ABSTRACT

Nearest neighbour search is a fundamental geometric problem important in variety of applications including data mining, machine learning, pattern recognition, computer vision, graphics, statistics, bioinformatics and location based services. Applications of the nearest neighbour problem are particularly motivated by problems like reconstruction, visualization, and simplification for geometry processing applications. With such a wide ranging impact, it becomes critical to make algorithms for finding nearest neighbours as efficient as possible.

Indexing of multi-dimensional data has been the focus of a considerable amount of research effort over many years but no generally agreed paradigm has emerged to compare with the impact of the B-Tree, for example, on the indexing of one-dimensional data. At the same time, the need for efficient methods is ever more important in an environment where databases become larger and more complex in their structures. This thesis describes the HCQT framework which has an efficient storage method which makes answering nearest neighbour query faster.

Mapping spatial data to one dimension, thus enabling one-dimensional access methods to be exploited, has been suggested earlier but for the most part interest has been confined to the Peano curve. The possibility of using other curve namely the Hilbert curve, whose characteristics differ from those of the Peano curve, has been used. The modified encoding and decoding algorithms from spatial data to one have been proposed. The results are very much encouraging for transportation network.

Data are indexed using a region quadtree which remains compact, regardless of the volume. In this thesis we design and implement a working file store which is underpinned by the principle of mapping spatial data to one dimension by Hilbert curve. The implementation has entailed developing algorithms for mapping data to one dimension and, most importantly, developing algorithms to facilitate the querying of data in a flexible way.
The implementation is competitive with the best approximate nearest neighbour searching. We report on preliminary testing of the implementation, which provides very encouraging results. We also undertake a brief exploration of the application of Hilbert curve to the indexing of spatial data. This framework can be used any dataset without making any changes in query processing.