

# Search in high dimensions: some surprising results

Bradford G. Nickerson

Faculty of Computer Science, University of New Brunswick  
Fredericton, N.B., Canada, E3B 5A3  
bgn@unb.ca

May 22, 2007

## Abstract

Some recent results on high dimensional search are presented. We define  $d$ -dimensional search is high when  $d$  approaches  $\log n$ , for  $n$  = number of points or objects to be searched. For example, the  $\ell$ -level  $k$ -range is reported to have orthogonal range time  $Q(n, d) = O(\log n + A)$  for  $A$  = number of points or objects in range. Our results show that the  $\ell$ -level  $k$ -range requires  $Q(n, d, \ell) = O((2\ell)^{(d-1)}(\log N + A))$  time for orthogonal range search, making it impractical for range search, even for relatively low  $d$ . For  $d$ -dimensional point data, the venerable  $k$ -d tree is found to be competitive with the Patricia trie adapted for  $d$ -dimensional search. For large  $d$ , we present a technique based on the pyramid technique that we call the PKD-tree. The PKD-tree shows good performance in testing with uniformly distributed random data points ( $n \leq 1,000,000$  and  $d \leq 100$ ) and with 68,040 32- $d$  data points from a colour histogram dataset.

We adapted the pyramid technique to implement a  $k$ -nearest neighbour algorithm called the decreasing radius or DR pyramid technique. Results indicate that for uniformly distributed random data, the DR pyramid and BBD-tree algorithms are comparable. For  $d \geq 16$ , we discovered that a naive (brute force) search was faster than six other algorithms for  $k$ -nearest neighbour search. The talk presents some observations about why efficient search in high dimensions is challenging.