowledge of the domain through their features and strategy for pattern discovery. Any mining process is flexible i.e. should not have any assumption about the data mining. Finally the method satisfy the requirement 5 i.e. discover the interesting patterns based on the data.

3.2 Video Classification

Classification is a way to categorize 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

3.3 Video Clustering

Clustering in data mining is a discovery process that groups a set of data. It maps a data item into one of several clusters, where clusters are natural grouping for data items based on similarity metrics. Traditional clustering algorithms can be categorized into two main types: Partitioned and hierarchical clustering. Video clustering has some differences with convential clustering algorithms, due to the unstructured nature of video data, preprocessing of video data by using image processing g or computer vision techniques is required to get structured format features, Another difference in video clustering is that the time factor should be considered while the video data is processed.

3.4 Indexing

The amount of multimedia content has increased in the recent year due to users needs like news, entertainments , games, medical ,education ,cinema etc .Multimedia is the combination of audio ,video, motion, sound and images. A lot of research work has been carried out on image indexing and image retrieval by many

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

researchers. It is important to organize the above data to make easy access to the large number of users. The storing and retrieval of multimedia is not an easy work, integrating of that item in standard format is also not simple. For this easy and efficient access of those files, in order to build image dictionary (indexing) in an efficient manner. The objective of image indexing is to reduce the overhead of the user. In multimedia creating of index for a particular type of data is not sufficient, we need different level of knowledge for different level of data. This result leads to a different level of quality indexing is available. The complexity of the multimedia content Retrieval is more complex than retrieval of textual information. Most of the video retrieval system uses still image as input to search the needed video content. It is also used for identifying the specified scene in the video content. Any video is represented changing the image continuously based on time. The user given input query retrieval done either stored order or time order.

4. IMAGE PROCESING FUNCTIONS

4.1 Image pre-processing

In video file, images are segmented as frames, that contains lot of impurities (such as noise) it is eliminated during the preprocessing steps. Finally we get the needed image based on the user's requirement

4.2 Feature extraction

Here image feature is set based on the users relevant of information. Many data analysis software are used for this purpose like MAT LAB, Numpy etc.

4.3 Training of images

After extracting the image features like texture and matrix conversion the pixel values are trained in the database by labeling the features of the images. The matrix conversion is done by giving intensity at each point x, y and RGB values are found [A matrix will be formed having M rows and N columns. Then the images are labeled in the database. So, it can be retrieved from the database easily. These labeling is done by the features of the image. Now, the image is stored in the database

4.4 Retrieval of Image

In the final phase we get the image based on the users query. Various techniques are used to retrieve the relevant image.

.

4.5 These Results Are Extracted By The Following Process

Step1: First, the image is given as the input to the from the camera.
Step2: Initially, this image is a raw image where it contains noise.
Step3: Then, the Features like Texture, Color and Shape are extracted by the RGB values
Step4: These features values and database values are matched. If it matched there will be fast retrieval of the image is done. The content of the image is also retrieved.

5. SEQUENTIAL INDEXING TECHNIQUES

Sequential Indexing technique similar to Binary search technique. Five sets of Video files are taken After segmenting the image, Grey value represents the value of the difference between the two adjacent image grey values. By assuming a threshold for the grey value the duplicate images are found out. Using file handling method the duplicate files are eliminated .After eliminated duplication the grouping frames are taking as input for matrix based indexing technique. In each frame the histogram value is calculated in the the following manner. This value is stored for retrieval of user input query

```
For i = 0 to matrix1 height
For j = 0 to matrix1 Width
  Histoval 1 = histoval 1 +histo (i,j)
        Loop
        Loop
        For i = 0 to matrix2 height
        For j = 0 to matrix2 Width
Histoval 2 = histoval 2+histo (i,j)
                    Loop
         Loop
```

After getting the values of matrix1, matrix2 values are sorted (ascending order) and the total frames values is splitted in three equal parts. The given frame values of matrix 1 and martrix2 is compared with middle element of the stored values. If the values are match hen a matching element has been found the corresponding frames are returned. If the values are not matches

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the middle element values, then the process is repeats on the other part of the middle elements.

5.1 Benefits Of The Sequential Scheme That They Propose

Searching Time reduced because instead of checking all the records we have to split it into number of segments and made search with appropriate segments of frames.

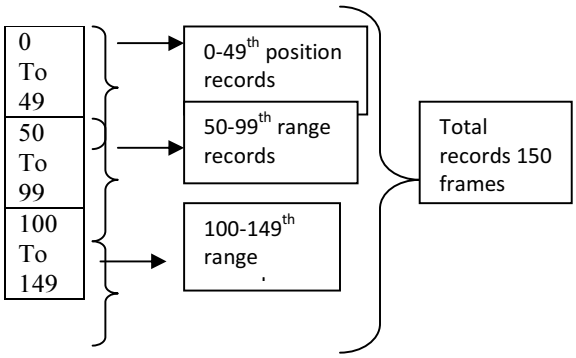

Figure 1. Flow of the sequentail index search

5.2 Drawbacks Of The Sequential Scheme That I Established By Practical Implementation

Time must be calculated including the indexing time and searching time. But sequential search scheme establish the method for search the frame only. When we include the indexing time (Training phase) with the search time, the method take much time than that of our cell based technique.

6. EXPERIMENTAL SETUP AND RESULTS

6.1 Training Phase Algorithm

- Select Frame
- Find pixel values of frame
- Select 2 matrixes diagonally with pixel values as dynamic as per the original height and width of the frame
- Read the matrix1 and matrix2 pixel values and convert to histogram format and store it in database
- Find the input parameters like frame no, frame name, pathname, matrix value1, matrix value2, index time, Sorted order Values of Matrixvalue1, Matirx value 2 are stored as trained inputs in the database
- Select next frame, continue with step2

6.2 Frame Retrieval Or Recognition Phase

- Select Input frame
- Find pixel values of frame
- Select 2 matrixes diagonally with pixel values as dynamic as per the original height and width of the frame
- Read the matrix1 and matrix2 pixel values and convert to histogram format and store it in database
- Find the input parameters like frame no, frame name, pathname, matrix value1, matrix value2, index time , Sorted order value of Matrix1, Sorted value of Matrix2.
- Compare the parameters with the existing database. But the comparison is based on the setting the threshold
- Matrix1 histo val (threshold) = matrix 1 histoval (input frame)+(Total pix val of matrix1/2)
- Matrix2 histo val (Threshold) = matrix 2 histoval (input frame)+(Total pix val of matrix2/2)
- Frames that are related to those threshold values are the output frames.

Table 1.Framecount Value of Cartoon Video Frame

Frmcnt	milliseconds	category
25	1174	Cartoon
50	1185	Cartoon
75	1219	Cartoon
100	1222	Cartoon
125	1300	Cartoon
150	1375	Cartoon

Table 2.Framecount Value of Cricket Video Frame

Frmcnt	milliseconds	category
25	1187	Cricket
50	1198	Cricket
75	1220	Cricket
100	1250	Cricket
152	1390	Cricket
125	1391	Cricket

Journal of Theoretical and Applied Information Technology
30th September 2014. Vol. 67 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 3.Framecount Value of Debate Video Frame

Frmcnt	milliseconds	category
25	1347	Debate
50	1364	Debate
75	1379	Debate
100	1398	Debate
125	1406	Debate
150	1407	Debate

Table 4.Framecount Value of News Video Frame

Frmcnt	milliseconds	category
25	1486	News
50	1495	News
75	1503	News
100	1547	News
125	1547	News
150	1563	News

Table 5.Framecount Value of Songs Video Frame

Frmcnt	milliseconds	category
25	1450	Songs
50	1453	Songs
75	1468	Songs
100	1498	Songs
125	1501	Songs
150	1547	Songs

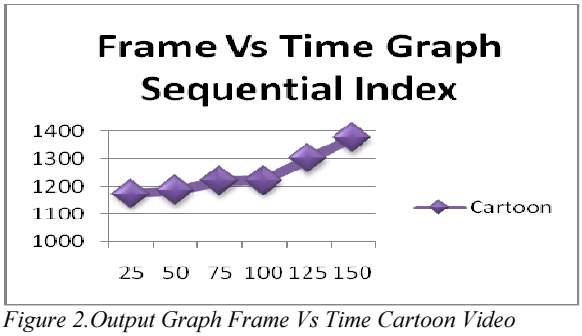

Figure 2.Output Graph Frame Vs Time Cartoon Video Frames

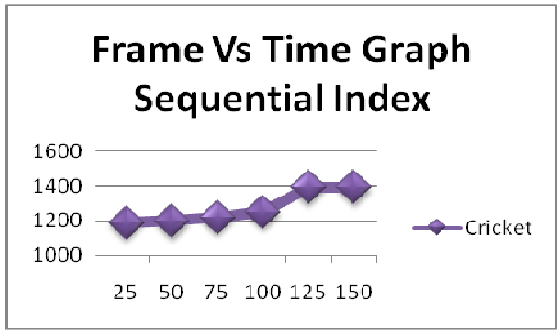

Figure 3 .Output Graph Frame Vs Time Cricket Video Frames

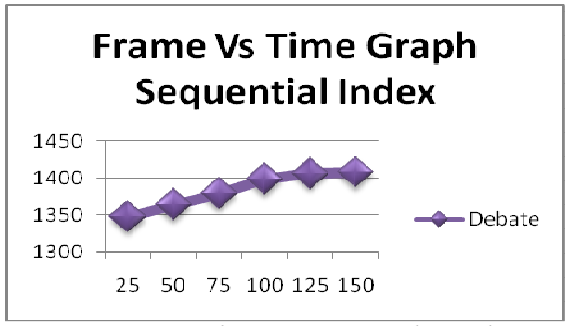

Figure 4.Output Graph Frame Vs Time Debate Video Frames

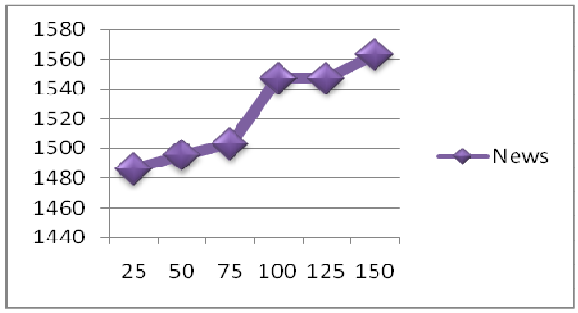

Figure 5.Output Graph Frame Vs Time News Video Frames

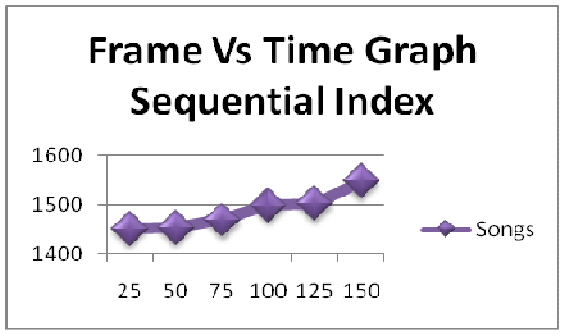

Figure 6.Output Graph Frame Vs Time Songs Video Frames

Journal of Theoretical and Applied Information Technology
30th September 2014. Vol. 67 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

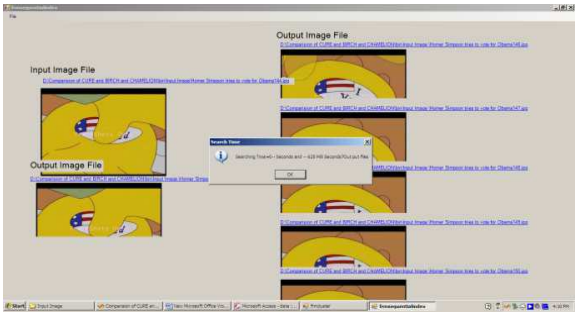

Figure 7. Matrix Based Sequential Indexing Output For Cartoon Video File

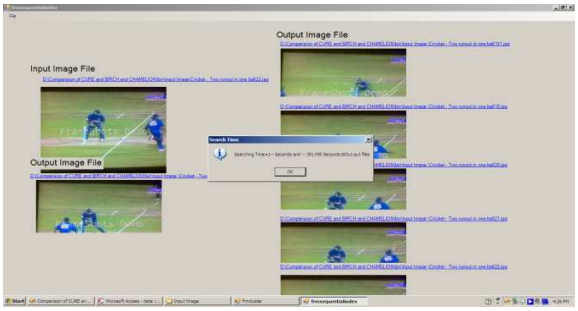

Figure 8. Matrix Based Sequential Indexing Output For Cricket Video File

Figure 9. Matrix Based Sequential Indexing Output For Debate Video File

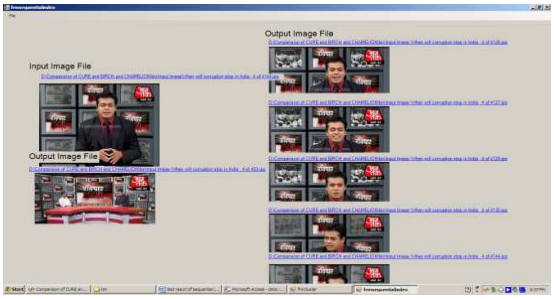

Figure 10. Matrix Based Sequential Indexing Output For News Video File

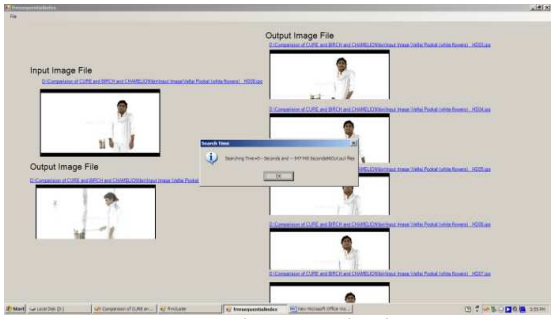

Figure 11. Matrix Based Sequential Indexing Output For Song Video File

7. CONCULSION

Recent advances in computing and storage increasing the usage of multimedia content in various fields. The proposed method improve the system performance, and it retrieve the more number of relevant image based on the user input query .The process reduces the searching time and improve the overall retrieval performance. The result of our proposed method video information retrieval using sequential indexing technique algorithm suitable for content-based video indexing and retrieval

REFERENCES

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

[2]Xu Chen, Alfred o.Hero, III, Fellow, IEEE, andSilvio Savarese (2012). Multimodal Video Indexing and Retrieval Using Directed Information, *IEEE transaction of Multimedia*, Vol 14, No.1, Feb-2012, 01-14.

[3] 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.

[4] D.Saravanan, A.Ramesh Kumar, "Content Based Image Retrieval using Color Histogram", International journal of computer science and information technology (IJCSIT), Volume 4(2), 2013, Pages 242-245, ISSN: 0975-9646

[5] Bahjat safadi, Geprges Qenot(2011). *Re-Ranking by local Re-Scoring for video indexing and retrieval*, in the proc.of International conference on information

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and knowledge management, Glasgow, United Kingdom.

[6] D.Saravanan, Dr.S.Srinivasan(2013) "Video information retrieval using :CHEMELEON Clustering" International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume-02, Issue 01, January –February 2013, Pages 166-170

[7] D.Saravanan, 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.

[8] D.Saravanan, Dr.S.Srinivasan (2010). *Indexing ad Accessing Video Frames by Histogram Approach*, In the Proc. Of International Conference on RSTSCC 2010, 196-1999.

[9] Ajay divakaran, koji miyahara, kadir A.peker, Regunathan Radhakrishnan, and Ziyou Xiong (2004), *Video mining using combinations of unsupervised and supervised learning techniques*

[10] Essam A.El-kwae, Mansur.R.Kabuka (2000). Efficient Content based indexing of large image database, *ACM transaction on information systems*, Vol 18, No.2 .April 2000, 171-210.

[11] D.Saravanan, A.Ramesh Kumar(2013),"Content Based Image Retrieval using Color Histogram", International journal of computer science and information technology (IJCSIT), Volume 4(2), Pages 242-245