

# An Empirical Analysis of Techniques for Constructing and Searching K- Dimensional Trees

---

## Authors

**Douglas A. Talbert and Doug Fisher**

---

## Presenters

Manish Pandya

Hemin Desai

Chirag Shah

---

# Overview

---

- Introduction to the problem
- Key contribution
- Relevant prior work
- Methodology
- Results
- Our opinion

# Problem

---

- ◆ The kd-tree nearest neighbor searches with respect to several independent variables is costly.
- ◆ Two fundamental problems
  - Multiple techniques of kd-tree construction
  - Identifying various ways to reduce the search and its subsequent computations

# Key Contributions

- ◆ Analysis of kd-tree construction
  - Split point selection algorithms
  - Maximum number of data items in a leaf
- ◆ Analysis of kd-tree search techniques
  - Pruning (strong and weak)
  - Stopping search early
  - Reduced distance calculations

# Related Work

- ◆ Basic kd-tree ( multi-dimensional binary search trees ) construction algorithm
  - Selection of a split point
  - Partitioning the data
  - Repeating until the size of each node is less than or equal to the threshold value

contd...

# Related Work

## ◆ Search Algorithm

- Starting at the root, use the values of search vector to follow the branches of kd-tree to an initial leaf.
- Compute the distance to each data item at the leaf; store the closest item and its distance.
- Continue searching the tree in a depth-first fashion
- Whenever a leaf is encountered, compute the distances to each data item at the leaf and update the nearest-neighbor information

# Methodology

## ◆ Construction

- Different leaf sizes – 1, 2, 5, 10, 20, 50
- Various split point algorithms
  - \* median
  - \* arithmetic mean
  - \* harmonic mean
  - \* inner quartile mean

contd...

# Methodology

-- The split attribute is the one with the broadest range at each node.

## ◆ Searching

-- Pruning techniques

\* Strong pruning

\* Weak pruning

\* Stopping search early

\* Reduced distance calculations

\* Hybrid pruning

# Results

- ◆ Arithmetic mean results in the lowest number of nodes examined during search for both weak and strong pruning
- ◆ Arithmetic mean results in the lowest cost in number of single dimension comparisons
- ◆ Hybrid pruning results in a lower cost in terms of dimensions than strong pruning alone
- ◆ With increasing leaf size, the cost in dimensions is less for hybrid pruning than weak pruning.
- ◆ For both weak and hybrid pruning, using reduced distance calculations lessens the cost in dimensions.

# Our Opinion

- ◆ This paper focuses only on the analysis of techniques for constructing and searching kd-trees.
- ◆ The authors have contributed significantly in improving the efficiency and performance gains for finding the nearest neighbor, however the fundamental algorithms for kd-tree construction and searching were already developed.
- ◆ We rate this paper 7 on a scale of 1 to 10.