

The development of database system for route recommendation based on sensor data

Yoshihiro Ito^{*}, Yoshitaka Nakamura^{**}, Yoh Shiraishi^{**} and Osamu Takahashi^{**}

^{*} Graduate School of Systems Information Science, Future University Hakodate, Japan

^{**} School of Systems Information Science, Future University Hakodate, Japan
{g2111003, y-nakamr, siraisi, osamu}@fun.ac.jp

Abstract - In recent years, we have become able collect granular sensor data through many sensors attached to a sensor network in the real world. This technology has been used for urban sensing. It is expected to achieve new personal navigation on the basis of sensor data. However, to recommend routes on the basis of sensor data, we need to handle road data and sensor data, but it is difficult for only the application side to handle these data. Therefore, a database system has been developed that can map sensor data to the road data by using spatial interpolation and provide path planning results to a navigation application. In addition, the effectiveness of the proposed system is discussed.

Keywords: sensor network, urban sensing, database system, spatial interpolation, route recommendation

1 INTRODUCTION

In recent years, miniaturization and upgrading of the sensor devices have been miniaturized and upgraded further. Because of this, we can gather environmental information in the real world easily. These data are measured at specified time intervals and stored in a database for applications. For experiment, there have been attempts that gather environmental information in urban areas. This is called urban sensing [1].

Urban sensing is expected to realize new personal navigation based on sensor data. For example, by sensing local crowded and high temperature places preliminarily, it is possible to direct people to a convenient route that avoids such places [2].

To recommend route on the basis of sensor data, we need to map sensor data to road data. At this time, sensors need to be searched for in the sensors in the vicinity of the road. However, if no sensor is in the neighborhood, it is difficult to find the route based on sensor data. In addition, if sensor data and road data are managed in different data format, it is difficult for the application side to inquire and process data.

For these reasons, in this study, we developed a database system that can map sensor data to the road data by using spatial interpolation and provide results of route recommendation to a navigation application. In addition, we discuss the effectiveness of our proposed system.

2 RELATED WORK

This section discusses research related to our study. Specifically, we discuss the research of a sensor database system that provides interpolated sensor data for application in section 2.1. In addition, we discuss research on urban sensing in section 2.2 and research on navigation and path planning in section 2.3.

2.1 Sensor database system

To use sensor data collected by a sensor network, they must be processed. There is the technology that can do this processing in the field of database. This technology considers the sensor network as a virtual database to which it sends the query. There is some research into this technology [3]. This technology needs to understand the inquiries between the application and sensor network. Specifically, a function is required that accesses sensor network and feeds back the answer to an inquiry to the application user. This system is called the sensor database system [4].

There has been research into a sensor database system to feed back interpolated sensor data. Iwai et al. [5] and Ishii et al. [6] suggested a system that generates mesh structured sensor data by using spatial interpolation and estimate sensor data in each areas. However this system does not estimate sensor data on roads necessary for route recommendation.

In addition, Arai et al. [7] suggested the database system for time series interpolation query. When sensor data which application requires does not exist, this system can estimate sensor data at arbitrary time by using cubic spline interpolation. However this system does not estimate sensor data anywhere else.

2.2 Urban sensing

We provide examples of research into urban sensing. Takagi et al. [8] and Tobe and Kurata [9] gathered environmental information in urban areas by using a sensor network. In these studies, they used a fine-grained sensor network within a 600m radius in front of Tatebayashi City Station in Gunma, Japan. This sensor network consists of leaf node and sink node. The leaf node gathers sensor data and transmits them to the sink node. The sink node transmits sensor data to database.

In addition, CitySense [10] is a sensor network that monitors meteorological phenomenon and aerial pollution in urban area. This system expands a sensor node's battery life by connecting the sensor node to a street lamp. By this approach, City Sense is used for long-range research. For example, it can monitor environmental full-time "Public health" means how healthy people are.

2.3 Path planning

2.3.1. Basics of path planning

In this section, we discuss basic elements of path planning. Solving the optimum route problem requires a graph structure like that in Figure 1.

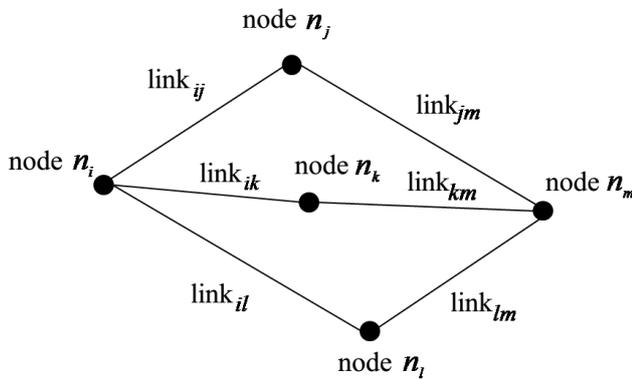


Figure 1: Graph structure

Each node is part of network that having information of node adjacent to them. In such structure, nodes are regarded as intersection, and the link as roads. This expresses the road network data. In addition, optimum routes can be searched for by giving the road distance to link as weight. Therefore, by changing one weight to another and paths can be planned on the basis of various parameters.

2.3.2. Related work on path planning

Architecture called pgRouting can search for optimal routes over the database [11]. This architecture can search for optimal routes by using a database table consisting of road data. This table needs a fixed-record format. It also needs "road id", "starting node", "end node", and "cost". However, pgRouting needs a table of road data. Therefore, when using sensor data as cost of the road data, a table must be created that records the road data for path planning, and updates it constantly during path planning. Applications have trouble executing these processes alone.

In addition, Endo and Tamura [12] suggested system using unusual WebAPI for navigation on the basis of sensor data. This system obtains sensor data by using this WebAPI, and calculates sensor data on roads, and searches route on the basis of these sensor data. This system is similar to our target system. The difference is that this system does not use special interpolation.

3 REQUIREMENTS OF PROPOSED SYSTEM

In this study, we aim to develop a system to recommend routes on the basis of sensor data. Therefore, we discuss its requirements below.

First, sensor data must be mapped to road data as cost. At this point, sensors need to be searched for in the neighborhood of the road. It makes no sense at all that there is sensor node on all roads. Therefore, an interpolate method is need to map sensor data to the road. In addition, our proposed system needs to calculate cost for route recommendation, to manage sensor data and road data at once and process inquiries from the application.

4 PROPOSED METHOD

4.1 Summary of proposed system

On the Basis of these considerations, we propose a database system to solve these problems. The database system needs the following functions.

- Map sensor data to road data
- Calculate sensor data-based cost of each road
- Consolidate road data and sensor data
- Process route recommendations
- Process Inquiries

These functions are divided into two sections namely the Data Management Section which management various data and Database Management Section which recommends routes of the basis of sensor data (Figure 2).

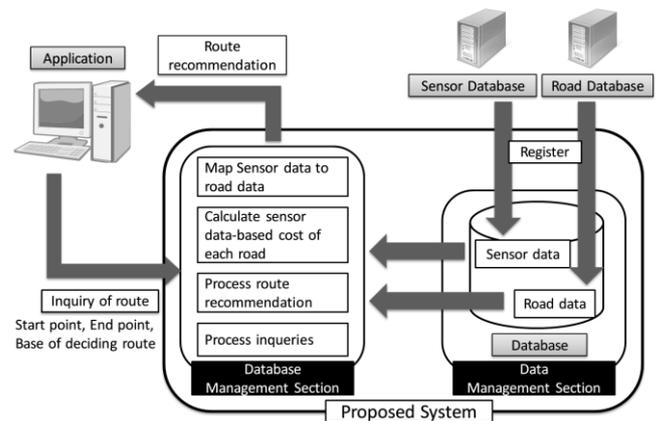


Figure 2: Summary of proposed system

The Data Management Section receives and manages sensor data and consolidates both road and sensor data. Road data are managed in road segment that separate each road from intersections. The Database Management Section receives sensor and road segment data and map sensor data to road segments. In addition, the Database Management Section handles and calculates the cost of road segment from mapped sensor data. Also, the Database Management Section processes inquiries and recommends route.

4.2 Management of road and sensor data

In this section, we discuss how to manage sensor and road data. Road data are managed as road segments that separate each road with intersections. Road segments are managed in the database as information of links described in section 2.3.1. Road segments are managed in the Road Segment Table which has the following data format (Table 1).

Table 1: Format of Road Segment Table

Attribute Name	Detail
Road ID	Identification numbers of road segment
Starting Node	Starting node numbers of road segment
End Node	End node numbers of road segment
Distance	Length of road segment (metric data)
Geometry	Location information of road segment (Line data)

The sensor data are managed in the Sensor Node Information Table and Sensor Data Table which has following data format (Table 2 and 3).

Table 2: Format of Sensor Node Information Table

Attribute Name	Detail
Sensor Node ID	Identification numbers of sensor node
Sensor Type	Types of sensor in sensor node
Geometry	Location information of sensor node (Point data)

Table 3: Format of Sensor Data Table

Attribute Name	Detail
Time	Identification numbers of sensor node
Data	The data which sensor node collected

Each of the Sensor Data Table named ‘‘Sensor Type_Sensor Node ID’’. In this way, we can access to Sensor Data Table by obtaining the Sensor Node ID and the Sensor Type from the Sensor Node Information Table.

The Data Management Section manages sensor data and road data as above, and the Data Management Section processes route recommendation by referring to all data in the Data Management Section.

4.3 Mapping sensor data to road data

Various factors that determine the environment of urban areas include buildings, parks and trees all of which affect the road environment. For example, when using temperature data, part of a road in the shadow of building is cooler than a part that is not. Therefore, to recommend the right route, sensor data need to be mapped on the road in a fine-grained manner. We explain how to do this.

First, we divide a road segment on the basis of its length. This is called a ‘‘divided road segment’’ (Figure 3). The number of divisions x is decided by the length of longest road segment in the road network (max_length) and defined width in meters is called ‘‘divide width’’ (d). Specifically, the number that is divided (i) is calculated as follows:

$$d \cdot (i-1) < x \leq d \cdot i \quad (d \cdot i \leq max_length, i \geq 1) \quad (1)$$

If the value of i is 1, the road segment stays undivided. In addition, the relationship between the road segment and divided road segment is as follows:

$$R_A = \{div_1, div_2, div_3, \dots, div_i\} \quad (2)$$

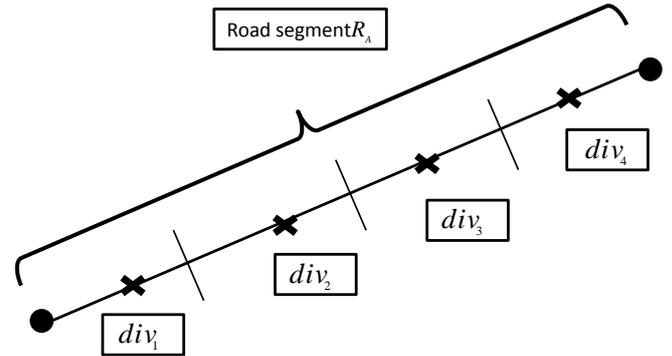


Figure 3: Divided road segment

By mapping sensor data to divided road segment generated by this method, we can know the environmental information on the road segment. However, depending on the situation of the sensor installation state, no sensor is considered to be near the road segment. In this case, possibly no sensor data are on the road segment, and this road segment can not be used for route recommendation. In our proposed method, we solve this problem by using special interpolation. Specifically, we estimate sensor data on a road segment by using Inverse Distance Weighting (IDW) [13]. IDW is the method most often used by GIS analysts. It estimates unknown measurements as weighted average over the known measurements at nearby points, giving the greatest weight to the nearest points. By using IDW, we can know the environmental information on divided road segment from a sensor near them.

When using IDW, nearby points must be found. The proposed method solves this problem by apply the Variable Radius Method [14]. The Variable Radius Method is the method used to draw data points by describing a circle with a radius of decided distance around an unknown point and locate data points in this area. When the required number of data point does not exist in the area of the circle, the radius is expanded to search for more data points again (Figure 4).

With the above method, we explain the method for sensor data estimation on a divided road segment ($div_1, div_2, \dots, div_i$). First, calculate the midpoint of divided road segments and set these midpoints as interpolate points. Next, define the search range in meters, and search for sensor node by using the Various Radius Method. Specifically, this method draws sensor node point by describing the circle with a radius of the search range introduced before around midpoint of a divided road segment. When no sensor node is in the area of circle, double the search range and search for sensor nodes again. It becomes possible to map sensor data on all divided

road segments by processing above method to the midpoint of divided a road segment.

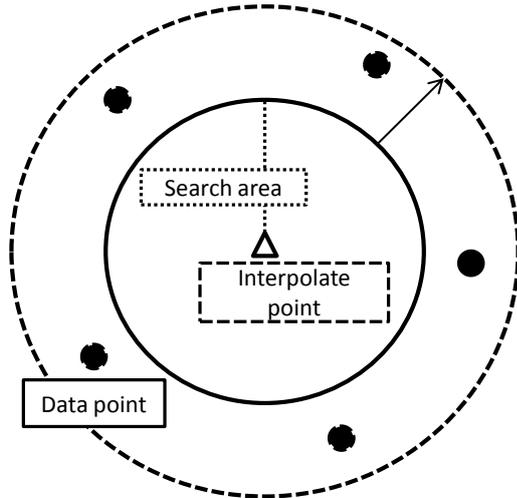


Figure 4: Variable Radius Method

4.4 Calculation of cost from sensor data

The method proposed in section 4.3, map sensor data on divided road segments. However, the cost needs to be calculated on each road segment for route recommendation. Therefore, we adapt the Dijkstra method which is the path planning algorithm because this algorithm excels in efficient and application, and define the cost-calculating formula for sensor data-based route recommendation. This formula calculates the cost of each road segment by using sensor value and the value which is decided on the basis of users feel comfortable (For example, using temperature, its value defined as 25 degrees Celsius). We explain this calculation formula below.

First, road segment R_A maps sensor data as shown in Figure 5.

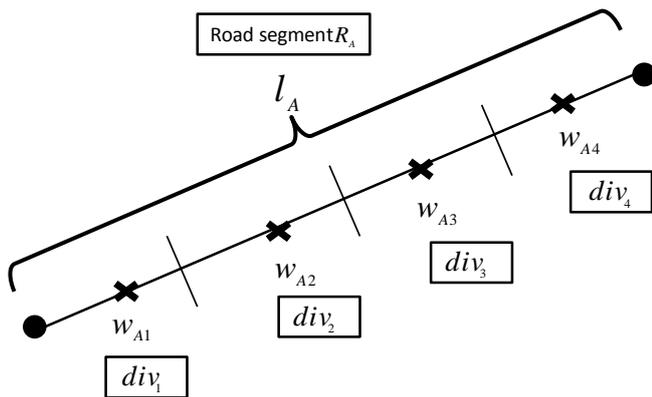


Figure 5: Parameter of road segment

w_{ik} is the estimated sensor data on divided road segment, and l_A is the length of the road segment R_A . In addition the division number of road segment is n , and the value which is decided on the basis of users feels comfortable as t_d . We find the sensor data based cost of each road segment by using these parameters. When sensor data and route length are considered, cost C_A is calculated by formula (3), and when

only sensor data are considered, calculate cost C_A by formula (4).

$$C_A = \left\{ \sum_{k=1}^n |w_{Ak} - t_d| \right\} \cdot l_i \quad (3)$$

$$C_A = \{ \max(w_{Ak}) - t_d \}^2 \quad (1 \leq k \leq n) \quad (4)$$

Formula (3)'s cost of road segment increases as the difference between t_d and w_{ik} increases. Therefore, by using this formula, we can select a convenient route for the user. In addition, when the length of a road segment is multiplied by the difference between w_{ik} and t_d , cost C_A can consider the route length. On the other hand, Formula (4) attaches importance to sensor data. It does not consider the route length but calculates cost of road segment from sensor data only. Next, we describe formula (3) the "sensor and distance base method" and formula (4) the "sensor base method".

4.5 Flow of processing

Figure 6 shows the processing flow with the above method.

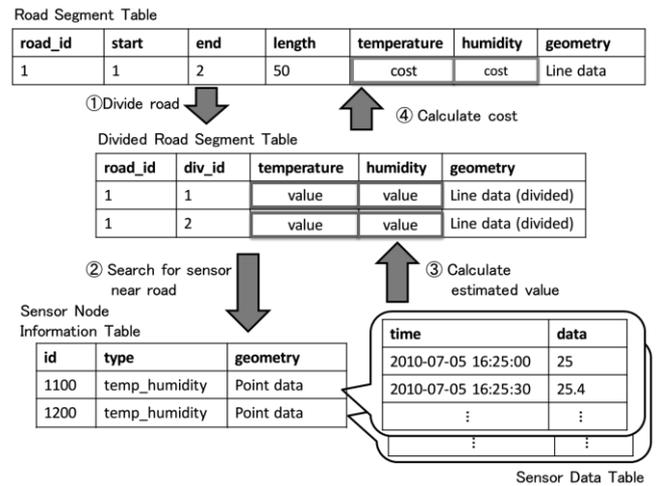


Figure 6: Flow of processing

First, the proposed method refers to the length of a road segment from the Road Segment Table and divides the road segment on the basis of the length. These results are stored in the Divided Road Segment Table (1). Next, to map interpolated sensor data onto a divided road segment, the method compares location information of the divided road segment and location information of sensor node, and pick up Sensor Node ID from Sensor Node Information Table by using the Various Radius Method (2). Therefore, by using Sensor node ID, the method searches for and obtains sensor data from Sensor Data Table, calculates interpolated sensor data on a divided road segment by using IDW and inserts these data into the Divided Road Segment Table (3). Finally, it calculates the cost for route recommendation by using the sensor and distance base method or sensor base method (4). After processing these methods, the proposed method recommends a route by referring to the cost in the Road

Segment Table and applying Dijkstra method. This data structure can control different types of sensor.

5 EXPERIMENT AND DISCUSS

5.1 Experiment environment

To verify the effectiveness of the proposed system, we conducted an experiment by using one computer (CPU: 3.2GHz, Memory: 4GB, HDD: 500GB OS: Windows 7 Professional). We developed a database system that had the function we proposed on this computer. Also, this database system is implemented by Java Servlet and uses a database-management system of PostgreSQL8.4, and it uses PostGIS1.5 an advanced function of PostgreSQL to manage geometry data.

5.2 Experiment for evaluation

In our experiment, we used sensor data gathered by the sensor network in Tatebayashi City, Japan. This sensor network gathers temperature and humidity data and a 600m radius in front of the station, as shown in Figure 7.



Figure 7: Distribution of sensor nodes

In our experiment, we used data gathered in summer, 2010. From these, we choose sensor data on clear days: August 23, 24 and 28, 2010. In addition, we used sensor data in August 21, 2010 because its temperature data on each point were high and low.

In our experiment, attempted to map and estimate sensor data. Specifically, we measured processing times elapsing from the beginning of the calculation to estimate the first to the last sensor data. In addition, we measured estimation accuracy in accordance with different times of day and evaluated the recommended routes. We discuss the experiment results below.

5.2.1. Route recommendation

We describe the result of route recommendation. In this experiment, we generated routes by using sensor and distance base method and sensor base method, and compared the generated and shortest routes. We used temperature data from 16:00 August 21, 2010. Figure 8 shows the results. Also, the value of temperature at which users feel comfortable (t_d) is defined as 25 degrees Celsius.

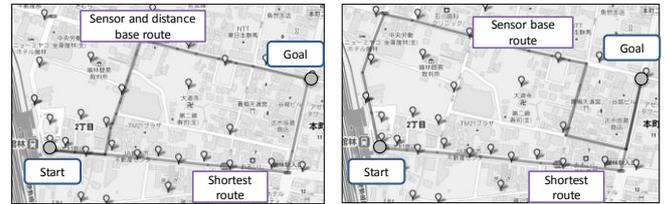


Figure 8: Generated route (Left: Sensor and Distance Base Method Right: Sensor Base Method)

To analyze these generated routes, we show a temperature gradient of road segment estimated by the proposed system in Figure 9.



Figure 9: Temperature gradients of road segments

From this temperature gradient of a road segment, we define t_d as 25 degrees Celsius, and the proposed system was expected to recommend the coolest route. Figures 10, 11 and 12 show estimated temperature gradients of each route.

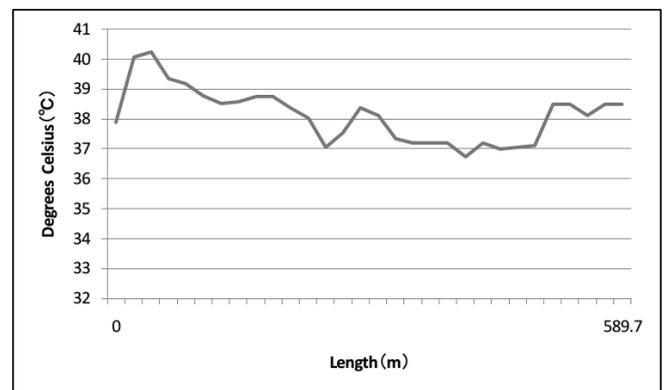


Figure 10: Estimated temperature gradient of shortest route

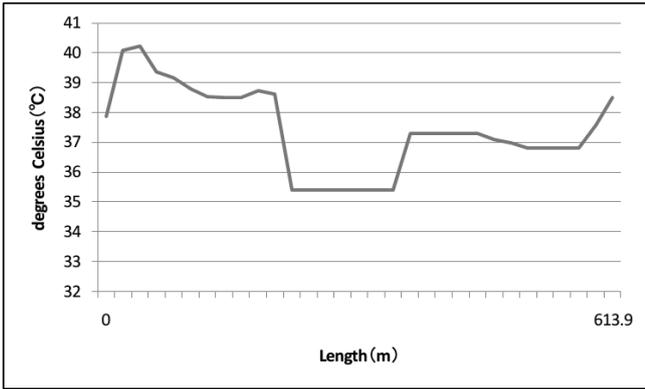


Figure 11: Estimated temperature gradient of sensor and distance base method route

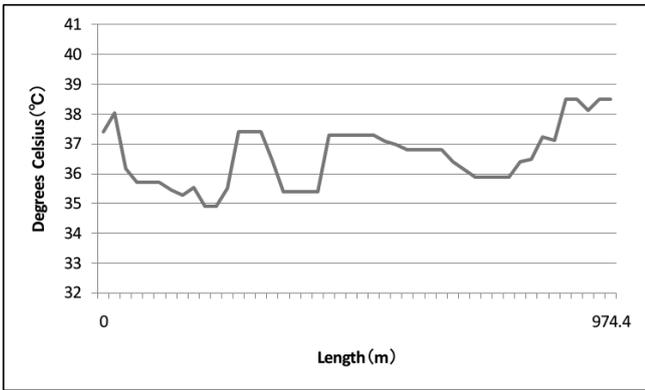


Figure 12: Estimated temperature gradient of sensor base method route

5.2.2. Processing time for estimating sensor data

We describe the processing time to calculate estimated value by using IDW. To use IDW, we define the value of the search range for the Variable Radius Method. We measure the processing time of value estimation by changing the search range. Specifically, we define the search ranges as 50, 100, 150, 200, 250, 300, 400, 450, 500, 1000 and 2000 meters, and measure processing time elapsing from the beginning of the calculation to estimate the first to the last sensor data of each range. These results are shown in Figure 13. Also, divide width of road segment (d) is defined as 25 meters.

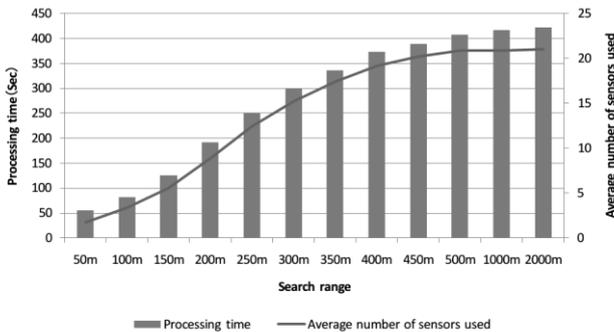


Figure 13: Processing time

According to these results, processing time increases as the search range expands, and the average number of sensor nodes used to calculate the estimated value at an interpolate point also increases. Processing time increases directly with the average number of sensor nodes used to calculate the estimated value, but when the search range is 500 meters, all sensor nodes are used to calculate all points, Neither processing time or the number sensor nodes significantly change after this.

5.2.3. Estimation accuracy

We describe results of an estimation accuracy experiment for the proposed system. This system has difficulty to analyzing estimation accuracy because it estimates sensor data on unknown points. For this reason, we set the place of a sensor node that has actual measured value in Tatebayashi City's sensor network as the interpolate point and estimated sensor value on that point by using proposed method. We tested this method on all points at which a sensor node existed, and calculated the difference between estimated and actual values. This difference was calculated by relative error and derived from the following calculation in percentage terms (E: estimated value, A: actual value):

$$\frac{|E - A|}{A} \cdot 100 \quad (5)$$

First, we measured estimation accuracy by changing the search range. Search ranges were the same as those given in 5.2.2', and we calculated relative error for all sensor node points by using formula (5) and calculated their mean. In this experiment, we used temperature data from 14:00 August 23, 24 and 28, 2010. Figure 14 shows the results.

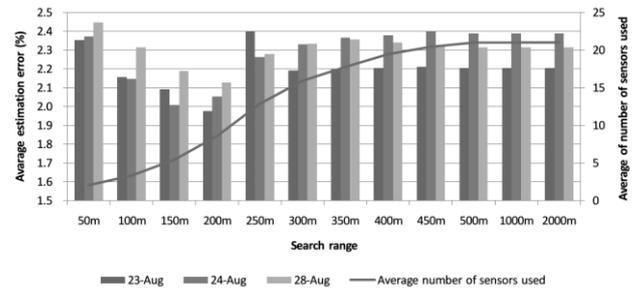


Figure 14: Change in estimated error by changing search range

As shown in Figure 14, as the search range expanded from 50 to 200 meters, estimation error decreased gradually. However, on the other hand, estimation error is increased from 250 to 500 meter. That difference may have been caused by using sensor data far from the interpolating points for estimation calculation. Next, we measured the estimation accuracy for temperature in accordance with different times of day. The times of day were 8:00, 14:00 and 20:00, and we calculated the average estimation error for each. Figure 15 shows the results.

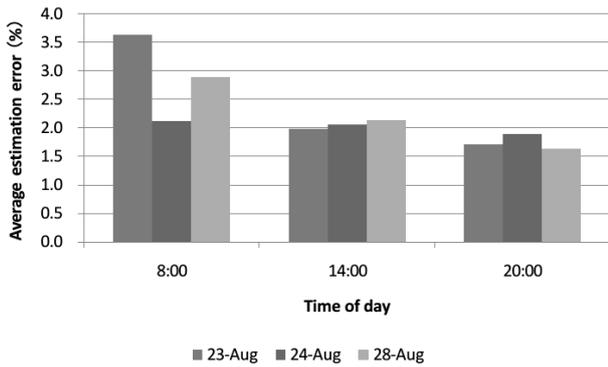


Figure 15: Estimation error by time of day

A 3% estimation error represented a degrees Celsius error. From this reason, Figure 15 shows the estimation error is small overall. However, at 8:00, estimation error was considerably larger on 14:00 and 20:00. To analyze this, we checked the estimated error of each estimated point at 8:00. Figures 16, 17 and 18 show the results.

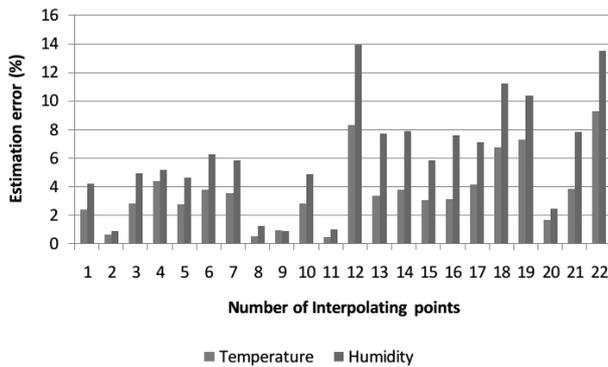


Figure 16: Estimation error at 8:00 Aug-23

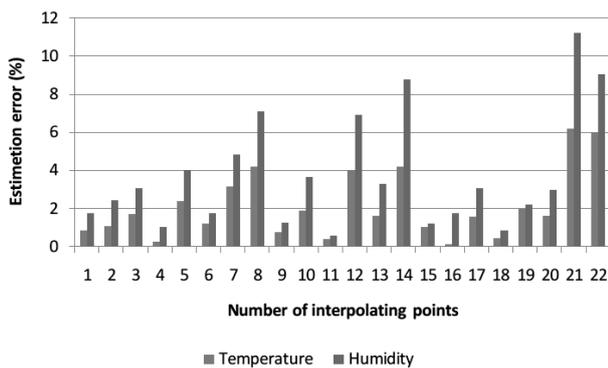


Figure 17: Estimation error at 8:00 Aug-24

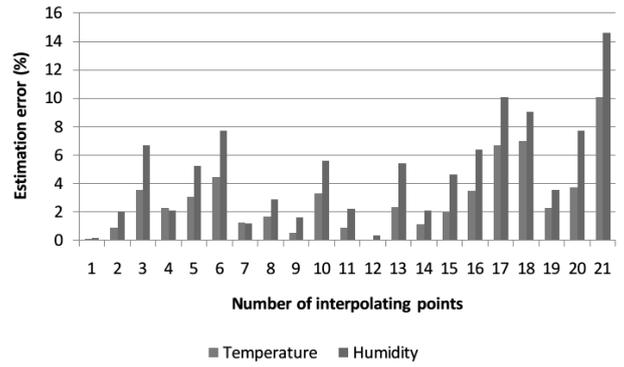


Figure 18: Estimation error at 8:00 Aug-28

In this experiment, we calculated estimation errors for temperature and humidity. These results show the estimation errors for some interpolating points are large. In addition, this figure shows the degree of estimation error for temperature tends to be the same as the degree of estimation error for humidity. For example, we found that at places at which large estimation error was measured for temperature, a large estimation error was similarly measured for humidity. From these results, the estimation accuracy apparently depends on a sensor node's installation location and situation. Therefore, we analyzed the interpolating points at which estimation error for temperature was over 5% at 8:00 August 23. Figure 19 shows the results. Sensor nodes with a cross have no data or error data, so they were not used to calculate the estimation.

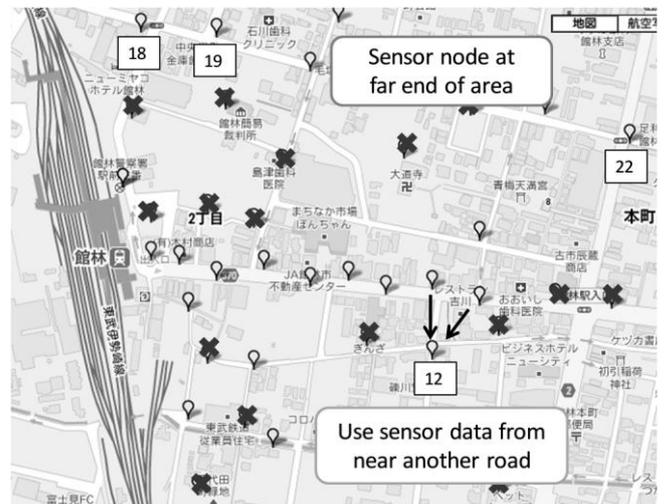


Figure 19: Interpolated point that measured large estimation error at 8:00 Aug-23

at 8:00 August 23, the installation location of sensor nodes the IDs of which ID is 18, 19, 22 and 12 had an estimation error over 5%. First, the standout feature of the interpolating point that had large errors was that they existed in the far end of the sensor network area. This applies to location of sensor node 18, 19 and 22. This reason is considered to be the lack of sensor nodes in far end of the area near these three. On the other hand, sensor node 12 exists near the center of the sensor network, but its estimation error was large. The reason is considered to be

that no sensor node for which data are available is near sensor node 1900, so it uses sensor node near another road to calculate estimations.

In another day's results, large error classifiable into these two patterns was measured at interpolating points. Therefore, the solution is needed to these problems.

5.2.4. Discussion of experiment results

In this section, we discuss results of the above experiment. In the evaluation of route recommendation, we confirmed that the proposed system recommend routes the way we assumed by using our proposed method and estimated sensor data.

However, from the evaluation of estimation accuracy, we found that some interpolating points have large estimation error. This error is possibly due to location and situation of interpolating points. This error affects the accuracy of recommended routes. To solve this problem, errors must be reduced at such points. One possible solution to this problem is participatory sensing [15]. By using participatory sensing, we can gather sensor data not only from a fixed sensor network but also from mobile devices. Therefore, estimation accuracy may be improved by gathering sensor data near point that have large estimation error by using participatory sensing.

In addition, from the evaluation of processing time, we determined that it takes time to estimate sensor data correctly. To solve this problem, processing must be speed up by improving database structure. Furthermore, the proposed system processes estimations and records this processing at constant intervals, and if an inquiry comes from the application, proposed system executes the route recommendation process only.

6 CONCLUSION

In this study, we developed a database system that can estimate sensor data on roads by using spatial interpolation and recommend routes on the basis of this sensor data. In addition, we evaluated processing time of the proposed system, estimation accuracy of sensor data and recommended routes generated by the proposed system.

In the future, to improve estimation accuracy of sensor data, we will consider a database system dealing with participatory sensing data. To achieve this system, we will consider a method to deal with mobile sensing data. Moreover, we will consider the way to use another sensor data.

REFERENCES

- [1] Takahiro Ono, Hiroki Ishizuka, Kanoko Ito, Yasuyuki Ishida, Shohei Miyazaki, Oru Mihirogi, Yoshito Tobe, "UScan : Towards Fine-Grained Urban Sensing, Int'l Workshop on Real Field Identification, 2007.
- [2] Keisuke Kanai, Takahiro Ono, Ishizuka Hiroki, Kanoko Ito, Yoshikazu Sawa, Oru Mihirogi, Yoshito Tobe, proposal of an Urban Information Sensing System, IPSJ SIG Technical Report UBI, No.18, pp.141-148, 2008. (*in Japanese*)
- [3] Shigeru Ando, Yosuke Tamura, Yoshito Tobe, Masaki Minami, "Sensor Network Technology", Tokyo Denki University Press, 2005. (*in Japanese*)
- [4] Yoh Shiraishi, "Database Technologies for Sensor Networks", Information Processing Society of Japan, Vol.47, No.4, pp.387-393, 2006. (*in Japanese*)
- [5] Msayuki Iwai, ThePVILOJANAPONG Niwat, Hiroki Ishizuka, Nakamura Yoichi, Keisuke Kanai, Yoh Shiraishi, Yoshito Tobe, "TomuDB: A Database Management system for Distributed Urban Sensors", IEICE Technical Report USN, Vol.108, No.138, pp.13-18, 2008. (*in Japanese*)
- [6] Nayuta Ishii, Yoh Shiraishi, Hiroki Ishizuka, Yoshiro Tobe, "An Integration Method of Heterogeneous and Multi-resolutional Sensor Data Utilizing Spatial Interpolation", Vol.49, No.SIG7 (TOD37) , pp.44-58, 2008. (*in Japanese*)
- [7] Kenji Arai, Yoh Shiraishi, Osamu Takahashi, "Design and development of a sensor database system for time series interpolation query", The 72nd National Convention of IPSJ (3), pp293-294, 2010. (*in Japanese*)
- [8] Atsuhiko Takagi, keiji Sugo, Yasuyuki Ishida, Tatsuya Morita, Takeshi Iwamoto, Hideyuki Kurata, Yoshito Tobe, "A Practical Implementation of Micro-Climate Networked Sensing : Case at Tatebayashi City, Gumma", IEICE Technical Report USN, Vol.109, No.248, pp.13-18, 2009. (*in Japanese*)
- [9] Yoshito Tobe, Hideyuki Kurata, "A Practical Implementation of Micro-Climate Networked Sensing - Case at Tatebayashi City", Information Processing Society of Japan, vol.51, No.6, pp692-699, 2010. (*in Japanese*)
- [10] R. Murty, G. Mainland, I. Rose, A. R. Chowdhury, A. Gosain, J. Bers, and M. Welsh : CitySense: A Vision for an Urban-Scale Wireless Networking Testbed. In IEEE International Conference on Technologies for Homeland Security, 2008.
- [11] pgRouting Project, "pgRouting",HTML Available at "<http://www.pgrouting.postlbs.org>".
- [12] Tatsuro Endo, Yosuke Tamura, "Research of Sensor Networks by Using Web API- : Development of Navigation System by Using Sensor Data", The 70th National Convention of IPSJ (3), pp23-24, 2008. (*in Japanese*)
- [13] P.A.Longley, M.F.Goodchild, D.J.Maguire, D.W.Rhind, "Geographic Information Science", John Wiley & Sons, Ltd, 2001.
- [14] Zhang Changping, "Analysis of Spatial Data by Using Geographic Information System", KOKON, 2001. (*in Japanese*)
- [15] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, M. B. Srivastava. "Participatory sensing", World Sensor Web Workshop, ACM Sensys 2006.