Dr. Shekhar’s prior supports include those from NSF and U.S. Department of Defense.

**Intellectual Merits:** Results from recent projects were published in 17 peer-reviewed articles documenting creative, original and potentially transformative concepts in the field of spatio-temporal computing and knowledge discovery. Space limitation does not permit a description of all the results. Thus, we summarize a few illustrative examples next.

- **Interesting sub-path discovery.** Given a ST dataset and a path in its embedding spatiotemporal framework, the goal of the interesting sub-path discovery problem is to identify all the sub-paths that exhibit interesting properties defined by an interest measure function. This problem is important to understanding climate change, e.g., finding time periods of precipitation decreases. It is challenging due to the massive volume of data, the varying length of sub-paths and non-monotonicity of interestingness. Previous approaches find interesting unit sub-paths (e.g., time points) or locations. In contrast, our work find long intervals or sub-paths with arbitrary length. We model the computational structure as a Grid-based Directed Acycalic Graph (G-DAG). We propose a novel algorithm, namely, the Row-wise Traversal (after leaf-evaluation) with Column Pruning (RTCP) which brings dramatically down the memory cost for G-DAG traversal in the preliminary approach while also reducing CPU cost. Experimental evaluation on both synthetic and real datasets show that the RTCP algorithm is always the fastest in computational time among all the proposed algorithms. Initial results of this work has been published at ACM SIGSPATIAL Conference on GIS [8], and full results has been accepted with major revision by IEEE Transactions on Knowledge and Data Engineering (TKDE) [7].

- **An inter-disciplinary survey on change footprint patterns and discovery techniques.** Given a definition of change and a dataset about spatiotemporal (ST) phenomena, ST change footprint discovery is the process of identifying the location and/or time of such changes from the dataset. Change footprint discovery is fundamentally important for the study of climate change, the tracking of disease, etc. Methods for detecting change footprints have emerged from a diverse set of research areas, ranging from time series analysis and remote sensing to spatial statistics. Researchers have much to learn from one another, but are stymied by inconsistent use of terminology and varied definitions of change across disciplines. Existing reviews focus on discovery methods for only one or a few types of change footprints (e.g., point change in a time series). To facilitate sharing of insights across disciplines, we conducted a multi-disciplinary review of ST change patterns and their respective discovery methods. We developed a taxonomy of possible ST change footprints and classified our review findings accordingly. This work allowed us to identify gaps in the research that we consider ripe for exploration, most notably change pattern discovery in vector ST datasets. A paper summarizing these work has been published at the Wiley Interdisciplinary Review: Data Mining and Knowledge Discovery (DMKD) [6].

- **Spatial decision tree for geographical classification.** Given a raster spatial framework, as well as training and test sets, the spatial decision tree learning (SDTL) problem aims to nd a decision tree model that minimizes classification errors as well as salt-and-pepper noise. The SDTL problem is important due to many societal applications such as land cover classification in remote sensing. However, the SDTL problem is challenging due to the spatial autocorrelation of class labels, and the potentially exponential number of candidate trees. Related work is limited due to the use of local-test-based decision nodes, which can not adequately model spatial autocorrelation during test phase, leading to high salt-and-pepper noise. In contrast, we proposed a focal-test-based spatial decision tree (FTSDT) model, where the tree traversal direction for a location is based on not only local but also focal (i.e., neighborhood) properties of the location. Experimental results on real world remote sensing datasets showed that the proposed approach reduced salt-and-pepper noise and improved classification accuracy. Preliminary results have been published at IEEE International Conference on Data Mining [1].

The results from our recent projects show that they could contribute to advancing computer science research in a broader range. For instance, the G-DAG representation and the RTCP algorithm proposed in our work [7] can be viewed as a general framework for partitioning and traversing the embedded space of raster datasets (e.g., images, LiDAR data). This contribution complements the research in data-partitioning based techniques (e.g., spatial clustering) and could potentially solve pattern discovery problems in areas such as image and video processing. In addition, our contributions also exhibit their potentials to transform the general area of scientific research. The interdisciplinary survey [6]we developed provided a way to bring together research ideas from various domains and enabled cross-fertilization of techniques. Researchers in need of finding a suitable change detection method may benefit from our work by learning techniques developed by other domains and further improve them for their own applications.
**Broader Impacts:** The above projects made significant contributions towards (a) creating globally competitive STEM workforce, (b) improving STEM education, (c) enhancing research infrastructure, and (d) improving involvement of underrepresented groups in computer science research. Contributions towards development of competitive workforce in STEM areas is evident from the job placements of our alumni. Doctoral students trained with the fundings from the above projects have been hired not only by leading companies like Oracle (B. George), SAS (P. Mohan), and Microsoft (M. R. Evans), but also by government laboratories research such as National Geo-spatial Agency (J.Kang, PhD). Overall the fundings from recent projects helped in training of over 10 graduate students during the award period.

Using the results obtained in our projects, we contributed towards improving STEM education by enhancing the curriculum of several graduate level courses in the areas of spatiotemporal computing. The graduate level course, titled “Spatial databases (CSci 8715)”, focuses on various aspects spatio-temporal computing and knowledge discovery and is attended by about two dozen graduate students. Dr. Shekhar also presented several invited lectures and keynotes at major conferences such as the Int’l Symposiums on Spatial and Temporal Databases.

Finally, the projects have engaged underrepresented students in research and education. In summer 2012, Ms. Rahni Sumler, an undergraduate student from North Carolina A&T University (a HBCU) had a six-week internship on spatiotemporal data analytics for understanding climate change, as part of the education program of NSF expeditions in computing project. As a report, Rahni presented a poster at the 2nd Workshop on Understanding Climate Change from Data. Ms. Reem Y. Ali joined group in Fall 2012 and is working on cloud computing approach to persistent change window discovery on Earth science data. Mr. Dev Oliver, a Ph.D. candidate, passed his thesis proposal exam in March 2013 on the topic of Spatial Network Activity Summarization.

Apart from creating STEM workforce and improving STEM education, we have participated in the development of softwares. For instance, we participated in the development of CrimeStat, a spatial statistical program for the analysis of crime incidence locations by the National Institute of Justice [2].

**References**


