International Journal of Engineering Research and General Science Volume 3, Issue 4, Part-2, July-August, 2015
ISSN 2091-2730

A New Approach To Graph Based Object Classification On Images

Sandhya S Krishnan,Kavitha V K
P.G Scholar, Dept of CSE, BMCE, Kollam, Kerala, India
Sandhya4parvathy@gmail.com

Abstract: The main idea of graph based image modeling is that the regions of the image, which contain similar properties, are denoted by graph vertices, and the relations between different regions are denoted by graph edges. The vertex and edge attributes usually describe the characteristics of that region and the relation between regions respectively. A simple approach to keep the structural and topological information of an image is to use digital image representation techniques; for instance, quad trees, etc. By modeling images as graphs, the task of image classification becomes one of classifying graphs. Quad trees have been used for representing images in the form of trees. In this paper we propose an algorithm that discovers the subgraphs present in a graph represented by a stream of labeled nodes and edges.Our algorithm is efficient and is easily tuned by the user to produce interesting patterns from various kinds of graph data.

Keywords: Graph,Vertex,Edge,Attribute,Digital image representation,Quad tree,Subgraph

1. INTRODUCTION

Data in multiple domains can be naturally modeled as graphs [1] since graphs are general and powerful data structures that can be used to represent diverse types of objects. Several authors have developed graph-based techniques and methods for satisfying the need of converting large volumes of data into useful information [2]. The frequent approximate subgraph (FAS) discovery is an example of such techniques [3,4,5]. These techniques have become important topics in mining tasks where the mined patterns are detected taking into account distortions in the data.

It is common to model complex data with the help of graphs consisting of nodes and edges that are often labeled to store additional information. Representing these data in the form of a graph can help visualise relationships between entities which are convoluted when captured in relational tables.Graph mining ranges from indexing graphs, finding frequent patterns, finding inexact matches, graph partitioning and so on. The subgraphs are categorized into various classes, and the approaches of graph based data mining strongly depend on the targeted class.

Graph invariants provide an important mathematical criterion to efficiently reduce the search space of the targeted graph structures in some approaches. Furthermore, the mining measures define the characteristics of the patterns to be mined similar to conventional data mining.Knowledge Discovery in Databases (KDD) is a systematic and automatic process to analyse and discover hidden knowledge (or patterns) from databases. The general process of KDD commences with the acquisition of relevant data, followed by preprocessing,feature extraction, patterns discovery and finally communication of the identified patterns to the user. A large number of different techniques have been developed to perform KDD. The work described in this paper is focused on classification, specifically the classification of image data (the term image mining will be used throughout this paper to indicate the application of data mining techniques to image data, and the term image classification to indicate the application of classification techniques to image data).

The basic outline of this paper is as follows.Section 2 provides some basic concepts and related works.The graph-based image representation is presented in Section 3. The framework for image classification is explained in Section 4 and the experimental results in a real image collection are discussed in Section 5. Finally, conclusions of the research and some ideas about future directions are exposed in Section 6.

2 BACKGROUND

In this section, we start by providing the background knowledge and notation used in the following sections.

2.1 Basic Concepts

This work is focused on simple undirected labeled graphs; henceforth, when we refer to graph we assume this type of graph. Before presenting their formal definition, we will define the domain of labels.

Definition 1 (labeled graph) A labeled graph can be represented as $G = (V, E, L_V, L_E, u)$, where V is a set whose elements are called vertices, $E \subseteq V \times V$ is a set of edges; L_V *and* L_E are vertex and edge labels respectively.

Definition 2 (sub graph) Let $G1 = (V_1, E_1, L_{V1}, L_{E1}, u_1)$ and $G2 = (V_2, E_2, L_{V2}, L_{E2}, u_2)$ be two graphs, we say that $G1$ is a subgraph of $G2$ iff $V1 \subseteq V2$, $\forall u \in V1$, $u_1(v) = u_2(v)$, and $E1 \subseteq E2$, and $\forall (u,v) \in E1$, $u_1(u,v) = u_2(u,v)$. In this case, we use the notation $G1 \subseteq G2$.G2 is also a super graph of G_1.

Definition 3 (Adjacency Matrix) Adjacency Matrix is a 2D array of size V x V where V is the number of vertices in a graph. Let the 2D array be adj[][], a slot adj[i][j] = 1 indicates that there is an edge from vertex i to vertex j. Adjacency matrix for undirected graph is always symmetric.

2.2 Related work

Graph classification has received lots of attention because of its wide range of applicability to real-world data such as biological data, chemical compound data, and semi structured data to name a few.In most of these approaches a graph is represented using various descriptors and a classification model is built using statistical or machine learning techniques. Karypis et al.[6] use similar approach by generating frequent sub structure based descriptors using a frequent subgraph mining algorithm and selecting the best substructures to define the feature vectors, which is subsequently used by Support Vector Machine (SVM) classifiers to build the classification model.

Han et al. [7] showed that frequent closed graphs based descriptor space is a better approach than the frequent subgraphs based descriptors, and it generates typically lower dimensional feature vectors. A number of methods have been proposed in recent years using cyclic patterns [8], acyclic, path and tree fragments [9] to define the descriptor space and to generate features. With this feature based representation any classification technique can be used for the classification task.

3.PROPOSED FRAMEWORK

We take a graph based approach to object classification.Let G=(V,E) be an undirected graph with vertices V and edges E. Generally speaking, image mining aims to discover implicit patterns among image databases. The fundamental issue of image mining is how to use low-level (primitive) representations to extract high-level hidden patterns with efficiency and effectiveness. Representing images as graphs can maintain the structural information of images.

There are a number of techniques for graph based image representation. The main idea of graph based image representation is that the regions of the image, which contain similar properties, are denoted by graph nodes, and the relations between different regions are denoted by graph edges. The node and edge attributes usually describe the characteristics of that region and the relation between regions respectively. The flowchart of the Graph based image classification framework is illustrated in Figure 1.

A quad-tree [10] is a widely used tree structure for representing images. The fundamental idea behind the quad-tree is that any image can be divided into four quadrants. Each quadrant can be further split into four equal-sized sub quadrants (NW, NE, SW and SE), and so on. The quad-tree decomposition is based on recursively subdividing the sub-images until the imposed limit is met. In this paper,we use a quad-tree to represent each image.

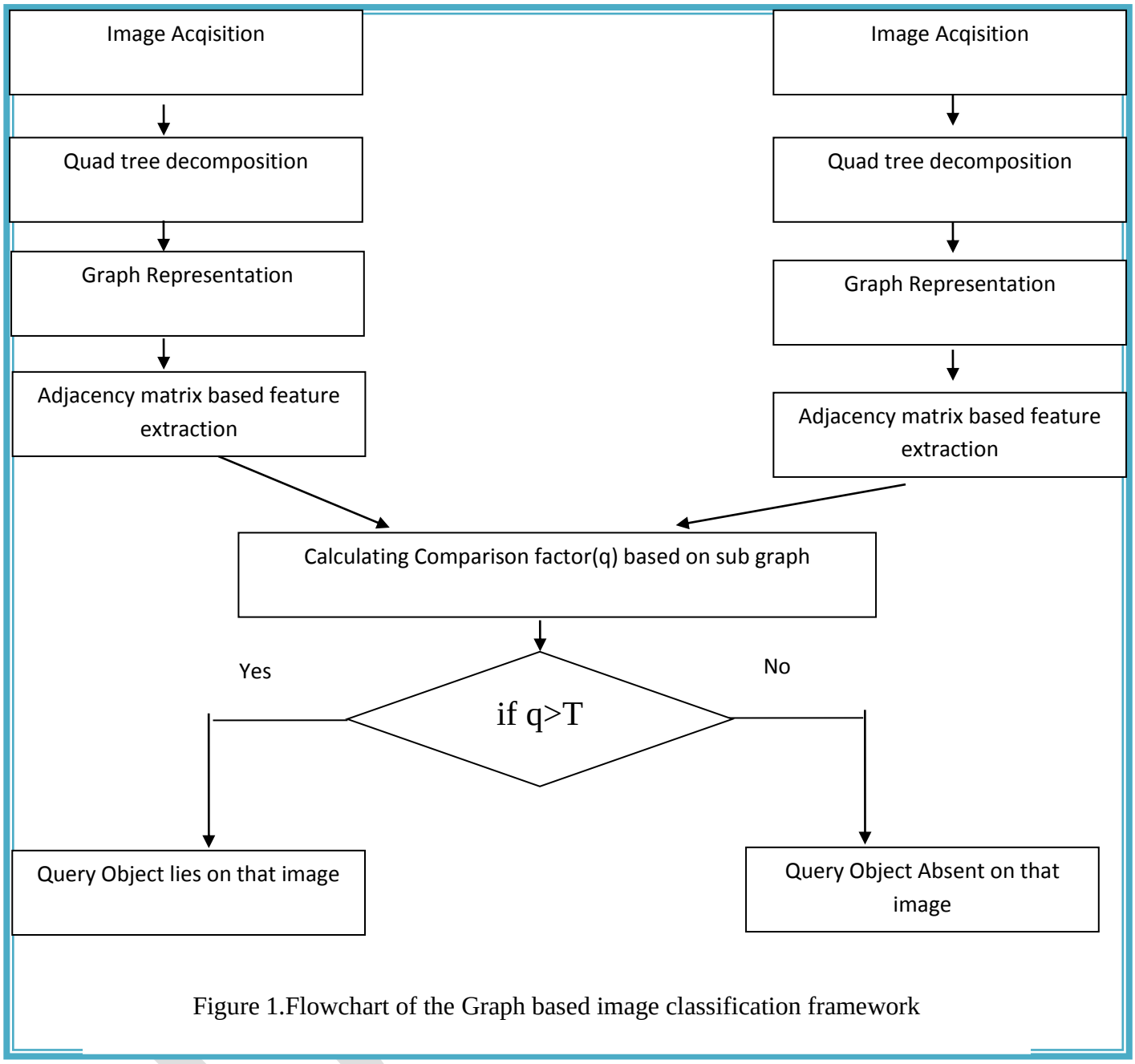

Figure 1.Flowchart of the Graph based image classification framework

The algorithm mainly comprises of the following steps:

- Initial step is the Quad tree decomposition of the acquired image .
- After that the graph is constructed based on the Adjacency matrix(adj[][]). An entry adj[i][j]= 1 in the respective matrix indicates that there is an edge from vertex i to vertex j.Otherwise the entry is 0.
- Calculation of comparison factor(q) based on the subgraph is the decision making step.
- If the comparison factor(q)>threshold(T) then the Object lies on that image,else the Object is absent on that image.

4.EXPERIMENTAL RESULTS

The graph based Object classification method is tested on the acquired sample image and Figure 2 shows the output of the proposed system. In subgraph mining the goal is to develop algorithms to discover frequently occurring subgraphs in the graph database. Although there are many efficient and scalable frequent pattern mining algorithms exist for itemset mining and sequence mining, our method achieve an efficient and scalable algorithm for subgraph mining.

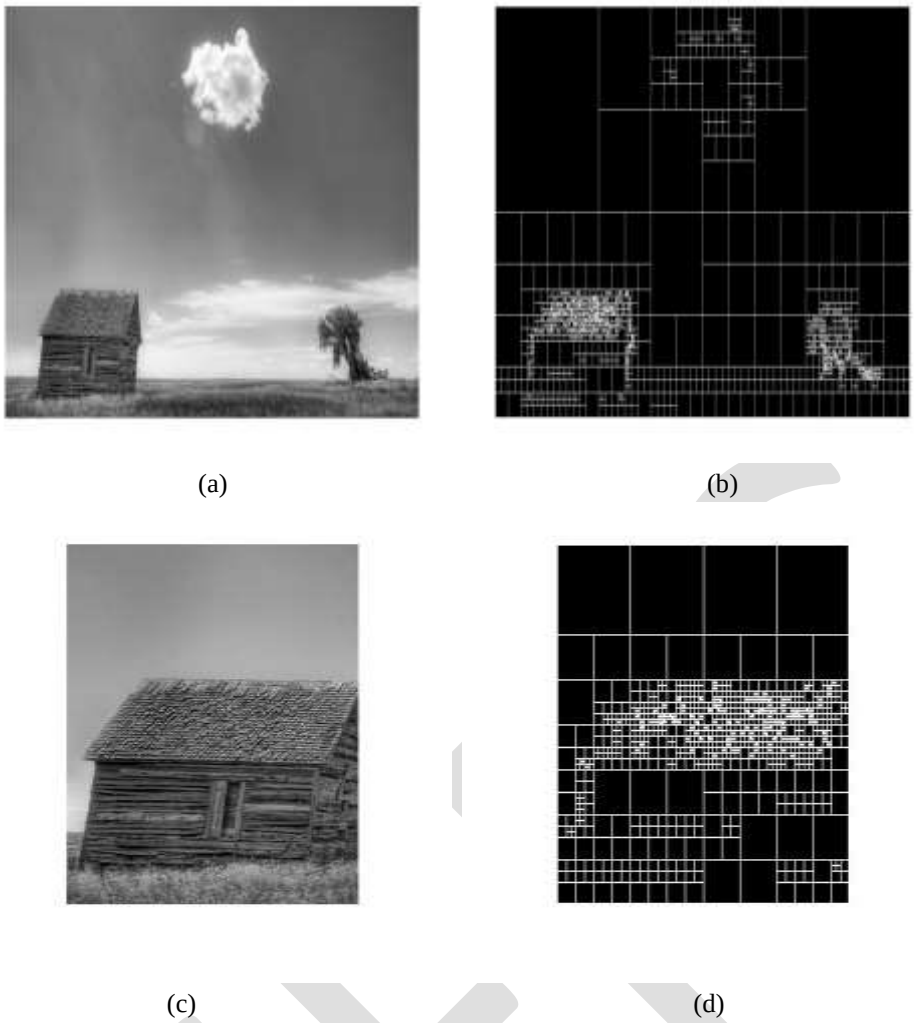

Figure 2.(a) Original image, (b) Quad tree of the Original image, (c) Query image, (d) Quad tree of the Query image

5. CONCLUSION

This paper presents an efficient approach for graph based Object Classification on Images. We presented an algorithm to find the pattern present in an image. We have shown how the proposed method finds relevant pattern using the comparison factor of the image. Future work will focus on making the presented approach more meaningful for the underlying application in which complex object can be classified.

REFERENCES:

[1] J. Friedman, T. Hastie, and Robert Tibshirani. Additive logistic regression: A statistical view of boosting. *The Annals of Statistics*, 38(2):337–374, April 2000.

[2] A. Berger. The Improved Iterative Scaling Algorithm: A gentle Introduction. Technical report, Carnegie Mellon University, 1997.

[3] A.L. Berger, V.J. Della Pietra, and S.A. Della Pietra. Maximum Entropy Approach for Natural Language Processing. Compu*tational Linguistics.*22(1) P: 39 – 71,1996.

[4] H. Mannila, and P. Smyth. Prediction with Local Patterns using Cross-Entropy. *In Proc. of the fifth ACM SIGKDD*, P: 357 – 361, ISBN:1-58113-143-7, 1999.

[5] K. Nigam, J. Lafferty, and A. Maccallum. Using Maximum Entropy for Text Classification. *IJCAI-99 Workshop on Machine Learning for Information*, 1999.

[6] M. Deshpande, M. Kuramochi, Nikil Wale, and G. Karypis, Frequent Substructure-Based Approaches for Classifying Chemical Compounds, *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 8, pp.1036-1050, Aug., 2005.

[7] C. Liu, X. Yan, H. Yu, J. Han, and P. S. Yu. Mining Behavior Graphs for Backtrace of Noncrashing Bugs. *In Proc. of SIAM Int. Conf. on Data Mining* (SDM'05), 2005.

[8] T. Horvath, T. Grtner, and S. Wrobel. Cyclic pattern kernels for predictive graph mining. *In Proc. of the 10th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 158-167, 2004

[9] N. Wale and G. Karypis. Acyclic Subgraph-based Descriptor Spaces for Chemical Compound Retrieval and Classification. *In Proc of IEEE International Conference on Data Mining* (ICDM), 2006.

[10] R.A. Finkel, J.L. Bentley, Quad trees: a data structure for retrieval on composite keys, Acta Informatica 4 (1974) 1–9.

[11] Y. Jia, J. Zhang, J. Huan, An efficient graph-mining method for complicated and noisy data with real-world applications, Knowledge and Information Systems 28 (2011) 423–447.

[12] C. Jiang, F. Coenen, Graph-based image classification by weighting scheme, in: Proceedings of the Artificial Intelligence, Springer, Heidelberg, 2008, pp. 63– 76.

[13] C. Jiang, F. Coenen, R. Sanderson, M. Zito, Text classification using graph mining-based feature extraction, Knowledge-Based Systems 23 (2010) 302–308.

[14] C. Jiang, F. Coenen, M. Zito, Frequent sub-graph mining on edge weighted graphs, in: Proceedings of the 12th international conference on Data warehousing and knowledge discovery, Springer-Verlag, Berlin, Heidelberg, 2010, pp. 77–88.

[15] A. Jiménez, F. Berzal, J.C.C. Talavera, Frequent tree pattern mining: a survey,Intelligence Data Analysis 14 (2010) 603–622. Statics of Large Networks, 2008.