JCSE Journal of Computing Science and Engineering

Regular Paper
Journal of Computing Science and Engineering,
Vol. 15, No. 1, March 2021, pp. 47-57

A Method to Measure the Degree of the Favorite Location Visiting of Mobile Objects

Dong Yun Choi
Samsung Electronics, Suwon, Korea
dy345959@gmail.com

Ha Yoon Song*
Department of Computer Engineering, Hongik University, Seoul, Korea
hayoon@hongik.ac.kr

Abstract
To understand the mobility of humans or things, it is necessary to measure the degrees of location visits in everyday mobility. In this paper, we discuss measures that can present human preferences to certain locations based on location data and analysis. From raw positioning data and the concept of location clusters, which are sets of positioning data representing location areas, several measures can be deduced. First, the location point and location area can be separated because visiting a pin point location is different from visiting a certain area. Second, the number of visits to a location and the duration of a visit to a location have different meanings. Third, the rank of the location visited is sometimes more meaningful than the absolute counts. In consideration of these aspects, we established six basic measures and two derived measures. The actual calculation of each measure requires raw positioning data to be processed. The raw positioning data were collected by volunteers over several years of their everyday lives. All measures for multiple volunteers were generated and analyzed for verification. The processing of raw positioning data to generate measures requires a vast number of calculations, like big data processing. As a solution, we implemented a generation process using the programming language R; GPGPU technology was utilized to derive numerical results within areas on able time limit with considerable speed-ups, because an undesirably large amount of time was required to process measures with CPU-only machines.

Category: Information Retrieval / Web

Keywords: Human location preference; Location visiting frequency inside area; Rank of location visiting frequency; Rank of area visiting frequency; Location visiting duration inside area; Rank of location visiting duration; Rank of area visiting duration; Positioning data analytics

I. INTRODUCTION

Human positioning data can be easily acquired due to the recent advancements in positioning devices. Global positioning system (GPS) receivers as well as smartphones and positioning mechanisms such as Wi-Fi-based positioning are good examples of the recent advancements that have been made. Accordingly, positioning data sets of human and things can be easily obtained and applied to many academic and industry fields. Mobility patterns are of

Open Access http://dx.doi.org/10.5626/JCSE.2021.15.1.47 http://jcse.kiise.org

Received 09 April 2020; **Revised** 13 December 2020; **Accepted** 14 March 2021
*Corresponding Author

 pISSN: 1976-4677 eISSN: 2093-8020

particular interest, and models of the mobility of humans or things can have wide areas of application. In addition to one position or one location (pinpoint location) of the mobility model, a set of positions or locations is also of interest, since humans or things tend to move or reside within a specific area. Therefore, a set of positions representing a specific area of human mobility comprises a specific location cluster, and the location cluster is notable for human mobility. Of course, there can be several location clusters for an individual mobility model. With this context, a study was conducted to determine latent location clusters in a mobility model, extrude the transitions between clusters, and construct a human mobility model with location clusters and transitions between clusters using a Markov chain [1].

Then, additional questions arise: How long does a person ordo objects reside in a cluster? Further, how frequently does a person visit a cluster? These questions lead to the concepts of visiting duration and visiting frequency for a specific area represented by a cluster. All of these questions are related to establishing measures for preferred location visit. For location based services (LBS), the measure of visiting frequency or duration is one of the bases for improved LBS. LBS includes navigation, location search, location-based advertisements, infotainment, senior or disabled person care, disaster situation control, finance, logistics, shopping, games, public transportation services, and soon. We will present six measure for capturing the preferred location visits of mobile objects, which is relatively simple but effective, and which will provide basic statistics for LBS.

Section II presents the definitions of the measurements for location visits. Section III describes the actual process of applying the measures to a real positioning dataset; an analysis of the measure results is also presented. Section IV deals with the acceleration of the calculation of measures. GPGPU (general-purpose computing on graphics processing units) technology is required to calculate our measures, since we are dealing with a huge amount of data; otherwise, we are unable to apply our measure to real-world data because of the excessive calculation time. In Section V, we conclude this paper by discussing possible future research topics.

II. DEFINITIONS OF MEASUREMENTS FOR LOCATION VISITS

A. Related Research and Background for Measurements

Various research results have been obtained in the field of human mobility pattern.

From the view of the theory of individual human mobility patterns, a result in a field of complex physics can be found in [2]. Another aspect of related research is concentrated on geopositioning acquisitions such as GPS and the treatment of erroneous positioning data. It is needed to determine the probability distribution of human mobile speed, and this topic is discussed in [3]. The exponential probability distribution can be a candidate for the probability distribution of human mobile speed in a certain speed range. A software-based approach involving positioning error detection and correction can be found in [4]. Without access to the underlying hardware system of geopositioning system, positioning errors can be detected and corrected probabilistically. The factors that affects the location selection were discussed specifically in terms of human personality, and these can be found in [5-8]. In these studies, human personality is represented in terms of the Big Five Factors, and the effects of the Big Five Factors to ward location preference were studied. In the field of applications,

Recommendation systems for location-based service can be found in [9-11]. There are the studies most related to our research [12, 13]; [12] deals with the movement of tourists in an urban area. [13] shows the different behavior of the first visitors and the frequent visitors with the metrics of mean and variance. However, we need to identify more sophisticated measures in order to describe tourist mobility in detail.

Based on these research backgrounds, we attempted to find any measure that could be used for location visit or location preference. However, tour knowledge, there have been no definitions or studies regarding the measure of location preference. The most similar result can be found in a study on text processing [14], where the importance of words in a document was measured based on the frequencies of occurrence of the words. In our research, the importance of location is measured and regarded as a more complicated topic due to the probabilistic nature of human movement while word counting is a kind of deterministic approach. The major motivation of proposing metrics for location visiting preference is that there are no such metrics, to our knowledge, and we need to establish metrics to fill the gaps between existing results toward the application of location-based service.

Therefore, we attempted to establish measures for location visiting preference. We decided to define a total of six basic measures. The criteria behind the definitions of the measures are:

- The measures must be simple enough to calculate easily and instinctively.
- The measures must reflect both the visiting frequency and the visiting duration.
- The measures can separate the pin point location from the location area (location clusters) since mobility for a location is different from that for an area.
- Meaningless mobility, such as purely transient mobility, must be differentiated from meaningful mobility. This is done by a clustering method.

According to these criteria, our definitions are:

- Location visiting frequency inside area (LVFIA)
- Rank of location visiting frequency (ROLVF)
- Rank of area visiting frequency (ROAVF)
- Location visiting duration inside area (LVDIA)
- Rank of location visiting duration (ROLVD)
- Rank of area visiting duration (ROAVD).

LVFIA, ROLVF, LVDIA, and ROLVD are based on positions, e.g., longitude and latitude, while ROAVF and ROAVD are based on location clusters. We must distinguish the frequency of location visit from the duration of location visit. For example, a person can frequently drop by cafe to take out a cup of coffee frequently, while the person can stay at office or a restaurant for a while. The former is related to the frequency of location visit and the latter is related to the duration of the location visit. This leads to another consideration: a person can sit down and work at a specific place for a day while another person can work while moving around a specific area, e.g., a large factory or a university. These phenomena imply that a case of a specific location area must be dealt with separately from a case of a location area or location clusters.

Therefore, we define combinations of count and duration vs. location point and location cluster. LVFIA and ROLVF are related to location point and count is related to location point. LVDIA and ROLVD are related to the duration time at the point. ROAVF and ROAVD are associated with location area (location cluster), and ROAVF is measured by a count at the location cluster. Meanwhile, ROAVD is measured by the duration of stay at the location cluster.

B. Location Visiting Frequency inside Area

LVFIA has arguments of position and cluster, and it is represented by frequency. It refers to the frequency of a visit to a specific position regarding a cluster to which the position belongs. LVFIA for a specific position $position_i$ is calculated using Eq. (3), where the same latitude and longitude stand for pinpoint location $position_i$.

LVF_i (location visit frequency) = Count of a specific location $position_i$ with the same latitude, longitude (1)

CVF_i (cluster visit frequency) = Total number of positioning data in a cluster including $position_i$ (2)

$$LVFIA_i = \frac{LVF_i}{CVF_i} \tag{3}$$

C. Rank of Location Visiting Frequency

ROLVF for a position $position_i$ is the logarithmic value to an inverse ratio for the total number of positions inside every cluster. The ROLVF of $position_i$ can be calculated as shown in Eq. (5).

TLVF (total location visit frequency) = Total number of location positions (of all cluster) (4)

$$ROLVF_i = log\frac{\mathrm{TLVF}}{LVF_i} \tag{5}$$

D. Rank of Area Visiting Frequency

ROAVF for a cluster $cluster_i$ is a logarithmic value to the inverse density of the cluster. $ROAVF_i$ can be calculated as shown in Eq. (6).

$$ROAVF_i = log\frac{\mathrm{TLVF}}{CVF_i} \tag{6}$$

E. Location Visiting Duration inside Area (LVDIA)

The following three measures are used for the duration of stay instead of the frequency of stay. LVDIA for a specific position and a cluster to which the positioning belongs is the duration of a stay for the position. *LVDIAi* for a specific position $position_i$ is calculated as shown in Eq. (9)

LST_i (location stay time) = Stay time at a specific location $point_i$ (7)

CST_i (cluster stay time) = Total stay time of every location position in the cluster having $position_i$ (8)

$$LVDIA_i = \frac{LST_i}{CST_i} \tag{9}$$

F. Rank of Location Visiting Duration

ROLVD for a position $position_i$ represents a rank for a position inside a cluster to which $position_i$ belongs, similar to ROLVF. $ROLVD_i$ can be calculated using Eq. (11).

TLST (total location stay time) = Total staying time of all location positions inside every existing cluster (10)

$$ROLVD_i = \frac{\mathrm{TLST}}{LST_i} \tag{11}$$

G. Rank of Area Visiting Duration (ROAVD)

ROAVD for a cluster $cluster_i$ represents the rank of duration time for the cluster. $ROAVD_i$ can be calculated as expressed in Eq. (12).

$$ROAVD_i = log\frac{\mathrm{TLST}}{CST_i} \tag{12}$$

Table 1. Sample of raw data for individual positioning dataset

Date	Time	UNIX time	Latitude	Longitude
2013–05–15	19:26:43	1368613603	37.55561833	126.9233483
	19:26:44	1368613604	37.55561833	126.9233483
	19:26:45	1368613605	37.55563333	126.9233367
	19:26:46	1368613606	37.55564667	126.923325
	19:26:47	1368613607	37.55565833	126.9233117
	19:26:48	1368613608	37.555670000	126.9232983
	19:26:49	1368613609	37.55568167	126.9232817
	19:26:49	1368613609	37.55569167	126.9232683
	19:26:51	1368613611	37.555700000	126.9232533
	19:26:52	1368613612	37.55570667	126.923240
	19:26:53	1368613613	37.55571333	126.9232283
	19:26:53	1368613613	37.555720000	126.923220
	19:26:54	1368613614	37.55573167	126.923215
	19:26:56	1368613616	37.55574500	126.923215
	19:26:56	1368613616	37.55575500	126.923210

H. Derived Metrics

Some combinations of the six basic measures mentioned in this section are found to be meaningful. Two combinations of measures, $LVFIA_i \times ROLVF_i$ and $LVDIA_i \times ROLVD_i$, can be additionally established. These combinations for a specific $position_i$ imply the importance of the position $position_i$ since they are the multiplication of rank and frequency and the multiplication of rank and visiting duration. The higher these combinations are, the more significant $position_i$ is for a person among all location positions. Based on these derived metrics, we can identify if the location point is less important among all meaningful location points even though the point has a high rank.

III. APPLYING MEASUREMENTS TO REAL POSITIONING DATA SET

A. Positioning Data Collection and Analysis

First, we need to identify the location clusters. From the method used to establish the individual mobility model [1], location clusters can be identified. Positioning data collected by volunteers are transformed into individual mobility models, and location clusters can be determined from individual mobility models. In this study, positioning data was collected using smartphones such as the Vega Racer, Vega Racer2, Vega Secret Note, Galaxy S2, and Galaxy S3LTE [15], Galaxy S4, Galaxy S5, Galaxy Note1, Galaxy Note 2 [16], iPhone3Gs, iPhone4, iPhone5, and iPhone6 [17]. Data were collected using an app installed called SportTracker [18] or a similar app. GPS receivers such as the Garmin [19] Garmin Edge800 [20], and Garmin Edge810 [21] were also used. Among the positioning data components, we used time, latitude, and longitude information. Table 1 lists the sample positioning data. The detailed information for data collection is presented in Table 5, including the duration of data collection, devices for data collection, and the total number of data for eight participants.

The raw positioning data of one volunteer can be mapped onto real maps as parts of positioning data, as shown in Fig. 1, and a part of the identified location clusters is shown in Fig. 2 with cluster numbers. Fig. 2 shows locations clusters obtained from the process illustrated in [1], which is mainly based on the Expectation-

Fig. 1. Mapping Raw Positioning Data.

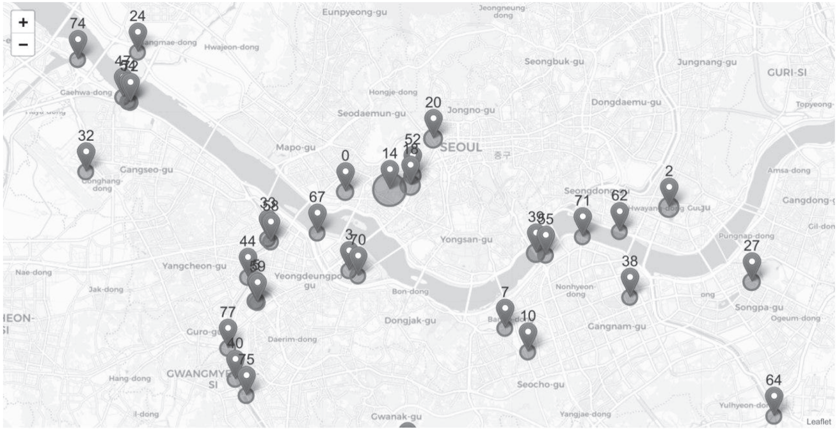

Fig. 2. Mapping Individual Mobility Model with Location Clusters.

Table 2. Summary of individual mobility statistics: frequency of visit for participant 1

C	Latitude	Longitude	Count of point	LVFIA	ROLVF	LVFIA× ROLVF
9	37.672608	126.79287	1,131	0.36308	4.90111	1.7795055
4	37.475975	126.93847	819	0.31744	5.22388	1.6582805
9	37.672577	126.79283	967	0.31043	5.05777	1.5701015
9	37.672465	126.79271	318	0.10208	6.16991	0.6298666
4	37.475990	126.93833	219	0.08488	6.54289	0.5553857
5	37.736325	126.86303	357	0.09090	6.05423	0.5503850
9	37.672502	126.79275	215	0.06902	6.56133	0.4528689
19	37.550055	126.92495	3,433	0.11281	3.79078	0.4276619
5	37.736327	126.86303	261	0.06646	6.36745	0.4231996
2	37.550055	126.92495	3,433	0.11119	3.79078	0.4215117
4	37.475983	126.93846	130	0.05038	7.06443	0.3559600
9	37.672622	126.79293	141	0.04526	6.98321	0.3160940
19	37.550053	126.92491	1,989	0.06536	4.33658	0.2834527
2	37.550053	126.92491	1,989	0.06442	4.33658	0.2793764
9	37.672695	126.79278	117	0.03756	7.16979	0.2692990
19	37.549922	126.92464	1,743	0.05727	4.46860	0.2559574
2	37.549922	126.92464	1,743	0.05645	4.46860	0.2522765
5	37.736340	126.86308	142	0.03615	6.97614	0.2522568
5	37.736328	126.86305	124	0.03157	7.11168	0.2245606
9	37.672432	126.79272	82	0.02632	7.52525	0.1980965
9	37.672483	126.79272	80	0.02568	7.54994	0.1938991

Maximization clustering method. Our measures show that a human visit to a specific location has a far different pattern than visits to clusters. For example, a meaningful location can be found with LVFIA×ROLVF and LVDIA×ROLVD, even though there are fewer visits to the location. This means that only frequent visits do not guarantee the importance of a location. From this location information and the measures defined in Section II, six measures can be calculated, as presented in Tables 2-4.

B. Location Visiting Frequency inside Area (LVFIA)

The large value of LVFIA indicates that specific location data can be found more often in the corresponding cluster of the location data. In other words, LVFIA is proportional to the number of occurrences at a specific location in a location cluster of a person. This includes the cluster number expressed as C, the latitude and longitude of a location point, the count of a location point, LVFIA, ROLVF, and LVFIA×ROLVF. Table 2 lists the largest counts of the location points of 3433 in clusters 19 and 2. This means that specific locations inside both cluster 19 and cluster 2 are the most favorable locations for volunteer 1, accounting for 3433 visits, and the location point is in the intersection of clusters 19 and 2. In addition, cluster 19 has an LVFIA of 0.11281, while cluster 2 has an LVFIA of 0.11119. Cluster 2 has a smaller LVFIA than cluster 19. This is due to the property of LVFIA. Referring to Eq. (3), the denominator is the total positioning data inside a cluster. Table 4 indicates that cluster 2 has a cluster point count of 30,874, and while cluster 19 has a cluster point count of 30,430. Therefore, cluster 19 has a larger LVFIA than cluster 2, meaning that cluster 19 is smaller than cluster 2.

C. Rank of Location Visiting Frequency

A smaller ROLVF indicates more occurrences of location data in total clusters. Because of the use of the log function for the ROLVF definition, the smaller ROLVF is, the more frequent visits are made to a specific location cluster (location area). Therefore, the better the rank, the smaller the ROLVF. The nominator is the total number of positions of all location clusters. Not every location position in the positioning dataset means that a human meaningfully visits that location position. This is because transient positions exist, and such transient positions must be separated from meaningful visits to a location position or location clusters. In other words, a location position which is not included in location clusters does not indicate a visit to the location since it stands for meaningless transient location points. Such transient locations must be distilled out by a clustering mechanism, as shown in [1]. Table 2 presents the ROLVF. Clusters 19 and 2 have the same ROLVF of 3.79078, which is the best-ranked location for participant 1.

D. Position Duration (LVDIA)

Table 3, similar to Table 2, presents information on

Table 3. Summary of individual mobility statistics: frequency of visit for participant 1

C	Latitude	Longitude	Total stay at position (min)	LVDIA	ROLVD	LVDIA×ROLVD
24	37.47087500	126.9359800	567	0.51498	6.81082	3.50748
14	35.26804000	129.0786717	501	0.38747	6.93458	2.68694
15	35.16300500	129.1623517	552	0.31011	6.83763	2.12043
22	34.73405500	127.7204267	145	0.23577	8.17445	1.92731
9	37.67260833	126.7928733	1,131	0.27747	6.12033	1.69825
17	37.52116667	127.1012117	549	0.23203	6.84308	1.58785
23	37.18339667	128.4656300	146	0.19185	8.16758	1.56697
9	37.67257667	126.7928333	966	0.23699	6.27802	1.48787
21	36.42675000	127.4182250	193	0.16959	7.88849	1.33785
16	34.89524500	127.5162350	208	0.15975	7.81364	1.24826
21	36.42676500	127.4182900	166	0.14586	8.03920	1.17267
16	34.89444500	127.5161217	191	0.14669	7.89891	1.15875
16	34.89523500	127.5162050	191	0.14669	7.89891	1.15875
15	35.16303833	129.1610650	246	0.13820	7.64585	1.05667
24	37.47106833	126.9359350	141	0.12806	8.20242	1.05044
24	37.47107500	126.9359350	137	0.12443	8.23120	1.02422
17	37.52117000	127.1006867	326	0.13778	7.36429	1.01469
22	34.73401500	127.7200500	70	0.11382	8.90269	1.01331
13	35.87916500	128.8103500	198	0.11849	7.86292	0.93169
10	37.50190500	127.0264950	417	0.1269	7.11810	0.90330

LVDIA, ROLVD, and LVDIA×ROLVD as well as total stay at each position. The major difference between the two tables is the stay time at a location vs. the count of total visits to a location. Similar to LVFIA, LVDIA is based on the stay time at a location inside a cluster. As listed in Table 3, a location in cluster 9 has the longest stay time of 1,131 with an LVDIA of 0.27747, while a location in cluster 24 with an LVDIA of 0.51498 has a stay time of 567. It is clear that the concentration of stay time at a location occurs regardless of the absolute stay time.

E. Rank of Location Visiting Duration

ROLVD is the rank of a stay time at a certain location for the total clusters, similar to ROLVF, which involves the frequency of visits. As listed in Table 3, cluster 16 has two locations with the same ROLVD of 7.89891, and the total stay at that location is 191 minutes. This means that the location points with the same ROLVD are equivalently ranked among the meaningful locations, and are thus equally visited by participant 1, since ROLVD deals with total clusters. Participant 1 stays at a location in cluster 9 for 1,131 minutes, which has the highest rank of 6.10233. Even though cluster 0 is the most frequent area for participant 1, the participant's stay at a location in cluster 9 shows the highest concentration of stay.

F. Rank of Area Visiting Frequency

Table 4 presents the cluster number, count of cluster points, ROAVF, stay time at a cluster, and ROAVD. ROAVF is the rank of location area (cluster) visit frequency.

G. Rank of Area Visiting Duration

ROAVD is the rank of location area (cluster) visit duration. Table 4 lists the stay time of a cluster and ROAVD; it is not always related to ROAVF. For example, comparing cluster 4 and cluster 5, the ROAVF of cluster 5 is 3.656340215 and the ROAVF of cluster 4 is 4.076426592, while the ROAVD of cluster 5 is 4.204292406 and the ROAVD of cluster 4 is 3.924875788. Regarding the frequency of visit vs. the duration of visit, the two rank values show inverse results. Even if a person visits a location area frequently, that person does not always stay there for a long time; for example, the location could be a bus stop or a subway station that the person must pass through during their daily activities.

H. Derived Measures

Table 2 also presents LVFIA×ROLVF, which is a derived measure. LVFIA×ROLVF concurrently considers the importance of area and that of a location position. A larger LVFIA×ROLVF can be achieved with a high frequency of visits to a position inside a certain cluster, and with a high rank of the position among all positions in all clusters. For example, clusters 2 and 19 have the same count of points (3,433), but due to their low LVFIA of 0.11, LVFIA×ROLVF is a slow as 0.42 even with the relatively high ROLVF of 3.79. Therefore, from the view of clusters 19 and 2, the location point is less important among all meaningful location points, even though the point has a high rank.

Table 4. Summary of statistics for individual mobility of participant 1

Cluster number	Count of cluster point	ROAVF	Stay time at cluster	ROAVD
1	49,520	1.121839363	190,004	0.996387526
2	30,874	1.594301585	112,959	1.516407730
3	959	5.066080195	6,357	4.393876083
4	2,580	4.076426592	10,161	3.924875788
5	3,927	3.656340215	7,684	4.204292406
6	917	5.110863798	7,955	4.169631989
7	994	5.030234063	5,381	4.560558420
8	174	6.772915971	3,169	5.090016570
9	3,115	3.887986839	4,076	4.838316535
10	215	6.561333242	3,286	5.053761631
11	1,856	4.405792357	5,806	4.484540785
12	616	5.508724306	1,748	5.684960373
13	137	7.011990344	1,671	5.730010401
14	95	7.378094378	1,293	5.986467551
15	132	7.049169347	1,780	5.666819286
16	81	7.537522115	1,302	5.979531107
17	264	6.356022167	2,366	5.382231885
18	300	6.228188795	2,680	5.257615856
19	30,430	1.608787027	91,017	1.732386348
20	2,812	3.990320017	2,383	5.370884842
21	116	7.178381079	1,138	6.114160315
22	37	8.321053357	615	6.729565662
23	64	7.773088187	761	6.516554572
24	33	8.435463708	1,101	6.147213793
25	20,782	1.990128763	47,255	2.387874182
26	1,021	5.003433452	664	6.652905780

Table 5. Execute time (in minute): frequency of visit versus duration of stay

PN	Execute time frequency (min)	Execute time duration (min)	Dated from	Dated to	Device	Total data
1	42.32	74.89	Dec 2013	Jan 2015	Vega Racer iPhone6	817,220
2	119.30	1,032.33	May 2013	Jan 2015	Galaxy Note2 Garmin 62S	2,311,815
3	1.35	4.83	Oct 2014	Jan 2015	Galaxy S4 Garmin 62S	267,600
4	170.21	402.62	May 2013	Jan 2015	Galaxy S2 Galaxy S3LTE Galaxy S5 Galaxy Note1	1,862,480
5	31.92	48.03	Oct 2012	Jan 2015	Vega Racer2 Vega SecretNote	679,455
6	137.37	410.16	May 2013	Jan 2015	Galaxy Note2 Garmin Edge800	2,013,425
7	810.67	4,279.59	Nov 2011	Jan 2015	iPhone3 Gs iPhone4 iPhone5 Garmin Edge800, Garmin Edge810	5,748,285
8	1.60	2.27	Oct 2014	Jan 2015	Galaxy S4 Garmin 62S	159,455

IV. EXECUTION TIME FOR MEASUREMENT ANALYSIS

We used the GPGPU technique, which is used for general-purpose computing in graphics processing units to achieve a practical calculation time for the measures. In some cases, we required more than one day to calculate the measures on CPU, which is not practical at all, and the application of the GPGPU technique reduces the calculation time to a reasonable range. Among the various techniques, we used CUDA [22] on GPGPU. GPGPU is a kind of parallel processing technique that is suitable for our purpose, since our algorithm requires the execution of a lot of transcendental functions. For example, more than 10 million pieces of data for eight participants must be handled, as presented in Table 5. As expected, we required a significant amount of time to calculate the measures. For example, participant 7 provided a total of 5,748,285 positioning data, and it took 4,279.59 minutes to apply duration measures, which is unacceptable for real-world applications. For some of the participants with huge amounts of geopositioning data, we cannot manage to calculate the measures within a reasonable time range. Therefore, we need to enhance the speed of our calculation algorithm. GPGPU is a technique used to speed up the calculation of the measures.

Algorithm 1 shows our algorithm that uses GPGPU to calculate the measures, as referenced in step 4 below.

Algorithm 1 Calculation of Location Visiting Measure with GPGPU

Require: N, $CountofClusterNumber$
▷ N is number of positioning data inside clusters
▷ $CountofClusterNumber$ is the total number of clusters
1: Read raw positioning data
2: Read location clusters information ▷ Clusters number, Clusters center $< latitude, longitude >$, minimum, maximum, mean of Radius of clusters
3: Identify whether a positioning data is included in a location cluster ▷ CUDA on GPGPU
4: Ignore time-adjacent same point with same $< latitude, longitude >$
5: Count the number of every specific position
6: **for** $i = 1$ to N **do**
7: Calculate LVFIA, ROLVF, LVFIA $\times$ ROLVF ▷ equation (3), equation (5)
8: **end for**
9: **for** $i = 1$ to $CountofClusterNumber$ **do**
10: Calculate ROAVF ▷ equation (6)
11: **end for**
12: **for** $i = 1$ to N **do**
13: Calculate LVFIA$\times$ ROAVF, ROLVF $\times$ ROAVF ▷ equation (3)$\times$ equation (6), equation (5)$\times$ equation (6)
14: **end for**

Table 6. Execution time of frequency-based measures: CPU versus GPU

PN	CPU (min)	GPU (min)	Speed-ups	Cluster count	Total data	Data in clusters	Meaningful data rate
1	42.32	17.27	2.45	27	817,220	490,686	0.60043
2	119.3	40.16	2.97	34	2,311,815	893,308	0.38641
3	1.35	0.5	2.67	18	267,600	37,843	0.14142
4	170.21	50.35	3.38	45	1,862,480	1,333,815	0.71615
5	31.92	15.8	2.02	27	679,455	577,326	0.84969
6	137.37	47.37	2.9	43	2,013,425	968,095	0.48082
7	810.67	407.37	1.99	76	5,748,285	1,507,257	0.26221
8	1.6	0.95	1.69	5	159,455	130,855	0.82064
G	7,029.26	1,357.00	5.18	115	9,852,759	9,116,947	0.92532

This algorithm is used for the visiting frequency measures. The procedure of Algorithm 1 is as follows:

1. Read every piece of positioning data.
2. Establish clusters as shown in Section III, and read location cluster information, such as the cluster number; center of the cluster; and the minimum, maximum, and mean of the cluster radius.
3. Identify whether the positioning data is a member of a location cluster.
4. The positioning data is included regardless of whether the distance between the position and the center of the cluster is less than the cluster radius.
5. The distance between the two positions is calculated using the Haversine formula [23].
6. CUDA [22] on GPGPU is used to calculate the distance. For example, in our experience, 2,311,815 points×89 clusters are the expected number of times that the Haversine formula is applied for participant 2.
7. Ignore the non-mobile position. Adjacent positioning data in time are regarded as non-mobile if two time-adjacent pieces of positioning data have the same <latitude, longitude> pair.
8. Count the number of occurrences of each piece of positioning data in a cluster.
9. Calculate LVFIA, ROLVF, and ROAVF. N stands for the number of positioning data in a cluster. The Cluster Number Count is the total number of clusters.

In the case of an algorithm for visiting duration measures, lines 3, 4, and 5 are changed.

Table 5 lists the execution times needed to calculate measures as well as the properties of the positioning data set, such as the duration of data collection, data collecting device, and number of total data. The execute time for duration is typically larger than the execute time for frequency. For example, participant 7 has an execute time for frequency of 810.67 minutes and an execute time for duration of 4,279.59 minutes. This is because calculating duration measures requires more calculating steps to measure time.

Our GPU machine is a plain PC with Ubuntu as the OS, a i7-3770 CPU, 24 GB RAM, and CUDA on NVIDIA GeForce GTX980Ti GPU. The basic reason that Algorithm 1 takes too much time on the CPU lies in the use of the Haversine formula to calculate the distance between two geolocation points in the form of <latitude, longitude>. Even though the Haversine formula is a simple set of codes, it is composed of transcendental functions. For example, for the case of participant G listed in Table 6, it requires more than 100,000,000 distance calculations and more than 20 hours, 1,357 minutes, even with GPGPU. The CPU-based calculation took 7,029 minutes, which is more than 117 hours, and which is nearly meaningless for adapting these measures to real-world examples.

Therefore, our conclusion is that GPGPU is a requirement for operations involving such a vast number of location points. It is also necessary to reduce the execution time by several seconds in further works.

As listed in Table 6, the speed-up with the GPGPU is in the range of 1.69 and 5.18. Table 6 presents the results of frequency-based measures only as well as the execution time on CPU, execution time on GPU, speed-ups, count of clusters for each participant, total number of data, total data in clusters, and rate of data in clusters to total number of data. Not all the positioning data will be used; only the positioning data in clusters will be used for the calculation, and the ratio of usage is between 0.1414 of participant 3 and 0.9253 of participant G, as listed in

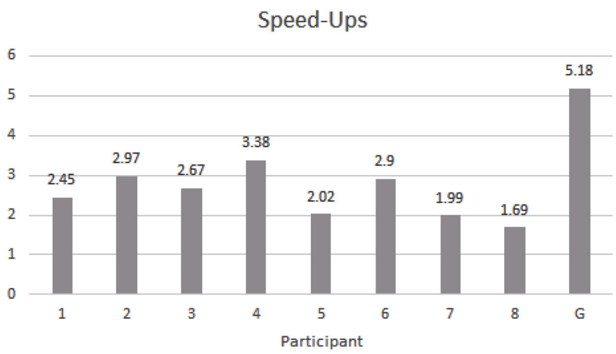

Fig. 3. Speed up of GPGPU versus CPU.

Table 6. The newly added participant G utilizes more than 95% of the total data, and we were nearly unable to obtain the measures by CPU. This is the main reason why we must use a GPU with such a huge number of positioning data.

With a higher rate of usage of positioning data, the participant tends to stay at location clusters; in other words, the participant does not like to move.

The numbers in Table 6 are graphically transformed in Fig. 3, where the x-axis represents the participant number and the y-axis represents the speed-up using GPU. Simply speaking, we can calculate the measures up to five times faster with the GPU.

V. CONCLUSION

Because of the need to establish metrics for the mobility of humans or things mobility, particularly in terms of individually preferred locations, we define several metrics of preferred location visits in terms of visiting frequency and staying duration. Another consideration was made to separately measure location position as well as location cluster, which stands for location areas. We also demonstrated our metrics to illustrate the visiting nature to meaningful locations. The demonstration was based on the application of defined metrics to raw positioning data, which were collected by volunteer participants over several years. LVFIA, ROLVF, and ROAVF were calculated for the visiting frequency of locations, while LVDIA, ROLVD, and ROAVD were calculated for the staying duration at certain locations. Comparing the two sets of metrics yields clearly different results.

The LVFIA and LVDIA results can be used to identify the importance of a certain location inside a cluster.

The rank of the location within all locations can be expressed by ROLVF and ROLVD, and the rank of a location area can be found using ROAVF and ROAVD. The resulting metrics were verified by each participant to check whether the correctness of location, the preferred locations, the rank of location visit, and so on were correctly derived. As a result, the participants reported high confidence in the metric.

We also have a problem of an excessive execution time needed for CPU-based calculations. We solved this problem by using GPGPU technology, and we achieved both speed-ups and reasonable execution times. Our conclusion is that GPGPU is a must for operations involving such a vast number of location points. Itisalsonecessarytoreducetheexecutiontimebyseveralseco ndsinfutureworks.

To our knowledge, there have been no such metrics for this purpose, and this research may be the first one providing location visiting metrics. We hope that our work will serve as a basis for other location-based researches and industrial applications.

Another question arises: what is the effect of positioning errors, which is unavoidable, on our metrics? First, we assume that each participant carries their own mobile devices so that the positioning errors dedicated to the same device are the same. Second, a side from the device-dependent errors, there could be geolocation errors due to weather, the circumstances of radio frequency, GPS signal jamming, and soon, regardless of specific location. In such cases, one of the possible solutions can be found in [4]. Without any knowledge of the underlying geolocation system, positioning errors can be corrected by the end-used positioning device itself. Some modern geopositioning systems, such as that presented in [24] which started in November 2018, show very high precision compared to previous geolocation systems. We estimate that such evolutions in geolocation systems, along with the combination of various geolocation systems, can guarantee that the geolocation data of human mobility will be sufficiently precise. Geopositioning error is one of the minor reasons why we introduced the concept of location area (location cluster). In cases where in positioning errors persist, the measures regarding location cluster will be applied instead of those regarding location position. Then, we will find that measures such as ROAVF and ROAVD, which are associated with location area (location cluster), became more meaningful. In such cases, the size of the cluster could be arrange of positioning errors. With such errors, it is difficult to identify measures of location position-based measures. However, if the size of the clusters is smaller than that of the positioning error range, it can be regarded that the location area stands for the location point. With a very high error rate of the positioning data, only measures for location clusters can be used. The detailed solution to cope with the positioning data error is another open problem to be solved in future research.

ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. NRF-2019R1F1A1056123). This research was supported by the WEF Global Partnership Program for the Fourth Industrial Revolution through KAIST KPC 4 IR funded by the Ministryof Science and ICT

REFERENCES

1. H. Y. Song, “Probabilistic space-time analysis of human mobility patterns,” *WSEAS Transactions on Computers*, vol. 15, pp. 222-238, 2016.
2. M. C. Gonzalez, C. A. Hidalgo, and A. L. Barabasi, “Understanding individual human mobility patterns,” *Nature*,

vol. 453, no. 7196, pp. 779-782, 2008.
3. H. Y. Song and J. S. Lee, "Finding a simple probability distribution for human mobile speed," *Pervasive and Mobile Computing*, vol. 25, pp. 26-47, 2016.
4. H. Y. Song and J. S. Lee, "Detecting positioning errors and estimating correct positions by moving window," *Plos One*, vol. 10, no. 12, article no. e0143618, 2015. https://doi.org/10.1371/journal.pone.0143618
5. S. Y. Kim, H. J. Koo, and H. Y. Song, "A study on influence of human personality to location selection," *Journal of Ambient Intelligence and Humanized Computing*, vol. 7, no. 2, pp. 267-285, 2016.
6. S. Y. Kim, H. J. Koo, and H. Y. Song, "A study on influence of human personality to location selection," in *Proceedings of the Korea Computer Congress,* Jeju, South Korea, 2015, pp. 392-394.
7. H. Y. Song and H. B. Kang, "Analysis of relationship between personality and favorite places with Poisson regression analysis," *ITM Web of Conferences*, vol. 16, article no. 02001, 2018. https://doi.org/10.1051/itmconf/20181602001
8. D. P. Schmitt, J. Allik, R. R. McCrae, and V. Benet-Martinez, "The geographic distribution of Big Five personality traits: patterns and profiles of human self-description across 56 nations. *Journal of Cross-Cultural Psychology*, vol. 38, no. 2, pp. 173-212, 2007.
9. J. Bao, Y. Zheng, and M. F. Mokbel, "Location-based and preference-aware recommendation using sparse geo-social networking data," in *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*, Redondo Beach, CA, 2012, pp. 199-208.
10. J. Christensen and I. Drejer, "The strategic importance of location: location decisions and the effects of firm location on innovation and knowledge acquisition," *European Planning Studies*, vol. 13, no. 6, pp. 807-814, 2005.
11. M. H. Park, J. H. Hong, and S. B. Cho, "Location-based recommendation system using Bayesian user's preference model in mobile devices," in *Ubiquitous Intelligence and Computing*. Heidelberg, Germany: Springer, 2007, pp. 1130-1139.
12. B. Mckercher and G. Lau, "Movement patterns of tourists within a destination," *Tourism Geographies*, vol. 10, no. 3, pp. 355-374, 2008.
13. B. McKercher, N. Shoval, E. Ng, and A. Birenboim, "First and repeat visitor behaviour: GPS tracking and GIS analysis in Hong Kong," *Tourism Geographies*, vol. 14, no. 1, pp. 147-161, 2012.
14. G. G. Chowdhury, *Introduction to Modern Information Retrieval*. London, UK: Facet Publishing, 2010.
15. Samsung Electronics, "Samsung Galaxy S3, GT-i9300, manual," 2013 [Online]. Available: https://www.devicemanuals.eu/samsung-galaxy-s-iii-gt-i9300-manual/532/.
16. Samsung Electronics, "Galaxy Note2," 2013 [Online]. Available: http://www.samsung.com/global/microsite/galaxynote/note2/index.html?type=nd.
17. Apple Inc., "iPhone," 2015 [Online]. Available: http://www.apple.com/iphone/.
18. L. Ferrari and M. Mamei, "Identifying and understanding urban sport areas using Nokia Sports Tracker," *Pervasive and Mobile Computing*, vol. 9, no. 5, pp. 616-628, 2013.
19. Garmin Ltd., "GPSMAP 62s," 2010 [Online]. Available: https://buy.garmin.com/en-US/US/p/63801
20. Garmin Ltd., "Edge 800," 2010 [Online]. Available: https://buy.garmin.com/en-US/US/into-sports/discontinued/edge-800/prod69043.html.
21. Garmin Ltd., "Edge 810," 2015 [Online]. Available: http://buy.garmin.com/en-US/US/into-sports/cycling/edge-810/prod112912.html
22. NVIDIA, "NVIDIA CUDA C programming guide," 2011 [Online]. Available: https://developer.download.nvidia.com/compute/DevZone/docs/html/C/doc/CUDA_C_Programming_Guide.pdf.
23. R. W. Sinnott, "Virtues of the Haversine," *Sky & Telescope*, vol. 68, no. 2, pp. 158-159, 1984.
24. Cabinet Office Japan, "Quasi-Zenith Satellite System (QZSS)," 2018 [Online]. Available: http://qzss.go.jp/en/.

Dong Yun Choi

Dong Yun Choi received his B.S. degree in Computer Engineering in 2016 and received his M.S. degree in Computer Engineering in 2018, respectively, both from Hongik University, Seoul, Korea. His research interests are in the areas of deep learning, location clustering, and next location prediction.

Ha Yoon Song

Ha Yoon Song received his B.S. degree in Computer Science and Statistics in 1991 and received his M.S. degree in Computer Science in 1993, both from Seoul National University, Seoul, Korea. He received Ph.D. degree in Computer Science from University of California at Los Angeles, USA in 2001. From 2001 he has worked at Department of Computer Engineering, Hongik University, Seoul, Korea and is now a full professor. In his sabbatical year 2009, he worked at Institute of Computer Technology, Vienna University of Technology, Austria as a visiting scholar and in the year 2018, he worked at Free University Berlin as a visiting scholar. Prof. Song's research interests are in the areas of data science, mobile computing, application of machine learning, location based service and human mobility modeling.