Spatial Cognition & Computation: An Interdisciplinary Journal

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hscc20

Spatio-Temporal Analytics for Exploring Human Mobility Patterns and Urban Dynamics in the Mobile Age

Song Gao

Department of Geography, University of California, Santa Barbara, CA, USA

Accepted author version posted online: 18 Nov 2014.

To cite this article: Song Gao (2015) Spatio-Temporal Analytics for Exploring Human Mobility Patterns and Urban Dynamics in the Mobile Age, Spatial Cognition & Computation: An Interdisciplinary Journal, 15:2, 86-114, DOI: 10.1080/13875868.2014.984300

To link to this article: http://dx.doi.org/10.1080/13875868.2014.984300

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the “Content”) contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms &
Spatio-Temporal Analytics for Exploring Human Mobility Patterns and Urban Dynamics in the Mobile Age

Song Gao
Department of Geography, University of California, Santa Barbara, CA, USA

In this research, we present a spatio-temporal analytical framework including spatio-temporal visualization (STV), space-time kernel density estimation (STKDE), and spatio-temporal-autocorrelation-analysis (STAA), to explore human mobility patterns and intra-urban communication dynamics. Experiments were conducted using large-scale detailed records of mobile phone calls in a city. The space-time path, time series graphs, vertical Bézier curves, STKDE, STAA, and related techniques in 3D GIS as well as statistical tests have been suggested for different spatio-temporal analysis tasks. We also investigated several statistical measures that extend the classic spatial association indices for spatio-temporal autocorrelation analysis. The spatial order of weighted matrix was found to have more significant effects than the temporal neighbors on influencing the autocorrelation strength of hourly phone calls.

Keywords: human mobility; mobile phone data; space-time GIS; spatio-temporal analytics; spatio-temporal visualization; space-time kernel density estimation; spatio-temporal autocorrelation analysis

1. INTRODUCTION

Despite the fact that humans have keen ability to discover patterns hidden in small-scale data, they may find it difficult to discover large-scale data that often vary over both space and time. Researchers have made great effort on spatial data mining and spatio-temporal visual analytics to raise the cognitive ceilings that often prevent the interpretation of large spatio-temporal datasets (Guo et al., 2006; Shaw, Yu, & Bombom, 2008; Andrienko et al.,...
In the Mobile Age, with the widespread use of location-awareness devices, it is possible to collect large-scale location-awareness datasets, such as mobile phone call data, GPS-enabled taxi trajectories, and social media data, to sense complex human movements and human-environment interactions.

The study of human mobility using emerging Big Geo-Data has been a hot topic in scientific research and many application domains, including but not limited to the analysis of social events, tourism, transportation and urban planning, land and real estate development, migration and health and epidemic studies (González et al., 2008; Calabrese et al., 2010; Song et al. 2010; Kang et al., 2012b; Hawelka et al., 2014; Liu et al. 2014; Pindolia et al., 2014). There are increasing mobile network operators in different countries and cities willing to share their datasets in collaborative projects, such as the Orange “Data for Development Challenge” in the Ivory Coast and Senegal (http://www.d4d.orange.com/en/home). At the same time, it might raise some ethical issues and privacy concerns. Several studies have been trying to investigate various models and techniques for providing important implications to protect individual location privacy (Ho & Ruan, 2011; de Montjoye et al., 2013).

With such data, it would be of great significance to explore and understand how cities function in short-term temporal scales compared with traditional long-term strategic planning in the new era of Big Data (Batty, 2013). For example, although human movements and activities may vary over time across different regions, the observed activity hotspots and information flow might exhibit a pattern of spatial dependence. Also, ignoring the temporal dimension would not be sufficient to discover underlying urban dynamics. For instance, urban governors might hope to understand patterns of human movements by observing the neighboring regions in previous time periods. In such space-time integration contexts, the spatio-temporal analytics should help to answer questions such as:

- Where are the spatio-temporal hotspots of human activities?
- Do human/vehicle movement flows exhibit different spatio-temporal patterns in contrast to overall trips?
- How do these patterns relate to the population distribution and urban land-use structure?
- How to conduct the spatio-temporal autocorrelation analysis?
- What is the impact of spatio-temporal granularity and uncertainty?

To cope with these research questions and problems, spatio-temporal data mining techniques and analytical workflows need to be studied. However, mining big geo-data and discovering knowledge of spatial-temporal relations,
patterns and trends about the real world process are not trivial. Issues on spatio-temporal data organization, computation and integration with visual representation and human cognition still challenge researchers to make great efforts with interdisciplinary knowledge.

There exists a quantity of literature in which space-time integration and quantitative methods for spatio-temporal data analysis have been facilitated and discussed (Dijst, 2013; Long & Nelson, 2013). After examining seven distinct spatio-temporal data types and related work, Goodchild (2013) argued that space–time geographic information system is unlikely to emerge in the near future; and instead, attention should focus on the systematic study of integration.

Richardson (2013) addressed some trends and research challenges of developing real-time space-time functions and capabilities, such as temporal scale, space-time models, and frameworks of spatio-temporal analysis operations to transform geographic data into meaningful information, which are also main issues to be addressed in this study. Stewart et al. (2013) introduced a spatio-temporal conceptual model for scheduling individual activities using an ontological approach. More discussions on space-time integration in Geography and GIScience can be found in the special forum of Annals of the Association of American Geographers [Vol. 103, No. 5, 2013]. In short, very few existing studies present systematic spatio-temporal analytical frameworks and workflows for integrating emerging Big Geo-Data with computational and statistical approaches to guide urban knowledge discovery in practice.

To this end, this article aims at proposing a spatio-temporal data processing and analytical framework, which can be applied not only in exploring dynamic mobility and intra-urban flow patterns based on mobile data, but also in other human and social science research from the emerging big geospatial datasets and computer techniques. Issues on identifying appropriate data structures for different spatio-temporal pattern discovery will be discussed with this framework.

This article is structured as follows: In Section 2, we briefly discuss some related work and propose a spatio-temporal analytical framework that includes spatio-temporal visualization (STV), space-time kernel density estimation (STKDE), and spatio-temporal autocorrelation-analysis (STAA) for exploring human mobility and urban dynamic patterns. The methodology, technical implementation of these analytics will be presented in detail. Then we apply the framework to analyze amounts of geo-referenced mobile phone call records in a city to reveal the spatio-temporal patterns hidden in such big geo-data and further, to understand the complex urban dynamics. The data processing, experiments, main findings, and discussions are presented in Section 3 and 4. We conclude with a summary and directions for further research in Section 5.
2. SPATIO-TEMPORAL ANALYTICAL FRAMEWORK

Modelling human mobility patterns and understanding dynamic urban structures based on a large amount of GPS sensors, mobile devices, persons, vehicles, and street networks have become a hot topic in many fields such as urban planning, transportation, GIScience and computer science (Jiang & Claramunt, 2004; Ratti et al., 2006; González et al., 2008; Kang et al., 2012a, 2012b; Liu et al., 2012; Yuan et al., 2012; Gao et al., 2013b, 2013c; Shen et al., 2013; Yue et al., 2014). In general, the mining and analyzing processes of such spatio-temporal data require combined qualitative-quantitative approaches that involve data extraction and analytics, statistical inference, and geovisualization.

Recently, Long and Nelson (2013) conducted a thorough review on quantitative methods for analyzing movement data which is one prominent type of spatio-temporal data. They classified those quantitative methods into seven groups: (1) time geography, (2) path descriptors, (3) similarity indices, (4) pattern and cluster methods, (5) individual–group dynamics, (6) spatial field methods, and (7) spatial range methods. Because of special data structure requirements and computational complexity, not all of the mentioned methods can be directly applied into different types of spatio-temporal datasets in practice. Identifying or creating appropriate space-time data structures for specified quantitative methods is still challenging.

In this research, we present a spatio-temporal analytical framework (Figure 1), which combines STV, STKDE, and STAA for understanding spatial-temporal patterns (both individual and aggregated) hidden in urban Big Geo-Data. As discussed before, there are not only these three types of spatio-temporal analytical methods available for exploring patterns. However, these three provide relatively simple and efficient ways to extend spatial analysis functions and spatial statistics from 2D space into 3D space and could be integrated well with GIS environments. Each of the proposed analytics has different characteristics and data-format requirements. In the processing, the raw data were converted into appropriate data structures for various analytical purposes. In the following part, we will discuss the roles of different spatio-temporal analytics for the presented research.

2.1. Spatio-Temporal Visualization Techniques for Trajectory and Flow

By utilizing the power of human vision, previous studies have demonstrated the effectiveness of geovisualization in spatial data exploration and knowledge discovery (MacEachren & Kraak, 2001; Kwan, 2004; Guo et al. 2005; Andrienko et al., 2008). The understanding of urban spatial structures can benefit the studies of visualizing individual space-time behaviors (Kwan,
We can use both individual-based visualization and aggregation-based visualization to explore dynamic patterns in urban studies. However, the representation of dynamic human activities and movements over both space and time is one of the major challenges in geocomputation and geovisualization.

Hägerstrand’s time-geometry conceptual framework provides an excellent integrated representation of human movements in space and time (Hägerstrand, 1970). But the space-time cube idea was not applied so widely until the development of GIS-based implementations and analytical discussions about space-time relationships, interactions and uncertainties (Miller, 2005; Shaw, Yu, & Bombom, 2008; Chen et al., 2011; Nakaya, 2013) moved forward, as well as the opportunity to explore potential human activities in both physical and virtual spaces (Yu & Shaw, 2008).

For the individual-based movement representation, a space-time path (3D polyline) was created to connect time-ordered sequence of locations \( <X, Y, T> \) of one person in a 3D-GIS environment which consists of a two-dimensional horizontal geographic plane and one vertical dimension of time (Figure 2). It can be used for visual exploration of continuous spatial and temporal movement patterns. Several analytical models and measurements have been developed by Miller (2005) such as space-time prism, composite path-prisms, stations, bundling and intersections to further analyze the complex spatio-
temporal relationships among human activities and interactions under special space-time constraints.

But when numerous trajectories were collected in the datasets, it was hard to visualize and interpret. Different approaches for generalization and aggregation of massive movement data have been introduced such as traffic-oriented view and trajectory-oriented view (Andrienko & Andrienko, 2008). Spatial clustering approaches have been developed for discovering places of interest and similar routes from human movement trajectories (Palma et al., 2008; Giannotti et al., 2011). In addition, sequence alignment method, which was originally employed by biochemists to analyze DNA sequences, has been applied in analyzing the space-time sequential aspects of human activities (Wilson, 1998; Shoval & Isaacson, 2007; Mavoa et al., 2011; Kwan et al., 2014). It could help the identification of individual-based spatio-temporal recurring trends and group-based similarity patterns.

In the urban context, the aggregation of massive human movement trajectories by origins and destinations (OD) can be utilized to understand the dynamic OD-flow patterns among traffic analysis zones (TAZs) or other polygonal divisions of region in different temporal scales. Traditional flow mapping is used for representing the amount and the direction (with arrow symbol) of from-to movements of human or things among regions in a 2D space, such as migration and goods trade (Tobler, 1987). Some graph layout optimization algorithms and aggregation strategies have been suggested to minimize the edge crossings between flow symbols (Phan et al., 2005; Andrienko & Andrienko, 2011).

Here we introduce another approach of using vertical Bézier curves in 3D-GIS environment for interactive visual exploration of information or movement
flows between places. The main advantages of such an approach lie in the integration of 3D visualization techniques which support interactions between 3D geometry objects and OD-flow values in multiple time snapshots or in a continuous animation.

A Bézier curve is defined by a set of control points \( P_0 \) through \( P_n \), where \( n \) represents called its order (\( n = 1 \) for linear, \( n = 2 \) for quadratic which is used in our work, etc.). Bézier curves have been widely used in computer graphics and geometry designs (Farin, 1996). We develop an algorithm to approximate the quadratic Bézier curves. As shown in Figure 3, the first point \( P_0 \) and the last point \( P_2 \) are used to represent the centroids of two regions (i.e., the origin and the destination of each flow) and the intermediate control points are interpolated by the functions as follows:

\[
B(t)_{\text{longitude}} = P_0 + t(P_2 - P_0) \tag{1}
\]

\[
B(t)_{\text{latitude}} = \text{Slope}(P_0, P_2) \cdot B(t)_{\text{longitude}} + b \tag{2}
\]

\[
B(t)_x = 2(1 - t)t*\text{Dist}(P_0, P_2)/2 + t^2*\text{Dist}(P_0, P_2) \tag{3}
\]

\[
B(t)_y = 2(1 - t)t*H \tag{4}
\]

Figure 3: Drawing and visualizing vertical Bézier curves in 3D-GIS environment (ArcScene).
\[ B(t)_h = \sqrt{B(t)_x^2 + B(t)_y^2} \]  \hspace{1cm} (5)

\[ t = \frac{k}{N} | k \in (1, 2, \ldots, N - 1) \]  \hspace{1cm} (6)

The parameter \( t \) is determined by the step \( k \) and the number of interpolate points \( N \) to be inserted between each OD pair. To control the shape of a curve, we took a point \( P_1 \) on the perpendicular bisector of the chord linking \( P_0 \) and \( P_2 \). The height \( H \) represents the amount of the flow. \( \text{Dist}(P_0, P_2) \) is the Euclidian distance between the two points. After getting all the control points \( B(t) \), we connected all of them together to draw a vertical Bézier curve and then project a flow between regions in the 3D-GIS environment. We wrote a Python script to generate all Bézier curve controlling points based on the OD-flow matrix table and linked them in Esri’s ArcScene software for further interactive exploration of information flow or physical movement flow patterns.

2.2. Space-Time Kernel Density Estimation

As discussed above, the temporal information of movements in geographic space is important to detect the spatio-temporal trends of underlying human mobility. But with the increasing number of aggregated human/vehicle trajectories in urban space, the space-time path representation model will be hard to interpret because of the overlapping and cluttering issues. To solve this problem, an extension of kernel density estimation (KDE) (Silverman, 1986) was suggested. The KDE has been widely used in spatial analysis to characterize a smooth density surface that shows the geographic clustering of point or line features in 2D space.

To incorporate the time information, the space-time kernel density estimation (STKDE) can be taken as a generalization approach of the 2D-space KDE into the 3D space-time cube which can support the exploration of spatio-temporal patterns, clusters and changes. Such STKDE techniques have been used in several studies, such as crime clustering analysis (Brunsdon et al., 2007; Nakaya and Yano, 2010), trajectory data mining (Demšar & Virrantaus, 2010), publication citation analysis (Gao et al., 2013a), and dengue fever pattern discovery (Delmelle et al., 2014). The STKDE value of each voxel (volumetric pixel) in the three-dimensional space-time cube is estimated as:

\[ D(x, y, t) = \frac{1}{nh_s^2h_t} \sum_i K_s\left( \frac{x-x_i}{h_s}, \frac{y-y_i}{h_s} \right) K_t\left( \frac{t-t_i}{h_t} \right) \]  \hspace{1cm} (7)

where \( D(x, y, t) \) is the density estimation of each voxel based on the data in neighboring volumetric pixels; \( n \) is the number of point events, and \( h_s \) and \( h_t \) are the spatial and temporal neighboring bandwidths.
Each point in the neighboring pixels is weighted based on the proximity in both space and time to the voxel using kernel functions \((K_s \text{ and } K_t)\). In this study, the Epanechnikov kernel was used for multivariate probability density estimation within the bandwidths (Epanechnikov, 1969). Similar to the 2D spatial KDE, larger bandwidths may result in smooth surface, yet smaller bandwidths may result in the lack of trending patterns, so we needed to calibrate both spatial and temporal bandwidths of STKDE based on the experiments with actual datasets. In practice, the data-driven approach could be applied by considering the least squares cross-validation for kernel-bandwidth selection (Sheather & Jones, 1991). More discussions on bandwidth selection methods for kernel density estimation were presented in a review by Jones et al. (1996).

The results of STKDE are volume data, i.e., 3D-grids. Direct visualization of such STKDE would require four-dimensional space because of their volumetric data structure consisting of 2D geographic space, time and another one for the density estimation scalar. Such volume visualization is not very common in GIS but very popular in geophysics, geology, medical science, and in computer graphics (Kaufman, 1990). The three main approaches for volume visualization were discussed by Demsˇar & Virrantaus (2010): (1) direct volume rendering by assigning color and transparency to voxels; (2) isosurface that is the equivalent of isoline connecting points of equal value on a two-dimensional map; and (3) volume slicing by planes. We apply the volume slicing approach with color schema and transparency to the voxels regarding the consistency of KDE visualization in GIS.

### 2.3. Spatio-Temporal Autocorrelation Analysis

Analyzing spatio-temporal autocorrelation structures of human activities would be helpful to understand the urban dynamic patterns in space and time simultaneously. In statistics, autocorrelation can be taken as the correlation of a variable with a lagged specification of itself (Box et al., 2008). The temporal autocorrelation can be defined as the correlation of the same variable \(X\) between values at different time \(s\) and \(t\).

\[
R(s, t) = \frac{E[(X_t - \mu)(X_s - \mu)]}{\sigma^2}
\]

(8)

\(E\) is the expected value operator, \(\mu\) is the mean of the observation values and \(\sigma^2\) is the variance. The temporal autocorrelation can be used to explore the time-series autocorrelation patterns.

With regard to the spatial dependence, spatial autocorrelation (association) statistics have been used to analyze the degree of dependency among observations in a geographic space (Cliff & Ord, 1973). These measurements
can be divided into two categories: global indices and local indices. Classic
global indices of spatial autocorrelation include Moran’s I (1950), Geary’s C
(1954), and Getis–Ord’s General G (1992), while local indices of spatial
association (LISA) can be established by transforming the global indices into
corresponding local measurements based on different measures of similarity
(Anselin, 1995).

All of these spatial autocorrelation statistics require a spatial weights
matrix that reflects the intensity of the geographic relationship between
observations and their neighbors, e.g., the distance-to-neighbor matrix or the
binary matrix in which the element value is 0 or 1 determined by whether there
is a shared boundary between the observation location and neighbors.
As suggested by Hardisty and Klippel (2010), adding the temporal neighbors
into the weights matrix would be one approach to extend the traditional spatial
autocorrelation measurements. Griffith (2010) gave an overview of spatio-
temporal modelling techniques including autoregressive-integrated-moving
average models, space-time autoregressive models, geostatistical models, and
panel data models, as well as proposed spatial filtering models. All of those
models have been motivated by considering the spatio-temporal associations
simultaneously. Cheng et al. (2012) further extended both global and local
spatio-temporal autocorrelation analysis onto the road network data for
complex traffic analysis.

Here, we present three global measures of spatio-temporal association
regarding the Moran’s I, Geary’s C, and Getis-Ord’s General G:

\[
I_{st} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} [z_i(t) - \bar{z}_t][z_j(t + \tau) - \bar{z}_{t+\tau}]}{\sigma_t \sigma_{t+\tau} \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}} \tag{9}
\]

\[
C_{st} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} [z_i(t) - z_j(t + \tau)]^2}{2 \sigma_t \sigma_{t+\tau} \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}} \tag{10}
\]

\[
G_{st} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} z_j(t)z_j(t + \tau)}{\sum_{i=1}^{N} \sum_{j=1}^{N} z_i(t)z_j(t + \tau)} \tag{11}
\]

Where \(I_{st}, C_{st},\) and \(G_{st}\) can be taken as different formats of space-time cross-
correlation (or cross-product) models (Getis, 1991); \(Z\) is the target variable of
interest; \(i\) and \(j\) are indices of total \(N\) spatial units; \(w_{ij}\) is an element of the k-
order-neighbor spatial weighted matrix \((1^{st}, 2^{nd}, \ldots, k^{th}); \bar{z}_t\) and \(\bar{z}_{t+\tau}\) are the
means of variable \(Z\) within a time lag, while \(\sigma_t\) and \(\sigma_{t+\tau}\) are the variances. The
local measures of spatio-temporal autocorrelation can be derived by
decomposing a global measure into particular spatial neighboring units.
In the experimental section, we evaluate how different spatial and temporal neighbors (lags) affect the results of three spatio-temporal autocorrelation measures for the mobile phone call activities in a city.

3. DATA PROCESSING

In this research, the dataset contains a week of about 74,000,000 anonymized mobile phone call detail records (CDR) in a large city from a Chinese telecommunication operating company. The CDR data lists the information of caller, receiver, mobile base stations, date, time, duration et al. (Table 1). As shown in Figure 4, every time when a user (caller/receiver) made a call, he/she was geo-referenced to a corresponding mobile base station that has a unique longitude/latitude position. The coverage area of each mobile base station can be expressed as a Voronoi polygon for call activity analysis and termed as a “cell.”

<table>
<thead>
<tr>
<th>Caller</th>
<th>Receiver</th>
<th>Date</th>
<th>Start Time</th>
<th>End Time</th>
<th>Duration (seconds)</th>
<th>Base Station</th>
<th>Long</th>
<th>Lat</th>
</tr>
</thead>
<tbody>
<tr>
<td>serveNub1</td>
<td>oppNub1</td>
<td>2007-07-23</td>
<td>09:25:10</td>
<td>09:28:20</td>
<td>190</td>
<td>A</td>
<td>127.495</td>
<td>50.243</td>
</tr>
</tbody>
</table>

Figure 4: Spatial distribution of mobile base stations and an illustration of phone call interactions among cells.
this Voronoi partition, all phone calls within a given polygon are closer to the corresponding mobile base station than any other station. Generally, urban central regions have a higher density of mobile cells (the coverage area of each cell is smaller) than the outer suburb regions. Different data subsets have been extracted and processed for various research purposes such as human mobility modelling (Kang et al., 2012b), travel behavior studies (Yuan et al., 2012), population estimation (Kang et al., 2012a), and urban community network structure analysis (Gao et al., 2013b).

In this work, on one hand, for individual spatio-temporal mobility pattern analysis, the corresponding approximate movement trajectories of each mobile subscriber were created by connecting his/her series of geo-referenced call records in space and time. On the other hand, for urban dynamic structure analysis, phone call activities and phone users’ physical movement flows were aggregated based on the Voronoi polygons. More detailed spatio-temporal analysis and results are presented next.

4. EXPERIMENTS

4.1. Individual Spatial-Temporal Movement Patterns
In urban studies, the understanding of human movement patterns over time is very important in transportation planning and city management. Figure 5 illustrates the 3D STV of a person’s mobile tracking trajectories in space over one week. Each color represents a single mobile subscriber. Different shapes of 3D polylines reveal different space-time behaviors for specified users. For instance, the regular patterns at fixed locations (e.g., home and job places) with respect to the daily/weekly cycles can be indicated by the length of vertical segments of the space-time paths (Figure 5a). In addition, irregular movement patterns have been also found for certain type of individuals (Figure 5b). One

![Figure 5: Space-time visualization of mobile phone users’ trajectories: (a) regular movement patterns of three individuals; (b) irregular movement pattern of an individual.](image-url)
might guess that the user is a delivery employee or working for related professions.

The STV technique can help to simultaneously visualize the mobile trajectory patterns in space and time with an intuitive manner but may need the statistical analysis to extract more meaningful personal places of interests (POIs). Figure 6a displays the spatial distribution of a mobile user’s call-activity movement in a week. A visit was recorded when the user made a call on a cell. The visit probability was calculated as the ratio of visiting frequency in a given cell to the total number of visits across the whole cell network. It tells that this person has a higher probability (Cell A: 0.31, Cell B: 0.22; Cell C: 0.14) to frequently visit or stay in a few places, while the visiting probability of most cells is less than 0.05. The interpretation of time series graphs of phone calls on these visited places by the user offers the possibility of identifying the personal POIs based on their temporal signatures.

As shown in Figure 6b, the user often made phone calls at Cell A and Cell B during 19:00–24:00 throughout the whole week. One can infer that his/her home might be located nearby or inside the boundary of the two cells. Also, the temporal signature of Cell C may indicate a working place since the user is present there only during office hours (9:00–18:00) on weekdays but not on weekends. However, the detailed information about personal characterized POIs and actual frequent visited places may need further investigation on daily-activity trips of the mobile subscribers or the study of geographical contexts, since the CDR-based trajectories are only abstract presentations of the actual trajectories of individuals at the cell-tower coverage scale. Generally, urban core areas have a higher density of cell-tower coverage (the distance between two mobile base stations is about 100–500 meters), whereas it is approximately 2 kilometers or larger in the suburbs.

To identify the spatio-temporal hot-spots of individual phone call locations, we calculated the space-time kernel density estimation (STKDE) of each user’s call activities. Figure 7a shows an individual’s mobile phone call activities over a week in a space-time cube. The STKDE of this user’s call activities was calculated in a spatial resolution of 500 meters and a temporal resolution of 100 minutes. The different combinations of spatio-temporal bandwidths result in various visual representations. The selection of bandwidths needs several rounds of calibration by adjusting both spatial and temporal bandwidths to find an optimization that can help to uncover hidden patterns.

In practice, we tried different combinations and finally chose twice the spatial resolution (1 km) as the spatial bandwidth for both horizontal dimensions, and 500 minutes as the temporal bandwidth for the vertical dimension, which gave us a better visual representation for identifying patterns in the case study. The resulting density volume of 50*50*100 voxels for a user’s call activities is visualized in Figure 7b. One can find that this user is more likely to make calls at a fixed spatial region (red and orange voxels)
Figure 6: Spatial distribution and temporal signatures of individual frequently visited mobile cells: (a) Spatial display of cell visiting probability for a user (the larger circle represents higher probability); (b) The time series graphs show the temporal signatures of the user’s frequent visited places (mobile cells) based on the number of hourly calls over a week.
Figure 7: Visualizing georeferenced phone call activities and space-time density: (a) phone call events in space-time cube; (b) STKDE results for a specified mobile user in a week and the kernel density value is a normalized value.
across time (including Monday night, Thursday afternoon and evening, Saturday, and Sunday evening). This example shows that it is much easier to visually identify the spatio-temporal hot-spot patterns of call activities by using the STKDE approach.

4.2. Aggregated Phone-Call Interaction Patterns

One of the prominent characteristics of phone call data lies in its indication of human communications and spatial interactions among different places. The aggregated phone call interactions among mobile cells in different time represent the dynamic intra-urban communication landscape that cannot be captured using other traditional activity survey data. As shown in Figure 8, the vertical Bézier curves were created to represent the hourly phone-call flow across cells. The height of the arcs represents the relative volume of phone calls. The tall and narrow arcs show strong call communication within a nearby intra-urban space; yet, some long-distance curves across nonadjacent cells indicate strong call interactions among these spatially separated regions.

It is apparent to observe some patterns: (1) there are only several long-distance interactions at midnight while more communications are concentrated in the urban center; (2) very few call interactions exist from 3 a.m.–6 a.m.; (3) there are increasing urban central interactions in the period 6 a.m.–9 a.m. and keeping high-volume interactions across the core region till night; (4) the call interactions decrease after 9 p.m. Such information flow patterns are strongly related daily human activity rhythms and working-social connections, as well as geographical contexts including urban land-use types and spatial distributions of home-job locations (Ratti et al., 2006; Gao et al., 2013b).

To understand the dynamic “source–sink” structures of information landscape (Liu et al., 2012), we calculated the phone call net-balance-flow for each mobile cell by subtracting the outgoing call volume from the incoming call volume in each hour. Figure 9 shows the time-series plot of phone-call net flow among all cells. Each line represents the net flow pattern for a specific Voronoi mobile coverage area. In Figure 10, it is clear to see dynamic spatial distributions of the “source” areas (red color), which have more outgoing phone calls, and the “sink” areas (blue color), which have more incoming phone calls in different hours. The yellow cells mean that the net-balance call flow in that hour is zero. One can interactively interpret the call flow patterns in the 3D-GIS environment or sense the dynamic urban phone-call landscape under the animation mode.

4.3. The Spatio-Temporal Autocorrelation Patterns of Phone Calls

The study of spatio-temporal autocorrelation structure of mobile phone calls in urban space can help to understand the citizens’ mobile communication
Figure 8: Vertical Bézier curve 3D visualization of hourly phone-call flow patterns across cells in a day. (Green dots are locations of mobile base stations; each arc represents an OD flow linking two mobile cells).
Figure 9: Time-series plot of phone-call net flow among all mobile cells (each line represents a different mobile cell).

Figure 10: Visualizing the dynamic phone-call “source” areas (red), “sink” areas (blue) and “zero-balance” areas (yellow) in 3D-GIS environment.
patterns and urban structures. In order to investigate how the spatial autocorrelation structure changes throughout the day in this city, the phone-call volume was aggregated into the Voronoi cells by hour at first. Then, the Moran’s I local indicator of spatial association (LISA) (Anselin, 1995) was calculated for each cell every three hours (see Figure 11). The value of Moran’s I has been standardized to lie in \([-1, 1]\). If the index is larger than 0, the cell shows positive spatial autocorrelation with its neighbors; yet, it indicates the negative spatial autocorrelation of phone call patterns if the index is smaller than 0. The closer the value is approaching to 1 (or \(-1\)), the stronger the positive (or negative) spatial autocorrelation is. Examining the spatial structure of LISA in different time periods, we can clearly see that the spatial autocorrelation patterns of phone calls across all cells are very dynamic and

Figure 11: Local Moran’s indicators of spatial autocorrelation (LISA) analysis on phone calls in different time periods of a day (using the 1st-order-neighbor spatial weighted matrix).
heterogeneous. The central region (small cells) shows more diverse patterns than the outer suburb areas, where most spatially adjacent cells show similar values in the whole day. It might reflect the mixture land-use types of urban central areas and human’s convergence and divergence in this place with different frequencies and purposes of phone call behaviors in different time periods.

To identify a more stable autocorrelation structure, we apply the spatial statistic test of running 10,000 simulations of randomized permutations of neighboring cells to find the local significant spatial autocorrelation patterns (Anselin, 1995). As shown in Figure 12, we generated both Moran’s scatter plot and spatial distribution of the labelled mobile cells with IDs to visually and interactively identify the statistically significant local association cells (with a 0.05 significance level). By comparing the phone call volume of each cell with its first-order adjacent neighbors, there are four types of associations identified: (1) HH: observations in both the target cell and neighbors are high; (2) HL: high call volume in the target cell with low volume in neighbors; (3) LH: low call volume in the target cell with high volume in neighbors; (4) LL: the call volume in both the target cell and neighbors are low. Figure 12a illustrates the results of significant Moran’s I LISA in the period 3 a.m.–6 a.m.

There are continuous significant LL spatial associations in the urban central region and HH association structure in the southwest and northeast suburb regions which contain large residential housing referring to the Google Earth imagery in this city (The imagery with labels is not shown here as required by the data provider). But such spatial autocorrelation structure changes over time, for instance, the central area tends to have mixture patterns of all HH, HL, LH, LL association types in the period 9 a.m.–12 p.m. (See Figure 12b). The findings help the understanding of local citizens’ mobile communications at the spatial and temporal dimensions simultaneously. In addition, the spatio-temporal autocorrelation structure helps the evaluation of current land-use in the study area, e.g., which cells are “the hot regions” all the day or at a given time-period. It provides important implications for real estate development and business locators. The findings could help the local government and planning agency to make future strategic planning and support their decision making.

The spatial weighted matrix plays an important role in spatial autocorrelation analysis (Getis & Aldstadt, 2010). As shown in Figure 13, a larger-order of spatial adjacency tends to have a larger number of neighbors. Another important factor for identifying spatio-temporal autocorrelation structure is time granularity (e.g., half an hour, per hour, two hours and others). It inspired us to examine how the different combinations of spatial weights and temporal neighbors affect the STAA results.

Using the methodology introduced in Section 2.3, we implement the global Moran’s I like statistic of STAA with different spatial lags and time lags for
Figure 12: Moran’s scatter plot and spatial display of mobile cells with statistically significant local associations (HH: High-High; HL: High-Low; LH: Low-High; LL: Low-Low): (a) period 3 a.m.–6 a.m.; (b) period 9 a.m.–12 p.m.
hourly phone-call patterns across all cells. Examining the results reveals two key findings (See Figure 14a). First, the strength of global Moran’s I, such as spatio-temporal autocorrelation measure ($I_{st}$), for hourly phone calls is temporally dynamic and there is a positive-association peak between 6 a.m. – 7 a.m.. Second, the $I_{st}$ measure is more sensitive to the spatial order than the temporal neighbors. A higher order of spatial weights generally results in higher strength of spatio-temporal autocorrelation structure.

In addition, we implement the global Geary’s C like STAA measure ($C_{st}$) and Getis-Ord’s G like STAA measure ($G_{st}$) to identify the spatio-temporal autocorrelation of hourly phone-call patterns in different hours. Note that the $C_{st}$ statistic indicates a positive autocorrelation structure when the value lie in
Figure 14: Three global measures of spatio-temporal association with different combinations of spatial weights (spatial orders) and temporal neighbors (1 time-lag: 1 hour; 2 time-lag: 2 hours; 3 time-lag: 3 hours) for hourly phone-call patterns: (a) \( I_{st} \); (b) \( C_{st} \); and (c) \( G_{st} \) measures.
(0–1); yet, it is a negative autocorrelation when the value lie in (1–2). It is found that the hourly autocorrelation trends of $C_{st}$ measures are more similar to the $I_{st}$ measures (Figure 14b). But using the $G_{st}$ measure did not reveal the temporal dynamics of autocorrelation strength in our datasets. The $G_{st}$ statistic is also sensitive to the spatial-order of weighted matrix (Figure 14c).

5. CONCLUSIONS AND FUTURE WORK

In this article, we introduced a spatio-temporal analytical framework for exploring human mobility patterns and urban dynamics. The integration of spatial-temporal visualization, space-time density estimation and spatio-temporal autocorrelation analysis can not only help to represent spatio-temporal data visually and interactively but also offer quantitative analytics to identify the spatio-temporal patterns (such as spatio-temporal hotspots) in the mobile phone data. Our experiments have demonstrated that different spatio-temporal techniques have their potential advantages but also limitations. Expeditions on the study of spatio-temporal analytical techniques can contribute to the future development of space-time GIS.

For instance, the space-time path is good for visual exploration or an overview of individual regular (or irregular) movement patterns but might not be suitable for massive trajectories because of overlapping and cluttering problems in the space-time cube. The study also demonstrates that the user’s “home” and “working place” can be inferred at the cell-tower coverage scale based on statistical information of frequently visited mobile cells with the characteristics of personal temporal signatures on these places. However, because of the mobile phone data limitations on the spatial scale which heavily depends on the coverage size of a cell tower, it is difficult to tell which POIs at a finer spatial resolution are actually visited or frequented by a particular individual. It might need the integration of phone-call data with survey or interviews on these mobile subscribers for a holistic view of personal activities. In addition, we implement the STKDE to analyze the continuous spatio-temporal density structure for the georeferenced phone call activities and this technique can facilitate the identification of spatio-temporal patterns simultaneously.

Furthermore, to understand the phone-call dynamic spatial interaction structure, we introduce a novel 3D flow visualization approach of generating vertical Bézier curves in the 3D-GIS environment which also support interactive analysis of spatial information and thematic attributes. Moreover, we have investigated different statistical measures ($I_{st}$, $C_{st}$ and $G_{st}$) which extended the classic spatial association indices for the spatio-temporal autocorrelation analysis. The spatial order of weighted matrix was found to have more significant effects than the temporal neighbors on influencing the
autocorrelation strength of hourly phone-call volume across the whole study area.

The spatio-temporal analytical framework introduced in this paper can be also applied in other spatio-temporal datasets (e.g., infectious diseases, crimes, and GPS tracks) for facilitating knowledge discovery and decision support in urban informatics and social sciences. However, there are several related research issues that still require further work: (1) Although the integration of space-time visualization and spatio-temporal statistics can help understand the individual human mobility patterns and identify limited types of personal place of interests (e.g., home and working place), more fruitful semantic information of locational activities need further investigation on the socio-economic background of the mobile subscribers and geographical context analysis in this region. (2) The results of STKDE computation are sensitive to the selection of spatial and temporal bandwidths. More work is needed to better understand the impact of both factors and may find an optimal solution. (3) The spatio-temporal uncertainty issues in both STKDE and STAA should be studied. (4) Last but not least, the effectiveness of spatio-temporal visual analytics should be evaluated by human participants to find which representations are more appropriate with respect to certain types of spatio-temporal data for pattern identification and knowledge discovery.

ACKNOWLEDGMENTS

The author would like to thank Dr. Yu Liu at Peking University for providing data support and thank colleagues Dr. Krzysztof Janowicz, Dr. Phaedon C. Kyriakidis and Yingjie Hu at UCSB Geography for insightful discussions and suggestions on this work. Thanks to guest editors Dr. Christophe Claramunt and Dr. Kathleen Stewart and reviewers for their comments that helped to improve this manuscript.

REFERENCES


