International Journal of Engineering Inventions
e-ISSN: 2278-7461, **p**-ISBN: 2319-6491
Volume 2, Issue 2 (January 2013) PP: 35-37

Objects Clustering of Movie Using Graph Mining Technique

Ph.D. Student Israa Hadi, Assist. Prof. Dr. Saad Talib Hasson
University of Babylon- Iraq

Abstract:- This paper aims to develop graph mining techniques to discover patterns consisting of relationships between objects in movie film. An algorithm was designed to operate over graph data. Our approach includes a mechanism for many steps, first step discrete the movie film into number of frames (still image), second step apply segmentation technique to determine the objects in each frame and extract the features for each objects, third step construct data base consist row for each feature, fourth step construct a graph structure represent each frame, and fifth step apply adaptive graph mining algorithm to cluster the objects and determine the behavior of that objects.

Keywords:- Graph mining, data mining, Movie Graph mining.

I. INTRODUCTION

Data mining techniques can be classified according to different views such as what kinds of knowledge to work on, what kinds of databases to be mined, and what kinds of algorithms to be applied [14]. Graphs are a natural way to represent multi-relational data and are extensively used to model a variety of application domains in diverse fields. Often, in such graphs, certain sub graphs are known to possess some distinct properties and graph patterns in the proximity of these sub graphs can be an indicator of these properties. As a result, a number of approaches have been developed which have achieved promising results in uncovering interesting patterns in biological networks, social networks [1] and the World Wide Web [2]. A concise overview of the current work in in the field can be achieved by categorizing it according to the types of tasks, the types of graph data and the approaches in graph-based data mining. Much of data mining research is focused on algorithms that can discover number of features that discriminate data entities into classes or clusters. Our work is focused on data mining techniques to discover relationships between entities and determine its behaviors. Clustering, in data mining, is useful for discovering groups and identifying interesting distributions and behavior in the underlying data. The problem of clustering can be illustrated as partition the given N data points into k clusters such that the data points within a cluster are more similar to each other than data points in different clusters. Graph-based data mining (GDM) is the technique of finding new, useful, and understandable graph-theoretic patterns in a graph representation of data. There are two approaches to graph based data mining to frequent sub graph mining and graph-based relational learning. The are number of attempts in the graph mining like the AGM (Apriori Graph Mining) system which uses the Apriori level-wise approach [3]. FSG takes a similar techniques and further optimizes the algorithm for improved running times[4]. gFSG is a variant of FSG which enumerates all geometric sub graphs from the database[5]. gSpanuses DFS codes for canonical labeling and is much more memory and computationally efficient than the previous approaches[6]. FFSM is a graph mining system which uses an algebraic graph framework to address the underlying problem of sub graph isomorphism [7].

The Subdue system is a structural discovery tool that finds substructures in a graph-based representation of structural databases. This technique operates by evaluating potential substructures for their ability to compress the entire graph [8]. Subdue uses MDL-based compression heuristics, and has been applied to learning predictive as well as descriptive models. While learning descriptive models, Subdue can deal with both the single graph as well as the graph transactions category. While learning predictive models, Subdue mainly deals with the graph transactions category.

GBI uses a graph size-based heuristic; the graph size definition depends on the size of the extracted patterns and the size of the compressed graph.

II. PREVIOUS WORK

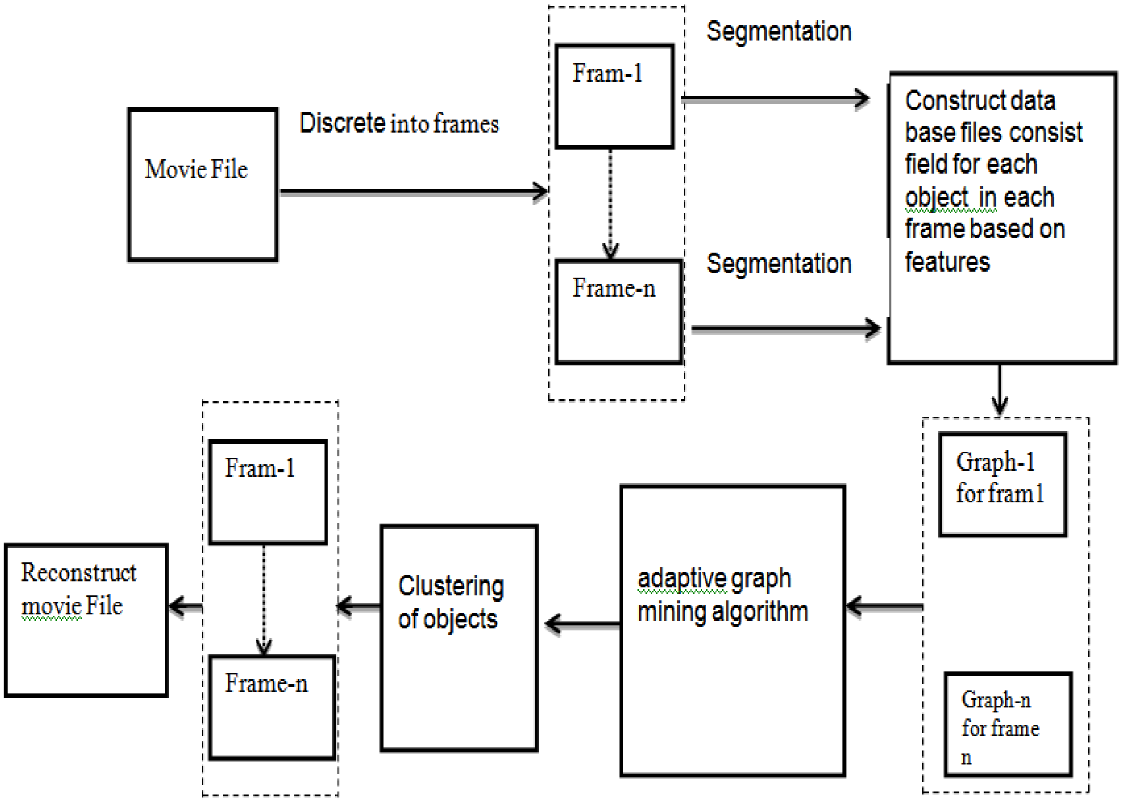
Kurmochi and Karyps (2001) developed the FSG system for finding all frequent subgraphs in large graph databases. FSG starts by finding all frequent single and double edge subgraphs. Then, in each iteration, it generates candidate subgraphs by expanding the subgraphs found in the previous iteration by one edge. In each iteration the algorithm checks how many times the candidate subgraph occurs within an entire graph. The candidates, whose frequency is below a user-defined level, are pruned. The algorithm returns all subgraphs occurring more frequently than the given level [11].

Yan and Han (2002) introduced gSpan, which combines depth-first search and lexicographic ordering to find frequent subgraphs. Their algorithm starts from all frequent one-edge graphs. The labels on these edges together with labels on incident vertices define a code for every such graph. Expansion of these one-edge graphs maps them to longer codes. Since every graph can map to many codes, all but the smallest code are pruned. Code ordering and pruning reduces the cost of matching frequent subgraphs in gSpan [12].

Inokuchi et al. (2003) developed the Apriori-based Graph Mining (AGM) system, which searches the space of frequent subgraphs in a bottom-up fashion, beginning with a single vertex, and then continually expanding by a single vertex and one or more edges [13].

III. THE PROPOSED SYSTEM

In generl, the proposed system consist of many steps as shown in figure(1)



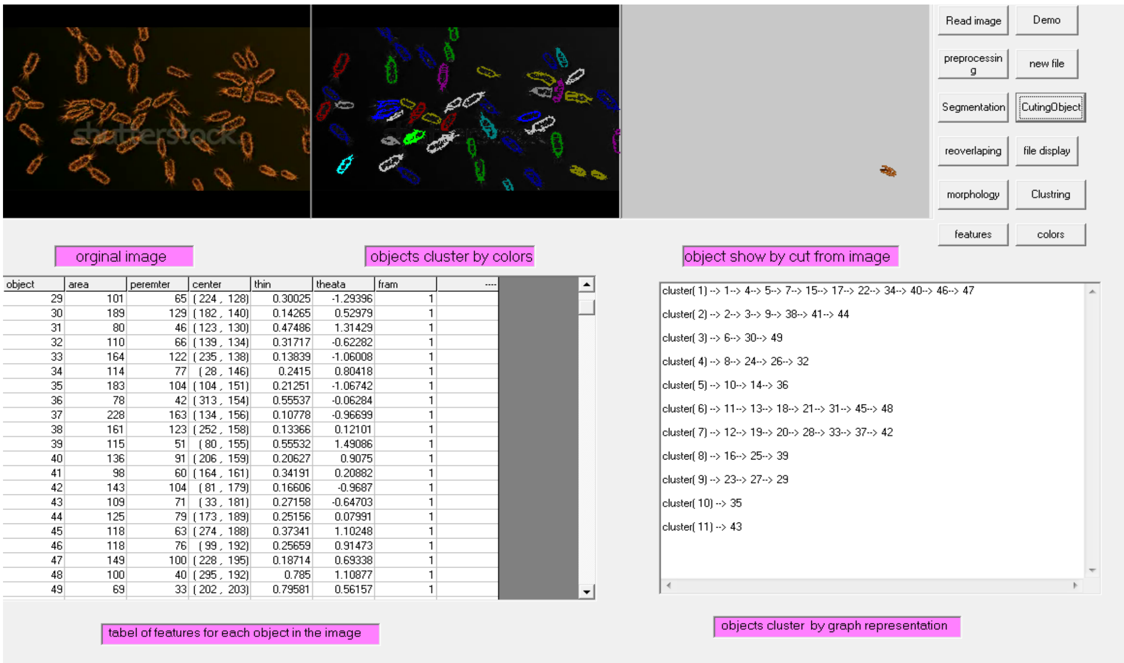
first step discrete the movie film into number of frames (still image), ***second step*** for each frame we apply segmentation technique to determine the objects in each frame and extract the features for each object such as area, orientation, center, perimeter, texture and aspect ratio, and use these features in the next step.

third step construct data base consist row for each object based on above features.

fourth step construct a graph representation which represent the relation between the objects in each frame , the graph representing each frame is a Feature Adjacency Graph (FAG). In FAGs , the objects in a frame are the nodes of the graph, and an edge exists between two nodes if there exist a similarity measurements between there features .

fifth step apply adaptive graph mining algorithm depending on Gspan algorithm to cluster the objects all frames and determine the behavior of that objects, and return back to maintain movie file to highlight that objects .

Figure 2 shows example of frame and part of database which is built to hold the features that extracted from each object in frames and the graph which represent the relations among these objects . These graphs will be passed to the adaptive graph mining algorithm to extract the similarities between objects in all frames and obtain the clusters.



Figure(2) Example of frame and part of data base and graph

IV. CONCLUSIONS

In this paper a technique for mining the data of movie film was presented. The algorithm clusters frame objects based on the features similarity among frame objects depending on separate each frame as still image and covert each frame into graph representation and using an adaptive graph mining algorithm. The results of applying this algorithm gives good clustering results with less computing costs because of the graph representation offer the spatial relation between objects, features similarity and the pattern (occurrence) of objects in many frames.

The algorithm deal with mining all movie objects based on different type of features. The algorithm is also can determine the behaviours of object and can maintain the movie film to highlight these objects.

REFERENCES

1) M. Mukherjee and L. B. Holder. Graph-based data mining for social network analysis. In Proceedings of the ACM KDD Workshop on Link Analysis and Group Detection, 2004.
2) A. Rakhshan, L. B. Holder, and D. J. Cook. Structural web search engine. In FLAIRS Conference, pages 319–324, 2003.
3) A. Inokuchi, T. Washio, and H. Motoda. An apriori-based algorithm for mining frequent substructures from graph data. In PKDD, pages 13–23, 2000.
4) J. Shetty and J. Adibi, Discovering Important Nodes through Graph Entropy: The Case of Enron Email Database. KDD, Proceedings of the 3rd international workshop on Link discovery, 2005, 74–81.]
5) M. Kuramochi and G. Karypis. Discovering frequent geometric subgraphs. In ICDM, pages 258–265, 2002.
6) [6] X. Yan and J. Han. gspan: Graph-based substructure pattern mining. In ICDM, pages 721–724, 2002.
7) J. Huan, W. Wang, and J. Prins. Efficient mining of frequent subgraphs in the presence of isomorphism. In ICDM, pages 549–552, 2003.
8) Kuramochi, M., & Karypis, G. (2001). Frequent subgraph discovery. In Proceedings of the First IEEE Conference on Data Mining.
9) M. Kuramochi and G. Karypis. An efficient algorithm for discovering frequent subgraphs. IEEE Trans. Knowl. Data Eng., 16(9):1038–1051, 2004.
10) Yan, X., & Han, J. (2002). Graph-based substructure pattern mining. In Proceedings of the International Conference on Data Mining.
11) Inokuchi, A., Washio, T., & Motoda, H. (2003). Complete mining of frequent patterns from graphs: Mining graph data. Machine Learning, 50, 321-254.