

An Analysis of Movement by Visualizing Spatio-Temporal Characteristics of Logdata ^a

Nahye Cho ^{b, *}, Youngok Kang ^c

^a This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2017S1A5B6066963)

^b Department Social Studies, Ewha Womans University, Seoul, South Korea; cho.nahye@gmail.com

^c Department Social Studies, Ewha Womans University, Seoul, South Korea; ykang@ewha.ac.kr

* Corresponding author

Keywords: Movement log data, Spatio-Temporal Analysis, STP, 3D DBSCAN

Abstract:

In this study, we visualized and analyzed log data in order to analyze the spatiotemporal characteristics of “moving” and “staying activities”. As a case study, we collected and preprocessed GPS log data generated by students participating in field activities. STP (Space-Time Path) was used to visualize movement logs. “Movement” and staying places were distinguished through density-based clustering, and the time “stayed” and activities performed at staying places were examined. The problem of over-measuring time at some staying places was examined. To resolve this, the 3D Density-Based Spatial Clustering of Application with Noise (DBSCAN) was used to more accurately measure the time spent at staying places. We propose 3D DBSCAN as methodology to accurately measure spatiotemporal data. We believe this method will remain effective even as this data becomes more numerous.

1. Materials and Methods

1.1 Data collection and processing

This study analyzed the log data which were collected during an field experience learning program which studied sustainable development at Yangdong Village. Yangdong Village is registered as a world heritage site because its Korean traditional-style houses, known as Hanok, have been well preserved from the Chosun era until the present day. However, there have been some changes in many buildings in that the existing houses were converted from residential to commercial use and buildings were converted from the traditional style to a modern-style residence. The students used the Collector for ArcGIS application on mobile devices to perform the research activities, and the results created by the research activities were saved in the cloud-based ArcGIS Online in real time. The Collector for ArcGIS application does not internally track and store the students’ movement logs. Therefore, this study used a multi-camcorder equipped with a GPS device to collect movement logs.

1.2 Visualization of Moving Logs Using STP

The STP technique was used to visualize the log data collected in this study. The STP technique uses the concept that time can be measured in the same way as distance (Vasiliev, 1997; Figure.1).

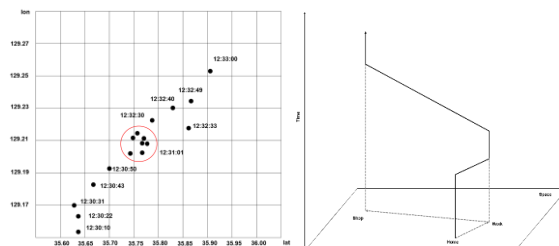


Figure 1. GPS tracking data example: The left shows GPS data that are regularly collected in 10-15 second units. The right shows a fixed time period at the same place.

1.3 Method for Finding Staying Places

To find the staying places, it is necessary to distinguish movement from stopping in the point data that were collected from the GPS device at fixed intervals. When movement-stopping was visualized as STP, the part parallel to the time axis was assigned to stopping, and the clusters derived from the cluster analysis of the point data can be classified as staying places. In order to distinguish movement and stopping and find staying places, many previous studies have used clustering methods. *This study used the density-based clustering method DBSCAN (Density-Based Spatial Clustering of Application with Noise) which has the advantage of having a data-specific clustering form.*

In this study, among the multiple GPS points included in a staying cluster, the first GPS log was defined as *gs*, and the last log was defined as *ge*. The time difference between *gs* and *ge* can be calculated to find the staying time. The time at a staying spot was found as shown in Figure.2.

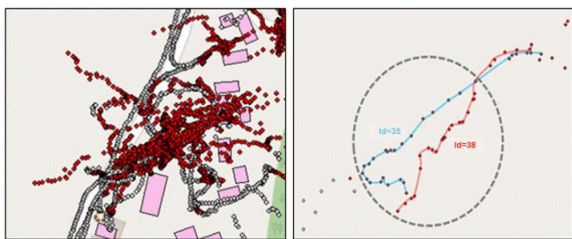


Figure 2. As the time interval for passing this point becomes larger, the time was over-measured.

3D DBSCAN is a method of calculating the distance between a point p and another point q considering the time dimension in 3D based on DBSCAN. That is, x, y, and the time information z are included in the factors in the R source code, and they are used in combination with the plot3d() function to be able to find clusters while considering time (Table.1).

```
// Implement 3D DBSCAN
> data_3d<- matrix(c(x, y, z), ncol=3)
> db<-dbSCAN(data_3d, eps = 0.0001, MinPts = 4)
> plot3d(x, y, z, col=db$cluster+1, pch=20, cex=2, main="Zone1, MinPts=4",
type = "p", size=5)
```

Table 1. 3D DBSCAN source code

2. Experiments and Analysis Results

2.1 Movement Log Visualizaiton

The log data were created during around 150 minutes of field experience learning by 8 teams consisting of 2-3 students. The STP visualization showed the characteristics of field experience learning with repeated moving and staying. The STP visualization showed not only the moving distance, moving range, and direction of movement but also staying places by creating straight lines in the vertical direction when the time of stay at the same point became long (Figure.3). Points where there are output data at a staying place can be defined as the places of observation. However, places that are staying places but have no output data are defined as the places of other activities.

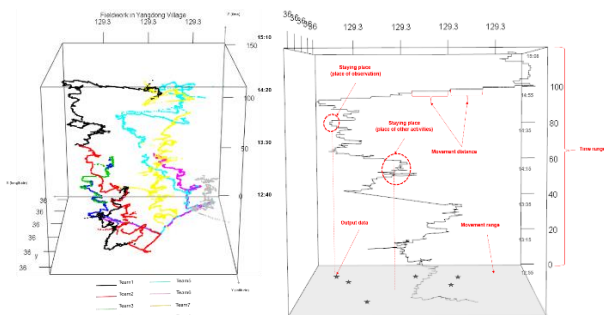


Figure 3. The left figure is visualization of all movement logs for each team and right figure is Diagram showing a generalization of the movement characteristics discovered through the visualization results.

2.2 Finding Staying Places(DBSCAN) and Time Analysis of Staying Places(3D-DBSCAN)

The most important element in using the DBSCAN method is setting the eps (minimum area) and minPts (minimum number of points in the area) for finding the clusters. The optimum values were found by cross-testing the eps and minPts values from the test data and using the kNNdistplot() function. In this study, staying places were found using the parameter conditions of eps=0.00005 and minPts=40 which can find the largest number of clusters and find the clusters which are closest to actual staying places in each of the experiment data. The results of performing DBSCAN clustering on the experiment data are shown in Figure.4(left). 100 clusters were found, and the number of clusters for each team was at least 4 and at most 27. When we compare the clustering results with the uploaded data and base maps, we find that 75 clusters out of a total of 100 were classified as places of observation, and 25 of the remainder were classified as places of other activities.

Figure.4(right) shows the clusters found by performing 3D DBSCAN analysis at the points where over-measurement occurred for each team. It can be seen that at the point where team 2 spent 18 minutes on observation activities, 10 minutes and 2 minutes were spent on different investigation activities, and the rest of the time was spent on movement.

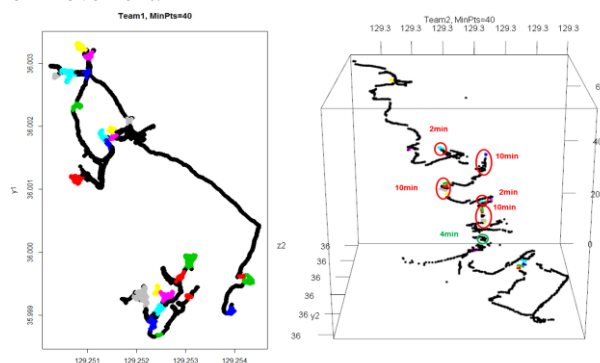


Figure 4. The left is an example of team 1's DBSCAN analysis result and the right is a calculation result of the time spent at staying places found through 3D DBSCAN.

2D DBSCAN Time and Activities		3D DBSCAN Time and Activities	
18 minutes	Observation activities	10 minutes	Observation activities
		2 minutes	Observation activities (investigating while walking slowly)
31 minutes	Other activities	10 minutes	Ice cream shop, resting
		2 minutes	Investigating surrounding buildings
		10 minutes	Ice cream shop, resting

Table 2. Time and activities at staying places analyzed by 3D DBSCAN.

3. Discussion

This study was able to visualize the characteristics of movement logs through STP and DBSCAN and find significant clusters. In order to overcome the limitations of existing clustering methods, this study was able to increase accuracy by using the 3D DBSCAN method. When the DBSCAN method is used, it has the benefit of being able to create cluster shapes that are suitable for the data's characteristics, and it finds clusters quickly and efficiently. However, there were errors caused by over-measuring the time spent. When 3D DBSCAN was used to find clusters while considering time in order to analyze spatial data clusters that include time information, it was found that clusters can be found more accurately. Therefore, we expect that it will be possible to use the method proposed in this study to analyze the characteristics of other data which include spatiotemporal information, such as geo-tagged SNS data, RFID and smartcard data, and sensor data.

4. References

- Vasiliev, I., 1997, "Mapping time." *Cartographica: The international journal for geographic information and geovisualization*, 34(2), 1-51.
<https://doi.org/10.3138/D357-234G-2M62-4373>