Foundations of Multidimensional and Metric Data Structures

Hanan Samet
University of Maryland, College Park

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Morgan Kaufmann Publishers is an imprint of Elsevier

MORGAN KAUFMANN PUBLISHERS

Contents

Foreword xv

Preface xvii

1 Multidimensional Point Data 1

1.1 Introduction 5
1.2 Range Trees 14
1.3 Priority Search Trees 19
1.4 Quadtrees 28
1.4.1 Point Quadtrees 28
1.4.1.1 Insertion 28
1.4.1.2 Deletion 31
1.4.1.3 Search 36
1.4.2 Trie-Based Quadtrees 37
1.4.2.1 MX Quadtree 38
1.4.2.2 PR Quadtrees 42
1.4.3 Comparison of Point and Trie-Based Quadtrees 47
1.5 K-d Trees 48
1.5.1 Point K-d Trees 50
1.5.1.1 Insertion 50
1.5.1.2 Deletion 52
1.5.1.3 Search 55
1.5.1.4 Point K-d Tree Variants 57
1.5.2 Trie-Based K-d Trees 70
1.5.2.1 PR K-d Tree 71
1.5.2.2 Sliding-Midpoint K-d Tree 72
1.5.2.3 Bucket PR K-d Tree and PMR K-d Tree 74
1.5.2.4 Path-Compressed PR K-d Tree 75
1.5.2.5 Path-Level Compressed PR K-d Tree 77
1.5.2.6[1] BD-Trees 79
1.5.2.7 Balanced Box-Decomposition Tree (BBD-Tree) 83
1.5.3 Conjugation Tree 88
1.6 One-Dimensional Orderings 90
1.7 Bucket Methods 95
1.7.1 Tree Directory Methods 96
1.7.1.1 K-d-B-Tree 98
1.7.1.2 Hybrid Tree 101
1.7.1.3 LSD Tree 102
1.7.1.4 hB-Tree 106
1.7.1.5 K-d-B-Tries 110
1.7.1.6 BV-Tree 116
1.7.1.6.1 Definition 117
1.7.1.6.2 Building a BV-Tree 120

1.7.1.6.3 Searching a BV-Tree . . . 122
1.7.1.6.4 Performance . . . 125
1.7.1.6.5 Summary . . . 127
1.7.1.7 Static Methods . . . 129
1.7.2 Grid Directory Methods . . . 130
1.7.2.1 Grid File . . . 131
1.7.2.2 EXCELL . . . 137
1.7.2.3 Linear Hashing . . . 140
1.7.2.3.1 Adapting Linear Hashing to Multidimensional Data . . . 140
1.7.2.3.2 Multidimensional Extendible Hashing (MDEH) . . . 146
1.7.2.3.3 Dynamic Z Hashing . . . 150
1.7.2.3.4 Linear Hashing with Partial Expansions (LHPE) . . . 152
1.7.2.4 Alternative Implementations of Linear Hashing . . . 152
1.7.2.4.1 Quantile Hashing . . . 153
1.7.2.4.2 Piecewise Linear Order Preserving (PLOP) Hashing . . . 154
1.7.2.4.3 Spiral Hashing . . . 156
1.7.2.4.4 Summary . . . 159
1.7.2.5 Comparison . . . 160
1.7.3 Storage Utilization . . . 163
1.8 PK-Trees . . . 164
1.8.1 Motivation . . . 165
1.8.2 Overview . . . 167
1.8.3 Definition . . . 169
1.8.4 Comparison with Bucket Methods . . . 170
1.8.5 Operations . . . 171
1.8.5.1 Representation and Searching . . . 172
1.8.5.2 Insertion . . . 172
1.8.5.3 Deletion . . . 178
1.8.6 Discussion . . . 181
1.9 Conclusion . . . 184

2 Object-Based and Image-Based Image Representations . 191

2.1 Interior-Based Representations . . . 193
2.1.1 Unit-Size Cells . . . 194
2.1.1.1 Cell Shapes and Tilings . . . 196
2.1.1.2 Ordering Space . . . 199
2.1.1.3 Array Access Structures . . . 202
2.1.1.4 Tree Access Structures . . . 203
2.1.2 Blocks . . . 204
2.1.2.1 Decomposition into Arbitrarily Sized Rectangular Blocks . . . 205
2.1.2.2 Medial Axis Transformation (MAT) . . . 208
2.1.2.3 Irregular Grid . . . 210
2.1.2.4 Region Quadtree and Region Octree . . . 211
2.1.2.5 ATree . . . 220
2.1.2.6 Bintree . . . 221
2.1.2.7 Adaptation of the Point Quadtree for Regions . . . 222
2.1.2.8 Adaptation of the K-d Tree for Regions . . . 223
2.1.2.9 X-Y Tree, Treemap, and Puzzletree . . . 225
2.1.2.10 Hexagonal Blocks . . . 230
2.1.2.11 Triangular Blocks . . . 231
2.1.3 Nonorthogonal Blocks . . . 232
2.1.3.1 BSPTree . . . 233
2.1.3.2 K-Structure . . . 237
2.1.3.3 Separating Chain . . . 242
2.1.3.3.1 Overview . . . 242
2.1.3.3.2 Layered Dag . . . 248
2.1.4 Arbitrary Objects . . . 254
2.1.4.1 Coverage-Based Splitting . . . 255
2.1.4.2 Density-Based Splitting . . . 260
2.1.5 Hierarchical Interior-Based Representations . . . 264

2.1.5.1 Image-Based Hierarchical Interior-Based Representations (Pyramids) . . . 266
2.1.5.2 Object-Based Hierarchical Interior-Based Representations (R-trees, Bounding Box Hierarchies) . . . 270
2.1.5.2.1 Overview . . . 271
2.1.5.2.2 Ordering-Based Aggregation Techniques . . . 275
2.1.5.2.3 Extent-Based Aggregation Techniques . . . 282
2.1.5.2.4 R*-Tree . . . 289
2.1.5.2.5 Bulk Insertion . . . 296
2.1.5.2.6 Bulk Loading . . . 299
2.1.5.2.7 Shortcomings and Solutions . . . 301
2.1.5.3 Disjoint Object-Based Hierarchical Interior-Based Representations (k-D-B-Tree, R+-Tree, and Cell Tree) . . . 304
2.2 Boundary-Based Representations . . . 312
2.2.1 The Boundary Model (BRep) . . . 315
2.2.1.1 Overview . . . 315
2.2.1.2 Edge-Based Methods Using Winged Edges . . . 317
2.2.1.3 Vertex-Based and Face-Based Methods Using Laths . . . 329
2.2.1.3.1 Lath Data Structures for Manifold Objects . . . 331
2.2.1.3.2 Lath Data Structures for Meshes with Boundaries . . . 336
2.2.1.3.3 Summary . . . 341
2.2.1.4 Voronoi Diagrams, Delaunay Graphs, and Delaunay Triangulations 346
2.2.1.5 Constrained and Conforming Delaunay Triangulations . . . 350
2.2.1.6 Delaunay Tetrahedralizations . . . 352
2.2.1.7 Applications of the Winged-Edge Data Structure (Triangle Table, Corner Table) . . . 353
2.2.2 Image-Based Boundary Representations . . . 355
2.2.2.1 Line Quadtree . . . 356
2.2.2.2 MX Quadtree and MX Octree . . . 357
2.2.2.3 Edge Quadtree . . . 359
2.2.2.4 Face Octree . . . 361
2.2.2.5 Adaptively Sampled Distance Fields . . . 362
2.2.2.6 PM Quadtrees . . . 365
2.2.2.7 PM Octree . . . 369
2.2.2.8 Bucket PM Quadtree and PMR Quadtree . . . 374
2.2.2.9. Sector Tree . . . 377
2.2.2.10 Cone Tree . . . 380
2.2.3 Object-Based Boundary Representations . . . 382
2.2.3.1 Strip Tree, Arc Tree, and BSPR . . . 382
2.2.3.2 Prism Tree . . . 386
2.2.3.3 HAL Tree . . . 389
2.2.3.4 Simplification Methods and Data Structures . . . 391
2.2.4 Surface-Based Boundary Representations . . . 399
2.3 Difference-Based Compaction Methods . . . 408
.2.3,1 Runlength Encoding . . . 409
2.3.2 Chain Code . . . 416
2.3.3 Vertex Representation . . . 417
2.4 Historical Overview . . . 423

3 Intervals and Small Rectangles . . . 427

3.1 Plane-Sweep Methods and the Rectangle Intersection Problem . . . 428
3Il .1 Segment Tree . . . 430
3.1.2 Interval Tree . . . 434
3.1.3 Priority Search Tree . . . 439
3.L4 Alternative Solutions and Related Problems . . . 443
3.2 Plane-Sweep Methods and the Measure Problem . . . 447
3.3 Point-Based Methods . . . 453
3.3.1 Representative Points . . . 453
•3.3.2 Collections of Representative Points . . . 459
3.3.2.1 Corner Stitching . . . 459

3.3.2.2 Balanced Four-Dimensional K-d Tree . . . 460
3.3.2.3' Grid File . . . 462
3.3.2.4 LSD Tree . . . 463
3.3.2.5 Two-Dimensional Representations . . . 463
3.3.3 Summary . . . 464
3.4 Area-Based Methods . . . 466
3.4.1 MX-CIF Quadtree . . . 466
3.4.1.1 Definition . . . 466
3.4.1.2 Operations . . . 468
3.4.1.3 Comparison with Other Representations . . . 473
3.4.2 Alternatives to the MX-CIF Quadtree . . . 474
3.4.3 Multiple Quadtree Block Representations . . . 480

4 High-Dimensional Data . . . 485

4.1 Best-First Nearest Neighbor Finding . . . 490
4.1.1 Motivation . . . 491
4.1.2 Search Hierarchy . . . 492
4.1.3 Algorithm . . . 493
4.1.4 Algorithm with Duplicate Instances of Objects . . . 499
4.1.5 Algorithm Extensions (*K*-Nearest, AT-Farthest, Skylines) . . . 502
4.1.6 Nearest Neighbors in a Spatial Network . . . 508
4.1.7 Related Work . . . 516
4.2 The Depth-First /sT-Nearest Neighbor Algorithm . . . 517
4.2.1 Basic Algorithm . . . 518
4.2.2 Pruning Rules . . . 521
4.2.3 Effects of Clustering Methods on Pruning . . . 529
4.2.4 Ordering the Processing of the Elements of the Active List . . . 534
4.2.5 Improved Algorithm . . . 538
4.2.6 Incorporating MAXNEARESTDIST in a Best-First Algorithm . . . 548
4.2.7 Example . . . 553
4.2.8 Comparison . . . 554
4.3 Approximate Nearest Neighbor Finding . . . 557
4.4 Multidimensional Indexing Methods . . . 566
4.4.1 X-Tree . . . 566
4.4.2 Bounding Sphere Methods: Sphere Tree, SS-Tree, Balltree, and SR-Tree . . . 567
4.4.3 Increasing the Fanout: TV-Tree, Hybrid Tree, and A-Tree . . . 572
4.4.4 Methods Based on the Voronoi Diagram: OS-Tree . . . 573
4.4.5 Approximate Voronoi Diagram (AVD) . . . 580
4.4.6 Avoiding Overlapping All of the Leaf Blocks . . . 585
4.4.7 Pyramid Technique . . . 587
4.4.8 Methods Based on a Sequential Scan . . . 592
4.5 Distance-Based Indexing Methods . . . 598
4.5.1 Distance Metric and Search Pruning . . . 600
4.5.2 Ball Partitioning Methods . . . 604
4.5.2.1 Vp-Tree . . . 604
4.5.2.1.1 Structure . . . 604
4.5.2.1.2 Search . . . 606
4.5.2.2 Vp^{sb}-Tree . . . 607
4.5.2.3 Mvp-Tree . . . 608
4.5.2.4 Other Methods Related to Ball Partitioning . . . 609
4.5.3 Generalized Hyperplane Partitioning Methods . . . 613
4.5.3.1 Gh-Tree . . . 613
4.5.3.2 GNAT . . . 616
4.5.3.3 Bisector Tree and Mb-Tree . . . 617
4.5.3.4 Other Methods Related to Generalized Hyperplane Partitioning . . . 622
4.5.4 M-Tree . . . 624
4.5.4.1 Structure . . . 624
4.5.4.2 Search . . . 626
4.5.5 Sa-Tree . . . 629
4.5.5.1 Definition . . . 630
4.5.5.2 Search . . . 633

4.5.6 kNN Graph . . . 637
4.5.7 Distance Matrix Methods . . . 643
4.5.7.1 AESA . . . 643
4.5.7.2 LAESA . . . 646
4.5.7.3 Other Distance Matrix Methods . . . 647
4.5.8 SASH: Indexing without Using the Triangle Inequality . . . 650
4.6 Dimension Reduction Methods . . . 662
4.6.1 Searching in the Dimension Reduced Space . . . 664
4.6.1-1 Range Queries . . . 664
4.6.1.2 Nearest Neighbor Queries . . . 665
4.6.2 Using Only One Dimension . . . 668
4.6.3 Representative Point Methods . . . 670
4.6.4 Transformation into a Different and Smaller Feature Set . . . 671
4.6.4.1 SVD . . . 671
4.6.4.2 Discrete Fourier Transform (DFT) . . . 676
4.6.5 Summary . . . 685
4.7 Embedding Methods . . . 685
4.7.1 Introduction ." . . . 687
4.7.2 Lipschitz Embeddings . . . 691
4.7.2.1 Definition . . . 691
4.7.2.2 Selecting Reference Sets . . . 692
4.7.2.3 Example of a Lipschitz Embedding . . . 693
4.7.2.4 SparseMap . . . 694
4.7.3 FastMap . . . 697
4.7.3.1 Mechanics of FastMap . . . 697
4.7.3.2 Choosing Pivot Objects . . . 698
4.7.3.3 Deriving the First Coordinate Value . . . 699
4.7.3.4 Projected Distance . . . 700
4.7.3.5 Subsequent Iterations . . . 701
4.7.3.6 Pruning Property . . . 703
4.7.3.7 Expansion . . . 707
4.7.3.8 Heuristic for Non-Euclidean Metrics . . . 708
4.7.3.9 Summary and Comparison with Other Embeddings . . . 709
4.7.4 Locality Sensitive Hashing . . . 711

Appendix A: Overview of B-Trees . . . 717
Appendix B: Linear Hashing . . . 729
Appendix C: Spiral Hashing . . . 735
Appendix D: Description of Pseudocode Language . . . 743
Solutions to Exercises . . . 747
References . . . 877
Reference Keyword Index . . . 947
Author and Credit Index . . . 953
Subject Index . . . 969