THE MAP MATCHING ALGORITHM OF GPS DATA WITH RELATIVELY LONG POLLING TIME INTERVALS

Jae-seok YANG
Graduate Student
Graduate School of Engineering
Seoul National University
San56-1, Shillim-dong, Gwanak-gu, Seoul,
151-742 Korea
Fax: +82-02-889-0032
E-mail: bluedge0@snu.ac.kr

Seung-pil KANG
Associate Professor
Graduate School of Engineering
Seoul National University
San56-1, Shillim-dong, Gwanak-gu, Seoul,
151-742 Korea
Fax: +82-02-889-0032
E-mail: kangsp@snu.ac.kr

Kyung-soo CHON
Full Professor
Graduate School of Engineering
Seoul National University
San56-1, Shillim-dong, Gwanak-gu, Seoul,
151-742 Korea
Fax: +82-02-889-0032
E-mail: chonks@snu.ac.kr

Abstract: Many map matching algorithms have been developed to match GPS points to a digital map in previous studies. But the previous studies assume short polling time intervals(about 1 second) of the GPS data. And the map matching algorithms of such studies are not appropriate for the GPS data with relatively long polling time intervals(about 2~5 minutes). In this paper, we will review the previous map matching algorithms and discuss the map matching algorithms which can be used under circumstances of relatively long polling time intervals.

Key Words: Map matching, GPS, Polling time

1. INTRODUCTION

Roles of the GPS in obtaining the traffic information will become significant more and more. To generate traffic information of better quality, the accuracy of GPS data and digital maps should be guaranteed.

Although the SA(Selective Availability) was removed on May 1, 2000, the positional error of GPS without additional error correction methods goes up to about 30 meters. In case of digital maps, various factors like errors in digitizing map and the distance from the road centerline to the both end of the road makes the positional error of nearly 20 meters.

As a result, the direct overlay of positional data obtained from GPS does not reconcile with a digital map. Therefore GPS data need to be corrected with various methods to match with a digital map and we call this procedure a map matching.

Algorithms of the map matching have been developed continuously and they can be classified into two categories roughly. First, map matching algorithms which consider only geometric relationships between GPS data and a digital map. Secondly, map matching algorithms which

consider not only geometric relationships but also the topology of the road network and the history of GPS data. It has been reported that the latter worked better mostly.

The first map matching algorithms can be classified again into the map matching algorithm using the distance of point-to-curve, one using the distance of curve-to-curve and one using the angle of curve-to-curve. Some past studies used the distance of point-to-point. But these vertex-based map matching algorithms are appropriate when one pursues simplicity rather than accuracy.

The second map matching algorithms use the result of map matching at time t-1 for the map matching of GPS data at time t. And for the selection of candidate segments which GPS data will be matched, the topology of the road network is inputted as a constraint. But these algorithms should be used under particular prudence. For example, if the result of map matching at time t-1 is wrong then the result of map matching after that time will be wrong also. Thus, it should be guaranteed that the result of map matching at time t-1 is exact to use these algorithms. Besides, if the vehicles with a GPS receiver follow abnormal routes(e.g. the left turn on the left turn restricted intersection) we cannot expect the right result of map matching because the normal topology respects traffic regulations.

Most of previous studies of map matching was under circumstances of very short polling time interval(about 1 second). The shorter the polling time interval is, the better the performance of the map matching algorithm is, because the availability of the GPS data history will be increased. But in practice, various problems restrict the shortening of the polling time interval. For example, there should be some telecommunication method to collect the GPS data of many persons on real time. And if the telecommunication is accomplished by the third telecommunication company, very short polling time will inevitably accompany with a vast cost as if you use your cellular phone for a long time, you will have to pay high cost.

Therefore, we will review the previous map matching algorithms and discuss the map matching algorithms which can be used under circumstances of relatively long polling time intervals(about 2~5 minutes). And the focus will be laid on the map matching algorithm considering not only geometric relationships but also the topology of the road network and the history of GPS data.

2. LITERATURE REVIEW

2.1 Discrimination of Vertex-based and Segment-based Map Matchings

Nodes, vertices, links and segments should be defined clearly before the literature review. Nodes and links are relatively well known. For example, an intersection is a node. A link is a section of road between intersections. White points in Figure 1 are nodes and there is only one link in Figure 1. When the real road network is digitized, a curve is described with a set of many straight lines. Vertices are points which separate these straight lines, and each straight line is a segment.

Figure 1. The Composition of the Road Network

The clear definition of the compostion of the road network helps us to understand previously studied map matching algorithms which can be roughly classified into two categories; vertex-based map matchings and segment-based map matchings. The Figure 2 shows these two map matching algorithms.
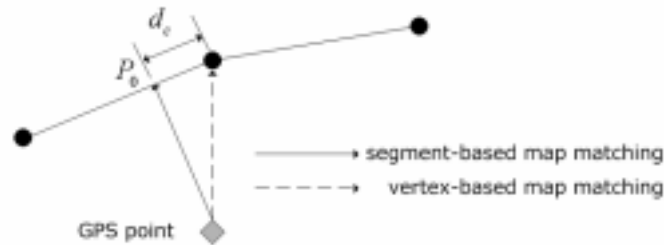

Figure 2. Vertex-Based and Segment-Based Map Mathcings

If the real position of the vehicle with a GPS receiver is P_0, the vertex-based map matching implies a positional error of d_e. While the segment-based map matching does not have this error. In fact, the Figure 2 shows map matchings under the ideal condition. There can be another positional error although a GPS point is matched to a segment properly(e.g. the real position of the vehicle can be the vicinity of P_0), but this kind of error is not dealt in this paper.

Sometimes vertex-based map matchings are used even though the weakness mentioned above. There can be many reasons for this; for the simplicity of the map matching process or the data structure which is not appropriate for the segment-based map matching. Nowadays, there is no reason to use vertex-based map matchings because the performance of computers and GIS packages is significantly improved.

2.2 Map Matchings Using the Geometric Relationship

The basic form of the previous map matching algorithms is the map matchings using the geometric relationships between the GPS points and the network. Noh and Kim(1998) classified such map matchings into three categories; the map matching using the distance of point-to-curve, one using the distance of curve-to-curve and one using the angle of curve-to-curve.
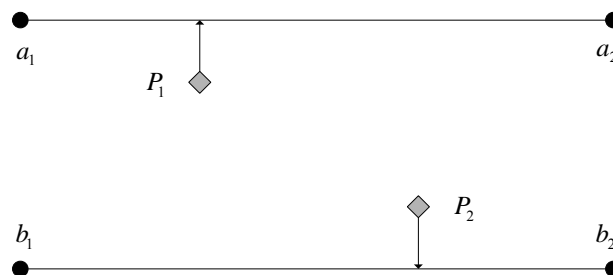

Figure 3. The Map Matching Using the Distance of Point-to-Curve

The map matching using the distance of point-to-curve matches a GPS point to the nearest segment. Because it is the simplest form of the map matchings, it has many problems. Figure 3 shows the example of the problem.

a_1 and a_2 are the vertice of the road A. b_1 and b_2 are the vertice of the road B. P_1 and P_2 are the two consecutive GPS points. If the GPS points are matched to the nearest segment, P_2 is matched to the road B while P_1 is matched to the road A. That is to say, the map matching concluded the unrealistic result that the car jumped from the road A to the road B. Map matching algorithms which use the information of only one point will frequently meet such a problem. To overcome this problem, the map matchings using the distance of curve-to-curve or the angle of curve-to-curve was developed.
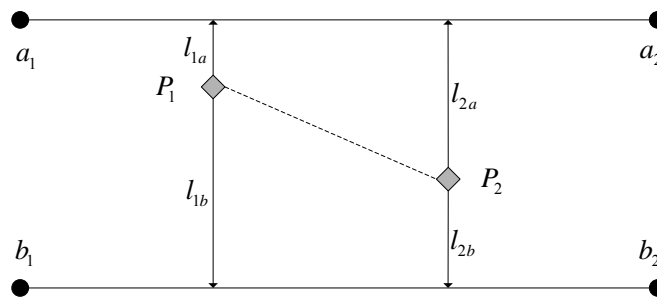
Figure 4. The Map Matching Using the Distance of Curve-to-Curve

The map matching using the distance of curve-to-curve matches two GPS points to the segment which has the shortest distance from the reference line. The reference line is the line which connect two GPS points. There are many methods of measuring the distance between the reference line and the segment. The simplest method is to sum up the distances between two end points. For example, P_1 and P_2 are matched to the road A because $(l_{1a} + l_{2a}) < (l_{1b} + l_{2b})$ in Figure 4.
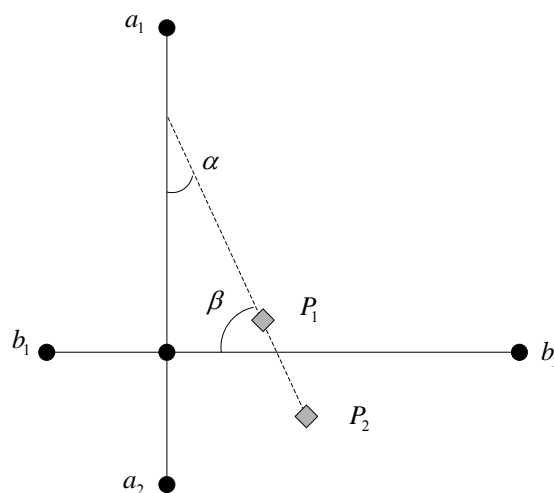
Figure 5. The Map Matching Using the Angle of Curve-to-Curve

The map matching using the angle of curve-to-curve matches two GPS points to the segment which has the smallest angle from the reference line. The reference line is the line which

connect two GPS points, too. For example, P_1 and P_2 are matched to the road A because $\alpha < \beta$ in Figure 5.

2.3 Map Matchings Using the Network Topology and the Data History

Most map matching algorithms using the network topology and the data history, which also use the geometric relationships, have similar main algorithm.

Greenfeld(2002) matched the GPS points to the nearest link using only the geometric relationships when the point is the first GPS point or the distance between the previous point(P_{t-1}) and the present point(P_t) is too long. On the other hand, when the distance between P_{t-1} and P_t is under a particular threshold, the algorithm evaluates the proximity and the orientation between the reference line which connects these two points and the segment(S_{t-1}) to which the previous GPS point(P_{t-1}) was matched. And P_t is matched to S_{t-1} when the evaluation criteria is met. When the evaluation criteria is not met, segments which are directly connected to S_{t-1} is evaluated through the same process.

White et al.(2000) also suggested a similar algorithm. When the result of the map matching of P_{t-1} is guaranteed, the result is referred to the map matching of P_t, and P_t is matched to one of the segments which is directly connected to S_{t-1}.

2.4 The Limitation of the Previous Map Matching Algorithms

First, the availability of the geometric relationships should be examined. The map matching algorithm using the distance of curve-to-curve or the angle of curve-to-curve are more refined algorithms than the map matching algorithm using the distance of point-to-curve. But under the circumstance of relatively long polling time intervals, the map matching algorithm using the distance of curve-to-curve is not appropriate because there can be many links between two consecutive GPS points.
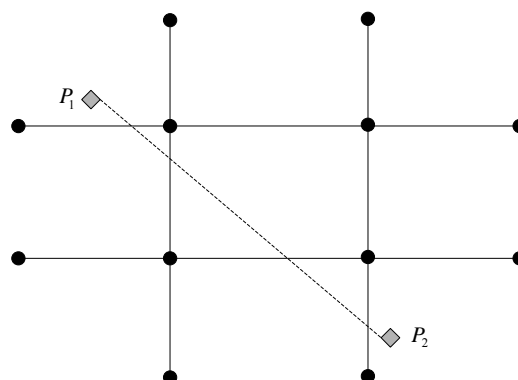

Figure 6. The Problem of Relatively Long Polling Time Intervals

Figure 6 shows such a problem, and it is not clear which segment should be evaluated with the reference line in those case. The map matching algorithm using the angle of

curve-to-curve has the same problem.

Secondly, the availability of the network topology and the data history is also limited for the same reason. For the map matching of P_t, to evaluate the segments which is directly connected to S_{t-1} as Greenfeld and White did, is not appropriate. In 2~5 minutes, the vehicle with a GPS receiver can reach the segment which is very far from S_{t-1}.

3. A NEW MAP MATCHING ALGORITHM

With relatively long polling time intervals of 2~5 minutes, the history of GPS data can not fully utilized like the previous studies which assume short polling time intervals of 1 second. So, the full utilization of the geometric relationships between the GPS points and the road network becomes more important. But the map matching algorithms using the distance of curve-to-curve or the angle of curve-to-curve are not appropriate with relatively long polling time intervals as mentioned above. Therefore, the map matching algorithm using the distance of point-to-curve will be used in this paper.

3.1 The Utilization of the Distance to the Nearest Link and Secondly Nearest Link

But the map matching using the distance of poin-to-curve has problem of abnormal result which was shown in Figure 3. To overcome this problem and to increase the accuracy of the map matching, not only the distance from a GPS point to the nearest link(will be abbreviated as NL), but also the distance from the GPS point to the secondly nearest link(will be abbreviated as 2NL) will be examined together. For example, when P_1 was matched to P_1' and P_2 was matched to P_2', the latter can be said to be more accurate, and this logic will be utilized in the new map matching algorithm.
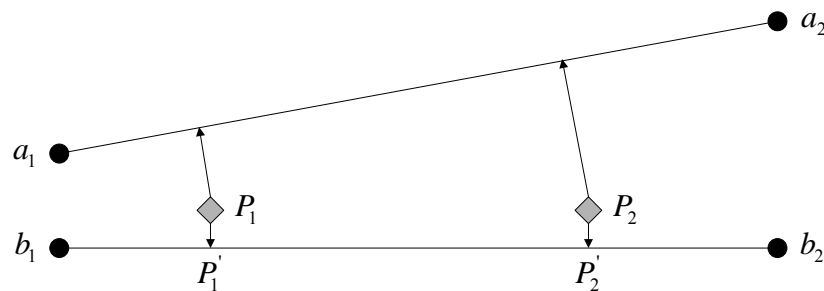

Figure 7. The Utilization of the Nearest Link and the Secondly Nearest Link

To include this logic to the map matching algorithm, the particular threshold is needed. If the ratio of the distance from a GPS point to 2NL to the distance from the GPS point to NL is larger than the threshold, the GPS point can be matched to NL. In this paper, the threshold is set up as 2. In other words, if the distance from a GPS point to 2NL is more than twice as large as the distance from the GPS point to NL, the GPS point is matched to NL. The threshold of 2 does not have some theoretical base. This threshold can be raised to faster the process of the map matching and can be lowered to increase the accuracy of the map matching.

It is also possible that the distance to 2NL is not long enough to match the GPS point to NL. For example, if the distance to NL is 10m and the distance to 2NL is 12m, the GPS point should not be matched to NL.

This problem occurs under two circumstances which are described below. One is the GPS point which is in the vicinity of an intersection and the other one is the GPS point which is between two nearly parallel links. P_1 shows the former case, and P_2 in Figure 8 shows the latter case in Figure 8.
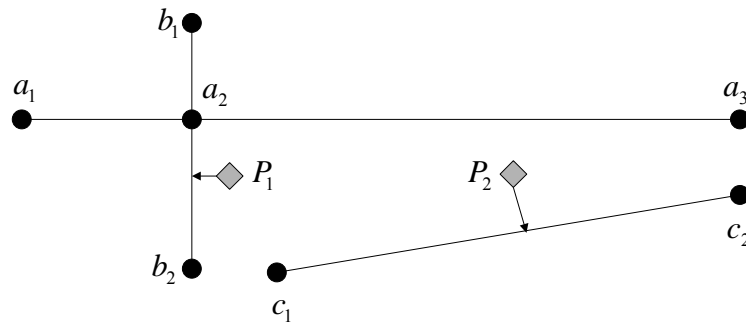
Figure 8. The Circumstances under which the Map Matching is Likely to Fail

3.2 The Problem in the Vicinity of an Intersection

This problem can be resolved by matching the GPS point to the nearest node. In other words, create a buffer with particular radius around a node. And if a GPS point is laid within the buffer, the GPS points is always matched to the node. Obviously, this process should be followed by the map matching using the NL-2NL relationship.

But, if the GPS point is matched to the nearest node, the result inevitably has some positional error which was shown in Figure 2. If we assume the polling time interval of 2 minutes and the vehicle speed of 20kph, the buffer of a 10m radius will generate the maximum positional error of ±1.5%. If the polling time interval becomes longer or the vehicle speed becomes faster, this error becomes lower. And such a low positional error does not matter mostly.

Now, the radius of the buffer should be determined. If there is an intersection on which two four-lane roads cross, the extent which is overlaid by the two roads can be completely covered with a buffer of 10m radius. But a GPS point has maximum error of about 10m. So, it is reasonable to set up the radius of a buffer as 20m. Obviously, this radius can be changed with the number of lanes which is connected to the intersection. If it is possible, the result of the map matching can be more accurate. But this paper assumes that there is no information about the number of lanes. Instead, the accuracy of the map matching will be evaluated respectively using the buffer of 10m radius and 20m radius in the field test.

3.3 The Problem between Two Nearly Parallel Links

This problem can be resolved by the utilization of the network topology and the data history. Before the description of the method, some notations have to be introduced to avoid

confusion.

P_i :   ith GPS point which is not map matched yet

P_i' :   ith GPS point which is map matched permanently

P_{i1}' :   ith GPS point which is map matched to NL temporarily

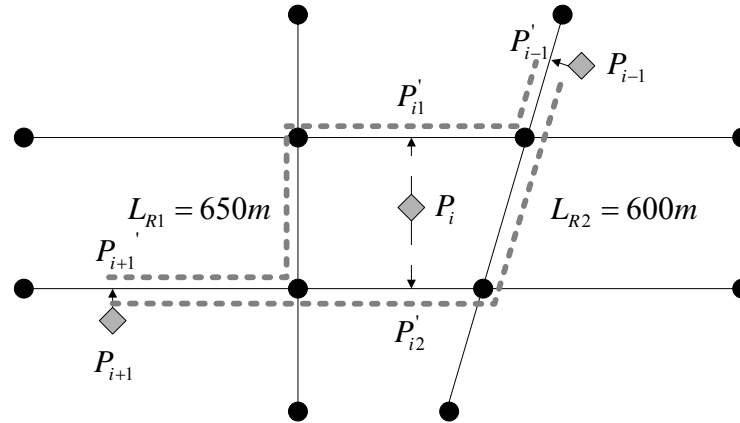P_{i2}' :   ith GPS point which is map matched to 2NL temporarily

Figure 9. The Utilization of the Network Topology and the Data History

The map matching of a GPS point which is between two nearly parallel links is done as follows. At first, the map matching of P_i is deferred and P_{i+1} is map matched to find P_{i+1}'. And then, the shortest path from P_{i-1}' to P_{t1}' and the shortest path from P_{t1}' to P_{i+1}' are searched respectively. And the sum of the distance of these two shortest path is named as L_{R1}. Similarly, L_{R2} is calculated with respect to P_{i-1}', P_{t2}' and P_{i+1}'. If $L_{R1} < L_{R2}$, P_i is matched to P_{t1}' and P_{t1}' becomes P_i' permanently, and vice versa.

Figure 9 shows this process. Because the distance from P_i to NL is 20m and the distance to the 2NL is 30m, the GPS point can not be matched to NL. In addition, no node have the buffer which can contain the GPS point. But the result of the calculation of L_{R1} and L_{R2} says that $L_{R1} < L_{R2}$, so P_i is matched to NL.

This methodology is based on the assumption that the driver do not choose the farther route when he move from P_{i-1}' to P_{i+1}'.

The shortest path above means that of physical distance, and the link-labeling Dijkstra algorithm(Noh and Nam, 1995) is used to find the shortest path. Using the physical distance means that this map matching algorithm assumes the all-or-nothing assignment. Needless to say, if one can use the travel time instead of the physical distance, the accuracy of this algorithm will be largely improved. But if it is hard to attain the travel time data on real time, to use the physical distance is recommended as the second best method.

Particular care should be taken to use P_{i-1}' and P_{i+1}'. If the P_{i-1}' or P_{i+1}' are the result of

incorrect map matchings, the result of the map matching of P_i which use P'_{i-1} and P'_{i+1} also can be incorrect. Therefore, the correctness of P'_{i-1} and P'_{i+1} should be guaranteed. And in this paper, if a GPS point was matched by the NL-2NL relationship or the node buffer, the result is treated as a correct result. When the accuracy of P'_{i-1} or P'_{i+1} can not be guaranteed, P'_{i-2} is used instead of P'_{i-1} or P'_{i+2} is used instead of P'_{i+1}. But this situation did not occur so many times in the field test.

There is one more problem in this process. The first GPS point does not have a previous GPS point and the last GPS point does not have a next GPS point. If they fall in the one of the node buffers fortunately, it does not matter. If not, the GPS point has no way but to be matched to NL despite the probability of an error.

There is still a problem. If the first GPS point was not map matched by the NL-2NL relationship or the node buffer, now the second GPS point will be under the same situation which was experienced by the first GPS point. In this case, the second GPS point is map matched with the same process which matched the first GPS point. GPS points after the second GPS point will be map matched with the same process if it is necessary.
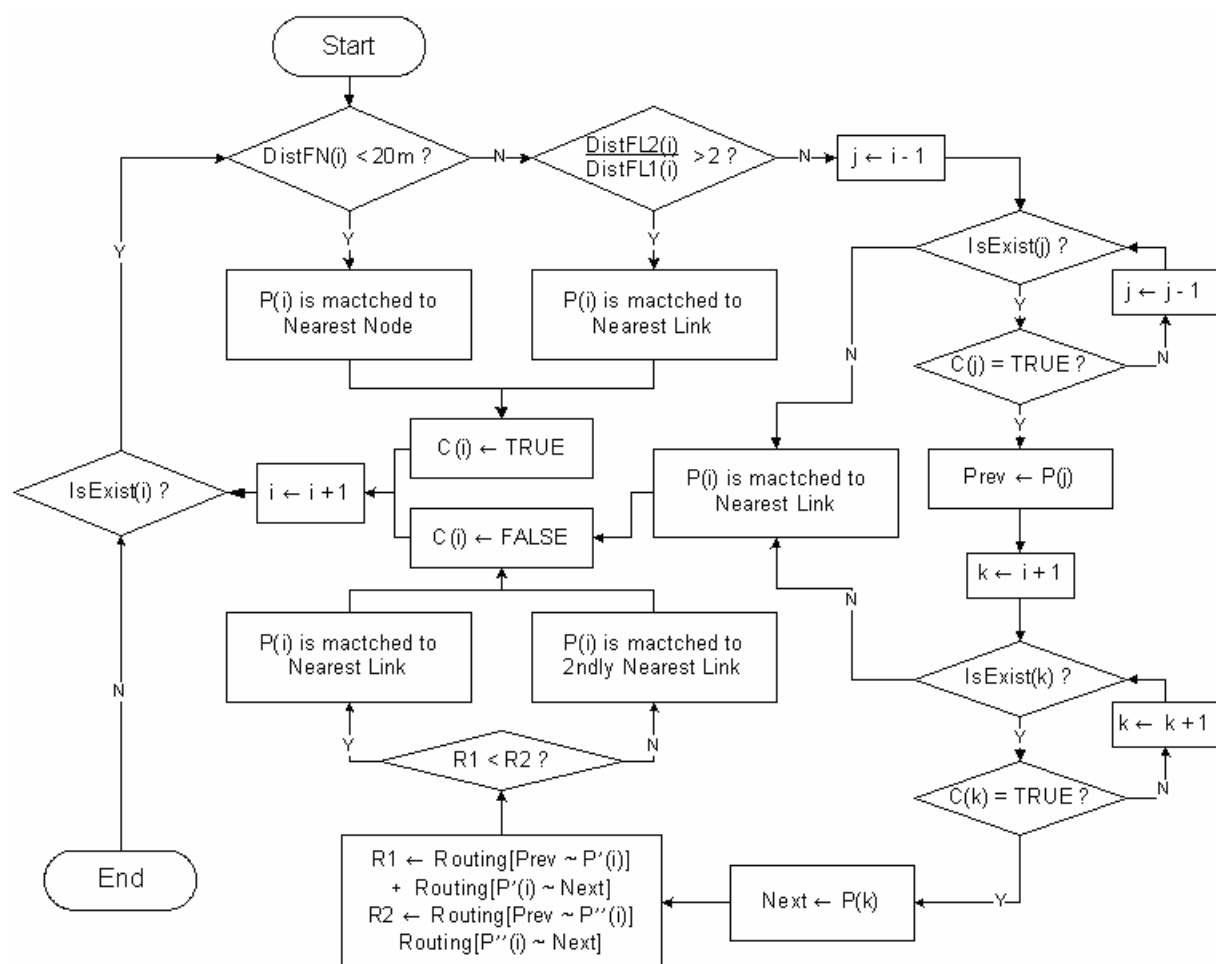

Figure 10. The Flowchart of the New Map Matching Algorithm

Figure 10 shows the flowchart of the new map matching algorithm which was described until now. It uses the notations below.

P(i):   ith GPS point
P'(i):   ith GPS point which is map matched to NL temporarily
P''(i):   ith GPS point which is map matched to 2NL temporarily
C(i):   guarantee index, P(i) can be used to find the shortest path if C(i)=TRUE
Prev: previous GPS point which is used to find the shortest path
Next: next GPS point which is used to find the shortest path
DistFN(i):   distance to the nearest node from P(i)
DistFL1(i):   distance to the NL from P(i)
DistFL2(i):   distance to the 2NL from P(i)
IsExist(i):   function which returns FALSE if ith GPS point does not exist
Routing[P(i)~P(j)]:   function which returns the distance of the shortest path between P(i) and P(j)

4. FIELD TEST

To evaluate the accuracy of the new map matching algorithm, totally about 9,500 GPS points was collected by the vehicle equipped with a GPS receiver(PDA: Compaq iPAQ H3630, GPS module: Royaltek RGM-2000) for three times. The study area was Gwanak-gu, Seoul. Table 1 and Figure 11 show the routes, time periods and numbers of GPS points of the field test.

Table 1. Time Period and the Number of Collected GPS Points

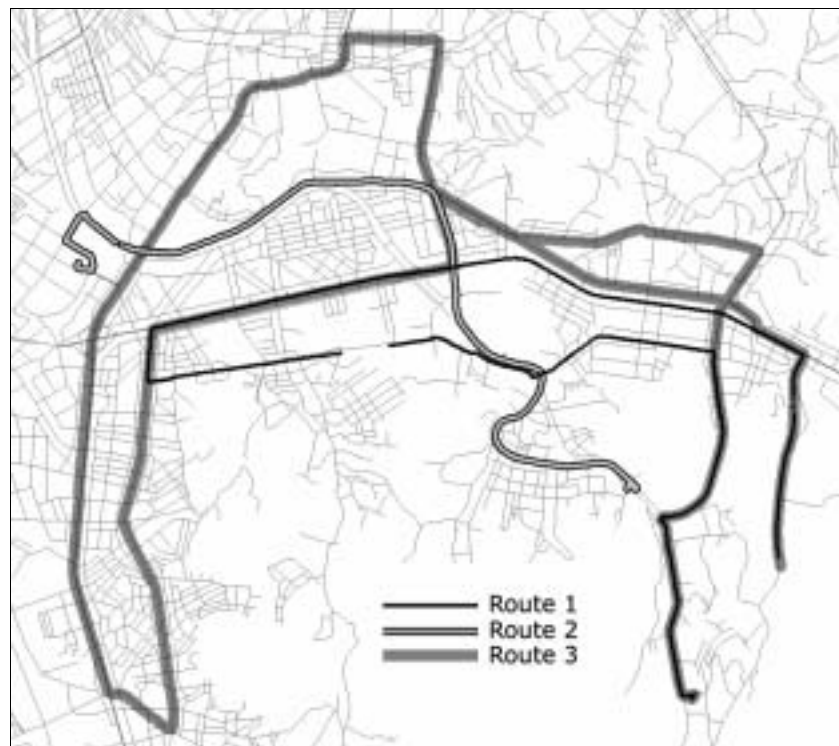| Route	Time period	Number of GPS points
1	04/09/19 00:11:37~00:49:38	2282
2	04/09/21 19:48:19~20:18:35	1817
3	04/10/22 00:06:21~01:20:23	4443

Figure 11. The Route for the Collection of GPS Points

Actually, the GPS points were collected every 1 second, but the focus of this paper is the map matching of GPS data with relatively long polling time intervals. So, whole GPS points were filtered every 2, 3, 4 and 5 minutes respectively. And these four groups were map matched by the new map matching algorithm respectively to evaluate the accuracy of the algorithm. Figure 12 shows the example of the GPS points which were filtered every 2 minutes.
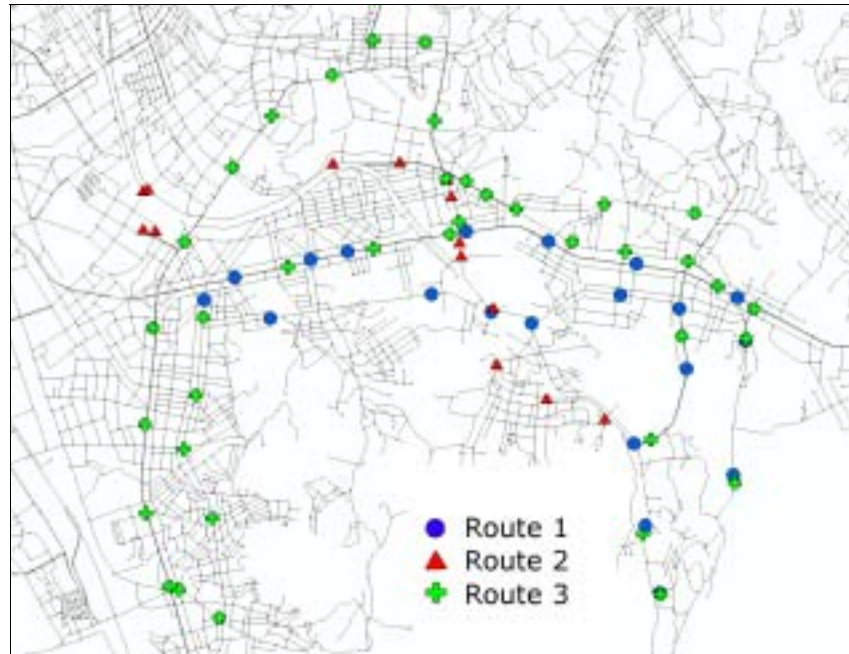

Figure 12. The GPS Points which were Filtered Every 2 Minutes

Before the exhibition of the result, the measure which can evaluate the accuracy of the map matching algorithm should be defined. The result of the map matchings using the NL-2NL relationship or the network topology and the data history can be said to be accurate when a GPS point is matched to the link on which the vehicle really existed at that time.

The result of the map matching using the node buffer should be viewed with particular care. Although a GPS point was matched to a node, it does not mean the GPS point was on that node accurately. Instead, it means that the GPS point was in the vicinity of that node. That is to say, it says that the GPS point was on the one of the links which is directly connected to that node. What matters is the fact that the map matching using the node buffer scarcely fails if we consider this result as a success. Therefore the distance from the matched node to the real position of the vehicle at that time can be more appropriate measure for the evaluation of the map matching using the node buffer. But, it is very hard to know the real position of the vehicle at particular time. So, in this paper, the map matching algorithm using node buffer is treated as successful if the GPS point was on the one of the links which is directly connected to the matched node even thought the weakness mentioned above.

Table 2 shows the result of the new map matching algorithm when the radius of the node buffer is set up as 10m. The type of map matching 1, 2 and 3 are respectively the map matching using the node buffer, the map matching using the NL-2NL relationship and the map matching using the shortest path. The figure before the slash(/) is the number of GPS points which was map matched successfully and the figure after the slash is the number of GPS points which was not map matched successfully.

Table 2. The Result of the Map Matching (the Radius of the Node Buffer is 10m)

Interval(min.)	Type of map matching			Number of total points
	1	2	3	
2	7/0	52/3	8/0	70
3	6/0	34/2	4/1	47
4	2/0	29/2	2/0	35
5	2/0	25/0	1/0	28

The number of GPS points which failed to properly map matched is 8 totally. And it was found that each of them was in the vicinity of an intersection. This result says that the radius of the node buffer was too short. Table 3 shows the result of the same map matching algorithm when the radius of the node buffer is set up as 20m.

Table 3. The Result of the Map Matching (the Radius of the Node Buffer is 20m)

Interval(min.)	Type of map matching			Number of total points
	1	2	3	
2	23/0	44/0	3/0	70
3	16/0	29/0	2/0	47
4	9/0	25/0	1/0	35
5	7/0	21/0	0/0	28

All the GPS points were map matched successfully. In addition, the number of GPS points which were matched using third type of map matching is decreased from 16 to 6 and the number of GPS points which were matched using first type of map matching is increased from 17 to 55 compared to the map matching with the 10m radius of buffer node. Because the third type of map matching requires more time than the first type of map matching, it can be said that the efficiency of the map matching algorithm also increased.

5. CONCLUSION

The new map matching algorithm which is appropriate for the GPS data with relatively long polling time intervals(2~5 minutes) was suggested in this paper. The process of the algorithm is as follows. First, a GPS point which is within the node buffer of particular radius is matched to that node. Secondly, compare the distance from a GPS point to the nearest link and the distance from the point to the secondly nearest link. If the ratio of these two distances is larger than the particular threshold, the GPS point is matched to the nearest link. Thirdly, a GPS point which can not map matched with both of the process above, the map matching of the GPS point(P_i) is deferred and the next GPS point(P_{i+1}) is map matched. And then, the result of the shortest path search using P_{i-1}, P_i and P_{i+1} is utilized to map match the P_i.

The field test which was accomplished in this paper shows the good performance of the new map matching algorithm. But, more field tests are needed to firmly verify the performance of the new map matching algorithm and to determine the efficient radius of the node buffer and the threshold of the ratio between the distance to the nearest link and the distance to the secondly nearest link from a GPS point.

REFERENCES

Chung, Y.S. *et al.* (2000) Classification of map-matching techniques and a development, **Journal of the Korean Society for Geospatial Information System, Vol. 8, No. 1**, 73-84.

Greenfeld, J.S. (2002) Matching GPS observations to locations on a digital map, **Proceedings of the 81th Annual Meeting of the Transportation Research Board**, Washington, DC, January 2002.

Noh, J.H. and Nam, K. S. (1995) Development of shortest path finding technique for urban transportation network, **The Journal of Korea Planners Association, Vol. 30, No. 5**.

Noh, S.H. and Kim, T.J. (1998) A comprehensive analysis of map matching algorithms for ITS, **Hongik Journal of Science and Technology, Vol. 9**, 303-313

White, C.E. *et al.* (2000) Some map matching algorithms for personal navigation assistants, **Transportation Research Part C, Vol. 8**, 91-108