Research Article

TRR
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

Transportation Research Record
1–15

DOI: 10.1177/0361198120980437
journals.sagepub.com/home/trr
SAGE

User Activity and Trip Recognition using Spatial Positioning System Data by Integrating the Geohash and GIS Approaches

Hafez Irshaid[1], Md Mehedi Hasan[2], Raed Hasan[3], and Jun-Seok Oh[4]

Abstract

Analyzing travel behavior in transportation networks within a city is significant to understand the user's activity and travel pattern in relation to making improved city plans for the future. Unlike the traditional travel diary survey, GPS data have helped researchers to analyze Big Data with enriched travel information in an automated way. The focus of this research was to identify user activity and travel pattern from GPS data logs. We proposed three different approaches, including Geohash clustering, the GIS-based approach, and Combined Geohash–GIS approach, for automatic user activity and trip recognition in a continuous and aggregate manner. We developed different individual models considering different dwell times for the above three approaches. We considered three different testing scenarios based on specified tolerance levels, including simple, moderate, and critical testing to identify trip only, activity only, and sequential activity–trip analysis. In comparison with other approaches, the Combined Geohash–GIS approach considering 5 min dwell time accurately classified data with about 95% accuracy. The proposed Combined Geohash–GIS approach could significantly enhance the efficiency and accuracy of GPS travel surveys by correctly recognizing user activity and trip patterns. This proposed combined approach could serve as a foundation for a future model system of full-scale travel information identification with GPS data.

Travel data related to transportation users' activity and trip information help in promoting transportation plans, projects, and policies in urban areas (*1*). The activity data are typically concerned with user behavior associated with the built-up environment of cities. The data help in detecting the movements along with the travel patterns of persons, goods, and information in a given or possible future environment (*2*, *3*). In recent years, smart cities have been characterized by their reliance on big and continuous user activity data. Recent advances in artificial intelligence and communication technologies are capable of collecting big travel data based on ubiquitous and location-aware smartphones (*1, 4–6*).

An individual's travel itinerary comprises a sequence of different activities and trips for a given time period (*3*). An activity could be defined as an event that occurs for a certain amount of time within the same socio-spatial environment, whereas a trip is defined as the movement of persons, goods, or information between different spatial–temporal environments (*2*). Activities are characterized based on the potential purposes of an individual's daily lifestyle (i.e., work activity, shopping activity, etc.). On the other hand, trips are the connecting events or the continuous sequence of stages between different activities by using different transportation modes. Activity and trip recognition of the travel itinerary is an important step for transport modeling and predicting travel behavior (*2*, *7*).

Over the last two decades, different methods and approaches have been practiced for predicting travel behavior and recognizing user activity and trip information. Some of the traditional methods used in previous studies include paper-based face-to-face interviews, activity/travel diaries, mail-back paper surveys, and so forth (*8*, *9*). However, those traditional surveys were limited in relation to short-term report-based study, including other

[1]General Motors Company, Milford, MI
[2]AECOM Technical Services Inc., Mt. Clemens, MI
[3]University of Samarra, Samarra, Saladin, Iraq
[4]Department of Civil and Construction Engineering, Western Michigan University, Kalamazoo, MI

Corresponding Author:
Hafez Irshaid, hafezkm.irshaid@wmich.edu

shortcomings such as trip under-reporting time inaccuracies, labor-intensive procedure, origin–destination location errors, and information bias resulting from subjective judgment and fatigue/forgetfulness (*1*, *10*, *11*). To overcome the above limitations, traditional paper-based surveys were replaced by online and computer-assisted programs, such as computer-assisted telephone interviews, computer-assisted self-interviews, web-based and internet surveys, and so forth (*12–14*). Although internet-based travel surveys are capable of saving time and money by incorporating automatic branching in questionnaires, this approach suffers from different shortcomings, such as sampling bias and representativeness of sample size, lower response rates, non-response and misreporting, and so forth (*8*, *15*, *16*).

In recent years, traditional transportation user data collection approaches have been replaced by automatic and digital methods following recent advances in information and communication technologies (ICT) (*6*, *17*, *18*). Global Positioning System (GPS) data logs are used for tracing human activity data through location sensors (*19*). To date, most people have used their smartphones, which were embedded with GPS systems that collect location points. GPS surveys are considered as a potential alternative to traditional surveys because of the accuracy in collecting spatiotemporal travel datasets along with the data validation approach by using in-app or prompted-recall applications (*20–24*). These opportunities, features, and collection of location data through GPS techniques and smartphone devices have prompted researchers to build different approaches that monitor activities through travel pattern analysis, such as rule-based algorithms, statistical methods, and machine learning methods (*19*, *25*). Rule-based algorithms were used in a variety of studies to detect the travel activity and trip pattern from spatial–temporal GPS datasets (*26–28*). For example, Bohte and Maat conducted a GPS-based travel survey to recognize the activity pattern through applying a rule-based algorithm by using different attributes, such as Geographic Information System (GIS) land-use data, home and workplace/school addresses, and so forth (*28*). In addition, different machine learning models were applied to detect activity and trip information from the GPS trajectory dataset, such as random forest (RF) (*29–31*), decision tree (*32*, *33*), neural networks (*21*), long short-term memory (*34*), and so forth. For example, Wu et al. performed a GPS-based study to automatically detect the user activity types through applying the RF model by using distance, speed, and acceleration datasets (*29*). Pereira et al. conducted a study to recognize user activity and trip pattern by applying the historical data-matching rules approach with the help of the GPS trajectories dataset along with user activity duration, point of interests, socio-demographics, and work hours' travel time data (*35*). In another study, Zhang et al. performed a GPS-based survey to recognize user activity through applying the sequential model-based clustering method by using visiting frequency, most frequently visited locations, distance between visited locations, and the relation between a location and its surrounding environment (*36*). In addition to GPS trajectories, the Light Detection and Ranging (LiDAR) dataset was used to predict traffic flow and user activities by analyzing real-time spatial–temporal information (*37*).

In addition, giant software companies in computer fields and other digital mapping companies have also created numerous location-based service applications by using Foursquare or Google Maps to understand transportation activity/travel patterns in real time (*1*, *38*, *39*). Furthermore, many applications have been applied for detecting spatial patterns of user travel activities through mapping techniques resulting from spatial positioning systems and locational data analysis (*40–43*). The GIS is one of the critical methods related to travel pattern behavior and transport modeling purposes, and analyzes the spatial relationship of transportation users (*44*, *45*). For example, Stenneth et al. explained the possibilities of determining transportation activities provided through GIS maps with GPS data by identifying speed and acceleration data (*46*). Domènech et al. developed a methodology to assess the effectiveness and spatial coverage of travel patterns in Spanish tourist cities through the GIS system (*47*). Also, Loidl et al. attempted to develop a relationship between activity and travel pattern through the GIS approach by applying geospatial data with geovisualization (*48*). In recent times, the Geohash technique was used in different sectors including spatial modeling studies for business (*49*), mobile sensing (*50*), spatial query (*51*), and spatiotemporal mapping (*52*). The Geohash method was used in the study of Singh et al. for the purpose of investigating the flow orientation of major activity regions in South Korea (*53*). In addition, the Geohash method was used in the study of Oh et al. (*54*), which is close to the idea of Singh's research, to assess spatial movement patterns of smart card transaction data for multi-modal transportation networks. However, to date, the Geohash method has not been used for detecting travel behavior data, especially for recognizing activity or trip information (*55*).

Research in travel surveys and recognizing activity patterns has come along a long way, and gained maturity in relation to different approaches and methodologies. However, the current methods suffer from different limitations, such as information bias, inconsistent datasets, sources of the data feed, small-scale datasets, labor-intensive characteristics, and so forth. Moreover, efforts have concentrated on one aspect of travel data detection (i.e., mode detection) at a time, rather than recognizing sequential activities and trips from a travel itinerary.

Therefore, this study aims to recognize the activity and trip with different thresholds of spatiotemporal change by applying the Geohash clustering approach and the GIS-based approach, forming a combined approach by integrating the Geohash and GIS systems. The above approaches were developed and implemented for activity only, trip only, and sequential activity–trip recognition with GPS data as a case study for the city of Kalamazoo, Michigan.

Data and Methodology

In this section, we discuss the major methods of data collection and the approaches that were applied in this study to recognize activity/trip information. The study area of this paper was chosen for the city of Kalamazoo, Michigan, which has an integrated urban land-use pattern and spatial variation in its transportation network.

Data Collection

This research considered data from the Transportation Research Center for Livable Communities' (TRCLC) Fitbit project. A comprehensive dataset for more than 60 users' daily activities for a 12-month period, ranging from January to December of 2018, was considered. The study took advantage of GPS survey data collected by smartphone application. Each respondent had an average of about 100,000 records for the 12-month study period. With missing values being eliminated from the records, the final sample consisted of nearly 10 million GPS records. Each record in the dataset represented a GPS signal captured by the Android GPS device and contains information on index, date and time (ET), latitude, longitude, altitude (m), speed (m/h), distance (m), and satellite information. We used GPS accuracy of 100 for user trip/activity recognition.

We developed an Android application using the Ionic 2 framework for the data collection process. The mobile application consisted of three parts: (1) GPS capturing process, (2) login and user authentication process, and (3) data verification process. GPS points were captured at every 1 s interval, and were stored in the database server for processing the data. Each GPS point (latitude and longitude) was stored along with the timestamp on when the point was captured for the associated user ID. Users received a notification once the server processed the data. After that, each user was asked to authenticate their data by verifying every activity and trip during the day, and the verified data were saved into the server.

Data Processing

The whole process required the advancement of an integrated work system consisting of four major components including development of mobile application to collect GPS data, maintenance of back-end server, database management, and classifier system development.

An activity/trip classifier was developed to classify different user activities and trips from raw GPS points. An activity/trip classifier was also designed to process the raw data in the database and extract the knowledge from the raw data. This classifier used GPS points and converted them into activity/trip segments by applying specific methods (e.g., GIS-based approach, Geohash approach, and Combined GIS and Geohash approach). The algorithm defined an activity/trip by validating the duration of a trip/activity in comparison with the dwelling time.

In relation to dwell time selection, most research has used a specific dwell time for separating activity and trips by utilizing it as a minimum duration for activities (*56*). Different thresholds were used by different researchers depending on the available GPS signal, such as more than 120 s (*57*, *58*), more than 180 s (*28*), more than 200 s (*42*) or more than 300 s (*59*). The threshold varies mainly depending on the characteristics of local activities (*56*).

For this study, dwelling time was defined as the minimum duration that the participant spent in a specific place, for example, visiting the supermarket for at least the value of the dwelling time. The dwelling time varies based on different variables, for example, socio-demographics, journey patterns, built-up environment factors, and so forth, for various activities including work, daily shopping, general daily activities, and sport/recreational activities (*57*, *60*, *61*). To determine the optimal dwell time for our study, we tested the pilot travel data based on a range of 1–15 min as the threshold time. The pilot data were collected from a sample group of 12 participants for the time period of 3 months, from October to December 2017. For the pilot study, we selected a different set of participants to incorporate the variation in their daily travel trajectories. We captured their GPS data and the model was set to predict the activity and trip events in a continuous manner. We observed their daily GPS trajectories and manually validated those against the model output. The model accuracy was tested for all of the considered thresholds of dwell time ranging from 1 to 15 min. After that, the threshold for dwelling time was selected for 5, 8, and 10 min based on the observation and accuracy of the pilot data that were gathered from the team members. In this study, if the duration was higher than the dwelling time in any particular place or boundary, it was defined as activity; otherwise, it was defined as trip. The dwelling time was considered only

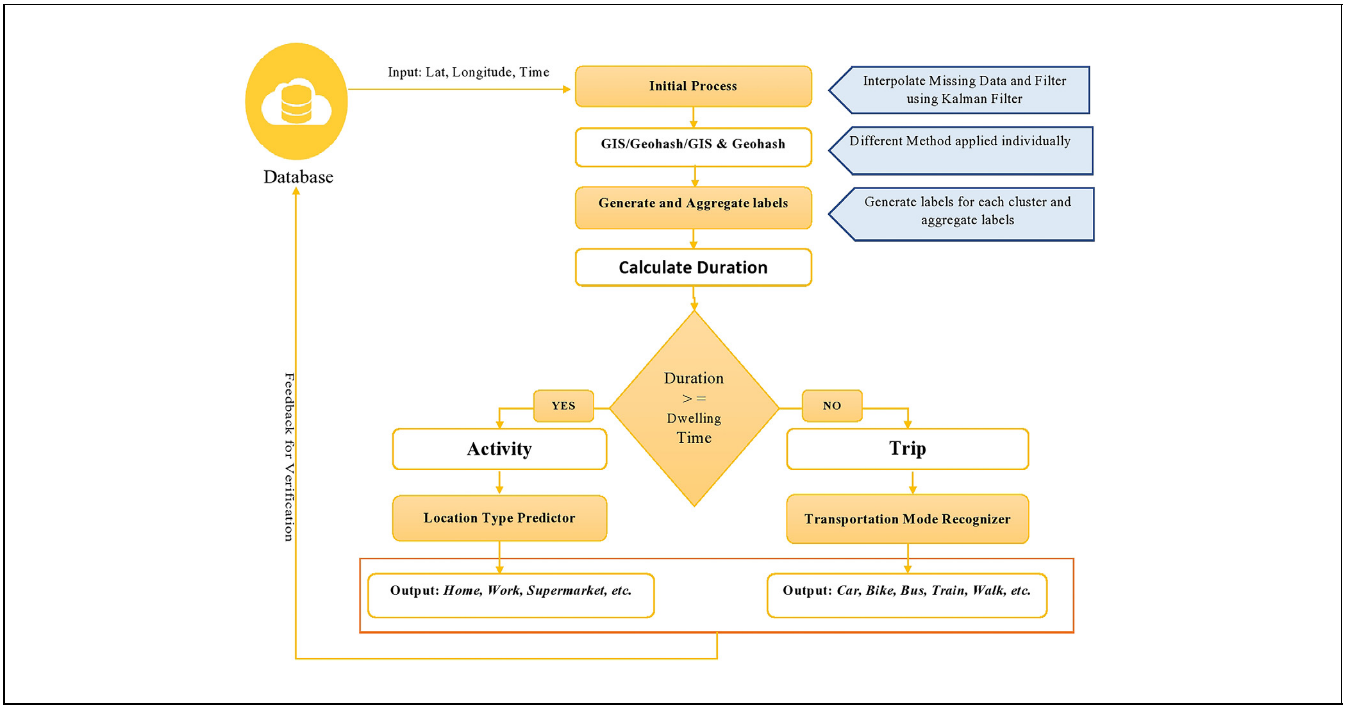

Figure 1. Schematic diagram of the study.
Note: GIS = geographic information system.

for activities; meaning, if the user drove for less than 5 min, the classifier identified this as a trip. The executive trips were merged into one trip from the source and the destination. After finding those segments, the activity was determined and verified by Foursquare API that returns the location types. Afterward, the classifier stored this information in the database and the user was notified to validate the output of the classifiers. We also evaluated the activity and trip recognition analysis regardless of the activity and trip types.

The detailed process of this study is shown in Figure 1 as a schematic diagram. The database contained data for user management along with storing user data, for example, user ID, registration code, socio-demographic profile, and so forth. The mobile application data stored location information, such as latitude, longitude, time captured, speed, and so forth. The MySQL database was used to store and combine the data and was hosted by the Google Cloud platform. After retrieving data from the database, it was smoothed and interpolated to fulfill the missing points because of the limitations of the mobile phone. The interpolation was performed for those segments that had a lot of missing data. These missing data were indemnified by measuring the distance between every two consecutive points; for example, if the distance was more than 1 mi, it was considered as a missing point. The Kalman Filter was applied to smooth and to interpolate the GPS data error and enhance the accuracy.

Methodology

In this research, we developed three different approaches to recognize user activity/trip from GPS data logs, including the Geohash clustering approach, the GIS-based approach, and the Combined Geohash–GIS Approach. For all of the approaches, we developed different models based on dwelling times of 5 min, 8 min, and 10 min, as described earlier in this section.

Geohash Clustering Approach. Geohash is a public domain geocoding system that encodes a geographic location into a short string of letters and digits (*62*). It maintains a hierarchical spatial data structure that subdivides space into buckets of grid shapes by using latitude and longitude points (*53*). In this research, we clustered the GPS points based on the Geohash approach. To cluster the adjacent GPS points, all points during a day were hashed using a Geohash algorithm. Increasing the number of Geohash string characters (precision) increased the neighboring points that incorporated themselves in the same cluster. Based on the geographic extent of our users' available data, we tested 5, 6, and 7-character precision at a spatial resolution level to cluster the adjacent points. Therefore, the analysis level of each cluster was considered for 4.9 km $\times$ 4.9 km area for 5-character precision, 1.22 km $\times$ 0.61 km area for 6-character precision, and 152.9 m $\times$ 152.3 m area for 7-character

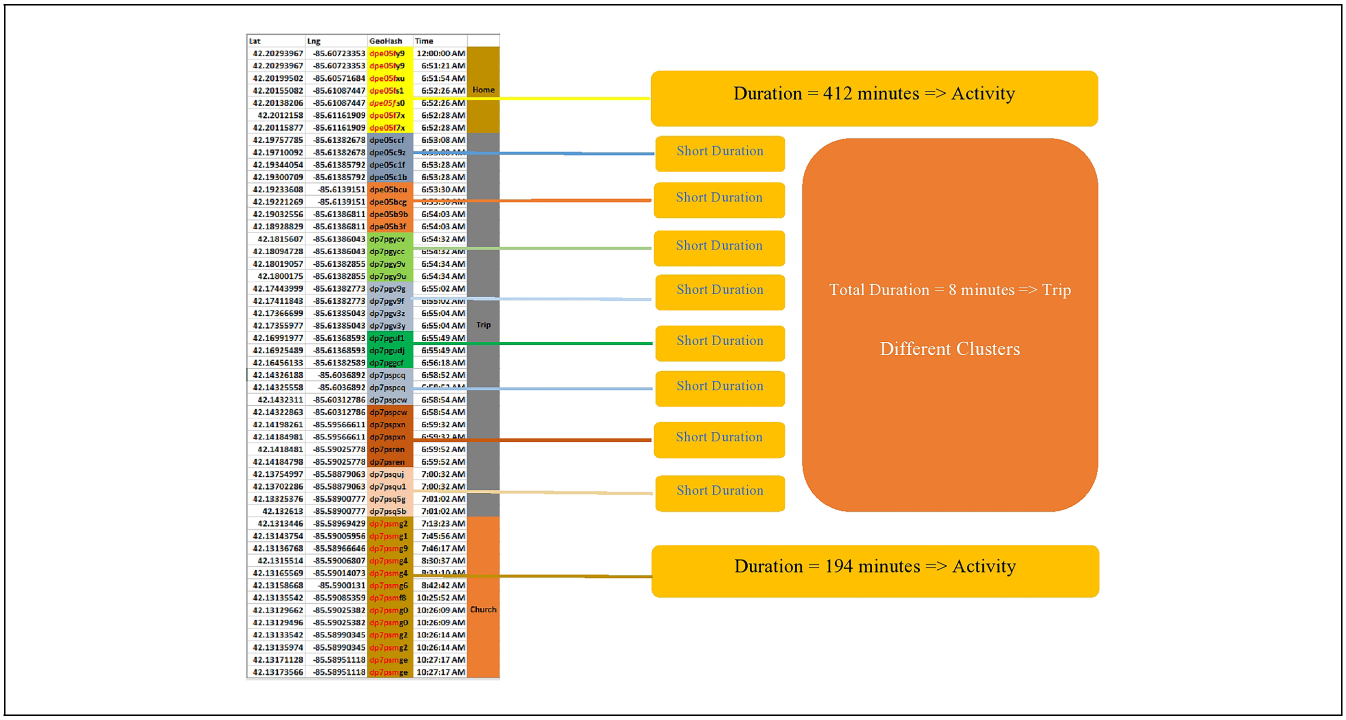

Figure 2. Geohash clustering example.

precision, where all GPS points located in this area were treated as a single Geohash cluster. Figure 2 shows an example of the Geohash clustering technique where different data were used including latitude, longitude, time, and activity type. The latitude and longitude points were converted and hashed into the Geohash format using a specific Geohash length. After, the clustering technique aggregated all similar points that had the same Geohash and added them to a cluster labeled by the Geohash string. Subsequently, the duration was calculated by subtracting the start and the end times of the corresponding trip/activity.

GIS-Based Approach. For this approach, we used the GIS-based boundary shapefile to detect and recognize a user's daily activity/trip. For this analysis we used the boundary shapefile from open street map, which is readily available on the Web. ArcPy code in the Python language environment was scripted for reverse geocoding purposes through using an "identity analysis tool" to retrieve the spatial boundary/address information from GPS points. Each of the GPS points in the user's trajectory returned the location information/address, which includes the mailing address, and later on these mailing addresses were aggregated to get the duration of each activity or trip. We also tested dwelling time as 5, 8, and 10 min to keep consistent analysis with other approaches.

The original GPS points in those trajectories were very noisy because of the density of high-rise buildings and street trees, which caused error in capturing accurate location. In this GIS-based method, we used the Kalman Filter to smooth the GPS data and further applied a 10 ft (3 m) buffer for polygon and line features to incorporate the outlier points. Table 1 shows the steps that were used in the GIS-based approach to predict the user activity and trip information.

According to Table 1, a total of 15 GPS points is pulled out of a user's travel trajectory. We used the reverse geocoding and obtained the mailing address (m) against each GPS point (p) for a specific time occurrence (t). If the spatial points for the user are identified inside the similar boundary (e.g., home, school, market, etc.) for more than the specific dwelling time, it is defined as an activity by the GIS-based approach. Similar mailing addresses are stored together, and the duration of the activity is finally calculated by observing the first and last point of any specific boundary feature. For example, GPS points ranging from p_1 to p_5 are observed as inside the similar boundary/address (m_1) for more than the specific dwelling time, therefore the event is considered as an activity with the duration between t_1 to t_5. On the contrary, if the spatial points for the user are identified in the address location for less than the specific dwelling time, it is defined as a trip by the GIS-based approach. The duration of the trip is finally calculated by observing the first and last point of those different boundary features. For example, GPS points of p_6, p_7, p_8, and p_9 are

Table 1. Example of User Activity/Trip Recognition Based on GIS-Based Approach

GPS trajectory point (p)	Time of occurrence (t)	Mailing address (m)	Output/Decision obtained from GIS-based approach	
p_1	t_1	m_1		
p_2	t_2	m_1	Activity	
p_3	t_3	m_1	(Stays inside same boundary for a specific dwell time)	
p_4	t_4	m_1	Duration from $\mathbf{t_1}$ to $\mathbf{t_5}$	
p_5	t_5	m_1		
p_6	t_6	m_2	Trip	
p_7	t_7	m_3	Trip	Trip
p_8	t_8	m_4	Trip	Duration from $\mathbf{t_6}$ to $\mathbf{t_9}$
p_9	t_9	m_5	Trip	
p_{10}	t_{10}	m_6		
p_{11}	t_{11}	m_6	Activity	
p_{12}	t_{12}	m_6	(Stays inside same boundary for a specific dwell time)	
p_{13}	t_{13}	m_6	Duration from $\mathbf{t_{10}}$ to $\mathbf{t_{15}}$	
p_{14}	t_{14}	m_6		
p_{15}	t_{15}	m_6		

Note: GIS = geographic information system; GPS = global positioning system; orange shading = same mailing addresses; other colors = different mailing addresses.

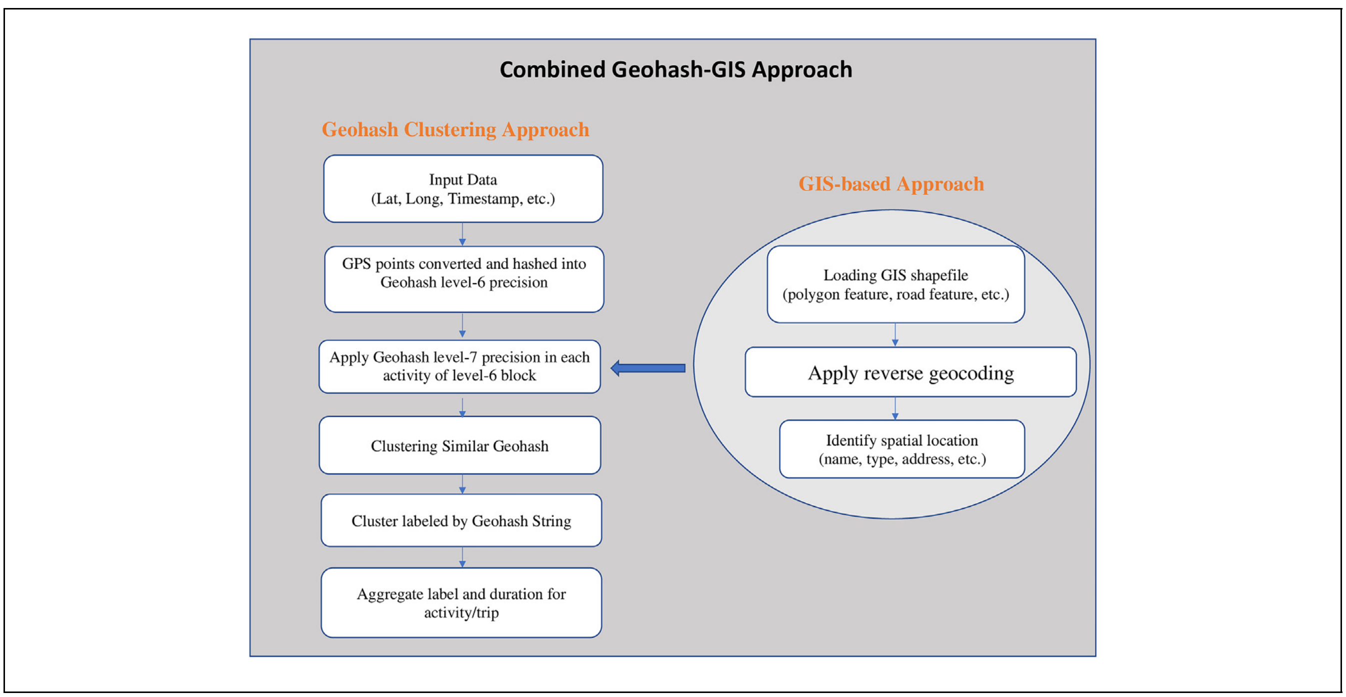

Figure 3. Schematic diagram for Combined Geohash and GIS-based approaches.
Note: GIS = geographic information system.

observed as instances that visit different boundaries/addresses for less than the specific dwelling time; therefore, the event is considered as a trip with the duration between t_6 to t_9. By following the above-mentioned steps, the activity/trip classifier aggregated all durations and generated the final outcome of activity or trip classification based on the GIS-based approach.

Combined Geohash–GIS Approach. The Combined Geohash–GIS approach was developed for this study by integrating the GIS-based and Geohash Clustering approaches. An algorithm was developed by combining the two approaches to achieve the best outcome for activity/trip detection. A schematic diagram for the combined approach is shown in Figure 3.

Geohash alone could not explain the internal activity/trip analysis within each cluster if the Geohash precision level is too big; consecutively, it generates multiple sub-sections for single trip/activity if minimal Geohash precision level is considered for analysis. Further, the GIS-based approach was not flexible enough to incorporate all of the points into the boundary shapefile because of the GPS data outliers. Therefore, the combined approach was developed to overcome these problems. A two-stage algorithm was developed to run this combined approach. For the first stage, a Geohash clustering algorithm was developed based on the approach methodology described. We used Geohash precision level-6 to hash the entire study area. This was followed by running Geohash precision level-7 in each block of level-6 to identify and check the inner activities among them. Once we identified the internal activities, the second stage of the algorithm was initiated by applying the GIS-based approach inside each internal hashed activity. In this stage, a reverse geocoding algorithm was applied with shapefile (e.g., object boundary, line features, etc.) information to accurately identify the user activity or trip information. By applying this combined approach, we were able to detect the inner activities/trips with the correct precision level, as well as validate the activity/trip by applying GIS-based boundary shapefiles. We were also able to reduce the outlier effect as Geohash considered those outliers by cluster aggregation at the final stage. Thus, the combined approach worked perfectly by minimizing the errors to identify the user activity/trip accurately.

Sample Dataset

For the testing dataset, we randomly collected 90 samples (90 different days/dates of user activity, sequentially as trip–activity–trip–activity–trip, etc.) from all users for the study period of 12 months. The testing dataset contained about 50,000 GPS data records. We considered three different testing scenarios based on specified tolerance levels, including simple, moderate, and critical testing, to identify trip only, activity only, and sequential activity–trip recognition analysis. From our sample data, we randomly collected 20% for simple testing, 30% for moderate testing, and 50% for critical testing. The standard and tolerance levels that were considered to identify the accuracy of activity/trip recognition based on three different testing scenarios are as below:

- Simple testing with 99.7% confidence bounds (tolerance range = ± 3E from true mean value)
- Moderate testing with 95% confidence bounds (tolerance range = ± 1.96E from true mean value)
- Critical testing with 68.3% confidence bounds (tolerance range = ± E from true mean value)

where $E = \frac{\sigma}{\sqrt{N}}$ .... σ is for standard deviation and N is sample size.

The true dataset was prepared based on the actual duration of activities and trips performed by users. The users' feedback data together with GPS tracking logs were displayed and carefully observed in a GIS map to identify the true data. The activity or trip duration usually varied because of the user's behavioral pattern for different activity types, which is somewhat problematic for calculating true mean value for the observed dataset. For example, in relation to mean value of trip duration, some of the trips were very long (e.g., 1 h or 2 h) and some were very short (e.g., 10–15 min). So, it was inappropriate to calculate the mean value by combining those long and short trips, which would make our results inaccurate. Therefore, we sorted the sample data with similar types of duration times and considered some group criteria (e.g., less than 15 min, 15–30 min, 31–45 min, 46–60 min, 61–80 min, more than 80 min, etc.) and calculated the mean value for that sample group.

The accuracy for a user's activity/trip recognition was calculated based on Equation 1 for different scenario *i*, as critical, moderate, and simple testing. Here, the accuracy means that the defined model accurately recognized the user trip or activity in such a way that the test/observe value is within 68%, 95%, and 99% confidence bounds of true mean value for critical testing, moderate testing, and simple testing, respectively.

$$\textbf{Accuracy in activity/trip recognition}(A_i) = \frac{\text{accurately classified data for scenerio } i}{\text{total data for scenerio } i} * 100 \quad (1)$$

In addition to testing data and accuracy measurement for activity/trip recognition, we also evaluated the approaches based on model training and prediction accuracy for the whole dataset. For training accuracy, we checked the misclassification error rate (MER) and calculated the accuracy based on Equation 2. The misclassification error rate was calculated to check whether the model accurately classifies the trip as trip, and activity as activity, or vice versa.

$$\textbf{Model Training Accuracy}, A_{\textbf{training}} = (1 - \textbf{MER}) * 100 \quad (2)$$

A comparison of different approaches was shown based on model prediction accuracy. The prediction accuracy was checked based on Equation 3 by calculating mean absolute percentage error (MAPE) through observing the absolute differences between the actual and predicted user activity duration.

$$\textbf{Model Prediction Accuracy}, A_{\textbf{prediction}} = 1 - \textbf{MAPE} \quad (3)$$

where

Table 2. Accuracy in Activity/Trip Recognition for Geohash Clustering Approach

		Activity/trip recognition accuracy (percentage)								
		Geohash-5			Geohash-6			Geohash-7		
		Dwell 5 min	Dwell 8 min	Dwell 10 min	Dwell 5 min	Dwell 8 min	Dwell 10 min	Dwell 5 min	Dwell 8 min	Dwell 10 min
Model accuracy ($A_{training}$)		78.11	77.71	78.74	89.23	85.71	83.63	80.91	80.92	77.08
Accuracy (A_i) of different testing scenarios based on sample dataset										
Critical testing	Sequential activity–trip	34.62	38.46	37.18	46.15	38.46	33.33	20.51	17.95	16.67
	Activity	53.85	46.15	43.59	51.28	43.59	41.03	17.95	15.38	12.82
	Trip	15.38	30.77	30.77	41.03	33.33	25.64	23.08	20.51	20.51
Moderate testing	Sequential activity–trip	50.68	52.05	50.68	57.53	56.16	49.32	52.05	50.68	43.84
	Activity	69.70	69.70	63.64	60.61	63.64	51.52	51.52	48.48	45.45
	Trip	33.33	35.90	38.46	53.85	48.72	46.15	53.85	53.85	43.59
Simple testing	Sequential activity–trip	75.18	75.64	73.08	80.77	75.64	75.64	67.95	67.95	67.95
	Activity	73.77	73.77	73.77	77.05	75.41	77.05	65.57	65.57	65.57
	Trip	76.32	75.00	72.37	80.26	75.00	75.00	68.42	68.42	68.42

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N} \left(\frac{\text{observed activity/trip duration - predicted activity/trip duration}}{\text{observed activity/trip duration}}\right) * 100 \quad (4)$$

where N *is the total number of observations*.

The model was validated based on the ground truth data, which was obtained by projecting the captured GPS trajectories of a travel itinerary on a heatmap using ArcGIS software. A high concentration of points in a location represents an activity, whereas the continuous discrete points represent a trip with a scattered pattern. Once the GPS travel trajectories are displayed in a GIS map, we verified the locations of the points to recognize the event, whether it is an activity or trip. We documented all of the events for the entire travel trajectory and stored them as ground truth data. In addition, we calculated the duration, start and end time for each activity/trip from the heatmap. After that, the model accuracy was calculated by comparing the output of the three approaches in respect to ground truth data.

Analysis and Numerical Results

The researchers analyzed the individual approach output, as well as the comparison between the three proposed approaches. We developed 15 different models based on different dwell times for the three approaches and analyzed the accuracy for user activity/trip recognition. We used dwell times of 5 min, 8 min, and 10 min to recognize user activity/trip for this research, as previously described.

Geohash Clustering Approach

Three different Geohash levels were applied for each of the dwell times based on GPS log data to predict the activity/trip duration. As a result, we developed nine individual models combining three Geohash character levels (character sizes 5, 6, and 7) for three different dwell times (5 min, 8 min, and 10 min). We computed the confusion matrix based on the output for each of the nine models to identify the misclassification between trip and activity. The confusion matrix showed the accuracy as to whether the predicted activity is actually an activity, or the predicted trip is actually a trip for the specific model. For example, in relation to considering the model of Geohash-5 with Dwell-5 min, $\begin{pmatrix} & \textbf{\textit{activity}} & \textbf{\textit{trip}} \\ \textbf{\textit{activity}} & 123 & 11 \\ \textbf{\textit{trip}} & 52 & 102 \end{pmatrix}$, the model accurately classified the activity as activity and trip as trip with an accuracy of 78.1%.

From Table 2, the Geohash-6 and Geohash-7 clustering showed better accuracy in comparison to the Geohash-5 clustering. The Geohash-5 clustering did not deliver a good outcome, as it covered a larger block of geographical area to cluster the GPS data where some of the trips/activities were overlooked and misclassified. The best accuracy (89.23%) was observed for the Geohash-6 clustering model with a dwell time 5 min.

Accuracy was tested for activity only, trip only, and sequential activity/trip recognition based on different

Table 3. Accuracy in Activity/Trip Recognition for GIS-Based Approach

		Activity/trip recognition accuracy (percentage)		
		Dwell-5 min	Dwell-8 min	Dwell-10 min
Model accuracy ($\boldsymbol{A}_{\mathbf{training}}$)		70.41	69.72	70.15
Accuracy ($\boldsymbol{A_i}$) of different testing scenarios based on sample dataset				
Critical testing	Sequential activity–trip	26.92	33.33	34.62
	Activity	35.90	38.46	41.03
	Trip	17.95	28.21	28.21
Moderate testing	Sequential activity–trip	26.03	31.51	35.62
	Activity	42.42	48.48	51.52
	Trip	10.26	15.38	20.51
Simple testing	Sequential activity–trip	63.50	51.28	51.28
	Activity	60.66	52.46	54.10
	Trip	65.79	51.32	51.32

Note: GIS = geographic information system.

character levels of Geohash clustering along with different dwell times. Table 2 shows the outcome of those nine different models based on different testing scenarios. Overall, Geohash-6 with dwell time 5 min showed the best result with about 50% accuracy for critical testing, 60% accuracy for moderate testing, and 80% accuracy for simple testing. As the critical testing (68% confidence interval from mean value) considered a narrow interval from mean value, it showed less accuracy in activity/trip recognition in comparison with other testing scenarios. It is evident from Table 2 that activity recognition performed good accuracy compared with trip recognition for all testing scenarios. Therefore, we could deduce that the Geohash-6 clustering with dwell time 5 min recognized the user activity successfully, with more than 80% accuracy with a 99% probability that the true value lies within the range of predicted values with three standard deviations.

GIS-Based Approach

We developed three different models to recognize activity/trip by applying the GIS-based approach considering 5 min, 8 min, and 10 min dwell times. Table 3 shows model accuracy (A_{training}) based on the misclassification error rate of training data. The GIS-based model with dwell time 5 min showed good accuracy (70.41%) in comparison with other GIS-based models. In an overall comparison, the classification accuracy was lower for GIS-based models in comparison with other models based on the Geohash clustering approach.

Likewise, for the GIS-based approach, the outcome accuracy was tested for the three different GIS-based models by applying different testing scenarios for activity only, trip only, and sequential activity/trip recognition. From Table 3, dwell time 10 min showed good accuracy to recognize user activity/trip for critical and simple testing in comparison with other dwell times. However, dwell time 5 min showed better accuracy for trip recognition based on simple testing. In general, dwell time 10 min worked better to identify activity/trip based on the GIS-based approach with an overall accuracy of about 50% with a 99% probability that the true value lies within the range of predicted values, which was less than the accuracy of the Geohash clustering approach.

Combined GIS and Geohash Approach

Three different models were developed based on a combined approach by using different dwell times. Table 4 shows the accuracy based on a confusion matrix, where a very good accuracy (above 90%) was observed to identify trip as a trip and activity as an activity based on all the different models of a combined approach, which was higher than previous models (based on the Geohash clustering and GIS approach).

From Table 4, all of the models based on the combined approach showed very good accuracy to recognize activity only, trip only, and sequential activity/trip for all testing scenarios in comparison with other approaches (Geohash clustering and GIS-based approach). Among them, dwell time 5 min showed the best accuracy (above 90%) for critical, moderate, and simple testing. On average, there is a 99% probability that the combined approach with dwell time 5 min could recognize the user activity successfully with 95% accuracy that the true value lies within the range of predicted values with three standard deviations.

Table 4. Accuracy in Activity/Trip Recognition for Combined Geohash–GIS Approach.

		Activity/trip recognition accuracy (percentage)		
		Dwell-5 min	Dwell-8 min	Dwell-10 min
Model accuracy ($A_{training}$)		94.10	92.12	92.01
Accuracy (A_i) of different testing scenarios based on sample dataset				
Critical testing	Sequential activity–trip	93.59	93.59	89.74
	Activity	87.18	87.18	79.49
	Trip	100.0	100.0	100.0
Moderate testing	Sequential activity–trip	91.78	87.67	89.04
	Activity	81.82	72.73	75.76
	Trip	100.0	100.0	100.0
Simple testing	Sequential activity–trip	91.97	91.44	91.44
	Activity	83.61	72.13	77.05
	Trip	98.68	98.68	98.68

Note: GIS = geographic information system.

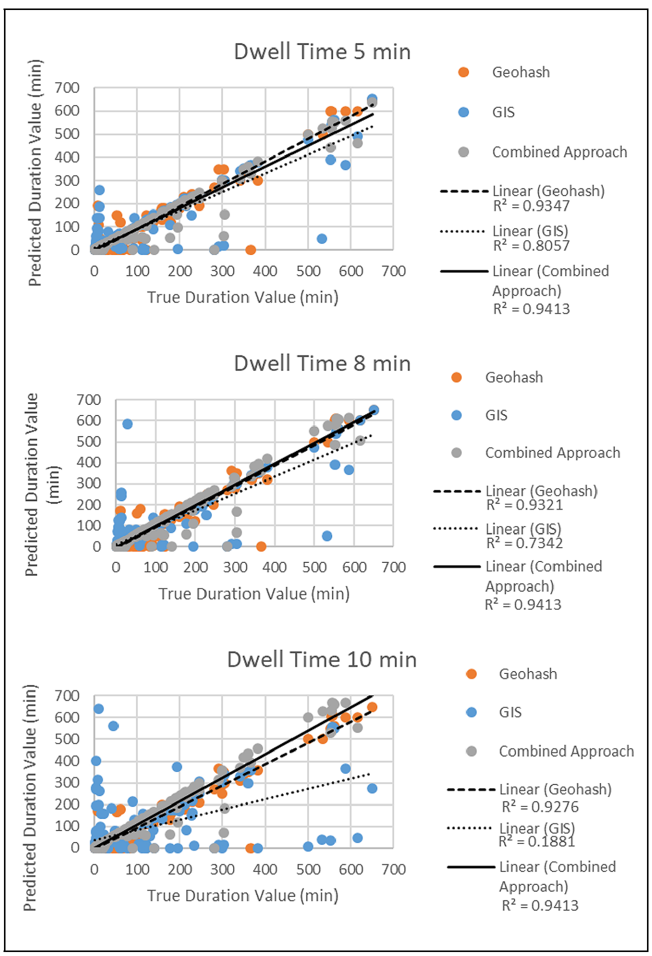

Figure 4. Accuracy comparisons for different approaches.
Note: GIS = geographic information system.

Discussion and Comparison of the Different Approaches

In this section, we talk about the predicted values (duration in minutes) of user activity/trip based on different approaches. Different models based on the Geohash-6 clustering approach, GIS-based and Combined Geohash–GIS approach were compared and analyzed for different dwell time scenarios. Geohash-6 was chosen in this comparison as it showed better accuracy from previous analyses of other Geohash precision levels in this study. As dwell time was considered as a key feature to identify the activity/trip in this analysis, we considered the model comparison based on this parameter.

We compared the predicted sequential activity–trip duration values with true duration values by developing a scatter plot diagram in relation to showing the relationship based on linear trend lines (Figure 4). Based on dwell time 5 min, the data relationship between true and predicted value was positive and strong with *r*-square value of above 0.8 for all of the approaches. Especially, the combined approach showed good accuracy with higher *r*-square value (above 0.9). Therefore, we can conclude that the combined approach explained more than 90% of data variability of the predicted activity–trip duration around its true mean value.

The MAPE was computed to show the actual deviation of predicted activity/trip duration from the true value. MAPE values are shown in Figure 5 for activity only, trip only, and sequential activity–trip accuracy based on different models for different approaches. All of the models based on the GIS-based approach

Different Criteria	Prediction Accuracy ($A_{prediction}$) based on MAPE								
	Geohash-6			GIS-based Approach			Combined Approach		
	Dwell 5 min	Dwell 8 min	Dwell 10 min	Dwell 5 min	Dwell 8 min	Dwell 10 min	Dwell 5 min	Dwell 8 min	Dwell 10 min
Sequential activity-trip	30.16	28.87	33.15	65.91	62.98	67.11	12.70	13.48	18.81
Activity	23.81	22.82	21.22	49.72	58.59	58.29	18.65	20.95	27.83
Trip	42.77	40.97	51.43	44.83	34.88	46.35	15.00	20.01	25.21

Figure 5. Accuracy in activity/trip recognition based on MAPE.
Note: GIS = geographic information system; MAPE = mean absolute percentage error.

(considering 5 min, 8 min, and 10 min dwell time) showed poor accuracy for predicting the activity/trip duration. The Geohash-6 clustering showed better accuracy in sequential activity–trip duration in comparison with GIS-based methods. However, the combined approach with dwell time 5 min showed the best accuracy, where the MAPE values were less than 15%.

This research also compared the accuracy in user activity/trip recognition by developing individual Receiver Operating Characteristics (ROC) for different models on different approaches. The diagnostic test based on ROC curves are shown in Figure 6, where the relationship between sensitivity and false-positive rate are explained for predicted activity/trip values by different models. The area under curve (AUC) was calculated and showed based on different dwell times for different approaches. The combined approach with dwell time 5 min showed the highest AUC value with 0.878. However, for the GIS-based approach, the ROC curve was under the curve, which shows very poor accuracy. The potential reason behind this poor accuracy could be the false prediction of an activity instead of a trip and vice versa for most of the time, as GPS points fluctuate in and out for the same location within an error range of 100 m (the phone's GPS accuracy). Because of the fluctuation of the GPS data, the mailing addresses of all points might not be the same, which leads to false results based on the GIS method. Inaccurate GPS points or outliers could be the primary reason behind the poor accuracy of the GIS approach. The outlier/inaccurate GPS points affect the reverse geocoding method (which is applied to retrieve the spatial boundary/mailing address information), which ends up retrieving a wrong mailing address, and leads to the misclassification of activity/trip information.

The Geohash approach shows the best computation time compared with the other approaches as this approach does not require any third-party database on the internet. The average calculation time for the Geohash approach is around 2 s for 100 GPS points. The GIS approach calculation requires an API call to a third-party database to obtain the reverse geocode and, therefore, the average calculation time for the GIS approach is 70 s for 100 points. The Combined Geohash–GIS approach is a mixture between both Geohash and GIS approaches. This approach requires a third-party API call for a portion of the points, and the average calculation time is 20 s. The rank of the three approaches in relation to calculation time is as follows: Geohash would be the quickest, then the Combined Geohash–GIS approach, then the GIS approach.

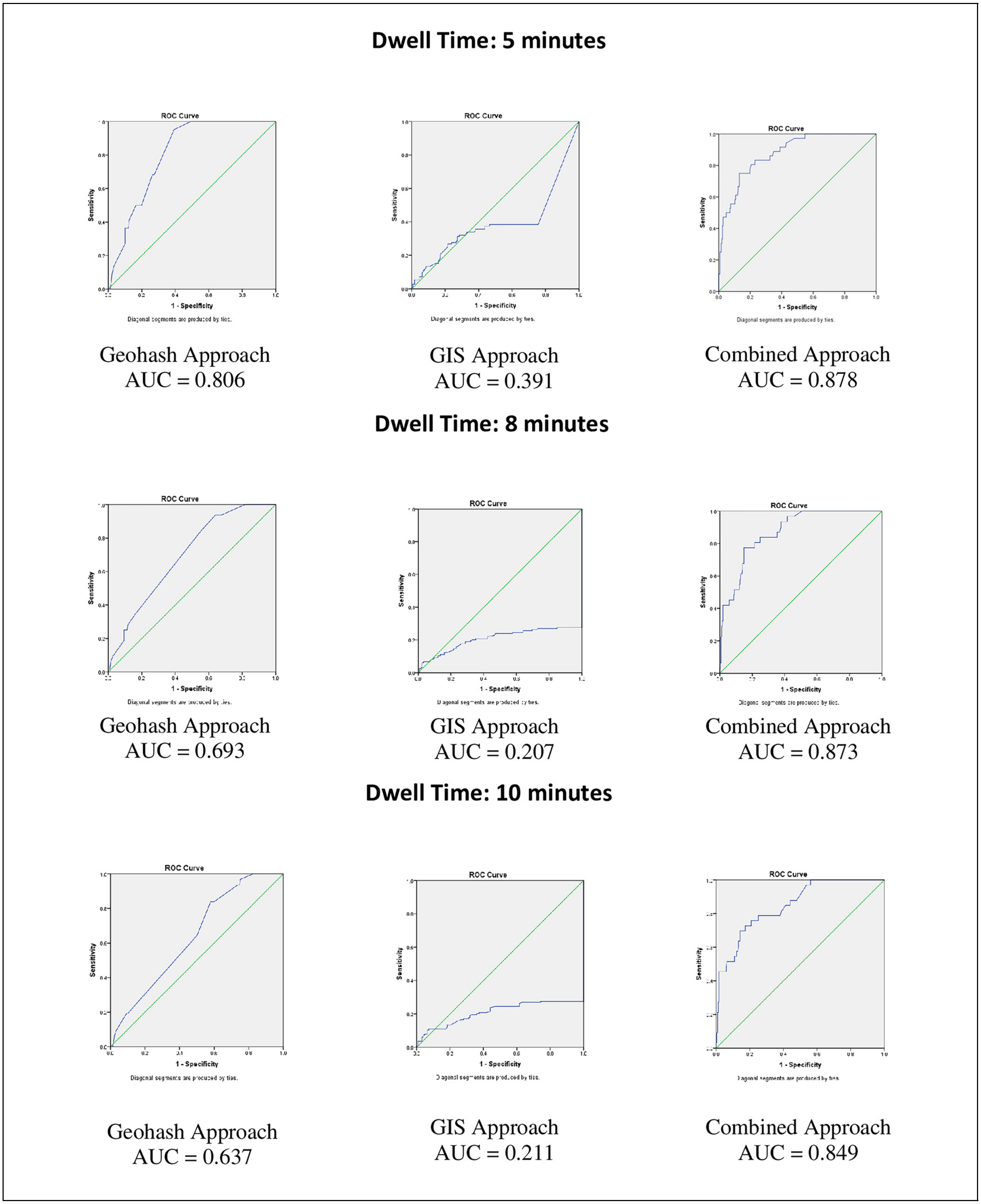

Figure 6. Accuracy comparisons based on ROC curves.
Note: GIS = geographic information system; AUC = area under curve; ROC = receiver operating characteristics.

Summary and Concluding Remarks

This research studied three different approaches that were developed and applied for user activity and trip recognition based on GPS log data. Based on Geohash clustering analysis, Geohash precision level-6 with dwell time 5 min showed good accuracy for user activity/trip detection. Based on the GIS-based approach, dwell time 10 min worked better rather than 5 min, as it was affected with data outlier problems. However, the best accuracy was observed for the Combined Geohash–GIS approach with dwell time 5 min. Therefore, we can conclude that our proposed combined approach could significantly enhance the efficiency and accuracy of GPS travel survey by correctly recognizing user activity and trip patterns. This proposed combined approach could serve as a foundation for a future model system of full-scale travel information identification with GPS data. The accuracy from the combined approach could contribute to the modeling and analyzing of travel behavior, and is readily applicable to a wide range of transportation practices by considering comprehensive travel information.

As a drawback of this study, the application could not collect IOS users' data. Therefore, the users' participation was less than expected. Moreover, we did not consider the trip types and activity categories in this work. In future research, we will focus on different trip and activity types together with their purposes in addition to activity/trip pattern recognition by applying the proposed Combined Geohash–GIS approach.

The technology of Big Data leads to new thought-provoking paradigms about scientific research, which could be useful in detecting transportation users' behavior associated with the built-up environment of cities. To manage and analyze the Big Data related to GPS trajectories of a user's activity/trip, this study's proposed combined approach could be considered as an efficient approach for accurately recognizing the user's pattern and behavior. The proposed approach is easy to replicate and could contribute to transportation and city planning research with a broader perspective by replacing the traditional survey method with automatic recognition of user travel patterns.

Author Contributions

The authors confirm contributions to the paper as follows: study conception and design: Hafez Irshaid, Md Mehedi Hasan, and Jun-Seok Oh; literature review and data collection: Md Mehedi Hasan, Hafez Irshaid, and Raed Hasan; analysis and interpretation of results: Md Mehedi Hasan, Hafez Irshaid, and Jun-Seok Oh; draft manuscript preparation: Md Mehedi Hasan, Hafez Irshaid and Raed Hasan. All authors reviewed the results and approved the final version of the manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was partially funded by the U.S. Department of Transportation's University Transportation Center program through the Transportation Research Center for Livable Communities (TRCLC) at Western Michigan University.

Data Accessibility Statement

Data belong to the Transportation Research Center for Livable Communities (TRCLC) at Western Michigan University.

References

1. Yazdizadeh, A., J. Patterson, and B. Farooq. An Automated Approach from GPS Traces to Complete Trip Information. *International Journal of Transportation Science and Technology*, Vol. 8, No. 1, 2019, pp. 82–100.
2. Axhausen, K. Definition of Movement and Activity for Transport Modelling. *Handbooks in Transport*, Vol. 1, 2007, pp. 329–343.
3. Wang, R. *An Activity-Based Trip Generation Model*. Dissertation report. University of California Transportation Center, 1997.
4. Cambridge Systematics. *NCHRP Report 716: Travel Demand Forecasting: Parameters and Techniques*. Transportation Research Board of the National Academies, Washington, D.C., 2012
5. Montini, L., S. Prost, J. Schrammel, N. Rieser-Schssler, and K.W. Axhausen. Comparison of Travel Diaries Generated from Smartphone Data and Dedicated GPS Devices. *Transportation Research Procedia*, Vol. 11, 2015, pp. 227–241.
6. Ta, N., M.-P. Kwan, Y. Chai, and Z. Liu. Gendered Space-Time Constraints, Activity Participation and Household Structure: A Case Study using a GPS-Based Activity Survey in Suburban Beijing, China. *Tijdschrift voor economische en sociale geografie*, Vol. 107, No. 5, 2016, pp. 505–521.
7. Castiglione, J., M. Bradley, and J. Gliebe. *SHRP 2 Report S2-C46-RR-1: Activity-Based Travel Demand Models: A Primer*. Transportation Research Board of the National Academies, Washington, D.C., 2015.
8. Shen, L., and P. R. Stopher. Review of GPS Travel Survey and GPS Data-Processing Methods. *Transport Reviews*, Vol. 34, No. 3, 2014, pp. 316–334.
9. Wolf, J. *Using GPS Data Loggers to Replace Travel Diaries in the Collection of Travel Data*. PhD thesis. School of Civil and Environmental Engineering, Georgia Institute of Technology, 2000.

10. Wolf, J., M. Oliveira, and M. Thompson. Impact of Underreporting on Mileage and Travel Time Estimates: Results from Global Positioning System-Enhanced Household Travel Survey. *Transportation Research Record: Journal of the Transportation Research Board*, 2003. 1854: 189–198.
11. Forrest, T. L., and D. F. Pearson. Comparison of Trip Determination Methods in Household Travel Surveys Enhanced by a Global Positioning System. *Transportation Research Record: Journal of the Transportation Research Board*, 2005. 1917(1):63–71.
12. Kim, Y., F. Pereira, F. Zhao, A. Ghorpade, P. Zegras, and M. Ben-Akiva. Activity Recognition for a Smartphone and Web-Based Travel Survey. *Computers and Society*, Cornell University, 2015.
13. Pan, B. Online Travel Surveys and Response Patterns. *Journal of Travel Research*, Vol. 49, No. 1, 2010, pp. 121–135.
14. Stopher, P., Y. Zhang, J. Zhang, and B. Halling. Results of an evaluation of TravelSmart in South Australia. 32nd Australasian Transport Research Forum ATRF 2009, Auckland, New Zealand. https://ses.library.usyd.edu.au/handle/2123/19549.
15. Eaton, D. K., N. Brener, L. Kann, M. Denniston, T. McManus, T. M. Kyle, A. M. Roberts, K. H. Flint, and J. G. Ross. Comparison of Paper-and-Pencil versus Web Administration of the Youth Risk Behavior Survey (YRBS): Risk Behavior Prevalence Estimates. *Evaluation Review*, Vol. 34, No. 2, 2010, pp. 137–153.
16. Dillman, D. A., J. Smyth, and L. Christian. *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. John Wiley & Sons, Newark, NJ, 2009.
17. Cottrill, C., F. Pereira, F. Zhao, I. Dias, H. Lim, M. Ben-Akiva, and P. Zegras. Future Mobility Survey: Experience in Developing a Smartphone-Based Travel Survey in Singapore. *Transportation Research Record: Journal of the Transportation Research Board*, 2013. 2354: 59–67.
18. Hasan, M. M. *Application of Artificial Intelligence and Geographic Information System for Developing Automated Walkability Score*. Dissertations. 3646. 2020. https://scholarworks.wmich.edu/dissertations/3646.
19. Wiehe, S. E., A. E. Carroll, G. C. Liu, K. L. Haberkorn, S. C. Hoch, J. S. Wilson, and J. D. Dennis. Using GPS-Enabled Cell Phones to Track the Travel Patterns of Adolescents. *International Journal of Health Geographics*, Vol. 7, 2008, p. 22.
20. Zhao, F., A. Ghorpade, F.C. Pereira, C. Zegras, and M. Ben-Akiva. Stop Detection in Smartphone-Based Travel Surveys. *Transportation Research Procedia*, Vol. 11, 2015, pp. 218–226.
21. Xiao, G., Z. Juan, and C. Zhang. Detecting Trip Purposes from Smartphone-Based Travel Surveys with Artificial Neural Networks and Particle Swarm Optimization. *Transportation Research Part C: Emerging Technologies*, Vol. 71, 2016, pp. 447–463.
22. Greene, E., L. Flake, K. Hathaway, and M. Geilich. A Seven-Day Smartphone-Based GPS Household Travel Survey in Indiana 2. Presented at 95th Annual Meeting of the Transportation Research Board, Washington, D.C., 2016.
23. Patterson, Z. The Itinerum Open Smartphone Travel Survey Platform. Technical Report. *Concordia University TRIP Lab*, Montreal, Canada, TRIP Lab Working Paper 2017-1, 2017.
24. Patterson, Z., and K. Fitzsimmons. MTL Trajet. Working Paper 2017-2. *Concordia University, TRIP Lab*, Montreal, Canada. 2017.
25. Dalumpines, R., and D. M. Scott. Making Mode Detection Transferable: Extracting Activity and Travel Episodes from GPS Data using the Multinomial Logit Model and Python. *Transportation Planning and Technology*, Vol. 40, No. 5, 2017, pp. 523–539.
26. Stopher, P. The Travel Survey Toolkit: Where to from Here? In *Transport Survey Methods: Keeping Up with a Changing World*, Emerald Group Publishing Limited, 2009, pp. 15–46.
27. Wolf, J., R. Guensler, and W. Bachman. Elimination of the Travel Diary: Experiment to Derive Trip Purpose from Global Positioning System Travel Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2001. 1768: 125–134.
28. Bohte, W., and K. Maat. Deriving and Validating Trip Purposes and Travel Modes for Multi-Day GPS-Based Travel Surveys: A Large-Scale Application in the Netherlands. *Transportation Research Part C: Emerging Technologies*, Vol. 17, No. 3, 2009, pp. 285–297.
29. Wu, J., C. Jiang, D. Houston, D. Baker, and R. Delfino. Automated Time Activity Classification Based on Global Positioning System (GPS) Tracking Data. *Environ. Health*, Vol. 10, No. 1, 2011, p. 101.
30. Shafique, M., and E. Hato. Travel Mode Detection with Varying Smartphone Data Collection Frequencies. *Sensors*, Vol. 16, No. 5, 2016, p. 716.
31. Ermagun, A., Y. Fan, J. Wolfson, G. Adomavicius, and K. Das. Real-time Trip Purpose Prediction using Online Location-Based Search and Discovery Services. *Transportation Research Part C: Emerging Technologies*, Vol. 77, 2017, pp. 96–112.
32. Lu, Y., and Y. Liu. Pervasive Location Acquisition Technologies: Opportunities and Challenges for Geospatial Studies. *Computers, Environment and Urban Systems*, Vol. 36, No. 2, 2012, pp. 105–108.
33. Oliveira, M., P. Vovsha, J. Wolf, and M. Mitchell. Evaluation of Two Methods for Identifying Trip Purpose in GPS-Based Household Travel Surveys. *Transportation Research Record: Journal of the Transportation Research Board*, 2014. 2405: 33–41.
34. Lin, C., K. Wang, D. Wu, and B. Gong. Passenger Flow Prediction Based on Land Use around Metro Stations: A Case Study, *Sustainability*, Vol. 12, No. 17, 2020, pp. 1–23.
35. Pereira, F., C. Carrion, F. Zhao, C. Cottrill, C. Zegras, and M. Ben-Akiva. The Future Mobility Survey: Overview and Preliminary Evaluation. In: *Proceedings of the Eastern Asia Society for Transportation Studies*, Vol. 9, 2013.
36. Zhang, Z., Q. He, and S. Zhu. Potentials of using Social Media to Infer the Longitudinal Travel Behavior: A Sequential Model-Based Clustering Method. *Transportation Research Part C: Emerging Technologies*, Vol. 85, 2017, pp. 396–414.

37. Zhao, J., H. Xu, H. Liu, J. Wu, Y. Zheng, and D. Wu. Detection and Tracking of Pedestrians and Vehicles using Roadside LiDAR Sensors. *Transportation Research Part C: Emerging Technologies*, Vol. 100, 2019, pp. 68–87.
38. Rashidi, T., A. Abbasi, M. Maghrebi, S. Hasan, and T. Waller. Exploring the Capacity of Social Media Data for Modelling Travel Behavior: Opportunities and Challenges. *Transportation Research Part C: Emerging Technologies*, Vol. 75, 2017, pp. 197–211.
39. Moore, M. *Tech Giants and Civic Power*. Centre for the Study of Media, Communication and Power. King's College London, 2016, pp. 1–85.
40. Cho, W., and E. Choi. *A GPS Trajectory Map-Matching Mechanism with DTG Big Data on the HBase System*, 2015.
41. Turner, S. M., W. L. Eisele, R. J. Benz, and J. Douglas. *Travel Time Data Collection Handbook*. FHWA, Report No. PL-98-035, 1998.
42. Gong, H., C. Chen, E. Bialostozky, and C. T. Lawson. A GPS/GIS Method for Travel Mode Detection in New York City. *Computers, Environment and Urban Systems*, Vol. 36, No. 2, 2012, pp. 131–139.
43. Oh, J. S.,Valerian K., M. M. Hasan, and Sepideh M. An Evaluation of Michigan's Continuous Count Station (CCS) Distribution. Publication MDOT No. OR 15-1872018. https://www.michigan.gov/mdot/0,4616,7-151-9622_11045_24249-486987–, 00.html.
44. Kamruzzaman, M., J. Hine, B. Gunay, and N. Blair. Using GIS to Visualize and Evaluate Student Travel Behavior. *Journal of Transport Geography*, Vol. 19, No. 1, 2011, pp. 13–32.
45. Hasan, M. M., and J. Oh. GIS-Based Multivariate Spatial Clustering for Traffic Pattern Recognition using Continuous Counting Data. *Transportation Research Record*, 2020.
46. Stenneth, L., O. Wolfson, P. S. Yu, B. Xu, and S. Morgan. Transportation Mode Detection using Mobile Phones and GIS Information. *Academia*, 2011. University of Illinois, Chicago. https://www.academia.edu/880679/Transportation_Mode_detection_Using_Mobile_Phones_and_ GIS_Information.
47. Domènech, A., and A. Gutiérrez. A GIS-Based Evaluation of the Effectiveness and Spatial Coverage of Public Transport Networks in Tourist Destinations. ISPRS *International Journal of Geo-Information*, Vol. 6, No. 3, 2017, p. 83.
48. Loidl, M., G. Wallentin, R. Cyganski, A. Graser, J. Scholz, and E. Haslauer. GIS and Transport Modeling—Strengthening the Spatial Perspective. ISPRS *International Journal of Geo-Information*, Vol. 5, No. 6, 2016, p. 84.
49. Suwardi, I. S., D. Dharma, D. P. Satya, and D. P. Lestari. Geohash Index Based Spatial Data Model for Corporate. *International Conference on Electrical Engineering and Informatics (ICEEI)*, Denpasar, 2015, pp. 478–483.
50. Lee, D., R. Moussalli, S. Asaad, and M. Srivatsa. Spatial Predicates Evaluation in the Geohash Domain using Reconfigurable Hardware. *Proc., - 24th IEEE International Symposium on Field-Programmable Custom Computing Machines*, FCCM 2016, 2016, pp. 176–183.
51. He, S. Spatial Query Processing for Location Based Application on Hbase. *IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, Beijing, 2017, pp. 110–114.
52. Deiotte, R., and R. La Valley. Comparison of Spatiotemporal Mapping Techniques for Enormous. *Etl and Exploitation Patterns*. Vol. IV, 2017, pp. 7–9.
53. Singh, P., K. Oh, and J.-Y. Jung. Flow Orientation Analysis for Major Activity Regions Based on Smart Card Transit Data. *ISPRS International Journal of Geo-Information*, Vol. 6, No. 10, 2017, p. 318.
54. Oh, K., K. Kim, A. Kim, Y. Lee, and J. Jung. Spatial Movement Pattern Analysis in Public. Transportation Networks in Seoul, 2017. https://agile-online.org/conference_paper/cds/agile_2016/posters/154_Paper_in_PDF.pdf.
55. Oh, J., S. Mattingly, A. Al-Fuqaha, K. Hyun, S. Lee, M. M. Hasan, R. Hasan, H. Irshaid, and K. Shirin. *Monitoring Daily Activities and Linking Physical Activity Levels Attributed to Transportation Mobility Choices and Built Environment*. TRCLC, Final Report, 2019.
56. Gong, L., T. Morikawa, T. Yamamoto, and H. Sato. Deriving Personal Trip Data from GPS Data: A Literature Review on the Existing Methodologies. *Procedia - Social and Behavioral Sciences*, Vol. 138, No. 0, 2014, pp. 557–565.
57. Buliung, R. N., M. J. Roorda, and T. K. Remmel. Exploring Spatial Variety in Patterns of Activity-Travel Behavior: Initial Results from the Toronto Travel-Activity Panel Survey (TTAPS). *Transportation*, Vol. 35, No. 6, 2008, pp. 697–722.
58. Liraz, S. P. Ships' Trajectories Prediction using Recurrent Neural Networks Based on AIS Data. *Naval Postgraduate School*, 2018.
59. Axhausen, K. W., S. Schönfelder, J. Wolf, M. Oliveira, and U. Samaga. Eighty Weeks of GPS Traces: Approaches to Enriching Trip Information. *Transportation Research*, November 2003, 2003.
60. Susilo, Y. O., and M. Dijst. Behavioral Decisions of Travel-Time Ratios for Work, Maintenance and Leisure Activities in the Netherlands. *Transportation Planning and Technology*, Vol. 33, No. 1, 2010, pp. 19–34.
61. Bayarma, A., R. Kitamura, and Y. O. Susilo. Recurrence of Daily Travel Patterns, Stochastic Process Approach to Multiday Travel Behavior. *Transportation Research Record: Journal of the Transportation Research Board*, 2008. 2021: 55–63.
62. Wikipedia Contributors. Geohash. In *Wikipedia*, the Free Encyclopedia, 2018. https://en.wikipedia.org/w/index.php?title=Geohash&oldid=864446668.