Dictionary Design Algorithms for Vector Map Compression

Shashi Shekhar, Yan Huang, Judy Djugash

Computer Science Department, University of Minnesota
[*shekhar*, *huangyan*, *djugash*]@cs.umn.edu

The enormous size of vector maps and limited storage available in hand-held devices motivate the need for data compression techniques. Compression techniques for vector maps can allow PDAs to carry larger subsets of vector maps or free-up memory for other datasets and can also reduce the communication cost of downloading new maps to the PDA, possibly over low-bandwidth wireless channels (e.g. beaming, cell phone modems). Compression schemes for vector maps should allow simple decoding schemes due to the limited computational resources of popular PDAs making dictionary-based compression techniques attractive as long as the error of approximation can be controlled.

A static dictionary does not consider the data distribution (slopes and offsets of segments in the curves) but rather uses a collection of line segments organized by a set of squares. If the distribution of line segments in a dataset does not match well with the dictionary entries, the errors of approximation can be large. In such a scenario, an alternative method which could incorporate the data distribution in the dictionary design would achieve better spatial accuracy than the static dictionary method. We propose the use of clustering techniques (e.g. K-mean clustering) to identify dictionary entries while minimizing errors of approximation for locations of spatial objects in the map. Vectors relative to the first node of a road or relative to the previous node of a road are feed into clustering algorithms. Clustering algorithms take as input a fixed number and generates that many clusters for the given dataset as output. The cluster centroids obtained becomes our dictionary. Based on this dictionary, we encode the vector dataset that we obtained earlier. Since each vector would now be assigned to a particular cluster, that vector would now be represented in terms of a reference to that cluster's centroid entry in the dictionary. We formally show that this proposed dictionary construction approach often yields a lower error of approximation than the error from conventional fixed dictionary techniques. Experimental results with a road map representing the major US Highways confirm the superiority of the proposed method in yielding lower errors of approximations for a fixed size dictionary [1].

In our future work we would like to experiment on different differential vector schemes. The general purpose clustering algorithm aims at minimizing the total square error. When a user tolerant accuracy level for each road segment is specified, the clustering algorithm needs to be modified to meet this requirement.

References

[1] S. Shekhar, Y. Huang, and J. Djugash. Dictionary Design Algorithms for Vector Map Compression. *University of Minnesota Technical Report.*

*Supported in part by the Army High Performance Computing Research Center under the auspices of the Department of the Army, Army Research Laboratory Cooperative agreement number DAAH04-95-2-0003/contract number DAAH04-95-C-0008.