Abstract

Subgraph matching is a challenging task, specially due to the presence of outliers most of the graph matching algorithms do not perform well in subgraph matching scenario. But subgraph matching is a very important task to be solved because any kind of visual object or data can be represented by graphs and then any part of that object can be detected or recognized by subgraph matching. On the other hand exact subgraph isomorphism has proven to be an NP-complete problem. So naturally, in graph matching community, there are lot of efforts addressing the problem of subgraph matching. And most of them work with approximate algorithms that try to get an inexact solution in approximate way. So, theoretically this thesis proposes algorithms for solving subgraph matching in an approximate and inexact way.

We consider the symbol spotting problem on graphical documents or line drawings from application point of view. This is a well know problem in the graphics recognition community. It can be further applied for indexing and classification of documents based on their contents. The structural nature of this kind of documents easily motivates one for giving a graph representation. So the symbol spotting problem on graphical documents can be considered as a subgraph matching problem. The main challenges in this application domain is the noise and distortions that can come during the usage, digitalization and raster to vector conversion of those documents. Apart from that graphics recognition nowadays is not any more confined within a limited number of images. So dealing a huge number of images with graph based method is a further challenge.

So in this thesis we have worked on efficient and robust graph representation to cope with the noise and distortions coming from the documents. On the other hand, we have also worked on different graph based methods and framework to solve the subgraph matching problem in a better approximated way, which can also deal with considerable number of images. Firstly, we propose a symbol spotting method by hashing the serialized subgraphs. The graph serialization allows one to create factorized substructures, which can be organized in hash tables depending on the structural similarities. The involvement of hashing techniques helps to reduce the search space substantially and speeds up the spotting procedure. Secondly, we introduce contextual similarities based on the walk based propagation on tensor product graph. These contextual similarities involve higher order information and more reliable than pairwise similarities. We use these higher order similarities to formulate the subgraph matching as a node and edge selection problem in the tensor product graph. Thirdly, we propose near convex region adjacency graph to eliminate the difficulties of tra-

ditional region adjacency graph representation. Fourthly, we propose a hierarchical graph representation by simplifying/correcting the structural errors to create a hierarchical graph of the base graph. Later these hierarchical graph structures are matched with some graph matching methods. Apart from that, in this thesis we have provided an overall experimental comparison of all the methods and some of the state-of-the-art methods. Furthermore, some dataset models are proposed to show some strengths of some algorithm.