James M. Kang

4-192 EE/CS Building, 200 Union Street SE, Minneapolis, MN, 55455 (585) 733-3159
e-mail: jkang@cs.umn.edu, homepage: www.cs.umn.edu/~jkang

EDUCATION

Ph.D. in Computer Science, Anticipated Summer 2010
Department of Computer Science, University of Minnesota, Minneapolis, MN
Thesis: Spatio-Temporal Data Mining of Vector Field Datasets
Advisor: Dr. Shashi Shekhar

M.S. in Computer Science, May 2005
Department of Computer Science, Rochester Institute of Technology, Rochester, NY
Thesis: A Query Engine of Novelty in Video Streams
Advisor: Dr. Ankur Teredesai

B.S. in Computer Science, May 2000
Department of Computer Science, Purdue University, West Lafayette, IN

RESEARCH INTERESTS

Spatio-Temporal Data Mining, Spatio-Temporal Databases, and Applications to Environmental Systems

AWARDS

- NSF Integrative Graduate Education and Research Traineeship (IGERT) Fellowship, 2010
- Highlighted as a success story at the NGA NARP Symposium, 2009
- Best Paper Award - 1^{st} Int. Workshop on Knowledge Discovery from Sensor Data (Sensor-KDD '07)
- NSF Integrative Graduate Education and Research Traineeship (IGERT) Fellowship, 2005 - 2007
- Selected as one of the best papers in 1^{st} Int. Work. on Spatial and Spatio-Temporal Data Mining, 2006
- Asia Pacific Exchange Leadership Award, Eastman Kodak Company, Rochester, New York, 2003

EXPERIENCE

Research

Spatial Database Group, University of Minnesota (Aug '05 - Present)
- Investigated flow data mining (ICDM '08, SSTD '09), spatial data mining (ICDM '07), query processing (ICDE '07) and image classification (SSTDM '06) techniques

Visiting Scientist, National Geospatial-Intelligence Agency, Reston, VA (Jun '09 - Aug '09)
- Identified new and interesting spatial and spatio-temporal patterns that showed a great interest to the intelligence community (PR#: 09-434)

Data Mining Research Group, Rochester Institute of Technology (Mar '04 - May '05)
- Investigated JPEG 2000 stream to annotate multimedia data (IS&T '05)

Teaching

Practice of Database Systems TA, University of Minnesota (Aug '07 - May '08)
- Managed homework problems, labs, and exams; and gave lectures in the class (Fall '07, Spr '08)

Data Mining TA, Rochester Institute of Technology (Aug '04 - Nov '04)
- Taught a recitation class for students regarding homework and projects

Professional

Eastman Kodak Company, Rochester, NY (Jun '00 - Feb '04)
- Managed business-to-business technical solutions between Eastman Kodak and suppliers located throughout Europe and Asia using SAP, XML, Web Methods, and Business Connector

JOURNAL ARTICLES

Publications

1. **J. M. Kang**, M. Mokbel, S. Shekhar, T. Xia, D. Zhang, Incremental and General Evaluation of Reverse Nearest Neighbors, *IEEE Transactions on Knowledge and Data Engineering* (TKDE), to appear.

2. V. Gandhi, **J. M. Kang**, S. Shekhar, Context-Inclusive Function Evaluation: A Case Study with EM-Based Multi-Scale Multi-Granular Image Classification, *Knowledge and Information Systems*, (KAIS), 21(2), 231-247, 2009.

3. B. George, **J. M. Kang**, S. Shekhar, Spatio-Temporal Sensor Graphs (STSG): A Data Model for the Discovery of Spatio-Temporal Patterns, *International Journal of Intelligent Analysis* (JIDA), 13(3), 457-475, 2009.

In Submission/Preparation

1. **J. M. Kang**, S. Shekhar, C. Wennen, P. Novak, A General Approach to the Discovery of Flow Anomalies, *IEEE Transactions on Knowledge and Data Engineering* (TKDE), submitted.

2. **J. M. Kang**, S. Shekhar, M. Henjum, P. Novak, W. Arnold, Discovering Teleconnected Flow Anomalies, *IEEE Transactions on Knowledge and Data Engineering* (TKDE), in preparation.

CONFERENCE ARTICLES

1. **J. M. Kang**, S. Shekhar, M. Henjum, P. Novak, W. Arnold, Discovering Teleconnected Flow Anomalies: A Relationship Analysis of spatio-temporal Dynamic (RAD) neighborhoods, *In the Symposium on Spatial and Temporal Databases* (SSTD '09), pp. 44-61, Aalborg, Denmark, July 8-10, 2009 (Selectivity of 1 out of 3).

2. D. Morgan, J. W. Kang, **J. M. Kang**, Mineable Data Warehouses, *In the 11th Int. Conf. on Enterprise Information Systems* (ICEIS '09), pp. 125-136, Milan, Italy, May 6-10, 2009, (Selectivity of 1 out of 9).

3. **J. M. Kang**, S. Shekhar, C. Wennen, P. Novak, Discovering Flow Anomalies: A SWEET Approach, *In the Eighth IEEE International Conference on Data Mining* (ICDM '08), pp. 851-856, Pisa, Italy, December 15-19, 2008 (Selectivity of 1 out of 7).

4. M. Celik, **J. M. Kang**, S. Shekhar, Zonal Co-location Pattern Discovery with Dynamic Parameters, *In the 7th IEEE International Conference on Data Mining* (ICDM '07), Omaha NE, October 28-31, 2007 (Selectivity of 1 out of 5).

5. **J. M. Kang**, M. Mokbel, S. Shekhar, T. Xia, D. Zhang, Continuous Evaluation of Monochromatic and Bichromatic Reverse Nearest Neighbors, *In the IEEE 23rd International Conference on Data Engineering* (ICDE '07), pp 806-815 Istanbul, Turkey, April 16-20, 2007 (Selectivity of 1 out of 7).

6. V. Misic, **J. M. Kang**, A. M. Teredesai, J. Sissel, Portable Image Archiving: Annotation, Search, and Data Retrieval, *IS&T Archiving Conference 2005* (IS&T '05), p. 114-118, Washington, D.C., April 26-29, 2005.

WORKSHOP ARTICLES

1. **J. M. Kang**, S. Shekhar, Discovering Flow Anomalies on a Smarter Water Infrastructure System, *Data Mining for Smarter Infrastructure Workshop, In conjunction with SDM*, 2010 (Invited Paper).

2. B. George, **J. M. Kang**, S. Shekhar, Spatio-Temporal Sensor Graph (STSG): A Sensor Model for the Discovery of Spatio-Temporal Patterns, *In the first SIG-KDD International Workshop on Knowledge Discovery from Sensor Data* (Sensor-KDD '07), San Jose CA, August 12, 2007 (Best Paper Award).

3. M. Celik, S. Shekhar, J. P. Rogers, J. A. Shine, **J. M. Kang**, Mining At Most Top-K% Mixed-drove Spatio-temporal Co-occurrence Patterns: A Summary of Results, *In the ICDE Workshop on Spatio-Temporal Data Mining* (STDM '07), Istanbul, Turkey, Apr 20, 2007.

4. V. Gandhi, **J. M. Kang**, S. Shekhar, J. Ju, E. D. Kolaczyk, S. Gopal, Using a Context Approach to Process Statistical Queries in Raster Data: An Extended Abstract, *In proceedings of the 1st International Workshop on Spatial and Spatio-temporal Data Mining* (SSTDM '06), in conjunction with ICDM '06, p. 371-376, Hong Kong, December 18, 2006 (Selectivity: 1 out of 3) (Selected as one of the best papers).

BOOK CHAPTERS

1. S. Shekhar, V. Gandhi, **J. M. Kang**, and M. Mokbel, Spatial Databases, *Handbook of Database Technologies (Ed. Markus Schneider and Joachim Hammer)*, CRC Press, to appear.

2. S. Shekhar, and **J. M. Kang**, Spatial Data Mining, (Field Editor: D. Papadis), *Encyclopedia of Database Systems*, Springer Publishers, isbn 978-0387355443, to appear.

3. V. Gandhi, **J. M. Kang**, and S. Shekhar, Spatial Databases, *Wiley Encyclopedia of Computer Science and Engineering (Ed. Benjamin Wah)*, John Wiley and Sons Inc., 2009, isbn 978-0471383932.

4. **J. M. Kang**, Voronoi Diagrams, *Ency. of GIS*, Springer Pub., 1232-1235, 978-0-387-30858-6, 2008.

5. **J. M. Kang**, Voronoi Terminology, *Ency. of GIS*, Springer Pub., pg 1241, 978-0-387-30858-6, 2008.

6. **J. M. Kang**, Distance Metrics, *Ency. of GIS*, Springer Pub., 245-246, 978-0-387-30858-6, 2008.

7. **J. M. Kang**, A. M. Teredesai, M. A. Ahmad, R. Gaborski, Cognitively Motivated Novelty Detection in Video Data Streams, *In the Multimedia Data Mining and Knowledge Discovery*, Springer Publishers, V. A. Petrushin & L. Khan (Eds.) 2006.

TECHNICAL REPORTS

1. **J. M. Kang**, S. Shekhar, C. Wennen, P. Novak, Discovering Flow Anomalies: A SWEET Approach, *Technical Report TR09-006*, Dept. of Comp. Sci., University of Minnesota, Minneapolis, USA, 2009.

2. V. Gandhi, **J. M. Kang**, S. Shekhar, J. Ju, E. D. Kolaczyk, S. Gopal, Context Inclusive Function Evaluation: A Case Study with EM-Based Multi-Scale Multi-Granular Image Classification, *Technical Report TR08-025*, Department of Computer Science, University of Minnesota, Minneapolis, USA, 2008.

3. V. Gandhi, **J. M. Kang**, S. Shekhar, Spatial Databases, *Technical Report TR07-020*, Department of Computer Science, University of Minnesota, Minneapolis, USA, 2007.

4. S. Mane, **J. M. Kang**, S. Shekhar, J. Srivastava, C. Murray, A. Pusey, Identifying Clusters in Marked Spatial Point Processes: A Summary of Results, *Technical Report 06-006*, Department of Computer Science, University of Minnesota, Minneapolis, USA, 2006.

5. **J. M. Kang**, A. M. Teredesai, V. Misic, R. Gaborski, Query Engine of Novelty in Video Streams, Master's Thesis, Department of Computer Science, B. Thomas Golisano College of Computing and Information Science, Rochester, NY, USA, 2005.

AVAILABILITY: Papers are available on my homepage at www.cs.umn.edu/~jkang

SERVICE

- Ext. reviewer: ICDM (06,07,08), SSTD (07,09), ICDE (07,08), Geoinformatica, TKDE, SIG-KDD 06
- Officer, New Hires Group, Eastman Kodak Company, Rochester, New York 2000 - 2004
- College of Science Student Council, Purdue University, West Lafayette, Indiana 1997
- Secretary, Asian American Association, Purdue University, West Lafayette, Indiana 1996

SELECTED PUBLICATIONS

- **J. M. Kang**, S. Shekhar, M. Henjum, P. Novak, W. Arnold, Discovering Teleconnected Flow Anomalies: A Relationship Analysis of spatio-temporal Dynamic (RAD) neighborhoods, *In the Symposium on Spatial and Temporal Databases* (SSTD '09), pp. 44-61, Aalborg, Denmark, July 8-10, 2009 (Selectivity of 1 out of 3).

Abstract

Given a collection of sensors monitoring a flow network, the problem of discovering teleconnected flow anomalies aims to identify strongly connected pairs of events (e.g., introduction of a contaminant and its removal from a river). The ability to mine teleconnected flow anomalies is important for applications related to environmental science, video surveillance, and transportation systems. However, this problem is computationally hard because of the large number of time instants of measurement, sensors, and locations. This paper characterizes the computational structure in terms of three critical tasks, (1) detection of flow anomaly events, (2) identification of candidate pairs of events, and (3) evaluation of candidate pairs for possible teleconnection. The first task was addressed in our recent work. In this paper, we propose a RAD (Relationship Analysis of spatio-temporal Dynamic neighborhoods) approach for steps 2 and 3 to discover teleconnected flow anomalies. Computational overhead is brought down significantly by utilizing our proposed spatio-temporal dynamic neighborhood model as an index and a pruning strategy. We prove correctness and completeness for the proposed approaches. We also experimentally show the efficacy of our proposed methods using both synthetic and real datasets.

- **J. M. Kang**, S. Shekhar, C. Wennen, P. Novak, Discovering Flow Anomalies: A SWEET Approach, *In the Eighth IEEE International Conference on Data Mining* (ICDM '08), pp. 851-856, Pisa, Italy, December 15-19, 2008 (Selectivity of 1 out of 7).

Abstract

Given a percentage-threshold and readings from a pair of consecutive upstream and downstream sensors, flow anomaly discovery identifies dominant time intervals where the fraction of time instants of significantly mis-matched sensor readings exceed the given percentage-threshold. Discovering flow anomalies (FA) is an important problem in environmental flow monitoring networks and early warning detection systems for water quality problems. However, mining FAs is computationally expensive because of the large (potentially infinite) number of time instants of measurement and potentially long delays due to stagnant (e.g. lakes) or slow moving (e.g. wetland) water bodies between consecutive sensors. Traditional outlier detection methods (e.g. t-test) are suited for detecting transient FAs (i.e., time instants of significant mis-matches across consecutive sensors) and cannot detect persistent FAs (i.e., long variable time-windows with a high fraction of time instant transient FAs) due to a lack of a pre-defined window size. In contrast, we propose a Smart Window Enumeration and Evaluation of persistence-Thresholds (SWEET) method to efficiently explore the search space of all possible window lengths. Computation overhead is brought down significantly by restricting the start and end points of a window to coincide with transient FAs, using a smart counter and efficient pruning techniques. Experimental evaluation using a real dataset shows our proposed approach outperforms Naïve alternatives.

- **J. M. Kang**, M. Mokbel, S. Shekhar, T. Xia, D. Zhang, Incremental and General Evaluation of Reverse Nearest Neighbors, *IEEE Transactions on Knowledge and Data Engineering* (TKDE), to appear.

Abstract

This paper presents a novel algorithm for Incremental and General Evaluation of continuous Reverse Nearest neighbor queries (IGERN, for short). The IGERN algorithm is general in that it is applicable for both continuous monochromatic and bichromatic reverse nearest neighbor queries. This problem is faced in a number of applications such as enhanced 911 services and in army strategic planning. A main challenge in these problems is to maintain the most up to date query answers as the dataset frequently changes over time. Previous algorithms for monochromatic continuous reverse nearest neighbor queries rely mainly on monitoring at the worst case of six pie regions, whereas IGERN takes a radical approach by monitoring only a single region around the query object. The IGERN algorithm clearly outperforms the state-of-the-art algorithms in monochromatic queries. We also propose a new optimization for the monochromatic IGERN. Furthermore, a filter and refine approach for IGERN (FR-IGERN) is proposed for the continuous evaluation of bichromatic reverse nearest neighbor queries which is an optimized version of our previous approach. The computational complexity of IGERN and FR-IGERN is presented, the correctness of IGERN and FR-IGERN are proved, and experimental analysis using synthetic and real datasets is shown.

REFERENCES

Dr. Shashi Shekhar
McKnight Distinguished University Professor
Department of Computer Science, University of Minnesota
4-192, EE/CS Bldg., 200 Union St. SE, Minneapolis, MN 55455
contact: stewart@cs.umn.edu
(612) 624-8307

Dr. Mohamed F. Mokbel
Assistant Professor
Department of Computer Science, University of Minnesota
4-192 EE/CS Building, 200 Union Street SE, Minneapolis, MN, 55455
contact: mokbel@cs.umn.edu
(612) 626-3025

Dr. Paige J. Novak
Associate Professor of Environmental Engineering
Department of Civil Engineering, University of Minnesota
122 Civil Engineering Building, 500 Pillsbury Drive S.E., Minneapolis, MN 55455
contact: novak010@tc.umn.edu
(612) 626-9846

Dr. Claudia Neuhauser
Vice Chancellor for Academic Affairs
Director of the Center for Learning Innovation
Director of Graduate Studies, Biomedical Informatics and Computational Biology
HHMI and Distinguished McKnight University Professor Morse-Alumni Distinguished Teaching Professor
University of Minnesota Rochester
300 University Square, 111 South Broadway, Rochester, MN 55904
contact: neuha001@umn.edu
(507) 281-7791

Dr. Vipin Kumar
Professor and Department Head
Department of Computer Science, University of Minnesota
4-192 EE/CS Building, 200 Union Street SE, Minneapolis, MN, 55455
contact: kumar@cs.umn.edu
(612) 624-8023

Dr. Jaideep Srivastava
Professor
Department of Computer Science, University of Minnesota
4-192 EE/CS Building, 200 Union Street SE, Minneapolis, MN, 55455
contact: srivasta@cs.umn.edu
(612) 625-4012