

Recommended Paper

A Sensor Data Collection Method for Multi-granularity Contour Lines Map Creations on an Overlay Network

JUN SHINOMIYA^{1,†1} YUICHI TERANISHI^{2,3,a)} KANAME HARUMOTO⁴ SHOJIRO NISHIO¹

Received: October 1, 2012, Accepted: July 3, 2013

Abstract: In this paper, we propose a novel sensor data collection method using an overlay network that forwards collection request messages to the sensor nodes which have sensor data needed to create interpolated contour lines maps. When there is an enormous number of sensor nodes, it is redundant to collect all sensor data from a target area since geographically close sensor data can be interpolated. Moreover, since the interpolation process requires heavy CPU load, the process may not be finished within a reasonable time period if too much sensor data arrived. On the other hand, there is a trade-off between the number of collected sensor data and the accuracy of the created contour lines map. In our proposed method, the number of collected sensor data can be limited to a constant number. At the same time, granularity of characteristic points in the contour lines maps can be specified by users in order to satisfy the various requirements. Our proposed method extends the hierarchical Delaunay overlay network (HDOV) and forwards the collection request messages according to the *feature amounts* that correspond to the characteristic points for each layer of the HDOV. Simulation results showed that our proposed method could create contour lines maps while satisfying the requested granularity of the characteristic points within the constant number of sensor data.

Keywords: sensor network, overlay network, sensor data aggregation, delaunay overlay network, contour lines map

1. Introduction

The importance of ubiquitous, context-aware services has grown significantly in the last decade. Context-aware services require contextual information derived from real-world observations like weather conditions, traffic congestions, the movement of shoppers in the market places, and so on. Such observations are also required for geographical information services in order to treat environmental problems such as global warming, air pollution, an increase in carbon dioxide, and so on. Real-time spatial data collection infrastructures are required for such services. The typical spatial data we treat here is sensor data which contains observation values generated by sensor nodes in sensor networks that observe real-time situations in the real world. There are many projects to construct a global scale sensor network that connects sensor nodes to the wide-area networks, including the Internet [1], [3], [4], [6], [11].

To address the scalability problem, some of these projects use a peer-to-peer architecture to manage widely distributed sensor nodes [1], [3], [6]. This architecture also reduces obstacles to joining to the network since it does not require the approval of a centralized organization. We also use the peer-to-peer archi-

ture to manage sensor nodes. For the simplicity, we assume one spatial data source corresponds to one node and constructs an overlay network by the peer-to-peer architecture. That is, one sensor corresponds to one node on an overlay network. The node can also be a database or archive of spatial data.

For the purpose of visualizing or analyzing the gathered sensor data, contour lines maps are useful for grasping characteristic points such as hot spots or the distribution tendency of a wide area. Contour lines maps divide a spatial area into partial areas that can be regarded as the same area using contour lines. To create contour lines maps from dispersed observation points, spatial interpolation is often used. The simplest method to create the maps is to collect all sensor data from the network. However, when all nodes respond to the data collection request, much of collected sensor data can be redundant since data from geographically close observation points can be interpolated. This causes an unnecessarily long response time caused by network traffic increase and a long data processing time on CPU if there are huge amounts of observation points.

Some existing methods for sensor data collection eliminate the redundancy and keep the accuracy [7], [13]. However, they cannot limit the number of collected sensor data and cannot guarantee that the interpolation is finished within a reasonable time period. They can cause an extraordinary delay since they try to reconstruct all characteristic points accurately regardless of the number of collected sensor data.

¹ Graduate School of Information Science and Technology, Osaka University, Suita, Osaka 565–0871, Japan

² National Institute of Information and Communications Technology, Koganei, Tokyo 184–8795, Japan

³ Cybermedia Center, Osaka University, Ibaraki, Osaka 567–0047, Japan

⁴ Graduate School of Engineering, Osaka University, Suita, Osaka 565–0871, Japan

^{†1} Presently with SONY Corporation

^{a)} teranisi@nict.go.jp

The initial version of this paper was presented at the SIGDPS held on Feb. 2012, which was sponsored by SIGDPS. This paper was recommended to be submitted to Journal of Information Processing (JIP) by the chairman of SIGDPS.

From the standpoints of real-time processing and usability, it is desirable to be able to create the contour lines map by limiting the number of sensor data as a system constraint. However, there is a trade-off relation between the collected number of sensor data and the accuracy of the created contour lines map. That is, the accuracy of the reconstructed contour lines map is limited when there is a constraint in number of sensor data.

It is difficult to keep both the tendency of wide areas and changes in small areas when the number of sensor data is not enough to reconstruct accurate contour lines maps. At the same time, the requirements for the contour lines map are different depending on the applications or situations. For example, users who require observing uncommon situations do not want to miss the unusual changes even in a small area. In this case, characteristic points with a small granularity must be detected. If a user need to grasp the weather of a wide area, the tendency of the entire area should be accurate, and changes in small areas are not important. In this case, the required granularity of the characteristic points is rather large. Thus, it is desirable to be able to specify the granularity of characteristic points if the number of collected sensor data is limited.

Therefore, in this research, we propose a sensor data collection method called “HDA-GC” (Hierarchical Delaunay-based Aggregation w/ Granularity of Characteristic), which can create contour lines map satisfying the limitation on the number of sensor data and the requested granularity of the characteristic points. Our proposal method uses a hierarchical Delaunay overlay network (HDOV) [8] as an overlay network, which can collect geographically uniform sensor data.

In our method, sensor data are periodically aggregated using the layered structure of the HDOV, and the number of collected sensor data by user request is limited to constant number. The number is assigned to each layer according to the requested granularity of the characteristic points. This mechanism creates the contour lines map that satisfies the user’s requirement using limited number of sensor data.

2. Related Works

2.1 HDOV

The article [8] proposes a hierarchical Delaunay overlay network (HDOV) construction method for wide area sensor data collection. Without investigating the surrounding nodes distribution, HDOV constructs a multi-layered Delaunay overlay network [10], whose node distribution is geographically uniform with a specified node density.

The Delaunay overlay network has features suitable for collecting wide-area sensor data. For instance, the average degree of each node is six, and linked nodes are more adjacent to each other than other unlinked nodes. These characteristics are independent from the number of nodes in the network. Therefore, neighbor node discoveries become scalable. In the Delaunay overlay network, the link structure is based on Delaunay triangulation, which corresponds to the dual graph of the Voronoi diagram (Fig. 1). The node’s link topology is decided according to the geographical coordinate of the nodes. Each node joined to the Delaunay overlay network has the geographical coordinates of the linked neigh-

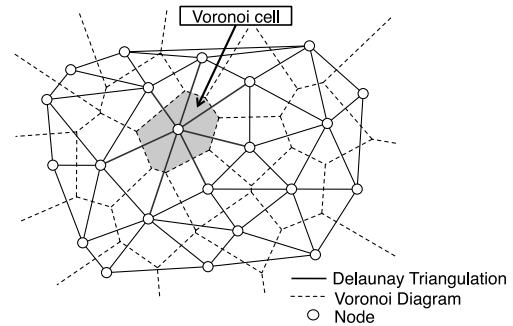


Fig. 1 Delaunay Triangulation and Voronoi Diagram.

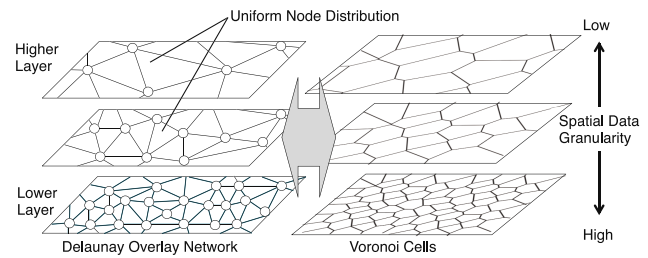


Fig. 2 A example structure of HDOV.

bor nodes. Node discoveries over far distances are performed by multi-hop forwarding to the closest nodes [2].

The HDOV selects upper layer nodes probabilistically. Each node decides whether it joins the upper layer or not probabilistically in order to construct an overlay network with a specified node density. The probability is decided by the size of the Voronoi cell of the Delaunay overlay network for each node and the requested node density for the layer. Since the node distribution of each layer is uniform in HDOV, the size of the Voronoi cell, which corresponds to the granularity of the sensor data responsible for the sensor node, becomes almost same in each layer. The higher the layer, the lower the granularity (Fig. 2).

The HDOV aims to collect sensor data with multiple levels of spatial granularity from a peer-to-peer sensor network. Therefore, using a certain layer of an overlay network cannot reconstruct an accurate distribution that requires a higher granularity than the corresponding spatial granularity. For example, a local downpour requires a very high spatial granularity. Using an upper layer overlay network cannot pick up the characteristic of the local downpour distribution. Hence, the spatial interpolation for the layer cannot create an accurate contour lines map. To create an accurate map, a lower layer overlay network with a higher spatial granularity should be used. However, in this case, redundant node selection occurs in other areas.

2.2 HDA-SN

The article [13] proposes a sensor data collection method called “HDA-SN” (Hierarchical Delaunay-based Aggregation w/ Segment Number) that can collect sensor data from distributed sensor nodes to create a contour lines map that reconstructs characteristic points avoiding redundancy. HDA-SN extends the HDOV for sensor data collection. In HDA-SN, sensor data are periodically aggregated according to the layered structure of the HDOV. When a query is issued, the query message is forwarded to the nodes that exist inside the characteristic points.

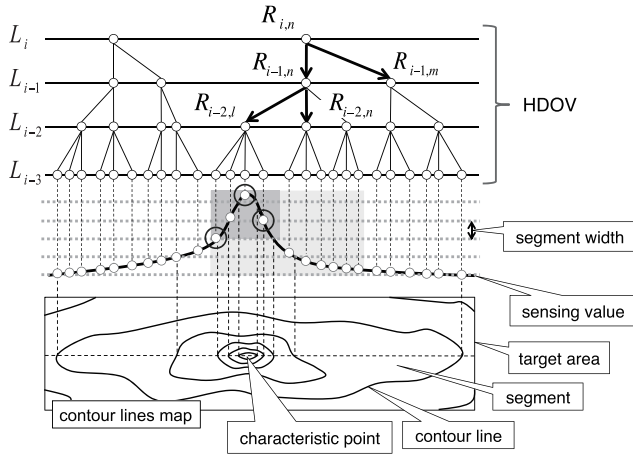


Fig. 3 Message forwarding mechanism used in HDA-RC.

In HDA-SN, each node in the HDOV periodically aggregates the maximum and minimum sensor data from lower layer nodes inside its Voronoi-cell.

A user can specify an area to create the contour lines map as a request. Hereafter, the area is represented as *target area*. HDA-SN forwards a query message that includes target area from top layer nodes to lower layer nodes of the HDOV. Each node forwards the query message only when the *number of segments* of the lower layer node is larger than a parameter S . It can be calculated by the minimum and maximum sensor data inside of the Voronoi cell of the node and the requested *segment width*. The term *segment* means the area that is surrounded by the contour line (see the lower part of Fig. 3). By means of this simple mechanism, the method can collect sensor data without missing characteristic points.

However, when a segment width is relatively smaller than the parameter S , HDA-SN forwards messages to the nodes that are not inside the characteristic point and redundant sensor data is collected. Therefore, to keep the performance, appropriate value for the parameter S should be selected according to the segment width. Furthermore, HDA-SN collects all sensor data for all characteristic points. If there are many characteristic points in the target area, sensor data will be collected boundlessly, which can cause network traffic increase and a long processing time.

2.3 HDA-RC

The article [12] proposes a sensor data collection method called “HDA-RC” (Hierarchical Delaunay-based Aggregation w/ Resolution of Characteristic) that can collect sensor data satisfying the user’s request and the limit number of collecting sensor data. HDA-RC also extends HDOV to aggregate the sensor data periodically. The limit number of collecting sensor data is assigned to each areas according to the user’s request.

In HDA-RC, the number of collecting sensor data specified by a system provider or a developer is assigned to the nodes on the top layer of the HDOV, whose Voronoi cell intersects with the requested target area. Between each hierarchy of HDOV, links are constructed. This link construction method in HDA-RC is described in Section 3.2. The link construction process of HDA-RC is same as our proposal method. Each node that has received

the message re-assigns the number of sensor data to the children nodes of the lower layer autonomously.

Thus, the limit on the number of collected sensor data is guaranteed.

The assignment of the number of collecting sensor data is based on two factors. The first is the number of geographically uniform sensor data, and the second is the number of collecting sensor data decided according to the user’s request. Uniformly distributed sensor data are necessary for the interpolation in order to reconstruct the tendency of the wide area, and they are collected by using the layered structure of the HDOV. A node n on level i of HDOV assigns the number of collecting sensor data ($R_{i,n}$) to the children nodes ($n.children_i$) depending on how the user wants to grasp the sensor data distribution, on the basis of the aggregated sensor data in advance (Fig. 3). The assignment of collecting sensor data is based on the following formula.

$$R_{i-1,m} = R_{i,n} \times \frac{(m.area \times m.var^{C/2} \times m.segments^C)}{\sum_{k \in n.children_i} (k.area \times k.var^{C/2} \times k.segments^C)}$$

m is a target node for assignment, and $m.area$ is the area size where the target area and the Voronoi cell of m intersect. $m.var$ is the variance of the sensor data inclination, which is calculated by the distance and observed value. $m.segments$ is the number of segments inside the $m.area$. It can be calculated by the minimum and maximum sensor data inside of the Voronoi cell of node m and the requested segment width. The sensor data contains maximum, minimum, var is aggregated according to the structure of HDOV by the periodical message exchange. In HDA-RC, a user specifies the control parameter C that corresponds to the user’s request to grasp the sensor data distribution. If C is positive, the number of sensor data is assigned more to the area where there are many segments and the variance is large. On the contrary, if C is negative, a lot of numbers are assigned to the area where there are little segments and variance is small. If $C = 0$, the number of sensor data is assigned uniformly since assignment is based on the area size of the Voronoi cell.

However, HDA-RC cannot deal with various requirements for grasping the characteristic points flexibly since it collects sensor data regardless of the granularity of the characteristic points. Therefore, HDA-RC may collect sensor data corresponding to characteristic points with granularity unnecessary for the application. In a limited number of sensor data, it is desirable to collect sensor data depending on the granularity of the characteristic points since it can be difficult to reconstruct both the tendency of the sensor data distribution and the small characteristic points accurately at the same time. It is difficult to clarify the relation between the parameter C and reconstructed contour lines map.

3. Proposal Method: HDA-GC

In this research, we propose a sensor data collection method called “HDA-GC” (Hierarchical Delaunay-based Aggregation w/ Granularity of Characteristic).

HDA-GC is an extension of HDOV. The same as HDA-RC, it assigns a specified number of sensor data according to the hierarchy of the HDOV. The key idea of HDA-GC is that it collects sensor data by avoiding sensors that correspond to characteris-

tic points smaller than the required granularity. To achieve this, each node maintains characteristic points inside the hierarchical Voronoi cells of the HDOV. Then, the number of sensor data is assigned only when the characteristic points with required granularity exist inside the Voronoi cell.

3.1 Assumed Environment

We assume that each sensor node is joined to an overlay network with the HDOV structure. The construction method of hierarchies follows the article [8]. We also assume that sensors always act correctly and the observed value in each sensor is accurate. That is, error data correction for broken sensors is out of the scope of this research. An assumed application is a geographical information system that users use to issue a request, and a corresponding contour lines map is generated by sensor data collected from the overlay network. User requests include the following:

- *target area*
- *segment width*
- *granularity*

The response from the overlay network is a set of pairs of observed value and its location to construct a contour lines map.

The *granularity* corresponds to the size of the characteristic point. In HDA-GC, it is the layer number of the HDOV. Therefore, the HDOV must have layers that correspond to the multiple granularity possibly requested from users.

After the collection of sensor data according to the user request, a spatial interpolation is applied. Any kind of spatial interpolation method such as Inverse Distance Weighted (IDW), Spline, and Kriging [9] can be applied to the collected sensor data.

In the proposed method, the service provider or developer can set the limit on the number of sensor data as a system constraint. The main difference between HDA-RC and HDA-GC is the aggregation and assignment methods. The remainder of this section describes how the aggregation and assignments are executed on HDA-GC.

3.2 Construction of the Links between the Hierarchies

HDA-GC (and HDA-RC) extends the HDOV to build links between the hierarchies at the time the Delaunay overlay structure is constructed. The links are used for the aggregation between hierarchies. In the remainder of the text, we represent layer level i of the HDOV as L_i . L_0 , which is the bottom level of the HDOV, consists of all sensor nodes. The number of levels in the HDOV is represented as H . The top level is L_{H-1} . A node that belongs to L_i ($i \geq 1$) calculates its own Voronoi cell region. The link is constructed for nodes that belong to L_{i-1} and exist inside the Voronoi cell of the node in L_i . A node that belongs to L_i discovers the nodes inside its Voronoi cell by using the Delaunay overlay structure of L_{i-1} . Then, the links between the node in L_i and the discovered nodes on L_{i-1} are established. This operation is performed between all levels of the hierarchy (Fig. 4).

3.3 Aggregation of Sensor Data

HDA-GC aggregates sensor data consists of observed value and other data by using the inter-hierarchy links on the HDOV pe-

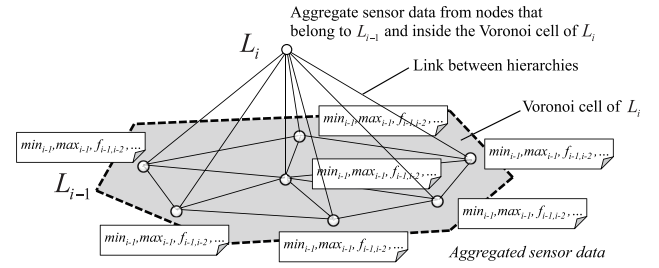


Fig. 4 Aggregation using links between hierarchies.

riodically. A node that belongs to L_i ($i \geq 1$) aggregates sensor data of the nodes on L_{i-1} . The sensor data aggregation is performed using exchanging messages by using the inter-hierarchical links as shown in Fig. 4. Hereafter, N_i represents a set of nodes that belongs to L_i .

In the remainder of the text, we use the following notations.

- *obs*: the observed value of a node
- *location*: location of a node (consists of lng(longitude) and lat(latitude))
- *vor_i*: size of Voronoi cell of a node on layer L_i
- *children_i*: children nodes on L_i . a set of nodes on L_{i-1} inside the Voronoi cell of a node in L_i
- *num_i*: total number of nodes inside Voronoi cell of a node on layer L_i .
- *min_i*: minimum observed value among *children_i*
- *max_i*: maximum observed value among *children_i*
- *ave_i*: average of observed value among *children_i*
- *center_i*: center location of *children_i*
- *f_{i,j}*: feature amount from L_i to L_j ($0 \leq j < i$) on a node

Each sensor data x of node n is represented as $n.x$ (e.g., *min_i* of n is represented as $n.min_i$).

Figure 5 shows the pseudocode of the aggregation of sensor data. The function *aggregateSensorData* (n, i) corresponds to the aggregation process of node $n \in N_i$. The function is executed by the function call from all nodes on the top layer of the HDOV. That is, the function call command of *aggregateSensorData* ($n, H-1$) is transmitted to the all of the nodes that belongs to N_{H-1} . Each node aggregates data recursively by this function.

A node $n \in N_i$ aggregates following sensor data:

$$\{obs, num_i, min_i, max_i, ave_i, center_i, f_{i,0}..f_{i,i-1}\}$$

The feature amount $n.f_{i,j}$, which is mainly used for assignment of the number of collecting sensor data, corresponds to an amount of a variance of observed values on n from L_i to L_j ($0 \leq j < i$). On a layer L_i , $n.f_{i,i-1}$ is calculated by using the set of $m.f_{i-1,i-2}$ where $m \in n.children_i$. Nodes $n \in N_i$ ($i > 1$) calculates r which is defined as below for each $m \in n.children_i$:

$$r = (n.max_{i-1} - n.min_{i-1}) \times (\sqrt{m.vor_{i-1}} / \sqrt{n.vor_i})$$

The value r corresponds to the range of the observed values for the difference of the width of Voronoi cells between layers L_{i-1} and L_i . If $m.f_{i-1,i-2}$, a feature amount of m in L_{i-1} , which is calculated recursively, has larger value than r , the difference ($r - m.f_{i-1,i-2}$) is added for $n.f_{i-1,i-2}$ as a feature amount of n in L_{i-1} . The average value of $m.f_{i-1,i-2}$ except which is larger than r


```

01 function aggregateSensorData(n, i)
02 /* initialize variables here */
03 if i = 0
04   n.num0 = 1, n.min0 = n.obs, n.max0 = n.obs,
05   ave0 = n.obs, n.center0 = n.location;
06 else
07   foreach m in n.childreni
08     aggregateSensorData(m, i - 1);
09     n.mini ← m.mini-1 < n.mini? m.mini-1 : n.mini;
10     n.maxi ← m.maxi-1 > n.maxi? m.maxi-1 : max;
11     n.numi ← n.numi + m.numi-1;
12     sum ← sum + m.numi-1 × m.avei-1;
13     lng ← lng + m.numi-1 × m.centeri-1.lng;
14     lat ← lat + m.numi-1 × m.centeri-1.lat;
15   end foreach
16   n.avei ← sum / n.numi;
17   n.centeri ← {lng: lng / n.numi, lat: lat / n.numi};
18   end if
19   if i = 1
20     n.f1,0 ← n.max1 - n.min1;
21   else
22     foreach m in n.childreni
23       r ← (n.maxi - n.mini) × √(n.vori-1 / √(n.vori);
24       if m.fi-1,i-2 > r /* there is a characteristic point in m.fi-1,i-2 */
25         n.fi,i-2 ← n.fi-1,i-2 + m.fi-1,i-2 - r;
26         cnt++;
27       else /* there is no characteristic point in m.fi-1,i-2 */
28         n.fi,i-1 ← n.fi-1,i-1 + m.fi-1,i-2;
29         sum ← sum + m.fi-1,i-2;
30       end if
31     end foreach
32     if n.fi,i-1 = 0
33       avef ← r;
34     else
35       avef ← sum / (n.childreni - cnt);
36     end if
37     n.fi,i-1 ← n.fi,i-1 + cnt × avef;
38     for (j ← i - 3; j ≥ 0; j--)
39       foreach m in n.childreni
40         n.fi,j ← n.fi,j + m.fi,j;
41       end foreach
42     end for
43   end if
44 end if
45 end function

```

Fig. 5 Pseudocode of sensor data aggregation.

is added to $n.f_{i,i-1}$ (lines from 22 to 37).

In addition, feature amounts under $L_i - 1$ plus $n.f_{i,i-2}$ calculated above are kept for each $n \in N_i$ (lines from 38 to 43).

By the above procedure, n on L_i can keep the feature amounts from L_i to L_j ($0 \leq j < i$), that is, variance of the observed value under L_i , ignoring variance of the observed values in the characteristic points smaller than the Voronoi cell of L_j . Thus, feature amounts represents whether there is a characteristic point inside a Voronoi cell or not in a layer of HDOV. If the feature amount is large, we can judge that there is a characteristic point inside the Voronoi cell in the layer.

Since Delaunay maintenance messages are exchanged periodically to maintain the link structure in the HDOV, the sensor data aggregation process can be executed periodically by extending these maintenance message.

3.4 Collection of Sensor Data for Contour Lines Map

The basic sensor data collection method of HDA-GC is same as HDA-RC. The difference is the calculation method of the number of sensor data ($R_{i,n}$ in Fig. 3) assigned to the children nodes while sensor data collection.

First, a system provider or a developer specifies M , the limit on the number of collecting sensor data. Then, M is assigned to nodes on the top level of the HDOV, whose Voronoi cell intersects with the target area. At this time, the assignment of M is based

on the two factors, the number of geographically uniform sensor data and the number of collecting sensor data, which is specified by the user's request, which are same as HDA-RC.

For wide-area interpolation, at least a certain number of uniformly distributed sensor data is required to create contour lines map. Since the system sets M , the number of uniformly distributed sensor data (denoted as uni) is decided within the range of M . When a user issues a query, the user specifies layer level g along with the target area. L_g is the layer in which the average size of the Voronoi cell corresponds to the required granularity. In the HDOV, the average geographical node density and size of the Voronoi cell for each layer can be controlled. Hereafter, the node density of L_i is denoted as D_i and the size of a target area is denoted as S_{req} . When a user issues a query for a certain target area, a set of geographically uniform sensor data is collected from the layer with level u which has the highest density that satisfies $D_u \leq uni/S_{req}$.

A node n ($n \in N_i, i > u$) assigns the number of sensor data $R_{i,n}$ to $n.children_i$ on the basis of the following formula.

$$R_{i-1,m} = R_{i,n} \times \frac{(m.area \times m.f_{i,g-1} \times m.segments)}{\sum_{k \in n.children_i} (k.area \times k.f_{i,g-1} \times k.segments)}$$

where m is a target node to assign number of sensor data. On the top level, $R_{H-1,n} = M - uni$. $m.area$ is the area size where the target area and the Voronoi cell of node m intersect. $m.segments$ ($m \in n.children_i$) is the number of segments inside of a Voronoi cell. $m.segments$ can be calculated by $m.min_i, m.max_i$ and the requested segment width.

When L_{i-1} , to which node m belongs, is lower than L_g , $R_{i,n}$ is assigned uniformly to the children nodes. Node n uses the normalized value from 0 to 1 for each feature amount of $n.children_i$.

When a node n ($n \in N_u$) assigns the number of sensor data to $n.children_u$, uni is assigned to collect geographically uniform sensor data in addition to the assignment of $R_{u,n}$. Node n assigns the quotient of $uni/(D_u \times S_{req})$ to $n.children_u$. If there is a remainder, node n adds 1 to the number of sensor data for $n.children_u$ with the probability $\{uni\%/(D_u \times S_{req})\}/(D_i \times S_{req})$.

While a node n ($n \in N_i, i < u$) assigns the number of sensor data to $n.children_i$, if the number of sensor data for geographically uniform sensor data is larger than 1, node n assigns the number of sensor data uniformly to the children nodes. If the number of sensor data for geographically uniform sensor data equals to 1, node n adds 1 to $R_{i,n}$ and assigns numbers on the basis of the above formula. At this time, if the assigned number of sensor data for node m ($m \in n.children_i, i < u$) is more than 1, node n forwards a message to node m . If the value equals to 1, node n returns $m.center_{i-1}$ and $m.ave_{i-1}$ in the aggregated information. When a node n ($n \in N_1$) receives a message, node n returns the observed value $n.obs$ of N_0 according to the assigned numbers.

Thus, a node n ($n \in N_i$) assigns the number of sensor data to $n.children_i$ on the basis of the aggregated sensor data including the feature amount.

4. Evaluations

We evaluated the effectiveness of the proposed method (HDA-GC) with simulations. The rest of the section describes the simu-

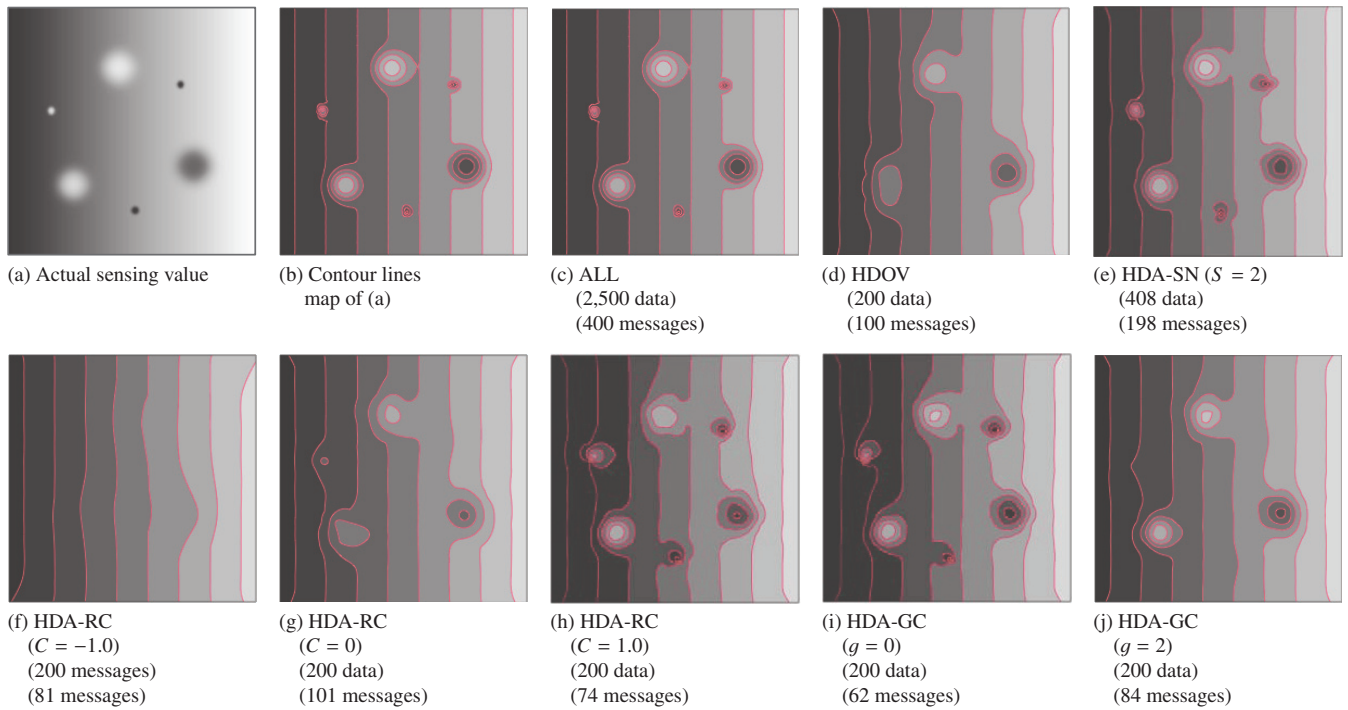


Fig. 6 Created contour lines map (segment width: 30).

lation setup and the results of the simulations.

4.1 Simulation setup

The simulation setup is as follows. The size of the area was 300×300 unit length and 2,500 nodes were located by grid distribution. A sample of an actual sensing value used for the simulation is shown in Fig. 6 (a) is used for the simulation. As shown in this figure, the sensor data distribution had steep peaks of different sizes and little noise in order to examine the features and accuracy of the proposed method. In this figure, the sensing value corresponded to the pixel value of a gray scale image 300 pixels $\times$ 300 pixels in size.

As a spatial interpolation, we applied the Kriging interpolation method [9] which can model both the wide area tendency and small area changes of the measured values according to the subordination feature of the observations.

In the simulation, we used the emulator function of PIAX [14] in which the Delaunay overlay network protocol [10] is implemented. The HDOV was constructed with five levels. The entire number of the sensor nodes was 2,500, which corresponded to the L_0 . L_1 included 400 nodes, L_2 included 100 nodes, L_3 included 25 nodes, and L_4 included 4 nodes, respectively.

As a comparison, HDOV, HDA-SN and HDA-RC were used. To evaluate the number of messages fairly, the sensor data of children nodes were aggregated in the HDOV evaluation. As a system constraint, the limit number of the sensor data collection is set as 200 for HDA-RC and HDA-GC.

4.2 Contour lines map

All figures in the Fig. 6 show the created contour lines maps when the requested segment width is 30. Figure 6 (b) is the optimal contour lines map based on the actual sensing value shown in Fig. 6 (a). Figure 6 (c) is the reconstructed contour lines map

by all sensor data (ALL).

Figure 6 (d) corresponds to the HDOV in which the number of collected sensor data is 200. Figure 6 (e) corresponds to the HDA-SN and parameter S was set as 2. Figure 6 (f), 6 (g) and 6 (h) correspond to HDA-RC in which the number of collected sensor data is 200 (of which the number of geographically uniform sensor data is $\{0, 100, 100\}$) and C is set to $\{-1.0, 0, 1.0\}$, respectively. Figure 6 (i), 6 (j) correspond to HDA-GC, in which the number of collected sensor data is 200 (of which the number of geographically uniform sensor data uni is 100) and g is set to $\{0, 2\}$, respectively.

In addition, the number of collected sensor data and the number of messages exchanged to deliver the request query to each node, which affects on the network traffic and response time, are shown in the caption of each figure.

ALL reconstructed the most accurate contour lines map. However, since it collected all sensor data, it exceeded the system constraint and collected 2,500 sensor data.

Apparently, the HDOV with 200 sensor data failed to reconstruct the original distribution. It failed to reconstruct not only the characteristic points with small granularity but also those with large granularity. To grasp characteristics with small granularity, more geographically uniform sensor data are needed. However, in that case, it cannot be guaranteed that the characteristic points with small granularity will be reconstructed since geographically uniform sensor data do not focus on the observed value. Moreover, a lot of redundant sensor data are collected in the outside of the characteristic points, and there is a possibility of exceeding the system constraint.

In HDA-SN, large and small characteristic points are both reconstructed accurately. But to keep the accuracy, the number of collected sensor data exceeds 200, which is a system constraint in this case (408 data).

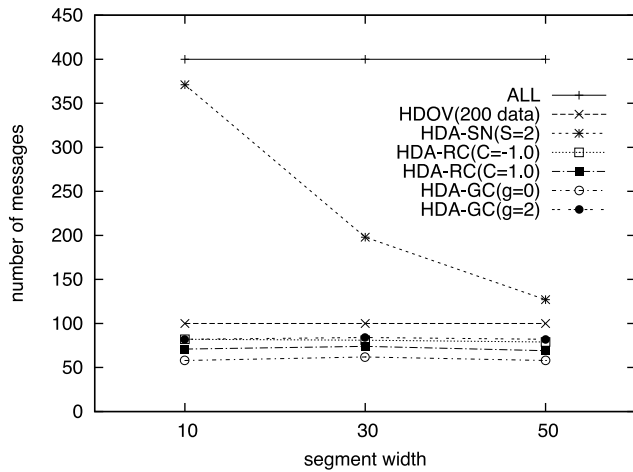


Fig. 7 Number of messages with variance of segment width.

In HDA-RC with $C = 0$, the contour lines map is similar to the case of HDOV with 200 sensor data since collected sensor data are geographically uniform. When C was negative, the contour lines map was reconstructed, omitting characteristic points. When C was positive, the contour lines map with characteristic points was reconstructed since it collected sensor data inside the area where the variance of sensor data inclination was large and there were characteristic points. However, since HDA-RC does not treat the granularity of characteristic points, it can always collect sensor data on small characteristic points even when a user only need to grasp more wide area tendency than small characteristic points.

In comparison, HDA-GC could control the granularity of the characteristic points by specifying g . In Fig. 6 (i), small characteristic points were reconstructed since the requested granularity is small ($g = 0$). In Fig. 6 (j), small characteristic points were omitted below the requested granularity ($g = 2$). In a limited number of sensor data, it is difficult to reconstruct both the tendency of a wide area and the changes in a small area. Since contour lines are drawn accurately to the entire area by omitting the characteristics with small granularity, it is effective in the case of grasping the wide area tendency.

4.3 Number of Messages

Figure 7 shows the number of forwarded messages with variance of segment width.

ALL and HDOV are assumed to aggregate the sensor data and the geographical location of children nodes. Therefore, they forward messages to the nodes that belong to the one upper layer that target nodes belong by flooding. Especially, ALL takes large load to the system since it has to forward messages to all nodes on L_0 .

In HDA-SN, generally the number of messages rises in proportion to the number of collected sensor data. Therefore, if many sensor data is collected to create accurate contour lines map, the number of messages are exchanged on the network is increased. As mentioned in the former section, when the segment width is small, too much sensor data can be collected in HDA-SN. In this evaluation, when the segment width is set as ten, 371 messages are forwarded. That is almost same as ALL.

On the other hand, HDA-RC and HDA-GC could limit the number of messages regardless of the segment width. In particular, HDA-GC can collect sensor data according to the user's request in the small number of messages following the number of sensor data specified by the system.

4.4 Signal to Noise Ratio (SNR)

To examine the features of HDA-GC, we used the signal to noise ratio (SNR) of a reconstructed contour lines map. The SNR is defined as following formula.

$$SNR = 10 \log_{10}(max^N / MNE)$$

where MNE (Mean to The N th power Error) is defined as following formula.

$$MNE = \frac{\sum |I_o - I_n|^N}{\sum 1}$$

max is the maximum value of segment values in the contour lines map. I_o is a value of the segment of the contour lines map that is reconstructed from the original distribution, and I_n is a value of the segment of the contour lines map by each method. That is, MNE is the average of the segment error margin for each pixel between I_o and I_n . When contour lines are drawn accurately to the entire area, the SNR indicates a high value at a small N . When a contour lines map is reconstructed without losing characteristic points, the SNR indicates a high value at a large N . The MNE is set to 0 if the segment value of I_o and I_n are the same and set to 1 if they are different.

Figures 8, 9 show the SNR of a reconstructed contour lines map when the number of sensor data is 200 and 400, respectively.

The result of ALL indicated the highest SNR at all N . Therefore, ALL reconstructed an accurate contour lines map. HDA-SN also shows high SNR at all N . However, as mentioned above, they collect sensor data exceeding the system constraint.

As shown in Fig. 8, HDA-GC ($g = 2$) indicated high SNRs at small N s compared with the other methods. That is, contour lines were drawn accurately to the entire area by omitting characteristic points with small granularity. Therefore, the tendency of the sensor data distribution was reconstructed more accurately by collecting sensor data according to the specified granularity. On the contrary, HDA-GC ($g = 0$) indicated high SNRs at large N . That is, the contour lines map is reconstructed without losing characteristic points. Compared with HDOV or HDA-RC ($C = 1.0$), HDA-GC ($g = 0$) indicated high SNRs. Therefore, HDA-GC decided the target nodes for the spatial interpolation effectively. When the number of sensor data was large, as shown in Fig. 9, the difference between HDA-GC ($g = 0$) and HDA-GC ($g = 2$) was smaller at small N s. This is because a sufficient number of geographically uniform sensor data was given and contour lines were drawn accurately to the entire area. Both HDA-GC ($g = 0$) and HDA-GC ($g = 2$) could grasp the tendency of the sensor data distribution, but HDA-GC ($g = 2$) had some superiority since it was effective to reconstruct the contour lines map by omitting the characteristics with small granularity for grasping the wide tendency. As well as Fig. 8, HDA-GC indicated a high SNR compared with the HDOV or HDA-RC. It is remarkable

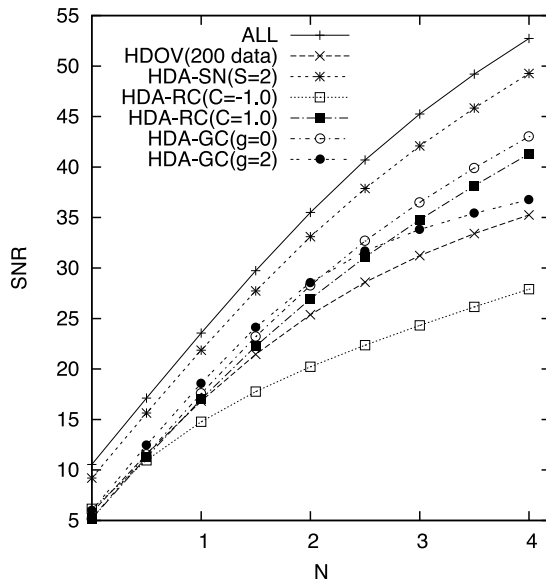


Fig. 8 SNR of Contour Lines Map (200 sensor data).

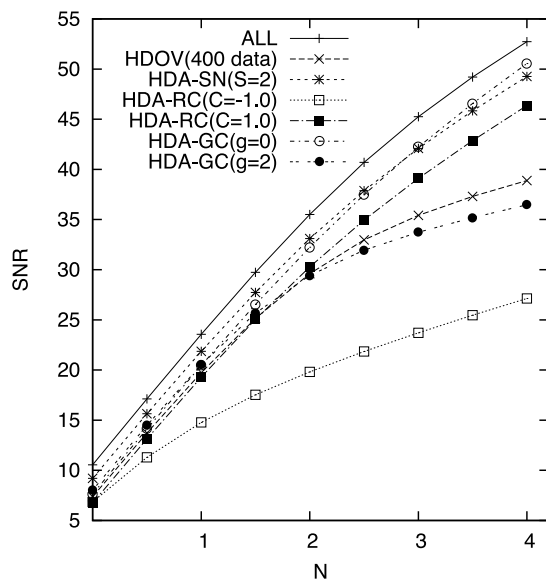


Fig. 9 SNR of Contour Lines Map (400 sensor data).

that HDA-GC ($g = 0$) shows higher SNR than HDA-SN when N is large. That is, we can say the HDA-GC decides the target nodes for the spatial interpolation effectively.

5. Discussions

The proposed method is effective from the viewpoint of the importance of reconstructing contour lines maps within system constraints. However, there are some points that we should consider to improve.

In the proposed method, sometimes it is difficult to identify the correspondence relation between each parameter. For example, the number of geographically uniform sensor data should be decided on the basis of the limited number of sensor data, sensor data distribution, the specified granularity of the characteristic points, and so on. Therefore, we must consider how to decide such parameters.

The proposed method requires forwarding messages to the

nodes with geographically uniform distribution. In this paper, HDOV was used for the sensor data aggregation and query forwarding. Since the HDOV has a convenient feature that can send to the nodes with specified geographical node density, it is suitable for our proposal. However, this feature may also be provided by other geographical overlay networks. For example, LL-Net [5] has an hierarchical overlay network structure that divides the entire space into multiple levels of four squares. Aggregation and collection algorithms for such different kind of geographical overlay networks should also be treated.

The proposed method only treats a single independent query. This can be applied to the relatively static observations like weather observations (1 to 5 minutes periodic updates on Live E! [4]), traffic observations (every 5 minutes on VICS [15]) and so on. However, more frequently updated sensors should be treated to keep up with the real time situations. In such cases, when the temporal changes are rather gentle, redundant sensor data collections can occur. Temporal interpolations might be applicable for such situations.

6. Conclusion

In this paper, we proposed a sensor data collection method (HDA-GC) for peer-to-peer sensor network that can create contour lines maps with system constraints on the number of collected sensor data. Our proposed method could create contour lines map on the basis of the granularity of the characteristic point specified by a user, using hierarchical Delaunay overlay network. We evaluated the effectiveness of our method by using a sensor data distribution with different sizes of characteristic points. The results showed that the proposed method created a contour lines map, satisfying the requested granularity of the characteristic points and the system constraint on the number of sensor data.

In this paper, we evaluated on a typical environment by simulations, but did not evaluate on the real sensor data. Evaluations using a real sensor data distribution that has complexity to show the effectiveness of our proposal will be our future work. Some features that was described in the former section, such as a derivation method for optimum parameters, will also be future work.

References

- [1] Aberer, K., Hauswirth, M. and Salehi, A.: Global sensor networks, Ecole Polytechnique Federale de Lausanne (EPFL), Technical Report LSIR-2006-001 (2006).
- [2] Araújo, F. and Rodrigues, L.: GeoPeer: A Location-Aware Peer-to-Peer System, *Proc. 3rd IEEE International Symposium on Network Computing and Applications (NCA'04)*, pp.39–46 (Aug. 2004).
- [3] Desnoyers, P., Ganesan, D. and Shenoy, P.: TSAR: A Two Tier Sensor Storage Architecture Using Interval Skip Graphs, *Proc. 3rd International Conference on Embedded Networked Sensor Systems*, pp.39–50 (2005).
- [4] Esaki, H. and Sunahara, H.: Live E! Project; Sensing the Earth with Internet Weather Stations, *Proc. International Symposium on Applications and the Internet (SAINT'07)*, p.1 (Jan. 2007).
- [5] Kaneko, Y., Harumoto, K., Fukumura, S., Shimojo, S. and Nishio, S.: A Location-based Peer-to-Peer Network for Context-aware Services in a Ubiquitous Environment, *Proc. Symposium on Applications and the Internet (SAINT'05) Workshops*, pp.208–211 (Feb. 2005).
- [6] Kanzaki, A., Hara, T., Ishi, Y., Yoshihisa, T., Teranishi, Y. and Shimojo, S.: X-Sensor: Wireless Sensor Network Testbed Integrating Multiple Networks, *Wireless Sensor Network Technologies for the Information Explosion Era Studies in Computational Intelligence*, Vol.278, pp.249–271 (2010).

- [7] Konishi, Y., Takeuchi, S., Teranishi, Y., Harumoto, K. and Shimojo, S.: A Data Collection Technique to Realize Rough Observation of Sensor Value Distribution in a P2P Environment, *Proc. Multimedia, Distributed, Cooperative, and Mobile Symposium*, pp.165–172 (July 2007).
- [8] Teranishi, Y., Takeuchi, S. and Harumoto, K.: HDOV: An overlay network for wide area spatial data collection, *Proc. ACM Symposium on Applied Computing (SAC'11)*, pp.506–513 (Mar. 2011).
- [9] Lloyd, C.D.: *Local models for spatial analysis*, CRC Press (2006).
- [10] Ohnishi, M., Nishide, R. and Ueshima, S.: Incremental Construction of Delaunay Overlay Network for Virtual Collaborative Space, *Proc. 3rd International Conference on Creating, Connecting and Collaborating through Computing (C5 2005)*, pp.77–84 (Jan. 2005).
- [11] Presser, M., Barnaghi, P.M., Eurich, M. and Villalonga, C.: The SENSEI project: Integrating the physical world with the digital world of the network of the future, *IEEE Communications Magazine*, Vol.47, No.4, pp.1–4 (2009).
- [12] Shinomiya, J., Teranishi, Y., Harumoto, K. and Nishio, S.: A Sensor Data Collection Method Under a System Constraint Using Hierarchical Delaunay Overlay Network, *Proc. International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP 2011)*, pp.300–305 (Dec. 2011).
- [13] Shinomiya, J., Teranishi, Y., Harumoto, K., Takeuchi, S. and Nishio, S.: An Examination of Sensor Data Collection Method for Spatial Interpolation on Hierarchical Delaunay Overlay Network, *Proc. International Workshop on Sensor Network Technologies for Information Explosion Era (SeNTIE 2010)*, pp.407–412 (May 2010).
- [14] Teranishi, Y.: PIAX: Toward a framework for sensor overlay network, *Proc. International IEEE CCNC Workshop on Dependable and Sustainable Peer-to-Peer Systems*, pp.1–5 (Jan. 2009).
- [15] Yamada, S.: The Strategy and Deployment Plan for VICS, *IEEE Communications Magazine*, Vol.34, No.10, pp.94–97 (1996).

Editor's Recommendation

The number of messages is reduced maintaining observation accuracy by using the geographical characteristic of an observed value, in order to collect observation information efficiently from the sensor connected by P2P network. The performance evaluation test is also settled, the completeness as a paper is high and a scientifically high contribution is accepted. Therefore, the recommendation from this SIG DPS is deserved.

(Chairman of SIGDPS Michiaki Katsumoto)



Jun Shinomiya received his M.E. degree from Osaka University in 2012. In 2012, he joined Sony Corporation. Currently, he is engaged in the Professional Solutions Group, Imaging Products and Solutions Sector of Sony Corporation. His research interest is distributed data management systems.



Yuuichi Teranishi received his M.E. and Ph.D. degrees from Osaka University, Japan, in 1995 and 2004, respectively. From 1995 to 2004, he was engaged Nippon Telegraph and Telephone Corporation (NTT). From 2005 to 2007, he was a lecturer of Cybermedia Center, Osaka University. From 2007 to 2011, He was

an associate professor of Graduate School of Information Science and Technology, Osaka University. Since August 2011, He has been a research manager and project manager of National Institute of Information and Communications Technology (NICT). He received IPSJ Best Paper Award in 2011. His research interests include technologies for distributed network systems and applications. He is a member of IEEE.



Kaname Harumoto received his M.E. and Ph.D. degrees from Osaka University, Japan, in 1994 and 1998, respectively. He is an associate professor of Graduate School of Engineering, Osaka University. He received IPSJ Best Paper Award in 2011. His research interests include technologies for distributed data management

and utilization. He is a member of IEEE and IEICE.



Shojiro Nishio received his B.E., M.E., and Ph.D. degrees from Kyoto University in Japan, in 1975, 1977, and 1980, respectively. He has been a full professor at Osaka University since August 1992, and was given a peculiar title “Distinguished Professor of Osaka University” in July 2013. He served as a vice president and

Trustee of Osaka University from August 2007 to August 2011. Dr. Nishio has co-authored or co-edited more than 55 books, and authored or co-authored more than 600 refereed journal or conference papers. He served as the Program Committee Co-Chairs for several international conferences including DOOD 1989, VLDB 1995, and IEEE ICDE 2005. He has served and is currently serving as an editor of several international journals including IEEE Trans. on Knowledge and Data Engineering, VLDB Journal, ACM Trans. on Internet Technology, and Data & Knowledge Engineering. Dr. Nishio has received numerous awards during his research career, including the Medal with Purple Ribbon from the Japanese Emperor in 2011. He is also a fellow of IEEE, IEICE and IPSJ, and is a member of five learned societies, including ACM.