An Innovative Video Summary and Pattern Mining using VSP Algorithm

N. Venkatesvara Rao, D.V.V. Prasad, D. Susitra

Abstract: ***This paper proposes a VSP-TREE algorithm that mines l associations from video. A tool is designed to annotate the Video sequences by giving appropriate values to the sequence and then these values are converted into two-dimensional datasets suitable for clustering. The datasets are clustered using innovative algorithm to form distinct group and known as summary candidate with user size, our system make summary by choosing important frame from candidate cluster and put them in original. A VSP-TREE Based Mining method is used find out frequent patterns occurrence in the video.***

Association mining algorithm used on clustered datasets with innovative method, Video sequential pattern Tree (VSP Tree) Structure to generate frequent patterns through efficient methodology called conditional search.

Keywords: Video, Clustering, Datasets, Sequence and Mining

I. INTRODUCTION

Information of Video vary with signals that not to be seen at a glance [3]. Automated summary are known for the user content.

Various application, video numbering. The research tool for video summary is video magnifier uniformly sampled video sequence at certain time intervals.

Our aim in summarization is to develop a tool for annotating the video sequence with numeric values for video sequences to generate dataset for integration clustering algorithm used to form cluster and summary creation according to user specified length.

To create summary, the following factor taken into account.

- Minimal Redundancies
- Convey useful Information

There are two requirements: (1) the summary contains all stories (2) the summary frames are concise. Based on these facts, [8] The Video sequences are annotated by assigning value creating datasets. Datasets are grouped into clusters using integration clustering algorithm. One set from every cluster according to user specified length and carry out mining

- Input: Sequence of Video and its length
- Output: Cluster

Revised Manuscript Received on October 21, 2019.

N. Venkatesvara Rao, Assistant Professor (selection grade) in Computer science and Engineering, Saveetha school of Engineering, Saveetha Institute of Medical Sciences and Technology

D.V.V. Prasad, Professor in Department of computer science and Engineering, SSN College of Engineering

D. Susitra, Associate professor in Electrical and Electronics Engineering, Sathyabama Institute of science and Technology

II. DESIGN

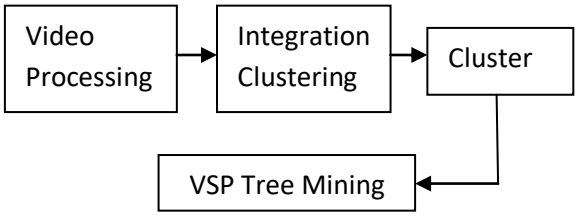

Fig. 1 Block diagram

Video pre-processing
(a) Videos are annotated
(b) Creating datasets

- Creating cluster for summary.
- Frequent pattern mining

Main steps in video summarization

Knowledge mining from large database has been recent studies and applications. Video mining used to find special pattern from video. However, in comparison with mining, it looks like video processing.

Video classified with content structure and without structure. The former are like movies and news scenarios is used to convey video content and no content. Editors usually edit videos with content structure information.

III. VIDEO PRE-PROCESSING

Knowledge extraction of video, one basic task is to convert given sequence into dataset, data items are explicitly specified with respect to problem. A Tool is developed for video annotation to process the video has been carried out for all video sequence by assigning numeric value to the sequence and store it in the database with field name data value, frame number and time interval for the sequence such as start and end time. The database can be further accessed through program to form sequence like 1,2,3,4,5,6 and 1,4,5,7,8,9 etc, algorithm used to create cluster from dataset obtained.

IV. INTEGRATION CLUSTERING ALGORITHM

Clustering is an unsupervised learning problem [9]. It is simply defined as "process of arranging objects into groups whose members are similar". A cluster is therefore a collection of objects, which are "similar" between them and are "dissimilar" to the objects belongs to other clusters. The data points belonging to same cluster assigned same label. When learning is unsupervised then the system has to discover its own classes i.e. the system clusters the data in

the database. The system has to discover subsets of related objects in the data set and then it has to find descriptions that describe each of these subsets. There are various approaches used to form clusters, one approach is to form rules which dictate member based on. Other approach is to build set functions that measure property of partions used as parameter In a broad definition, clustering of data include recognition and focusing on key dimensions of data and estimation of the correct number of clusters inherent to the data. Assuming that diverse clustering algorithm [12] can provide all the needed clusters, this project uses an Objective function to select the set of clusters.

Satisfying by minimizing the number of conflicting and unassigned data points, from the partitions created by different clustering algorithms. Our data value are processed by various clustering algorithm such as K means, Single Link and expectation maximization [13] applied on data of our video sequence and Clusters are given input as integration clustering algorithm to find optimal cluster.

4.1 Objective Function: Selection of Clusters

Given input S = {C1… CM}, we find T = {C1* … CK*}, the output clusters using cluster fitness function (objective function).

Minimize J (u) = ∑∑qij + (S-DPcount) /S

Subject to ui ∈ {0,1}

4.2 Refining Target Clusters

Once the elements from different partitions are combined, the elements can be represented using a representative point Z (J), usually chosen as the centroid of the points in the set . The energy of each element is E(J) = Sum (all points X(I) in cluster J) || X(I) - Z(J) |^2 .

4.3 Integration Algorithm

Input: Partitions (containing clusters) from different clustering algorithms [14]

Output: Target clusters

1. From the input set of elements, eliminate redundant element and empty element.
2. Let C= {c1, c2...cm}
3. Create conflicting data point (CDP) matrix among candidate element.
4. Let U = (u1, u2…um) vector element.
5. Find the sub matrix (I X j) where & j are the subscript indices of 1's in the U vector.
6. The Sum of all the values in the sub matrix (iXj) gives the no. of Conflicting Data Points among the included Clusters at any point of time.

CDP=∑∑qij

7. Unassigned Data Points (UDP) is determined as UDP = (S -DPcount) / S where S is the size of the Ci DPcount is count of Distinct data points that are Classified.
8. Determine the objective function J(U)= CDP +UDP where qij is e. The sub matrix formed using no. Of conflicting points among Clusters assumed to form Target cluster.
9. Optimize the objective function (Minimize J (U)), Iteratively by Varying position of 1's in U.
10. The position of 1's in the min J (U) Gives the Clusters to be included in the Target cluster.

V. SUMMARY CREATION

The cluster, which is obtained, using integration clustering algorithm is the required summary. In our database each data value will have frame number, start and end time of video sequence so without altering the sequence of the original video the exact summary can be displayed in a thumbnail form with the user specified length. Using the summary an association-mining algorithm can apply to find the frequently occurring pattern in the video.

VI. VIDEO ASSOCIATION MINING

Finding frequent pattern in database, was first introduced by Agrawal and Srikant [16], information in database are treated as transactions and contain set of items, find patterns with user specified minimum support is the number of data sequence that contain the patterns, mostly video viewed as a sequence and each transaction is represented as item sets in that sequence.

Support(s)=No. of video sequence / total no of sequence. pattern mining is of extracting certain sequential pattern when support exceeds a predefined minimal support threshold. Since the number of sequences can be very large, and users have different interests and requirements, to obtain interesting patterns, minimum support is pre-defined by users, using prune out those patterns of no interest, which

The studies exponential growth of candidate sequence and scanning database 'n' number of times are the drawback to be handled. When no of sequence increase or when they are large, the number of candidate sequences generated may grow exponentially not advisable to extract frequent patterns.

The innovative highly compressed VSP Tree structure is formed to handles the sequence elegantly by reducing the database scans twice, second, an efficient mining algorithm, VSP mine, is used for mining the complete frequent patterns from database.

Pattern mining focus on identification of sequences, which occur frequently in a time series, or sequence of data, where certain pattern could be identified, such as what might be the likely follow-up certain video sequence. A novel algorithm, VSP algorithm, is devised for efficient mining of frequent patterns from large dense datasets.

The basis mining algorithm is conditional search narrowing the search space and look for patterns with the same suffix, and count frequent events in the set of prefixes with respect to condition as suffix. Conditional search is an divide-and-conquer methods to avoid generating large candidate sets.

Video are converted into set of cluster, which is our video summary.

Of sequence like 1,3,4,5,6,7 , Let us denote D is a process of one consumer, and that each item in the process D is one transaction. The sequential correlations among D reflect the association among video and groups. Our aim to find the sequence of D.

To mine values of D [19]. We construct a Video sequential Tree (VSP Tree) Structure, gathering information about occurring patterns and mining set of frequent patterns called conditional search, efficiency achieved by (1) A large database is compressed into a condensed smaller data structure, VSP tree avoids costly, repeated database scans, (2) VSP tree mining adopts a patterns fragment method to avoid generation larger candidate sets and (3) A partioned based, divide-and-conquer method is used to decompose the mining task , which dramatically reduces the search space. Performance study shows that the VSP methodology is efficient and scalable for mining both long and short frequent patterns, and is about an order of magnitude faster than the conventional Apriori algorithm. As analyzed, the major costs in Apriori are generation of huge number of candidates and e repeated scanning databases is the bottleneck for apriori.

To overcome this problem, VSP Tree (video sequential pattern) and its VSP mine algorithm has been designed and implemented. First forms an effective data structure, VSP tree has been developed as and efficient algorithm for mining frquent pattern from a VSP Tree.

Video patterns mined using sequential pattern mining techniques, all previously proposed methods for sequential pattern mining are based on a sequential pattern version of Apriori heuristic. The main procedure is depicted in fig.2.

6.1 Algorithm 1 (CPP Methodology)

Input: Video input database VPS and support threshold ξ (o<ξ<=1)

Output: The set of ξ-patterns in VPS

1. Scan database once, find all frequent items.
2. Scan database, construct a VSP tree over.

Overview of work and Architecture

- Input Video
- Give numeric value to video
- Create database with respect to particular video
- Apply mining technique that mine frequent patterns from database with respect to user p specified threshold. The architecture of the whole process is given in fig 2

6.2 Algorithm 2 (Construction of VSP tree)

Input: Information of video

Output: A tree T

Method:

1. Create a node with root T
2. Every sequence database do

a) Retrieve items in S', let current node point to the root of T.

b) Make current node point to the new node, and

c) Return (T)

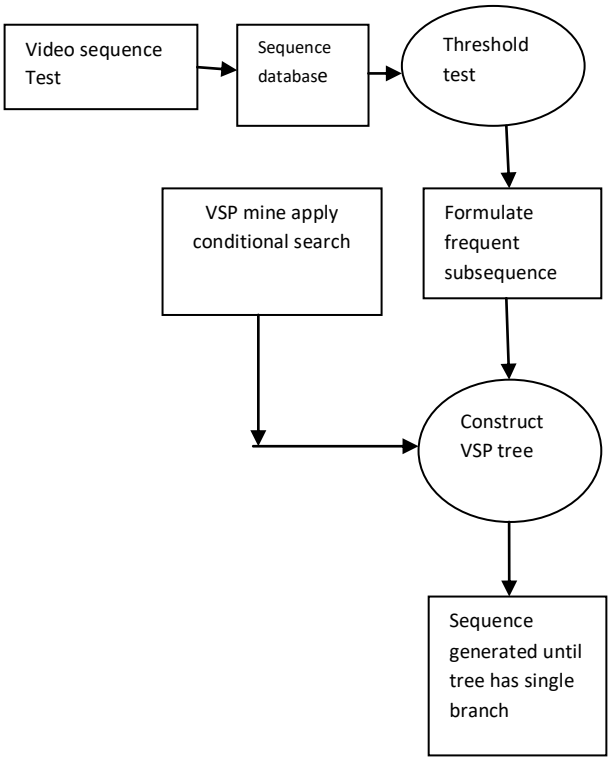

Fig. 2 VSP Tree and Mine Architecture

6.3 Algorithm 3 (VSP Mine)

Input: Tree T and support Threshold ξ

Output: The set of ξ-patterns

Method:

1. Return all the unique combinations of nodes in that branch
2. Initialize VSP =0. Every event in VSP tree T itself is a frequent pattern, insert them into VSP
3. For each event ei in VSP tree T,

(d) Return VSP

VII. EXPERIMENTAL RESULTS

7.1 Video Pre-Processing

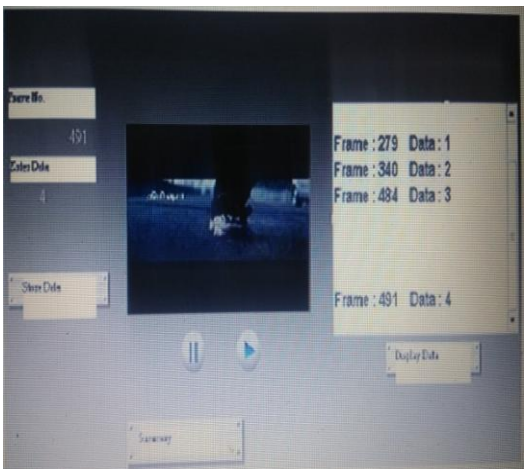

Fig. 3 Video Pre-Processing

1,2,3,4,5,6,D1
2,3,4,5,6,7,D2
1,3,4,5,6,8,D3
1,2,4,6,7,9,D1
1,2,3,4,5,6,D2
2,3,4,5,6,7,D3
1,3,4,5,6,8,D2
1,2,4,6,7,9,D3
2,3,4,5,6,7,D1
1,3,4,5,6,8,D1
1,2,4,6,7,9,D2
1,2,3,4,5,6,D
3,5,6,8,9,10,D
3,5,6,8,9,10,D
4,5,6,7,8,11,D
3,5,7,8,9,12,D
4,5,6,10,11,13,D
4,5,6,10,11,13,D
2,3,4,6,7,8,D
3,5,6,8,9,10,D
1,2,3,4,5,6,D
2,3,4,5,6,7,D
1,3,4,5,6,8,D
4,5,6,7,8,11,D
3,5,7,8,9,12,D
4,5,6,10,11,13,D
2,3,4,6,7,8,D
3,5,6,8,9,10,D
4,5,6,7,8,11,D
3,5,7,8,9,12,D
4,5,6,10,11,13,D
2,3,4,6,7,8,D
3,5,6,8,9,10,D
4,5,6,7,8,11,D
3,5,7,8,9,12,D
4,5,6,10,11,13,D
2,3,4,6,7,8,D
3,5,6,8,9,10,D
4,5,6,7,8,11,D
3,5,7,8,9,12,D
1,2,3,4,5,6,D

7.2 Integration Clustering

```
C:\j2sdk1.4.2_05\bin>javac integration.java
C:\j2sdk1.4.2_05\bin>java integration
Enter the input file name: sum
Input 'K' (int): 4
Pick up seeds Randomly or Specifically?
(R/S): R
Please specify the output filename:
POOJITHAA
-------- Cluster 1 (16 data) --------
Centroid: (3.0, 5.0, 6.0, 7.0, 8.0, 10.0)
-------- Cluster 2 (5 data) --------
Centroid: (4.0, 5.0, 6.0, 10.0, 11.0, 13.0)

-------- Cluster 3 (13 data) --------
Centroid: (1.0, 2.0, 3.0, 4.0, 5.0, 6.0)
-------- Cluster 4 (7 data) --------
Centroid: (1.0, 2.0, 4.0, 6.0, 7.0, 8.0)
```

7.3 Clustering Analysis

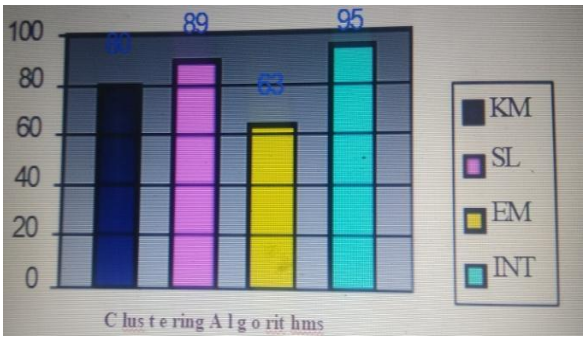

Fig. 4 Algorithm Comparison

7.4 Video Summary Results

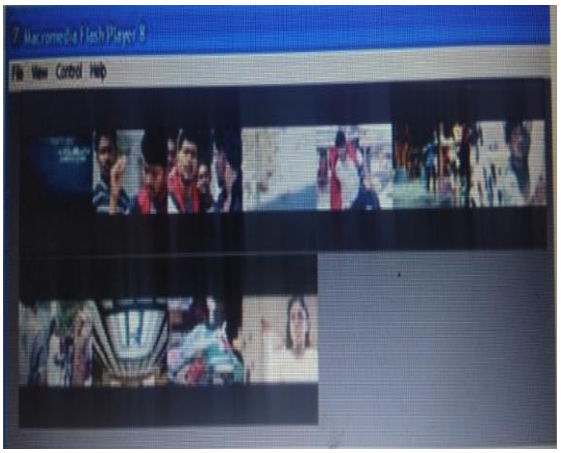

Fig. 5 Video Summary

7.5 VSP Tree Mining

An VSP tree mining algorithm is designed by modifying the existing FP-Tree algorithm by introducing new research issue like conditional search make the tree construction is very easy for removing redundant path through an optimized way of finding the patterns occurring in the video and also very efficient than other algorithms like Apriori.

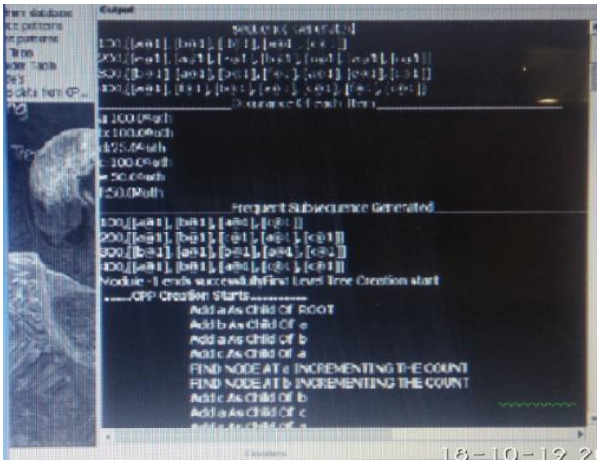

Fig. 6 VSP Tree Mining

7.6 VSP Mining Results

Final Discovered frequent pattern set, after mining the vsp tree is

Frequent pattern with respect to item '3'
3, 113, 213, 1213, 13, 123, 23
Frequent pattern with respect to item '2'
2, 12
Frequent pattern with respect to item '1'
1, 11, 12, 121.

VIII. CONCLUSION

In this paper, we created an video summary using integration clustering algorithm and with that finding out frequent pattern in the video with respect to specific value such as 1, 2 and 3 etc. our system in selecting useful clusters for summary creation, with specified length, we first present the information in visual and systematically related frames from clusters and address the important video content. The person can get chance to catch the information and construct needed summary. The program results prove that effectiveness of the algorithm is measured through execution and memory utilization.

REFERENCES

1. R.hart s. Pfeiffer, W.berg,"Efficient Video annotating", communication of the ACM. Vol.40,
2. M. kan.T.Kenndy " summarisation and partitions"- Springer-Verlag London UK -2001.
3. Y. X. " video summary decomposition", Engineering, Taipei, Taiwan, PP.3-142.
4. M.Mills" An efficient tool for video data", proc. Of ACM Computer Interface., may
5. A. Garmid, waref "Hierarchial summary on bio informatic data" IS&T/SPIE, vol.4676, 2002.
6. Christ "New information for news video library", Conf.Fl..Oct.,
7. A.Kantan and " Innovative representation on video", ACM Multimedia systems, vol. L, no. l, 1993.
8. X.Zhube and XWuer,"Mining useful association for efficient system", proc. of 18th IJCAI conf., aug., 2003.
9. H.C.Law, P.Tophey and Anil K.Jain, " Overall objective of data clustering", IEEE Transactions on computer Vision Patterens Recognition, June 2004.
10. Ana Freed and K.jain,"Clustering data using Robust technique"-IEEE Conference on patterens analysis and machine Intelligence.
11. Ujjwal and mitra,"New technique improving Performance on Clustering algorithm"- In IEEE Transactions on Patterens Analysis and Machine Intelligence - 2002.
12. M.kidi, Y.stakis, and M.giannis, "Algorithm on Clustering and validity".
13. Asosk Strehl and Jain .Ghosh. "A new approaches on Cluster grouping- framework for forming multiple part ions", Dec-2002.
14. Hanisha and Kamber sha," A novel Concept of Data Mining and Techniques.
15. Ramesh Agrawal R. and kara Srikant R (1995) "Mining For efficient
16. M.kidi, Y.stakis, and M.giannis, "Algorithm on Clustering and validity".
17. Asosk Strehl and Jain .Ghosh. "A new approaches on Cluster grouping- framework for forming multiple part ions", Dec-2002.
18. Hanisha and Kamber sha," A novel Concept of Data Mining and Techniques.
19. Ramesh Agrawal R. and kara Srikant R (1995) "Mining For efficient Engineering, Taipei, Taiwan, PP.3-142.