Bibliography

[ABH97] P. Apers, H. Blanken, and M. Houtsama. *Multimedia Databases in Perspective*. Springer, 1997.

[AFS93] R. Agrawal, C. Faoutsos, and A. Swami. Efficient similarity search in sequence databases. In *FODO'93, Chicago, IL, October 1993*, pages 69–84, 1993.

[AG87] A. Apostolico and C. Guerra. The longest common subsequence problem revisited. *Algoritmica*, 2:315–336, 1987.

[All96] James Allan. Incremental relevance feedback for information filtering. In *SIGIR*, pages 270–278, 1996.

[AM95] Sunil Arya and David M. Mount. Approximate range searching. In *Symposium on Computational Geometry 1995*, pages 172–181, 1995.

[AMN+98] Sunil Arya, David M. Mount, Nathan S. Netanyahu, Ruth Silverman, and Angela Y. Wu. An optimal algorithm for approximate nearest neighbor searching in fixed dimensions. *Journal of ACM*, 45(6):891–923, 1998.

[BBK01] Christian Böhm, Stefan Berchtold, and Daniel Keim. Searching in high-dimensional spaces: Index structures for improving the performance

of multimedia databases. *ACM Computing Surveys*, 33(3):322–373, September 2001.

[BC92] Nicholas J. Belkin and W. Bruce Croft. Information filtering and information retrieval: Two sides of the same coin. *Communication of the ACM*, 35(12):29–38, December 1992.

[Ben75] J. Bentley. Multidimensional binary search trees used for associative searching. *Communication of the ACM*, 18(9):509–517, 1975.

[Ben79] J. Bentley. Multidimensional binary search trees in database applications. *IEEE Transactions on Software Engineering*, 5(4):333–340, 1979.

[Bes95] S. N. Bespamyatnikh. An optimal algorithm for closest pair maintenance (extended abstract). In Joan Peckham, editor, *Proceedings of 11th ACM Symposium on Computational Geometry*, pages 152–161, 1995.

[BET95] Marshall W. Bern, David Eppstein, and Shang-Hua Teng. Parallel construction of quadtrees and quality triangulations. In *WADS '93, Proceedings of 4rd Algorithms and Data Structures, August 11-13 1997, Montral, Canada*, volume 709 of *LNCS*, pages 188–199. Springer Verlag, 1995.

[BFR78] R.L. Burden, J. Douglas Faires, and A.C. Reynolds. *Numerical Analysis*. Prindle, Weber & Schmidt, 1978.

[BGRS99] Kevin S. Beyer, Jonathan Goldstein, Raghu Ramakrishnan, and Uri Shaft. When is "nearest neighbor" meaningful? In Catriel Beeri and Peter Buneman, editors, *Database Theory - ICDT '99, 7th International Conference, Jerusalem, Israel, January 10-12, 1999, Proceedings*, volume 1540 of *Lecture Notes in Computer Science*, pages 217–235. Springer, 1999.

[BKK96] Stefan Berchtold, Daniel A. Keim, and Hans-Peter Kriegel. The x-tree: An index structure for high-dimensional data. In T. M. Vijayaraman, Alejandro P. Buchmann, C. Mohan, and Nandlal L. Sarda, editors, *VLDB'96, Proceedings of 22th International Conference on Very Large Data Bases, September 3-6, 1996, Mumbai (Bombay), India*, pages 28–39. Morgan Kaufmann, 1996.

[BKK97] Stefan Berchtold, Daniel A. Keim, and Hans-Peter Kriegel. A cost model for nearest neighbor search in high-dimensional data space. In *PODS'97, Proceedings of 16th ACM Symposium on Principles of Database Systems, May 1997, Tucson, AZ*, pages 78–96. ACM, 1997.

[BKSS90] N. Beckmann, H. P. Kriegel, R. Schneider, and B. Seeger. The r*-tree: An efficient and robust access method for points and rectangles. In *Proceedings of the 1990 ACM-SIGMOD International Conference on Management of Data, May 199a, Atlantic City, NJ*, pages 322–331. ACM, 1990.

[BM72] Rudolf Bayer and Edward M. McCreight. Organization and maintenance of large ordered indices. *Acta Informatica*, 1(3):173–189, 1972.

[BÖ97] Tolga Bozkaya and Z. Meral Özsoyoglu. Distance-based indexing for high-dimensional metric spaces. In Joan Peckham, editor, *SIGMOD 1997, Proceedings ACM SIGMOD International Conference on Management of Data, May 13-15, 1997, Tucson, Arizona, USA*, pages 357–368. ACM Press, 1997.

[BÖ99] Tolga Bozkaya and Z. Meral Özsoyoglu. Indexing large metric spaces for similarity search querie. *TODS*, 24(3):361–404, 1999.

[Bri95] Sergey Brin. Near neighbor search in large metric spaces. In Umeshwar Dayal, Peter M. D. Gray, and Shojiro Nishio, editors, *VLDB'95, Proceedings of 21th International Conference on Very Large Data Bases, September 11-15, 1995, Zurich, Switzerland*, pages 574–584. Morgan Kaufmann, 1995.

[BWY80] J. Bentley, B. Weide, and A. Yao. Optimal expected-time algorithms for closest point problems. *ACM Transaction on Mathematical Software*, 6(4):563–580, 1980.

[Cas96] K.R. Castelman. *Digital Image Processing.* Prentice-Hall, Inc., 1996.

[CF80] J. M. Chang and K. S. Fu. A dynamic clustering technique for physical database design. In Peter P. Chen and R. Clay Sprowls, editors, *Proceedings of the 1980 ACM SIGMOD International Conference on Management of Data, Santa Monica, California, May 14-16, 1980*, pages 188–199. ACM Press, 1980.

[Cha97] Timothy M. Chan. Approximate nearest neighbor queries revisited. In *Proceedings of the thirteenth annual symposium on Computational geometry, 1997 , Nice, France*, pages 352 – 358. ACM Press,New York, NY, USA, 1997.

[Chi94] Tzi-cker Chiueh. Content-based image indexing. In Jorge B. Bocca, Matthias Jarke, and Carlo Zaniolo, editors, *VLDB'94, Proceedings of 20th International Conference on Very Large Data Bases, September 12-15, 1994, Santiago de Chile, Chile*, pages 582–593. Morgan Kaufmann, 1994.

[CK95] Paul B. Callahan and S. Rao Kosaraju. Algorithms for dynamic closest pair and n-body potential fields. In *SODA, Proceedings of 6th ACM-SIAM Symposium on Discrete Algorithms*, pages 263–272, 1995.

[CNBYM01] Edgar Chávez, Gonzalo Navarro, Ricardo Baeza-Yates, and José L. Marroquín. Searching in metric spaces. *ACM Computing Surveys*, 33(3):273–221, September 2001.

[Com79] Douglas Comer. The ubiquitous B-tree. *ACM Computing Surveys*, 11(2):121–137, 1979.

[CP00] Paolo Ciaccia and Marco Patella. Pac nearest neighbor queries: Approximate and controlled search in high-dimensional and metric spaces. In *Proceedings of the 16th International Conference on Data Engineering, 28 February - 3 March, 2000, San Diego, California, USA*, pages 244–255. IEEE Computer Society, 2000.

[CPZ97] Paolo Ciaccia, Marco Patella, and Pavel Zezula. M-tree: An efficient access method for similarity search in metric spaces. In Matthias Jarke, Michael J. Carey, Klaus R. Dittrich, Frederick H. Lochovsky, Pericles Loucopoulos, and Manfred A. Jeusfeld, editors, *VLDB'97, Proceedings of 23rd International Conference on Very Large Data Bases, August 25-29, 1997, Athens, Greece*, pages 426–435. Morgan Kaufmann, 1997.

[CPZ98a] Paolo Ciaccia, Marco Patella, and Pavel Zezula. A cost model for similarity queries in metric spaces. In *Proceedings of the Seventeenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, June 1-3, 1998, Seattle, Washington*, pages 59–68. ACM Press, 1998.

[CPZ98b] Paolo Ciaccia, Marco Patella, and Pavel Zezula. Processing complex similarity queries with distance-based access methods. In Hans-Jörg Schek, Fèlix Saltor, Isidro Ramos, and Gustovo Alonso, editors, *Advances in Database Technology - EDBT'98, 6th International Conference on Extending Database Technology, Valencia, Spain, March 23-27,*

1998, Proceedings, volume 1377 of *Lecture Notes in Computer Science*, pages 9–23. Springer, 1998.

[Dam64] F. Damerau. A technique for computer detection and correction of spelling errors. *Communication of the ACM*, 7(3):171–176, 1964.

[Fal96] C. Faloutsos. *Searching Multimedia Databases by Content.* Kluwer Academic Publishers, 1996.

[FB74] Raphael A. Finkel and Jon Louis Bentley. Quad trees: A data structure for retrieval on composite keys. *Acta Informatica*, 4:1–9, 1974.

[FK97] C. Faloutsos and I. Kamel. Beyond uniformity and independence: Analysis of r-trees using the concept of fractal dimension. In *Proceedings of the 13th ACM Symposium on Principles of Database Systems, El Paso, Texas, USA, May 4-6, 1997*, pages 609–617. ACM, 1997.

[FSA+95] Myron Flickner, Harpreet S. Sawhney, Jonathan Ashley, Qian Huang, Byron Dom, Monika Gorkani, Jim Hafner, Denis Lee, Dragutin Petkovic, David Steele, and Peter Yanker. Query by image and video content: The qbic system. *IEEE Computer*, 28(9):23–32, 1995.

[FTAA00] H. Ferhatosmanoglu, E. Tuncel, D. Agrawal, and A. El Abbadi. Vector approximation based indexing for non-uniform high dimensional data sets. In *Proceedings of the 2000 ACM CIKM International Conference on Information and Knowledge Management, McLean, VA, USA, November 6-11, 2000*, pages 202–209. ACM, 2000.

[FTAA01] H. Ferhatosmanoglu, E. Tuncel, D. Agrawal, and A. El Abbadi. Approximate nearest neighbor searching in multimedia databases. In *Proceedings of the 17th International Conference on Data Engineering, April*

2-6, 2001, Heidelberg, Germany, pages 503–511. IEEE Computer Society, 2001.

[Fuk90] K. Fukunaga. *Introduction to Statistical Pattern Recognition.* Academic Press, 2nd edition, 1990.

[GG98] V. Gaede and O. Gunther. Multidimensional access methods. *ACM Computing Surveys*, 30(2):170–231, 1998.

[GPR97] Alberto Gobbi, D. Poppinger, and B. Rohde. Finding biological active compounds in large databases. In *ECSOC-1, Proceedings of First International Electronic Conference on Synthetic Organic Chemistry, September 1-30, 1997*, 1997.

[GSZ00] Claudio Gennaro, Pasquale Savino, and Pavel Zezula. A hashed schema for similarity search in metric spaces. In *Proceedings of First DELOS Network of Excellence Workshop, Information Seeking, Searching and Querying in Digital Libraries, Zurich, Switzerland, December 11-12*, 2000.

[GSZ01] Claudio Gennaro, Pasquale Savino, and Pavel Zezula. Similarity search in metric databases through hashing. In *Proceedings of MIR 2001 - 3rd Intl Workshop on Multimedia Information Retrieval October 5, 2001. Ottawa, Canada*, 2001. In conjunction with ACM Multimedia 2001.

[GSZ02] Claudio Gennaro, Pasquale Savino, and Pavel Zezula. D-index: Distance searching index for metric data sets. Technical report, IEI-CNR, 2002.

[Gut84] A. Guttman. R-trees: A dynamic index structure for spatial searching. In *Proceedings of the 1984 ACM SIGMOD International Conference on Management of Data, Boston, MA*, pages 47–57, 1984.

[Hal52] A. Hald. *Statistical Theory with Engineering Applications.* John Wiley & Sons, Inc., 1952.

[HB99] S. Hettich and S. D. Bay. The uci kdd archive. Irvine, CA: University of California, Department of Information and Computer Science, 1999. http://kdd.ics.uci.edu.

[HD80] P.A.V. Hall and G.R. Dowling. Approximate string matching. *ACM Computing Surveys*, 12(4):381–402, December 1980.

[HKR93] D.P. Huttenlocker, G.A. Klanderman, and W.J. Rucklidge. Comparing images using the hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):850–863, September 1993.

[HNP95] Joseph M. Hellerstein, Jeffrey F. Naughton, and Avi Pfeffer. Generalized search trees for database systems. In Umeshwar Dayal, Peter M. D. Gray, and Shojiro Nishio, editors, *VLDB'95, Proceedings of 21th International Conference on Very Large Data Bases, September 11-15, 1995, Zurich, Switzerland*, pages 562–573. Morgan Kaufmann, 1995.

[HPS71] P. Hoel, S. Port, and J. Stone. *Introduction to Probability Theory.* Houghton Mifflin Company, 1971.

[HS95] G.R. Hjaltason and H. Samet. Ranking in spatial databases. In *SSD'95, Portland, ME, August, 1995*, volume 951 of *LNCS*, pages 83–95. Springer Verlag, 1995.

[HSE+95] James L. Hafner, Harpreet S. Sawhney, William Equitz, Myron Flickner, and Wayne Niblack. Efficient color histogram indexing for quadratic form distance functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(7):729–736, 1995.

[IM98] P. Indyk and R. Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. In *Proceedings of the Thirtieth Annual ACM Symposium on the Theory of Computing, Dallas, Texas, USA, May 23-26, 1998*, pages 604–613, 1998.

[JMM95] H. V. Jagadish, Alberto O. Mendelzon, and Tova Milo. Similarity-based queries. In *Proceedings of the Fourteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, May 22-25, 1995, San Jose, California*, pages 36–45. ACM Press, 1995.

[Kai85] T. Kailath. *Modern Signal Processing.* Springer Verlag, 1985.

[KF92] I. Kamel and C. Faloutsos. Parallel r-trees. In *Proc. of the ACM SIGMOD Conf., June, 1992*, pages 195–204, 1992.

[KF93] I. Kamel and C. Faloutsos. On packing r-trees. In *CIKM'93, Proceedings of the Second International conference on Information and Knowledge Management, Washington, DC, November, 1993*, pages 490–499, 1993.

[Kle97] Jon M. Kleinberg. Two algorithms for nearest-neighbor search in high dimensions. In *STOC'97, Proceedings of the twenty-ninth annual ACM symposium on Theory of computing, 1997 , El Paso, Texas, United States*, pages 599–608, 1997.

[Knu98] D.E. Knuth. *Sorting and Searching*, volume 3 of *The Art of Computer Programming.* Reading, Massachusetts: Addison-Wesley, 2nd edition, 1998.

[Koh84] T. Kohonen. *Self-Organization and Associative Memory.* Springer-Verlag, 1984.

[KOR99] Eyal Kushilevitz, Rafail Ostrovsky, and Yuval Rabani. Efficient search for approximate nearest neighbor in high dimensional spaces. In

STOC'98, Proceedings of the thirtieth annual ACM symposium on Theory of computing, 1998Dallas, Texas, United States, pages 614 – 623, 1999.

[Lev65] V. Levenshtein. Binary codes capable of correcting spurious insertions and deletions of ones. *Problem of Information Transmission*, 1:8–17, 1965.

[MMLP97] Javed Mostafa, Snehasis Mukhopadhyay, W. Lam, and Mathew Palakal. A multilevel approach to intelligent information filtering: Model, system, and evaluation. *Transaction on Information Systems*, 15(4):368–399, October 1997.

[MS94] Masahiro Morita and Yoichi Shinoda. Information filtering based on user behaviour analysis and best match text retrieval. In W. Bruce Croft and C. J. van Rijsbergen, editors, *Proceedings of the 17th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval. Dublin, Ireland, 3-6 July 1994 (Special Issue of the SIGIR Forum)*, pages 272–281. ACM/Springer, 1994.

[Nav01] Gonzalo Navarro. A guided tour to approximate string matching. *ACM Computing Surveys*, 33(1):31–88, 2001.

[NBE+93] Wayne Niblack, Ron Barber, William Equitz, Myron Flickner, Eduardo H. Glasman, Dragutin Petkovic, Peter Yanker, Christos Faloutsos, and Gabriel Taubin. The qbic project: Querying images by content, using color, texture, and shape. In *Proceedings of Storage and Retrieval for Image and Video Databases (SPIE) 1993*, pages 173–187, 1993.

[NHS84] Jürg Nievergelt, Hans Hinterberger, and Kenneth C. Sevcik. The grid file: An adaptable, symmetric multikey file structure. *TODS*, 9(1):38–71, 1984.

[NW70] S. Needleman and C. Wunsch. A general method applicable to the search for similarities in the amnio acid sequences of two protein. *Journal of Molecula Biology*, 48:444–453, 1970.

[OS89] A.V. Oppenheim and R.W. Schafer. *Discrete-Time Signal Processing.* Prentice-Hall, Inc., 1989.

[PAL99] S. Pramanik, S. Alexander, and J. Li. An efficient searching algorithm for approximate nearest neighbor queries in high dimensions. In *ICMCS 1999, IEEE International Conference on Multimedia Computing and Systems, June 7-11 , 1999, Florence, Italy*, volume 1. IEEE Computer Society, 1999.

[PL99] Sakti Pramanik and Jinhua Li. Ab-tree: Angle based index tree for approximate nearest neighbor search. Technical Report, Michigan State University, Department of Computer Science, February 1999. http://www.cse.msu.edu/ pramanik/research/papers/AB-tree.pdf.

[PM97] A. Papadopulos and Y. Manolopoulos. Performances of nearest-neighbor queries in r-trees. In *ICDT'97, Proceedings of the 6th International Conference on Database Theory, 1997, Delphi, Greece*, pages 394–408, 1997.

[PMD01] Dimitris Papadias, Nikos Mamoulis, and Vasilis Delis. Approximate spatio-temporal retrieval. *ACM Transactions on Information Systems (TOIS)*, 19(1):53 – 96, January 2001.

[RL85] Nick Roussopoulos and Daniel Leifker. Direct spatial search on pictorial databases using packed r-trees. In Shamkant B. Navathe, editor, *Proceedings of the 1985 ACM SIGMOD International Conference on Management of Data, Austin, Texas, May 28-31, 1985*, pages 17–31. ACM Press, 1985.

[Rob81] John T. Robinson. The k-d-b-tree: A search structure for large multidimensional dynamic indexes. In Y. Edmund Lien, editor, *Proceedings of the 1981 ACM SIGMOD International Conference on Management of Data, Ann Arbor, Michigan, April 29 - May 1, 1981*, pages 10–18. ACM Press, 1981.

[RV97] Paul Resnick and Hal R. Varian. Guest editors' introduction to the special issue on recommender systems. *Communication of the ACM*, 40(3), 1997.

[Sam88] Hanan Samet. Hierarchical representations of collections of small rectangles. *ACM Computing Surveys*, 20(4):271–309, 1988.

[Sam95] H. Samet. The quad-tree and related hierarchical structures. *ACM Computing Surveys*, 16(2):187–260, 1995.

[SK83] D. Sankoff and J. Kruskal. *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison.* Addison-Wesley, 1983.

[SM83] Gerald Salton and Michael J. McGill. *Introduction to Modern Information Retrieval.* McGraw-Hill Book Company, 1983.

[Smi97] John R. Smith. *Integrated Spatial and Feature Image Systems: Retrieval, Analysis, and compression.* PhD thesis, Graduate School of Arts and Sciences, Columbia University, 1997.

[SO95] M. Stricker and M. Orengo. Similarity of color images. In *Storage and Retrieval for Image and Video Databases III SPIE Proceedings 2420*, pages 381–392, 1995.

[SRF87] Timos K. Sellis, Nick Roussopoulos, and Christos Faloutsos. The r+-tree: A dynamic index for multi-dimensional objects. In Peter M.

Stocker, William Kent, and Peter Hammersley, editors, *VLDB'87, Proceedings of 13th International Conference on Very Large Data Bases, September 1-4, 1987, Brighton, England*, pages 507–518. Morgan Kaufmann, 1987.

[SW85] Hanan Samet and Robert E. Webber. Storing a collection of polygons using quadtrees. *ACM Transactions on Graphics*, 4(3):182–222, 1985.

[TS96] Y. Theodoridis and T. Sellis. A model for the prediction of r-tree performance. In *PODS'96, Proceedings of the 15th Symposium on Principles of Databses Systems, June 1996, Montreal, Canada*, pages 161–171, 1996.

[TTSF00] C. Traina, A.J. Traina, B. Seeger, and C. Faloutsos. Slim-trees: High performance metric trees minimizing overlap between nodes. In *EDBT 2000, Proceedings of the 7th EDBT International Conference, March 2000, Konstanz, Germany*, pages 51–65, 2000.

[Uhl91] J.K. Uhlmann. Satisfying general proximity/similarity queries with metric trees. *Information Processing Letters*, 40(4):175–179, November 1991.

[Wat95] M. Waterman. *Introduction to Computational Biology.* Chapman and Hall, 1995.

[WFHC92] Steven P. Wartik, Edward A. Fox, Lenwood S. Heath, and Qi Fan Chen. Hashing algorithms. In William B. Frakes and Ricardo A. Baeza-Yates, editors, *Information Retrieval: Data Structures & Algorithms*, pages 293–362. Prentice-Hall, 1992.

[WJ96] D.A. White and R. Jain. Similarity indexing with the ss-tree. In *Proceedings of the 12th International Conference on Data Engineering, New Orleans, USA*, pages 516–523, 1996.

[WSB98] Roger Weber, Hans-Jörg Schek, and Stephen Blott. A quantitative analysis and performance study for similarity-search methods in high-dimensional spaces. In Ashish Gupta, Oded Shmueli, and Jennifer Widom, editors, *VLDB'98, Proceedings of 24rd International Conference on Very Large Data Bases, August 24-27, 1998, New York City, New York, USA*, pages 194–205. Morgan Kaufmann, 1998.

[YI99] A. Yoshitaka and T. Ichikawa. A survey on content-based retrieval for multimedia databases. *IEEE Transactions on Knowledge and Data Engineering*, 11(1):81–93, 1999.

[Yia93] P.N. Yianilos. Data structures and algorithms for nearest neighbor search in general metric spaces. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 311–321, 1993.

[Yia99] P.N. Yianilos. Excluded middle vantage point forests for nearest neighbor search. In *Sixth DIMACS Implementation Challenge: Nearest Neighbor Searches workshop*, January 1999.

[ZSAR98] Pavel Zezula, Pasquale Savino, Giuseppe Amato, and Fausto Rabitti. Approximate similarity retrieval with m-trees. *VLDB Journal*, 7(4):275–293, 1998.