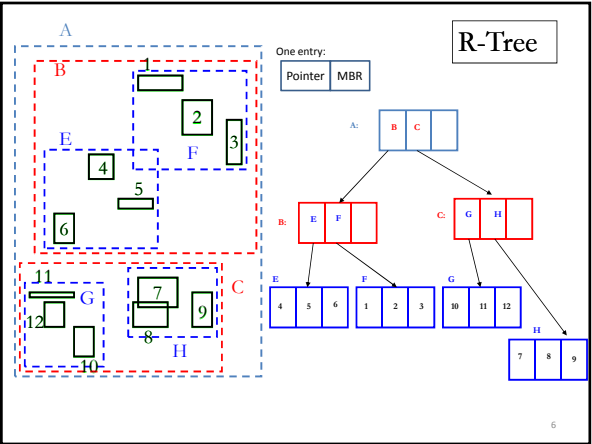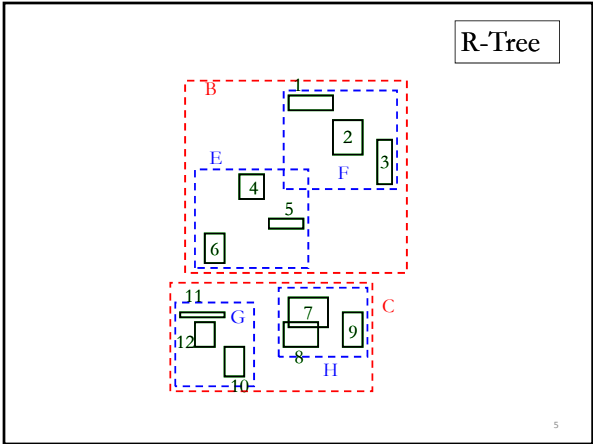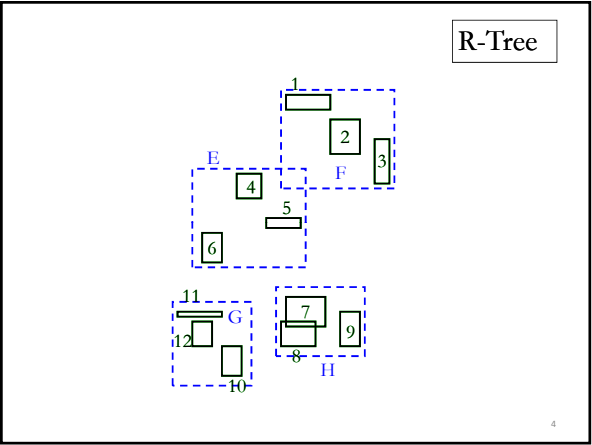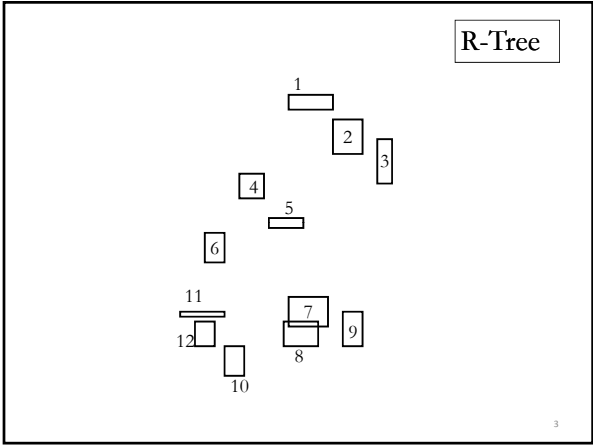# Nearest Neighbor Queries

Nick Roussopoulos, Stephen Kelly and Frédéric Vincent

Ling Hu
lingh@usc.edu

Outline

- Spatial Index – R-Tree
- NN query – Naïve solution
- A better solution – Branch-and-bound
- Can we do better?
- Experiment Results

2

R-Tree

3

R-Tree

4

R-Tree

5

R-Tree

6

Outline

- Spatial Index – R-Tree
- NN query – Naïve solution
- A better solution – Branch-and-bound
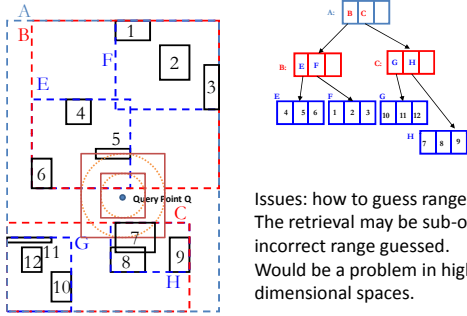- Can we do better?
- Experiment Results

7

Nearest Neighbor Search

- Retrieve the nearest neighbor of query point Q
- Simple Strategy:
  - convert the nearest neighbor search to range search.
  - Guess a range around Q that contains at least one object say O
    - if the current guess does not include any answers, increase range size until an object found.
  - Compute distance d' between Q and O
  - re-execute the range query with the distance d' around Q.
  - Compute distance of Q from each retrieved object. The object at minimum distance is the nearest neighbor!!!

8

Naïve Approach

Issues: how to guess range?
The retrieval may be sub-optimal if incorrect range guessed.
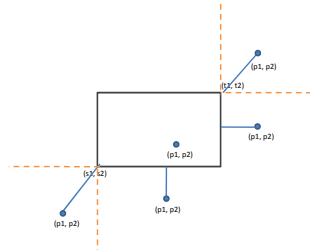Would be a problem in high dimensional spaces.

Outline

- Spatial Index – R-Tree
- NN query – Naïve solution
- A better solution – Branch-and-bound
- Can we do better?
- Experiment Results

10

MINDIST Property

- MINDIST is a lower bound of any k-NN distance

$$\forall o \in O, \mathrm{MINDIST}(Q,R) \leq \|(Q,o)\|$$

11

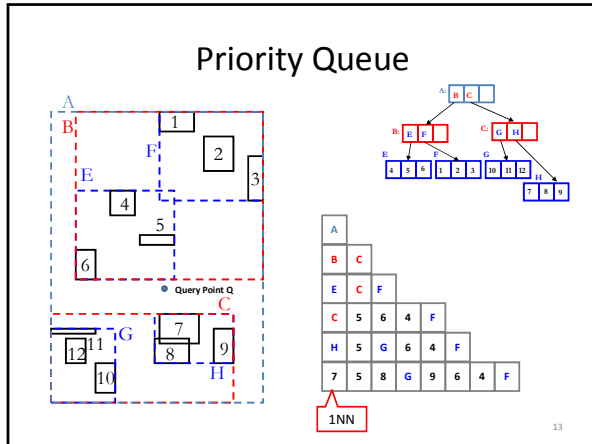## A Better Strategy for KNN search

- A sorted priority queue based on MINDIST;
- Nodes traversed in order;
- Stops when there is an object at the top of the queue; (1-NN found)
- k-NN can be computed incrementally;

I/O optimal

12

Priority Queue

13

Outline

- Spatial Index – R-Tree
- NN query – Naïve solution
- A better solution – Branch-and-bound
- Can we do better?
- Experiment Results

14

MBR Face Property

- MBR is an n-dimensional Minimal Bounding Rectangle used in R trees, which is the minimal bounding n-dimensional rectangle bounds its corresponding objects.
- MBR face property: Every face of any MBR contains at least one point of some object in the database.

15

MBR Face Property – 2D

16

MBR Face Property – 3D

Rectangle R

17
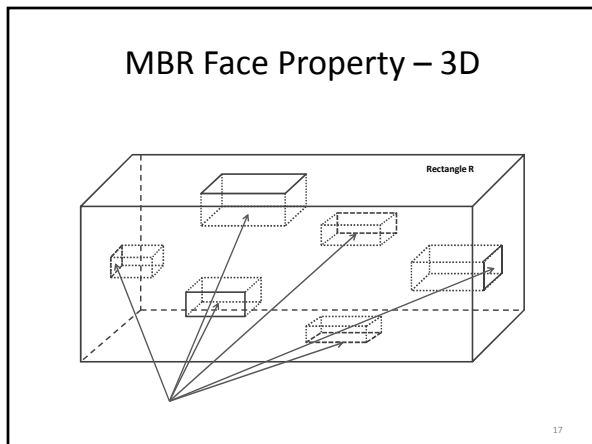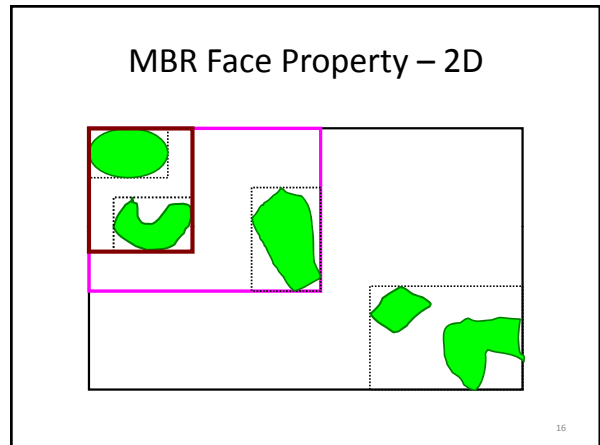
Improving the KNN Algorithm

- While the MinDist based algorithm is I/O optimal, its performance may be further improved by pruning nodes from the priority queue.

18

3

Properties of MINMAXDIST

- *MINMAXDIST(P,R)* is the minimum over all dimensions distances from *P* to the furthest point of the closest face of R.
- *MINMAXDIST* is the smallest possible upper bound of distances from the point *P* to the rectangle *R*.
- *MINMAXDIST* guarantees there is an object within the R at a distance to P less than or equal to it.

$$\exists o \in O, \|(Q,o)\| \leq \text{MINMAXDIST}(Q,R)$$

- MINMAXDIST is an upper bound of the 1-NN distance

19

MINDIST & MINMAXDIST

MINDIST(P,R) <= NN(P) <= MINMAXDIST(P,R)

20

MinDist & MinMaxDist – 3D

21

Pruning 1

- Downward pruning: An MBR R is discarded

If there exists another R' such that MINDIST(P,R)> MINMAXDIST(P,R')

22

Pruning 2

- Downward pruning: An object O is discarded

If there exists an R such that Actual_dist(P,O) > MINMAXDIST(P,R)

23

Pruning 3

- Upward pruning: An MBR R is discarded

If an object O is found such that MINDIST(P,R) > Actual_dist(P,O)

24

MINDIST vs MINMAXDIST Ordering

- MINDIST: optimistic
- MINMAXDIST: pessimistic

- Example: MINDIST ordering finds the 1-NN first

25

MINDIST vs MINMAXDIST Ordering

- Example: MINMAXDIST ordering finds the 1-NN first

26

Generalize to k-NN

- Keep a sorted buffer of at most k current nearest neighbors
- Pruning is done according to the distance of the furthest nearest neighbor in this buffer
- Example:

27

Outline

- Spatial Index – R-Tree
- NN Query – Intuitive Solutions
- Optimized NN Query – branch-and-bound
- Experiment Results

Database Size Scalability - Synthetic Data
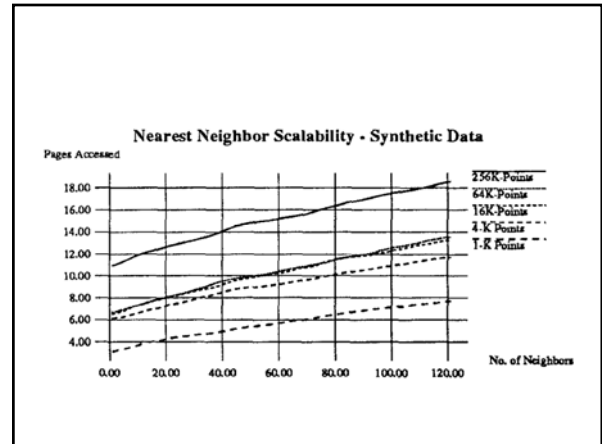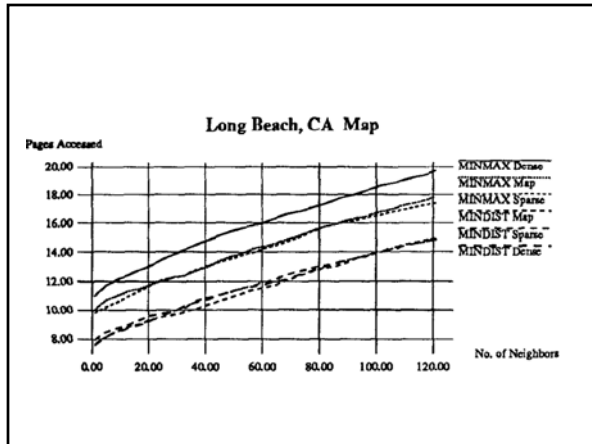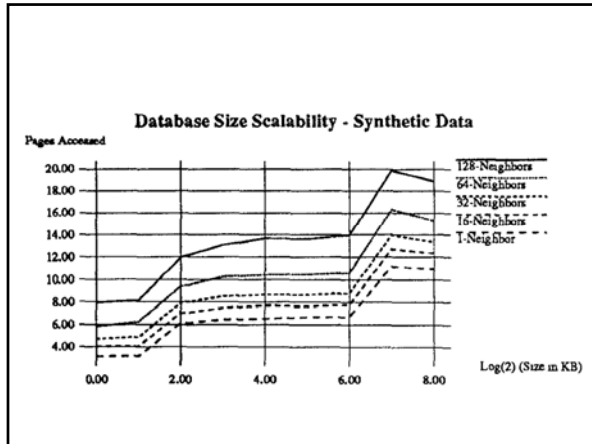
Key Insights

- # of pages accessed grows when k grews;
- The denser the dataset, the more page access;
- MinDist v.s. MinMaxDist: same in shape, but MinMaxDist has more I/O cost;
- In Dense area, MinMaxDist is bad;

Thanks & Questions ?