TRAVEL TIME ESTIMATION USING MOBILE DATA

Byeong-Seok YOO
Senior Consultant
SCM Business Group
Samsung SDS
159-9 Gumi, Bundang, Seongnam, Gyeonggi
463-810 Korea
Fax: +82-2-6484-0055
E-mail: ybs74@hotmail.com

Seung-Pil KANG
Associate Professor
Graduate School of Engineering
Seoul National University
San 56-1 Shillim-Dong, Gwanak-Gu, Seoul
151-742 Korea
Fax: +82-2-888-7144
E-mail: kangsp@snu.ac.kr

Chang-Ho Park
Professor
Graduate School of Engineering
Seoul National University
San 56-1 Shillim-Dong, Gwanak-Gu, Seoul
151-742 Korea
Fax: +82-2-889-0032
E-mail: parkch@snu.ac.kr

Abstract: Travel time information is one of the important information in transportation planning, transportation operation, and transportation management. By the rapid supply of cellular phones, cellular phone information is expected to be useful in traffic information field. In this study, travel speed and travel time estimation methodology using GPS and base station information of cellular phone is presented and the effectiveness and usefulness are validated by field data. Travel time of 6 main arterial roads is estimated in Cheongju. The estimated travel time using cellular base station is almost similar to the estimated travel time by GPS. However, the accuracy is low when the hand-off happens greatly or the coverage of cellular base station is large. Also, significance test between travel speed with passenger and travel speed without passenger is performed and it shows that there is a significant difference of travel speed between passenger existence and nonexistence.

Key Words: Travel Time, Cellular Phone, GPS, Map Matching, Shortest Path Algorithm

1. INTRODUCTION

Travel time information is one of the important information in transportation planning, transportation operation, and transportation management. Because of the latest remarkable development in route guidance field, real-time travel time information is considered as one of the most urgently required information.

Loop detector or beacon are widely used for estimating travel time, however these system need to be constructed, operated and managed with much budget.

By the rapid supply of cellular phones and the development of telecommunication, cellular phone information such as base station information or GPS phone's coordinates is expected to be useful in traffic information field. In the handset position detection technology in Korea, approximately accurate position can be obtained with GPS phone, but only positions of cellular base station can be acquired with normal cellular phone.

In this study, travel speed and travel time estimation methodology using GPS and base station information of cellular phone is presented and the effectiveness and usefulness are validated by field data.

2. LITERATURE REVIEW

2.1 Measurement System Development

Important transportation information in transportation planning and operating are volume, speed, and density etc and these information can be categorized into spot information and wide area information according to measuring site and can be divided into instantaneous information and interval information.

Technologies and equipments to generate transportation information have developed as 3 stages. In first stage, clocker counted and measured volume, speed, and travel time and so on. This stage is characterized that people manually measure the traffic information. Therefore, the data may contain observation error.

In the early 1960', loop detector introduced as a measuring instrument of traffic characteristics. After loop detector, various detectors developed such as radar detector, ultra sonic detector, infrared detector, and video image processor etc. Researchers have concentrate on estimating link travel speed using spot speed from various detectors. This stage can be represented as second stage.

Recently, mobile data such as GPS or cellular phone's location can be acquired. These mobile data is useful because they contain wide area location information not spot information. Especially, on 1 May 2000 the US government removed SA(Selective Availability) which is an intentional error term. This measure improved the accuracy of GPS. Also, with the rapid spread of cellular phones, cellular phone's location information is thought to be useful in generating traffic information. Researchers started to have an interest in estimating travel time from the mobile data source such as cellular phone and GPS. These mobile data source don't need to establish a detector or a detection facility. Besides, they can give large sample size.

2.2 Location Detection Solution

GPS is known as to be identified its XY coordinates with in about 30 meters and it is widely used as car navigation system.

In the case of cellular phone, there are several location identification techniques such as E-CGI(Enhanced Cell Global Identity), A-GPS(Assisted Global Positioning System), TDOA(Time Difference of Arrival), AOA(Angle of Arrival) and so forth. In these techniques only E-CGI and A-GPS methods can be used at detecting the location of cellular phones in Korea. E-CGI method can identify the base station that is serving a cellular phone. And the coverage of BS is known as over several hundreds meter to a few kilometers. On the other hand, A-GPS method can confirm a phone's location within about 50 meters.

2.3 GPS and Cellular Phone Study

Quiroga and Bullock(1999) presented methodology to estimate travel time and travel speed using GPS data. Using real data, they described the space-time diagram of a vehicle.

Chung and Shalaby(2004) presented trip reconstruction tool which identify the traveled route and used modes for GPS-based personal trip surveys.

Li et al.(2004) studied the day-to-day variability of the journey-to-work trips, including the departure time, route choice, and trip chaining behavior, using GPS-based disaggregate morning commute data.

Axhausen et al.(2004) examined the GPS data and said that GPS data can be used intense spatial and temporal investigation of daily life travel.

Yim and Cayford(2002) studied the transportation information collecting methods using GPS and cellular phone using simulation. They said that cellular phone tracking shows promise but needs further study to determine its ability to produce quantifiable travel time data.

Cayford and Johnson(2003) studied the impacting factor which influencing the confidence of traffic monitoring system using cellular phone. These parameters include the accuracy of the locations, the frequency with which location measurements are taken, and the number of locations available.

Lee(2003) studied the cellular base station data collected every 5 minutes and estimate travel speed. He said that travel speed is similar to 80% when estimated at every 15 minutes.

According to the above studies, the state of the art is in early stage and needs further study using real data in estimating traffic information such as travel time and travel speed. Also, there was little research on estimating travel time using cellular information.

3. TRAVEL TIME ESTIMATION PROCEDURE

General travel time estimation procedure is presented with a mobile data such as GPS coordinates or base station (BS) coordinates of cellular phone. There may be difference of the process between using GPS data and using cellular BS data. However, general procedure to estimate travel time may be same like Figure 1.

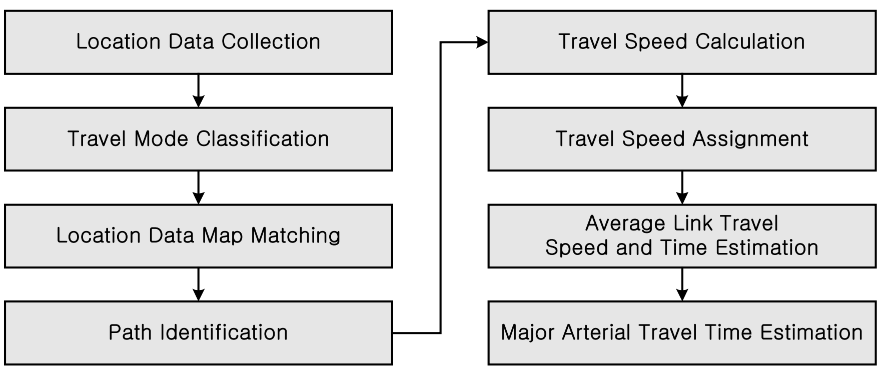


Figure 1. Travel Time Estimation Procedure based on Mobile Data

Location data are collected at every time interval. The collected coordinates would be checked the changes sequentially, and one can categorize that coordinate into moving or standing by the change of the coordinates.

Also, transportation mode can be identified by rule-based algorithm. For example, pedestrian walks below 5 kph speed. Bus or train would stops at stops or stations. Usually passenger car has a faster speed than bus. These characteristics can make a rule-based algorithm.

Once identified auto or taxi data that are moving, these coordinates data can be matched into a digital map according to the pre-defined map-matching algorithm. This process need to done if one wants to identify traveled links and distance. In this step, there is a difference between GPS and cellular BS data. Figure 2 shows the difference of map-matching and searched link between GPS and cellular BS. A probe vehicle moved from G1 to G4, namely real route is G1-G2-G3-G4. However, in case of cellular BS, G1 take a service from near BS that will be matched to C1. Like this G2 will be served from near cellular BS and will be matched to C2. Therefore, the path is C1-B-C2-C3-D-C4. This characteristic of map-matching leads the estimated speed between GPS and cellular BS to be different.

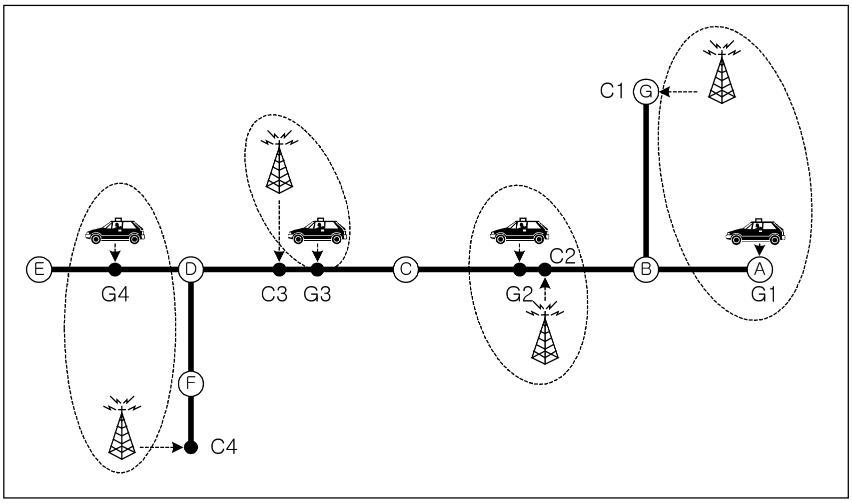


Figure 2. Map-matching between GPS and Cellular BS

After map-matching, the moved links can be identified using shortest path algorithm. If the time interval is relatively long, for example 5 minutes, the vehicle can move from a certain link to another link in the network. In this reason, shortest path algorithm needs to be applied. Common shortest path algorithm such as Dijkstra algorithm can be adjusted in this process.

Once the traveled links and distance are found, travel speed is calculated using travel distance and time interval. This travel speed is assigned to the traveled links as Figure 3. Each link can have another sample speed during some time period, for example 15 minutes or 1 hour etc.

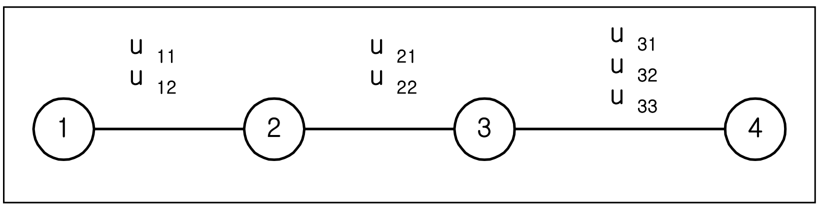


Figure 3. Travel Speed Assignment

The average speed of each link can be calculated using harmonic mean speed.

$$\overline{u_i} = \frac{1}{\frac{1}{m}\sum_{j=1}^{m}\frac{1}{u_{ij}}} \tag{1}$$

where, $\overline{u_i}$ = Average speed of link i
u_{ij} = j th vehicle speed of link i
m = Number of sample speed of link i

Using above average link speed and link distance, the travel speed of major arterial can be calculated like as follows.

$$\overline{u_L} = \frac{L_T}{\sum_{i=1}^{n}\overline{t_i}} = \frac{L_T}{\sum_{i=1}^{n}\left(\frac{L_i}{\overline{u_i}}\right)} \tag{2}$$

where, $\overline{u_L}$ = Representative speed of major arterial L
L_T = Total distance of major arterial L
L_i = Distance of link i
n = Number of links consisting of major arterial L

Using those speeds, one can calculate travel time of each link and major arterial. Consequently, the procedure of estimating travel time have steps such as travel speed calculation for individual vehicle, estimation of link travel speed and travel time, and estimation of arterial travel speed and travel time.

4. REAL DATA COLLECTION USING TAXI

The purpose of this research is to estimate travel speed using GPS coordinates and using cellular BS coordinates. Therefore, information of GPS coordinates and BS coordinates of cellular phones have acquired by establishing GPS equipments and cellular phones on taxies in Cheongju by the help of Openpass Inc. The data acquisition system is like as Figure 4.

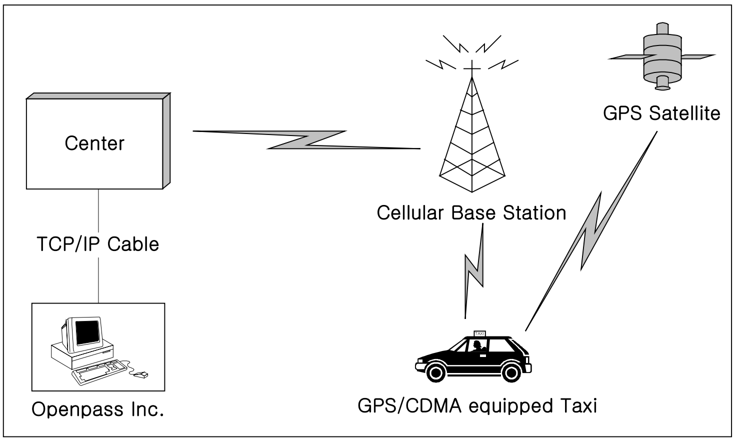


Figure 4. Data Acquisition System

The data is composed of GPS information and cellular BS information. GPS information is composed of vehicle ID, GPS coordinates, time, and passenger existence code. Cellular BS information consists of vehicle ID, cellular base station coordinates, and time. Because the data is gathered by means of taxi, the paired data is transmitted to center whenever passenger boards the taxi and after then the data is transmitted at every 5 minutes until the passenger alights taxi. In the case of non-passenger, data is transmitted at every 2 minutes.

Because taxi was used as a probe vehicle, travel mode classification process in Figure 1 doesn't need to be implemented in this study.

Real data were gathered for 31 days from January 13, 2004 to February 12, 2004 with the average 10 probe vehicles every day. Total 67,564 GPS coordinates were used in the study. Every GPS coordinates have cellular BSs in pairs.

Figure 5 describes the GPS coordinates distribution and Figure 6 represents cellular base station. Total 39 base stations are identified.

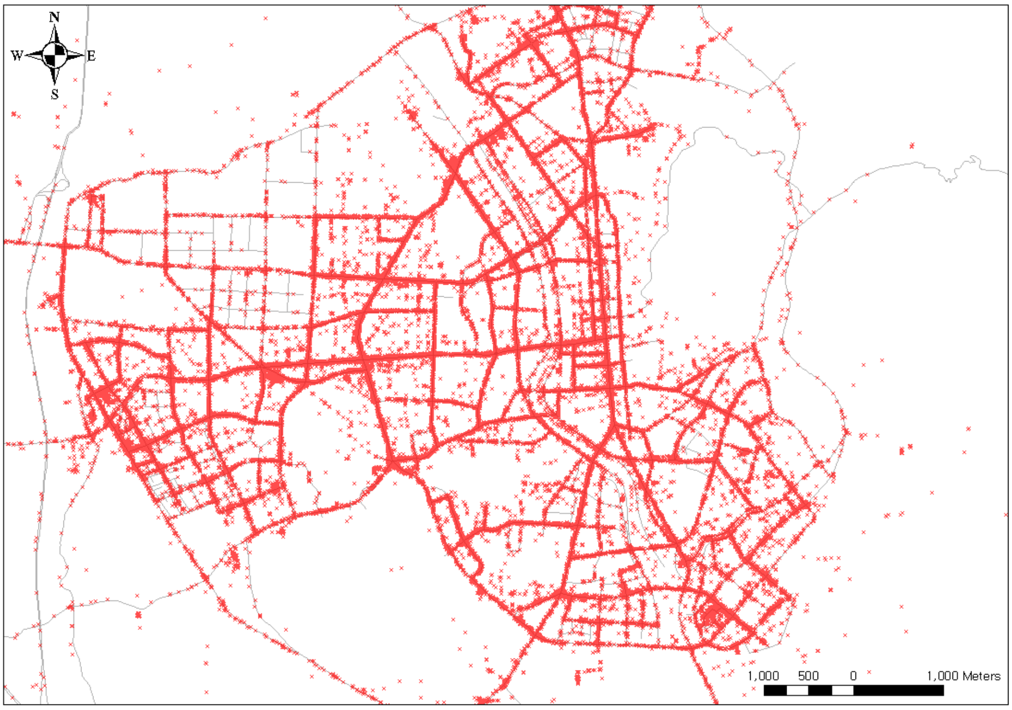


Figure 5. GPS Coordinates Distribution

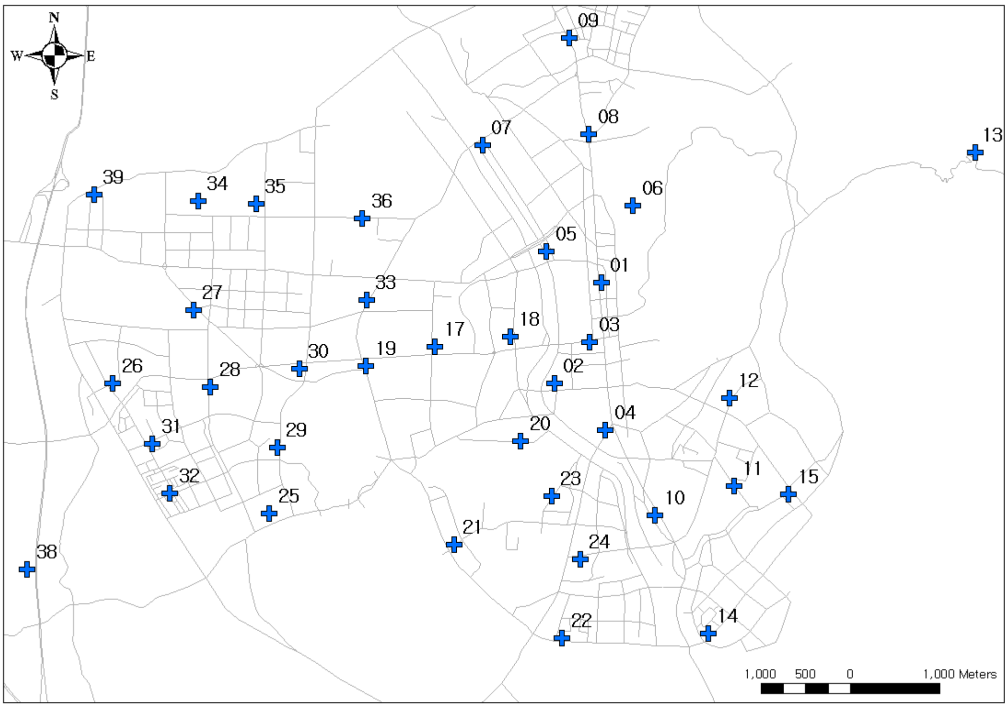


Figure 6. Cellular Base Station Distribution

5. RESULTS

5.1 Analysis Arterial

Link travel speed is calculated as mentioned earlier using GPS coordinates and cellular BSs. What is important is how well cellular phone estimate travel speed with only base station information. In order to check this, major 6 arterials like Figure 7 are selected and examined.

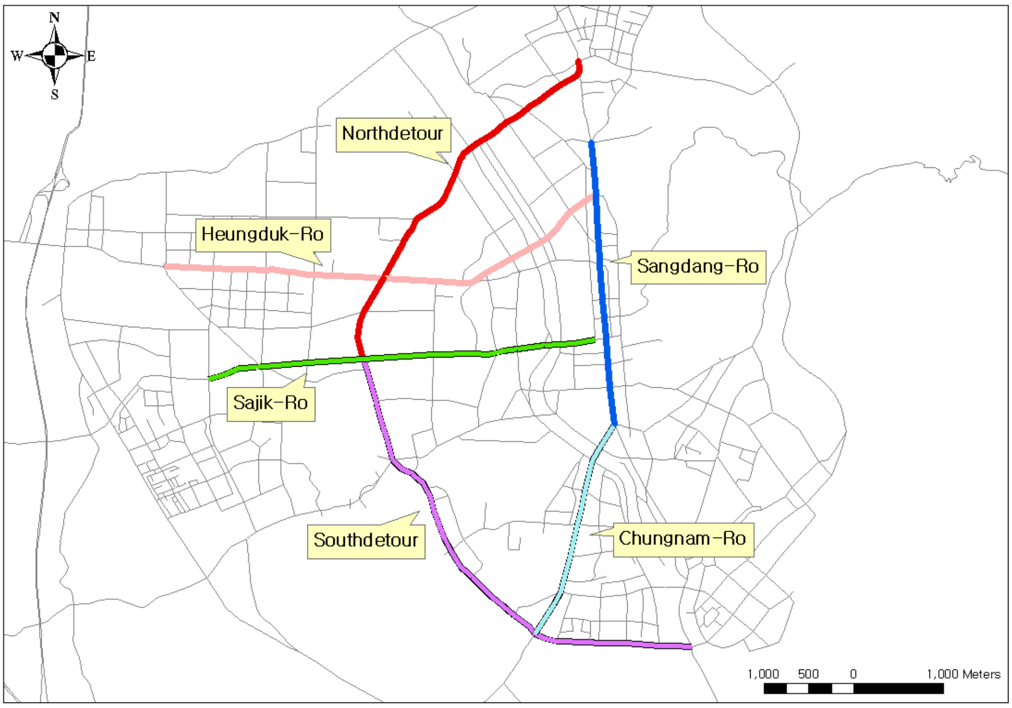


Figure 7. Interested Major Arterials

5.2 Map Matching and Shortest Path Search

Every GPS coordinates are matched to the nearest link and searched the path using Dijkstra shortest path algorithm based on the distance. And travel speed is calculated at every links.

Cellular base station data are followed same procedure like GPS. However, as mentioned earlier, GPS and cellular phone estimates different path and different speed.

5.3 Coverage Analysis of Cellular Base Station

The accuracy of estimated travel time depends on the cellular BS's coverage, namely BS's service area. If the coverage is small, the estimated travel speed will be close to the real travel speed of GPS and vice versa. In this study, GPS coordinates are assumed to be a true value and then the coverage of BSs is calculated as an average value as follows as Figure 8. The coverage is almost from 300 meter to 600 meter in CBD and is a few kilometers in suburbs.

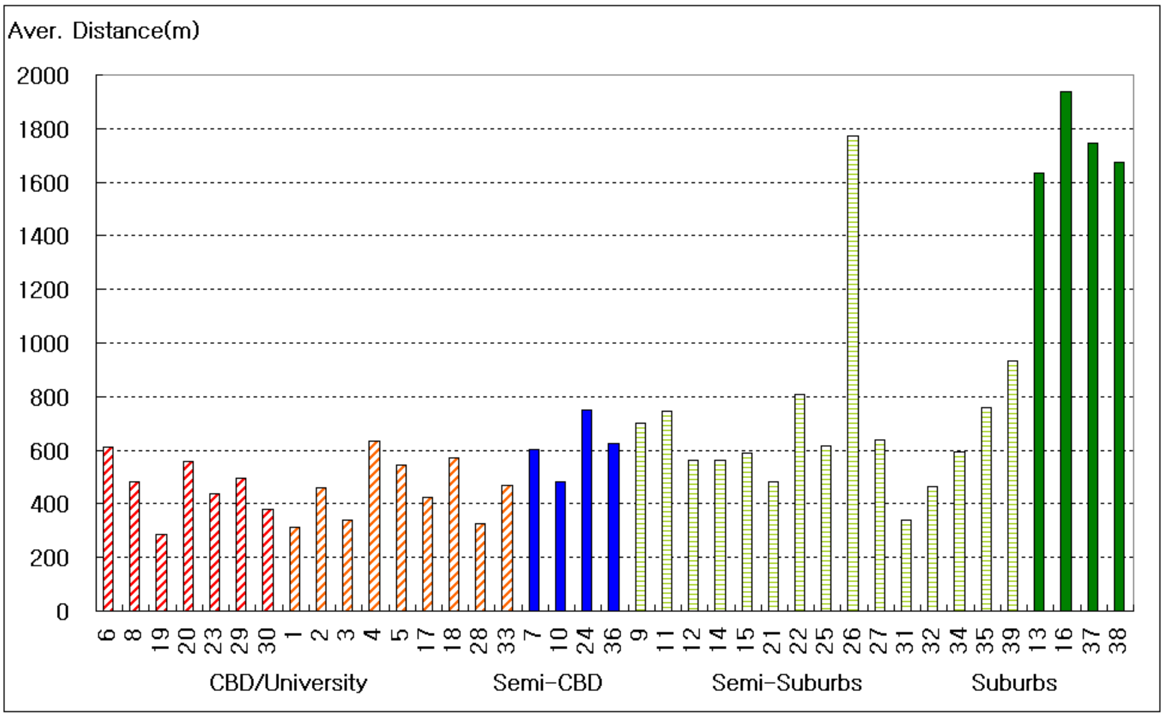


Figure 8. Averaged Coverage of Cellular Base Stations

5.4 Travel Speed Estimation and Results

To validate the travel speed estimated from cellular BS, the speed from GPS is assumed as true value. The travel speed is estimated from GPS and cellular BS at every hour respectively. Only passenger-boarding data is used in estimating travel speed because the travel speed is different by passenger existence. In this study, the result of morning peak from 7:00 to 9:00 is presented and compared.

Each links has sample size below 30. To compare two speeds over every links, F-test and t-test are implemented to check the speed distribution of two speeds. F-test should be carried out before t-test.

F-test show that almost links has same variance of speed except some cases. Therefore, the approximate t-test of Satterthwaite is implemented to test the average speed. The approximate t-test can be applied when the sample size is small and the variance of two populations is unknown. Satterthwaite's t-statistics is like as follows with n_0 degrees of freedom.

$$\frac{\left(\overline{X_1}-\overline{X_2}\right)-\left(\mu_1-\mu_2\right)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \tag{3}$$

Here, n_0 is the natural number close to the next formula.

$$\frac{\left(S_1^2/n_1+S_2^2/n_2\right)^2}{\frac{\left(S_1^2/n_1\right)^2}{n_1+1}+\frac{\left(S_s^2/n_s\right)^2}{n_2+1}}-2 \tag{4}$$

Namely, test statistics is $\overline{X_1}-\overline{X_2}$, and null hypothesis is $\mu_1=\mu_2$. Level of significance is α, null hypothesis is rejected at next conditions.

$$\overline{X_1}-\overline{X_2}>t_{n_0,\alpha/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}} \text{ or } \overline{X_1}-\overline{X_2}<-t_{n_0,\alpha/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}} \tag{5}$$

Not only statistical test, APE (Absolute Percent Error) will be checked to see the error ratio.

$$Absolute\,Percent\,Eror=\left|\frac{Speed_{BS}-Speed_{GPS}}{Speed_{GPS}}\right|\times 100 \tag{6}$$

Table 1 shows the result of statistical test as an example. T-test results show that almost every links have the same speed estimate between GPS-based speed and cellular phone BS-based speed.

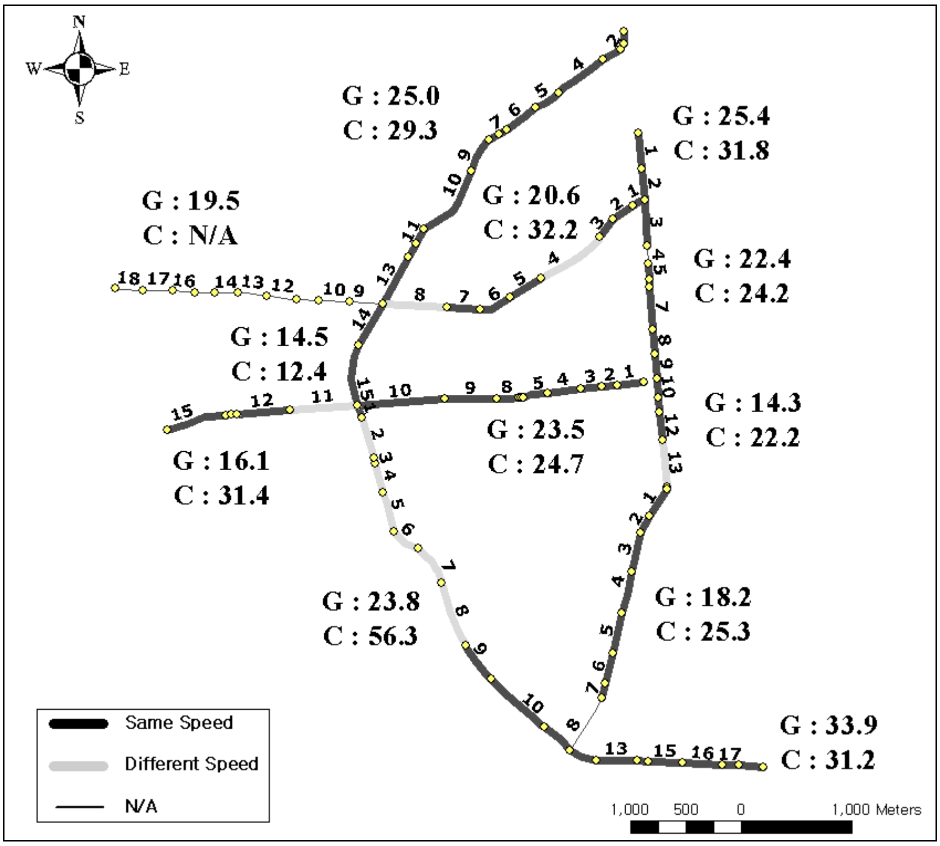
Figure 9 and Table 2 show the results of above statistical test at every 6 arterial links and show arterial travel speed per section. Except for some links, almost every links show that estimated speed by cellular BSs is the same speed by GPS coordinates. This means that cellular BS information may be a good information source to estimate arterial travel speed.

Table 1. Results of Estimated Like Travel Speed Example (Sajik-Ro)

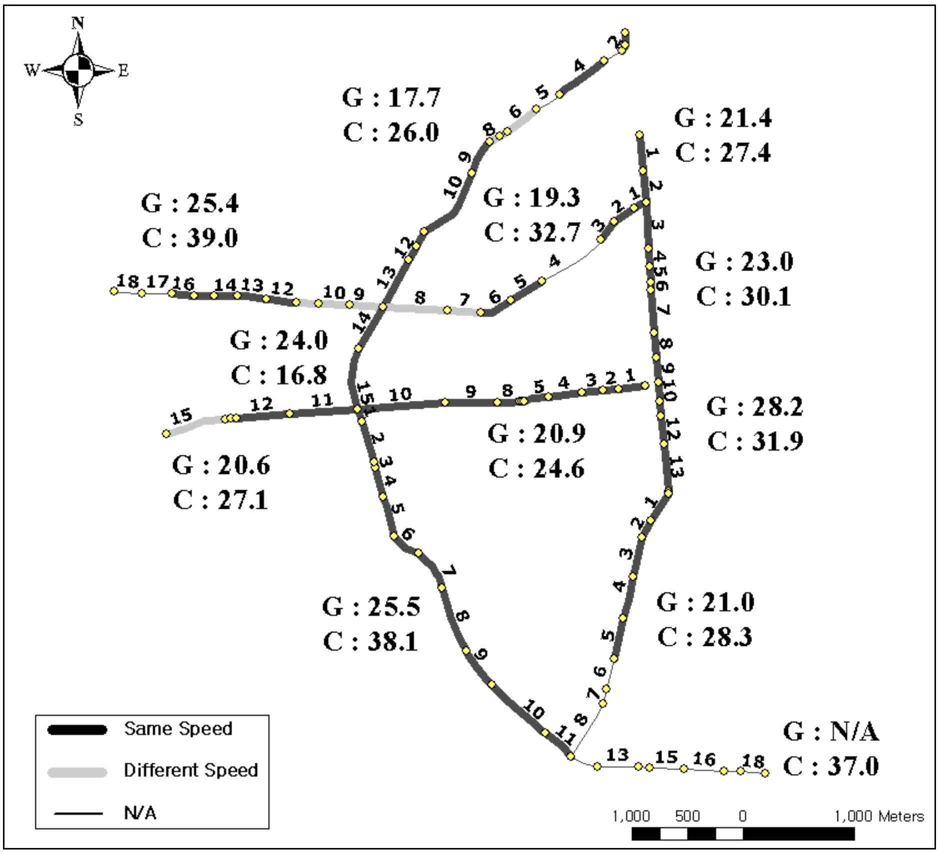
Link	Upward					Downward				
	GPS (kph)	Cell. P. (kph)	AE (%)	F-test	t-test	GPS (kph)	Cell. P. (kph)	AE (%)	F-test	t-test
1	21.4	21.3	0.3	Accept	Accept	20.4	20.7	1.1	Reject	Accept
2	20.2	20.7	2.2	Accept	Accept	17.5	21.0	19.8	Accept	Accept
3	29.5	21.9	25.8	Accept	Accept	20.7	21.6	4.0	Accept	Accept
4	28.6	21.9	23.6	Accept	Accept	20.7	21.6	4.0	Accept	Accept
5	27.3	26.7	2.1	Reject	Accept	22.8	28.0	22.9	Accept	Accept
6	31.5	26.7	15.2	Accept	Accept	23.8	28.0	17.7	Accept	Accept
7	31.5	26.7	15.2	Accept	Accept	23.0	27.1	17.4	Accept	Accept
8	31.5	26.7	15.2	Accept	Accept	23.0	27.1	17.4	Accept	Accept
9	25.0	26.7	6.8	Accept	Accept	20.7	27.1	30.9	Accept	Accept
10	19.4	26.7	37.9	Accept	Accept	21.0	26.1	24.4	Accept	Accept
11	13.1	28.4	116.7	Accept	Reject	20.3	20.8	2.5	Reject	Accept
12	16.9	27.1	60.1	Accept	Accept	21.4	21.4	0.4	Accept	Accept
13	18.3	35.0	91.5	Accept	Reject	18.5	33.1	78.9	Accept	Accept
14	19.6	40.7	107.8	Accept	Reject	19.5	55.2	182.7	Accept	Reject
15	20.0	40.7	104.0	Accept	Accept	20.7	55.2	167.0	Accept	Reject

Table 2. Result of Estimated Arterial Travel Speed

Arterial	Section	Link No.	Upward			Downward		
			GPS (kph)	Cell.P. (kph)	AE (%)	GPS (kph)	Cell.P. (kph)	AE (%)
South detour	1	1~11	23.8	56.3	137.0	25.5	38.1	49.7
	2	12~18	33.9	31.2	7.9	-	37.0	-
North detour	1	1~13	25.0	29.3	16.9	17.7	26.0	47.0
	2	14~15	14.5	12.4	14.0	24.0	16.8	30.1
Sajik-Ro	1	1~10	23.5	24.7	5.1	20.9	24.6	17.5
	2	11~15	16.1	31.4	94.7	20.6	27.1	31.3
Sangdang-Ro	1	1~2	25.4	31.8	25.2	21.4	27.4	28.2
	2	3~9	22.4	24.2	7.7	23.0	30.1	31.2
	3	10~13	14.3	22.2	55.9	28.2	31.9	13.2
Chungnam-Ro	1	1~8	18.2	25.3	38.7	21.0	28.3	35.0
Heungduk-Ro	1	1~8	20.6	32.2	56.7	19.3	32.7	69.5
	2	9~18	19.5	-	-	25.4	39.0	53.6



(a) Upward



(b) Downward

Figure 9. Results of Estimated Arterial Travel Speed

Heungduk-Ro shows that statistical test is not available from link 9 to link 18 in the direction of upward because the sample was never chosen. This sampling problem occurs to the South detour downward, too.

However, in the case of Heungduk-Ro downward, even though there are both samples of GPS and cellular BS, there are many links that shows different travel speed. It comes from the location of BS and path finding algorithm. Because there are little cellular base stations along Heungduk-Ro, the estimated speed from cellular BS is different from that from GPS coordinates. This problem may be improved by modifying map-matching method like that BS is matched to the major road and find path along major roads.

Besides, in case of South detour upward, there are many links that the speed from cellular BS is differ from the speed of GPS. This may be due to BS coverage problem. Figure 10 shows the coverage of BS 26. That BS 26 shows strange shape similar to dumbbell. This coverage can be translated like this. Even though the real GPS is located in the South detour, it is thought as to be located in coordinates near 26. These phenomena distort the estimated travel speed. Therefore regulating the cellular BS is need to be done by mobile communication company.

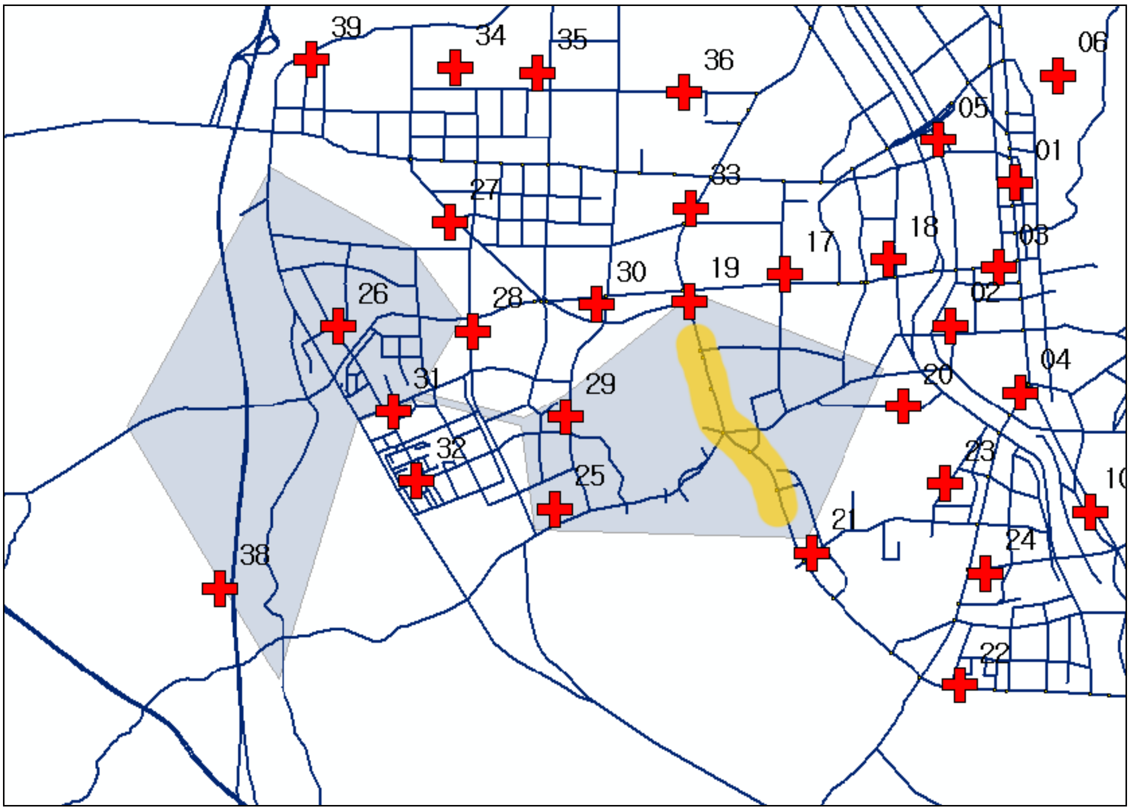


Figure 10. BS Coverage (26 BS)

5.5 Passenger's Impact on Travel Speed at Taxi

According to the characteristics of data acquisition system which collect GPS coordinates at every 2 to 5 minutes under passenger existence, travel speed when passenger is boarding a taxi can be estimated, and travel speed when passenger is not boarding a taxi can be estimated.

The speed of taxi is usually thought to be different between passenger existence and passenger non-existence. Taxi usually runs fast when passenger is riding a taxi, but after passenger is alighting, taxi driver runs slowly or stand on for a moments.

To confirm this phenomenon, F-test and z-test are applied. Table 3 shows the results of F-test. According to the F-test, there is no significant difference between variances.

Table 3. F-test result

	Passenger existence	Passenger Non-existence
Average	24.072	18.278
Variance	209.530	236.336
Sample No.	16110	42516
Degrees of Freedom	16109	42515
F ratio	0.8866	
F reject: One-tailed test	0.9786	

Because the variance is not different, z-test is applied to check whether the average speed is same or not. Table 4 shows the results of z-test. The test shows that the average speed is significantly different between passenger existence and non-existence.

Table 4. z-test result

	Passenger existence	Passenger Non-existence
Average	24.072	18.278
Known Variance	209.530	236.336
Sample No.	16110	42516
Hypothetical difference	0	
z statistics	42.5218	
z rejection value (One-tailed test)	1.6449	
z rejection value (Two-tailed test)	1.9600	

According to above analysis, we can conclude that there is a significant difference between passenger existence and non-existence. This result is very suggestive to traffic monitoring company. If taxi is used as a probe to estimate travel speed or travel time, it means that the sampled data should be divided into two categories – passenger existence and passenger nonexistence. In this study, this was possible because data transmission system connected to taxi meter.

6. CONCLUSION

In this study, travel speed and travel time estimation methodology using GPS and base station information of cellular phone is presented and the effectiveness and usefulness are validated by field data.

This study acquired coordinates and time information of GPS and base station by establishing GPS equipment and cellular phone on taxies that communicate with management center in every 2 to 5 minutes. And these data are matched on digital map by the map-matching algorithm. And the distance between two adjacent points is calculated by Dijkstra algorithm. Link travel speed and time is estimated in the end using those travel distance and time. Finally, travel time is estimated by both GPS positions and base station positions of cellular phone, and the travel time by GPS positions is supposed as true value and the travel time by base station of cellular phone is validated statistically.

In this study, travel time of 6 main arterial roads is estimated in the city of Korea, Cheongju. According to the result of statistical analyses, the estimated travel time using cellular base station is almost similar to the estimated travel time by GPS. It means that cellular base station information can be a good alternative to estimate link travel time.

However, the accuracy is low when the cellular phone is served from a distant base station rather than the closest one. And map-matching method influence the accuracy of speed.

Significance test between travel speed with passenger and travel speed without passenger is performed and it shows that there is a significant difference of travel speed between passenger existence and nonexistence. This imply that information of passenger existence connecting to taxi meter is required in the case taxi is used as a probe vehicle.

This study is early work of traffic information generation using cellular phone information, and it shows that estimated travel speed is meaningful for major roads. But this method couldn't estimate every links' travel time. GPS phones coordinates would offer more detailed and accurate travel time if the data acquisition is possible.

In this study, only taxi was used as a sampling mode, but information of all modes would be collected in the future. In such case, more algorithms should be studied such as travel mode classification algorithm, and sample duplication eliminating method when people are riding in a bus and so on.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to Ministry of Construction & Transportation and Engineering Research Institute SNU by which they are supported, for giving them the opportunity to carry out this paper.

REFERENCES

Axhausen, K.W., Schonfelder, S., Wolf, J., Oliveira, M., and Samaga, U.(2004) Eighty Weeks of GPS Traces: Approaches to Enriching Trip Information, **83th TRB Annual Meeting**.

Cayford, R. and Johnson, T.(2003) Operational Parameters Affecting Use of Anonymous Cell Phone Tracking for Generating Traffic Information, **82th TRB Annual Meeting**.

Chung, E.H. and Shalaby, A.(2004) Development of a Trip Reconstruction Tool to Identify Traveled Links and Used Modes for GPS-Based Personal Travel Surveys, **83th TRB Annual Meeting.**

Lee, C.J.(2003) Transportation information based on personal communication services base station data, master dissertation, Seoul National University

Li, H., Guensler, R., Ogle, J. and Wang, J.(2004) Using GPS Data to Understand the Day-to-Day Dynamics of the Morning Commute Behavior, **83th TRB Annual Meeting**.

Meaker, J. and Horner, M.W.(2004) An examination of the use of APRS data for enhancing transportation planning operations, **83th TRB Annual Meeting.**

Ormsby, T., Napoleon, E, Bruke, R., Groessl, C., and Feaster, L.(2001) **Getting to know ArcGis Desktop: Basics of ArcView, ArcEditor, and ArcInfo**, ESRI Press. California.

Quiroga, C.A. and Bullock, D.(1999) Travel Time Information using GPS and Dynamic Segmentation Techniques, **78th TRB Annual Meeting**.

Wolf, J., Hallmark, S., Oliveira, M., Guensler, R., and Sarasua, W.(1999) Accuracy Issues with Route Choice Data Collection using GPS, **78th TRB Annual Meeting**.

Yim, Y.B. and Cayford, R.(2002) Positional Accuracy of Global Positioning System and Cellular Phone Tracking for Probe Vehicles, **81th TRB Annual Meeting**.