The well-separated pair decomposition and its application for fast point of view based clustering

Norbert Bus
Consultant: Dr. Katalin Friedl

The well-separated pair decomposition is a well-known data structure to compactly represent all pairwise distances between points, it is an ideal tool for solving many distance related problems. Without giving a complete list such problems include finding the closest pair in a point set, calculating the approximate diameter of a point set, the all nearest neighbors problem, geometric spanners, N-body simulation and clustering. It was first introduced by Callahan and Kosaraju in 1995. The original work only considered Euclidean spaces. Recently there has been considerable work in extending this notion to more general metric spaces. In this study we review the metric spaces that are known to have a provably good well-separated pair decomposition, the latter meaning that mainly we are concerned about the size of the data structure since this shows how compactly it can store information about the distances occurring in the point set. Surprisingly the results show that even though there is a quadratical number of different distances in a point set a data structure of near linear size is sufficient. To the best of our knowledge the only metric spaces considered so far are the metrics induced by a set of unit disks or balls in Euclidean spaces and the so called doubling metric. Both metrics have many applications in the study of e.g., networks. Our focus is on the theoretical results concerning the existence of the well-separated pair decomposition and its size. We review the algorithms to construct it in the aforementioned metric spaces. One can extend the data structure in certain cases to handle dynamic operations, e.g., adding a new point. The results considering dynamic maintenance are omitted from the current work. Indeed the dynamic maintenance problems of the well-separated pair decomposition have only been considered for Euclidean spaces and for other metrics these are open questions. Finally we show several problems that are efficiently solvable using small sized well-separated pair decompositions, but less emphasis is given to these applications.