

Guest editorial: GeoStreaming

Mohamed Ali¹ · Farnoush Banaei-Kashani² ·
Chengyang Zhang³

Published online: 25 January 2017
© Springer Science+Business Media New York 2017

We are entering the era of “big data” thanks to the exponential growth and availability of structured and unstructured data, among which a large amount are real-time streaming data emitted from sensors, imagery and mobile devices. In addition to the temporal nature of stream data, various sources provide stream data that has geographical locations and/or spatial extents, such as geotagging twitter streams, mobile GPS location streams, spatial temporal image streams, and so on. On one hand, this amount of streamed data has been a major propeller to advance the state of the art in geographic information systems. On the other hand, the ability to process, mine, and analyze that massive amount of data in a timely manner prevented researchers from making full use of the incoming stream data. GeoStreaming refers to the ongoing effort in academia and industry to process, mine and analyze stream data with geographic and spatial information.

The purpose of this special issue is to showcase some of the recent developments and novel applications of GeoStreaming. The open call for GeoStreaming has attracted nine papers covering broad range of GeoStreaming technologies and applications. After two rounds of peer-reviews by a team of international experts, seven papers were selected to be included in this special issue.

We start this issue with two papers focused on system design for mobility data analysis [1, 2]. Trajectory analysis is of crucial importance in several fields as social analysis, zoology, climatology or traffic monitoring. Over the last decade, the number of mobile systems and

✉ Mohamed Ali
mhali@uw.edu

Farnoush Banaei-Kashani
FARNOUSH.BANAEI-KASHANI@UCDENVER.EDU

Chengyang Zhang
zhchengyang@gmail.com

¹ University of Washington, Tacoma, WA, USA

² University of Colorado Denver, Denver, CO, USA

³ Amazon, Seattle, WA, USA

devices recording their positions has grown significantly generating a deluge of spatial and temporal data to analyze. This increasing volume of data raises numerous issues in terms of storage, processing and extraction of information. Previous works considering movement analysis have been mainly oriented towards either archived data processing and mining or continuous handling of incoming streams. The first paper [1] introduces the design principles of a holistic approach combining real-time processing and archived data analysis to process mobility data “on the fly”. This solution aims to provide better results comparing to both purely offline and online approaches. This research considers distributed data and processing to be more efficient. The design principles are applied to maritime traffic analysis and a few representative examples are introduced to demonstrate the relevance of our approach. The second paper [2] proposes a framework towards efficient real-time managing and monitoring of mobile objects through distributed spatiotemporal streams processing on large clusters. A prototype implementation is rooted in a new stream processing model that overcomes the challenges of current distributed stream processing models and enable seamless integration with batch and interactive processing like MapReduce.

The third paper in this issue [3] discusses how emerging modern hardware can be leveraged to enable efficient and large-scale GeoStreaming. This paper argues that the key characteristics of the Location-Based Services (LBS) applications include a high rate of time-stamped location updates, and many concurrent historical, present and predictive queries. The commercial providers of LBS must support all three kinds of queries and address the high update rates. While they employ relational databases for this purpose, traditional databases are unable to cope with the growing demands of many LBS systems. Support for spatiotemporal indexes within these databases are limited to R-tree based approaches. Although a number of advanced spatiotemporal indexes have been proposed by the research community, only a few of them support historical queries. These indexing techniques, with support for historical queries, are unable to sustain high update and query throughput typical in LBS. Technological trends involving increasingly large main memory and growing processing core count offer opportunities to address some of these issues. Therefore, this paper presents several key ideas to support high performance commercial LBS by exploiting in-memory database techniques. Taking advantage of very large memory available in modern machines, the proposed system maintains the location data and index for the recent past in memory. Older data and index are kept in disk. An in-memory storage organization for high insert performance is presented along with a novel spatiotemporal index that maintains partial temporal indexes in a versioned grid structure. The partial temporal indexes are organized as compressed bitmaps. With extensive evaluation, this paper demonstrates that the proposed system supports high insert and query throughputs and it outperforms the leading LBS system by a significant margin.

Next, this issue includes two papers [4, 5] focused on efficient implementation of spatiotemporal operations on moving spatial objects. Many natural phenomena are intuitively represented as spatiotemporal data objects, or moving objects. For example, vehicles, rivers, hurricanes, low pressure systems, areas of high density of foliage, etc. align well with a geometric representation, and all change position or shape over time. Moving object models exist that represent real world objects as point, line, and region geometries that change continuously over time, leading to research into spatiotemporal analysis functionality over these objects. Models of moving objects are ideal for representing data streams that record the motion of spatial data over time. However, the implementation of operations to support spatiotemporal analysis over moving objects, particularly over moving regions, has proven

difficult. In the first paper, authors develop a mechanism to support the implementation of the set operations of intersection, union, and difference between pairs of moving regions. The mechanism builds on the Component Model of Moving Regions and the semantic specifications of its operations. Specifically, they develop a generalized method of computing an intermediate data structure from which the results of various operations are then derived. The mechanism utilizes well-known 2D and 3D operational primitives and achieves $O(n \lg n)$ time complexity using appropriate data structures.

The second paper extends the notion of temporal coverage operation that computes the duration that a moving object covers a spatial area. In particular authors extend this notion into temporal coverage aggregates, in which the spatial area covered for a maximum or minimum amount of time by a moving region, or set of moving regions, is discovered. They define the max temporal aggregate coverage operation and the min temporal aggregate coverage operation, and provide an algorithm to compute these operations, and show that it is correct. Finally, the algorithm is implemented in the open source, Pyspatiotemporalgeom library to verify the algorithm under a variety of test cases.

Finally, the last two papers of this issue [6, 7] are dedicated to spatiotemporal even and activity analysis. Messages published via social media sites, such as Twitter, Facebook, and Four-square hide a considerable amount of information about real world events. The timely identification of such events from this huge, unstructured, and noisy user-generated content plays an important role in increasing situation awareness and in supporting useful applications such as recommendation systems. Interestingly, a large number of these messages are enriched with location information, due to the recent advancements of today's location acquisition techniques. This, in turn, enables location-aware event mining, i.e., the detection and tracking of localized events such as sport events, demonstrations, or traffic jams, to name but a few. The main building blocks of a localized event are local keywords that exhibit a surge in usage at the event location. In the first paper [6], authors propose an approach that aims at extracting local keywords from a stream of Twitter messages by (1) identifying local keywords, and (2) estimating the central location of each keyword. This extraction procedure is performed in an online fashion using a sliding window over the Twitter stream. Additionally, they address the problem of spatial outliers that adversely affect a sound identification of local keywords. Spatial outliers occur when people far away from the location of an event use related keywords in their Tweets. The authors handle this problem by adjusting the spatial distribution of keywords based on their co-occurrence with place names that may refer to the location of an event. To ensure scalability, they utilize a hierarchical spatial index to gradually prune the geographic space and thus to efficiently perform complex spatial computations. Extensive comparative experiments are conducted using Twitter data. The analysis of the experimental results demonstrates the superiority of our approach over existing methods in terms of efficiency and precision of the obtained results.

The last paper [7] presents a system for online monitoring of maritime activity over streaming positions from numerous vessels sailing at sea. The proposed system employs an online tracking module for detecting important changes in the evolving trajectory of each vessel across time, and thus can incrementally retain concise, yet reliable summaries of its recent movement. In addition, thanks to its complex event recognition module, this system can also offer instant notification to marine authorities regarding emergency situations, such as suspicious moves in protected zones, or package picking at open sea.

Together these seven papers showcase various aspects of GeoStreaming. We hope that this special issue will be appealing to both experts and practitioners in this area.

References

1. Salmon L, Ray C (2016) Design principles of a stream-based framework for mobility analysis. *GeoInformatica* 1–25. doi:[10.1007/s10707-016-0256-z](https://doi.org/10.1007/s10707-016-0256-z)
2. Galić Z, Mešković E, Osmanović D (2016) Distributed processing of big mobility data as spatio-temporal data streams. *GeoInformatica* 1–29. doi:[10.1007/s10707-016-0264-z](https://doi.org/10.1007/s10707-016-0264-z)
3. Ray S, Blanco R, Goel SAK (2016) High performance location-based services in a main-memory database. *GeoInformatica* 1–30. doi:[10.1007/s10707-016-0278-6](https://doi.org/10.1007/s10707-016-0278-6)
4. McKenney M, Shelby R, Bagga S (2016) Implementing set operations over moving regions using the component moving region model. *GeoInformatica* 1–28. doi:[10.1007/s10707-016-0259-9](https://doi.org/10.1007/s10707-016-0259-9)
5. McKenney M, Frye R, Benchly Z, Maughan L (2016) Operations to support temporal coverage aggregates over moving regions. *GeoInformatica* 1–14. doi:[10.1007/s10707-016-0257-y](https://doi.org/10.1007/s10707-016-0257-y)
6. Gertz M, Abdelhaq H, Armiti A (2016) Efficient online extraction of keywords for localized events in twitter. *GeoInformatica* 1–24. doi:[10.1007/s10707-016-0258-x](https://doi.org/10.1007/s10707-016-0258-x)
7. Alevizos E, Patroumpas K, Artikis A, Voudas M, Pelekis N, Theodoridis Y (2016) Online event recognition from moving vessel trajectories. *GeoInformatica* 1–39. doi:[10.1007/s10707-016-0266-x](https://doi.org/10.1007/s10707-016-0266-x)



Mohamed Ali is an associate professor at the Institute of Technology, University of Washington, Tacoma. Mohamed's research interests include the processing, analysis and visualization of data streams with geographic and spatial information. For the past decade, Mohamed has been building commercial spatiotemporal data streaming systems to cope with the emerging Big Data requirements. In 2006, Mohamed and his colleagues at the database group at Microsoft Research ramped up the Complex Event Detection and Response (CEDR) project. Then, Mohamed joined the SQL Server group at Microsoft to productize the CEDR project. CEDR has shipped and brand-named as Microsoft StreamInsight. Since the first public release of StreamInsight, Mohamed has been advocating for realtime spatiotemporal data management everywhere; that is the use of StreamInsight in monitoring, managing and mining real time geospatial information across a diversity of verticals. These verticals include but are not limited to: online advertising, behavioral targeting, business intelligence, computational finance, traffic management, socialnetworking, homeland security, emergency and crisis management. In 2011, Mohamed started another journey at Microsoft Bing Maps where he became at the frontline with the Big Data challenge and where he battled various types of spatial search queries. In 2014, Mohamed joined the University of Washington, Tacoma where he leads the geospatial data science team at the Center for Data Science.



Farnoush Banaei-Kashani is currently an assistant professor at the Department of Computer Science and Engineering, College of Engineering and Applied Science, University of Colorado Denver. Previously, Dr. Banaei-Kashani was a research scientist at the Computer Science Department, University of Southern California (USC), where he also earned his PhD degree in Computer Science and MS degree in Electrical Engineering in 2007 and 2002, respectively. Dr. Banaei-Kashani is passionate about performing fundamental research toward building practical, large-scale data-intensive systems, with particular interest in Big Data" data-driven decision-making systems. He has published more than 40 referred papers and has received several awards, including a ACM Sensys 2016 Best Paper award and an IEEE CloudCom 2010 Best Paper award. He frequently serves as a program committee member of highly ranked database conferences (including ICDE, VLDB and ACMGIS), and has also co-chaired IWGS, the ACM geo-stream data processing workshop (a subsidiary of ACMGIS) over last few years. Dr. Banaei-Kashani's research has been supported by grants from both governmental agencies (NSF, NIH, DOT, DOJ and NASA) and industry (Google, IBM, Chevron and Intel).



Chengyang Zhang is currently a senior software engineer at Amazon.com leading the design and development work of GeoTech team that solves geospatial infrastructure problems for Amazon. Previously, Dr. Chengyang Zhang was a senior software engineer at Teradata and Microsoft where he worked on database query optimization, big data and cloud computing infrastructure. He earned his PhD degree in Computer Science in 2011 from University of North Texas. Dr. Chengyang Zhang is passionate about mapping problems in the practical industry domain into research issues in the area of geospatial databases, geostreaming data, spatial data mining, database query optimization and big data. He is particularly interested in bridging the domain of big data an spatial databases. He has published over 20 peer-reviewed papers and has received several patents. He frequently serves as organizing committee member of international conferences and also co-chaired IWGS, the ACM geo-stream data processing workshop (a subsidiary of ACMGIS) over last few years.