Project Proposal
DJ Oneil      Esten Rye

1. Problem Statement

Given: Spatial database with \( n \) objects and query point, \( q \).

Find: The \( k \leq n \) ranked nearest neighbors.

Objective: Use minimum bounding boxes, a polygon intersection method, and extensible heuristic model for faster output and more accurate approximations than previous methods.

Constraints: Spatial objects are stored in an R-Tree and there is already a fast distance function defined. Also heuristics will have to adhere to a template that we will define.

1.1 Contributions

- Surveys existing, state-of-the-art approaches to the kNN problem. In our method, \( k \) does not have to be predefined and the algorithm is incremental.
- Proposes new method that combines minimum bounding boxes and polygon intersection techniques.
- Incorporates an extensible heuristic model. This will allow the user to add further heuristics depending on the search type.

1.2 Related Work

The following works represent the state-of-the-art on kNN. None of these incorporate a general heuristic model or consider other ways besides retraversing or backtracking through the R-tree for faster candidate node searching.

  - Describes an incremental kNN algorithm on an R-tree.
  - The major contribution of this paper is that it uses a priority queue for best-first-search for smarter retraversal when \( k \) is not fixed.

  - Compares two algorithms for using an R-tree to get the \( k \) nearest neighbors; one uses minimum bounding boxes in an R-tree and the other adds the constraint that the minimum bounding boxes must be disjoint.

  - The multi-step algorithm is intended for cases where there is a high-dimensional search space. The idea is to create a smaller (more accurate) set of candidates during the filter step, and then finish the search in a refinement step.
The shortcomings of the state-of-the-art knn search (1996) are discussed. We will also consider using the following paper on polygon intersection for determining overlapping boxes in the R-tree. However, we might add to the R-tree constraints so we can use a simpler algorithm.

  - Proposes an approximation technique for spatial queries.
  - Discusses some of the drawbacks of current uses of histograms and/or sampling.

2. Methodology
We plan to do a paper survey to first illustrate where current algorithms are lacking, and second find where they can be improved. We will propose an improved algorithm and illustrate why we believe the changes or additions would be an improvement over the existing algorithm. If time allows, we will attempt to implement our proposed algorithm and compare its performance to the brute force method.

3. Next Steps
1. Review existing algorithms for areas of improvement.
2. Consider the shortcomings of the existing algorithms and develop an improved algorithm to address them.
3. Search for existing software packages that could be used or extended for our proposed algorithm.
4. Evaluate if it is feasible to complete an implementation in the time allowed.
5. If it is found to be feasible to implement:
   a) Implement the algorithm
   b) Compare its performance with other implementations.
6. If it is found to not to be feasible to implement:
   a) Propose which software packages could be extended
   b) Provide pseudo code for the algorithm.