Geosensor Data Abstraction for Environmental Monitoring Application[1]

Young Jin Jung and Silvia Nittel

Spatial Information and Engineering, University of Maine, USA

{yjung, nittel}@spatial.maine.edu

Abstract. Environmental observation applications are designed for monitoring phenomena using heterogeneous sensor data types and for providing derived and often integrated information. To effectively handle such a large variety of different sensors, both in scale and type and data volume, we propose a geosensor abstraction for large-scale geosensor networks. Our SGSA(Slope Grid for Sensor Data Abstraction) represents collected data in single grid-based layers, and allows for summarizing the measured data in various integrated grid layers. Within each cell, a slope vector is used to represents the trend of the observed sensor data. This slope is used as a simplifying factor for processing queries over several sensor types. To handle dynamic sensor data, the proposed abstraction model also supports rapid data update by using a mapping table. This model can be utilized as a data representation model in various geosensor network applications.

Keywords : Sensor data abstraction, Geosensor network, Slope grid, GIS, Surface model

1 Introduction

Environmental monitoring applications have become significant tools for analyzing nature's phenomena. The advances in wireless communication and sensor miniaturization technologies as well as small-form computing devices have significantly contributed to the enablement of environmental monitoring in the physical world [1]. To detect the conditions or events in a wide-area geographic space, a monitoring application requires a large-scale geosensor network [2] including various kinds of sensors. Today, many large, autonomous sensor platforms such as wind sensors or ocean buoys are deployed; in a few years, these traditional sensor environments will be integrated with the deployment of small-form sensor networks providing rapid rates of real-time sensor data of various type, scale and location.

[1] Published in Proc. 5th International Conference on Geographic Information Science (GIScience), Sept 23-26 2008, Park City, Utah, Springer LNCS.

Monitoring applications can provide useful information for environmental science (e.g. a habitat analysis or a micro-climate model for spatially limited area such as a vineyard or a greenhouse) and for a warning system of dangerous environmental events (e.g. forest fire and air pollution) [3]. For developing sensor network applications, various kinds of technologies including sensing, communication and computing are required [4]. Furthermore, the application of domain knowledge is also necessary in order to interpret and understand an environmental condition, based on the collected data [3].

Today, queries in sensor data applications are typically executed on the raw, collected sensor data stored in a database system. To process all sensor data without any data abstraction, query execution slows down significantly due to the large volume of real-time sensor data. In addition, it is difficult to compile useful information for answering queries such as "When will air pollution be reduced?" and "Where is a potential dangerously polluted area in the near future?" It requires a sensor data abstraction as a preprocessing step for interpreting the observed data and for processing a query in a central monitoring server. The sensor data in a wide area can be created in large volume, and streamed to a central server. In addition, it also can change dynamically. In order to handle the observed data, some requirements are posed for the data abstraction model itself, such as a rapid data processing, an effective update policy, and easy data access, etc.

In this paper, we designed SGSA (Slope Grid for Sensor Data Abstraction) for representing large-scale geosensor data in a central monitoring system. The proposed abstraction model is based on a grid-based technique for characterizing terrain surface area [5]. After collecting a sensor data type, the grid represents the data on each cell as a spatial slope that is described by min(), max(), and a slope direction. It can find the condition change in the area because the slope shows the data gradient on the cell. Users can understand the conditions by checking the slopes on cells in the grid as if you would see a simple contour map. Besides, it can support a rapid data update using a mapping table. This proposed model can be utilized as a data representation model in various geosensor network applications.

2 Related Work

The Environment Observation and Forecasting System (EOFS) is an application for understanding a situation and providing forecasting using a large-scale sensor network [6, 7]. The system supports centralized processing, handles huge data volume, and provides an autonomous operation, etc. Some usages examples of EOFS are the habitat monitoring of seabirds [8], the CORIE for monitoring the Columbia river [9], the Automated Local Evaluation in Real-Time(ALERT) [10], and the framework for in-situ sensor data processing [11]. The seabird habitat monitoring project focuses on habitat condition analysis. The goal of the CORIE project is to guide vessel transportation and forecasting. It utilized the 13 stationary sensor nodes and features a computationally intensive physical environment model. GLACSWEB project monitors the behavior of ice caps and glaciers for understanding the Earth's climate [12]. The in-situ sensor data processing system monitors continuously updated sensor data and the communication status in sensor network [11]. It can provide an alarm with the registered context aware model and rules. The PODS project monitors rare and endangered species of plants by using high-resolution cameras, a thermometer, and solar radiation sensors [13]. It describes a simple data display with colored dots, which shows the category values for each sensor. For example, green dots mean a normal condition. On the other hand, red indicates one of the extreme categories. It also considers data summarization techniques such as consolidating the data with Theissen polygons (Delaunay triangulation) [14] to show the general pattern of the data values as a map. GMT, the Generic Mapping tools, is available to present the measured data [15].

On the spatial resolution, some applications need the raw data of all sensing points [16, 17], whereas others need just a summary of all sensors' data, as those TAG [18]. An intermediate data summarization between these two cases is also utilized such as maps of temperature and relative humidity [7]. These applications can identify zones of interest such as hot and cold zones. To avoid information overload, spatial aggregation is also utilized for data summarization over subregions, pre-defined zones. For example, it uses a aggregation predicate such as "SELECT avg(volume) FROM Sensors GROUP BY region HAVING avg(volume) > threshold." [19] The spatial distribution of these summarizations provides a report of the data variability over the entire region. In order to recognize the conditions of the circumstance, monitoring applications including EOFS need a well-organized data representation model, because they have to rapidly process a huge volume of sensor data and interpret the transmitted observed data.

3 Slope Grid for Sensor Data Abstraction

In a large-scale geosensor network, sensor data abstraction is required for handling large volumes of data and making the information useful even though a large scale sensor network has not yet been utilized for practical environmental monitoring until recently [7]. The volume of sensor data can potentially be generated over a large geographic area, and we assume that the collected values dynamically change continuously. A grid is generally utilized to present the measured data or the condition on an area such as a graph and a terrain surface model in GIS [20]. Our slope grid is based on the tilted plane, which is useful to present a surface area [5]. When it present a surface area, a slope (the tilted plane) is better than the horizontal plane, because the slope can present the data elevation change as well as the height in each cell. When it is used to manage a large volume of moving object location data, this grid structure also shows good search performance by using a hash table as an index structure [21]. It focuses on the access method for updating and searching the location of moving objects. In this paper, we concentrate the summarization of environmental condition for analyzing a phenomenon with slope grid.

If an application stores the measured sensor data without any data spatio-temporal abstraction model, all data has to be searched for answering user queries. It requires more time to process the queries and to derive the information that users want to know. The objectives of providing a sensor data abstraction model are the frequent data updates, the geographic area covered, and a continuous amount data process, data representation and a compact size, etc.

To apply the slope for surface areas to the sensor data monitoring application, we simplify the presentation of the tilted plane to 9 directions derived from min() and a max() of 4 subcells in a cell. In order to present the continuously changed data motion, we use an update policy with a mapping table and a gradient count for checking the data change in a cell.

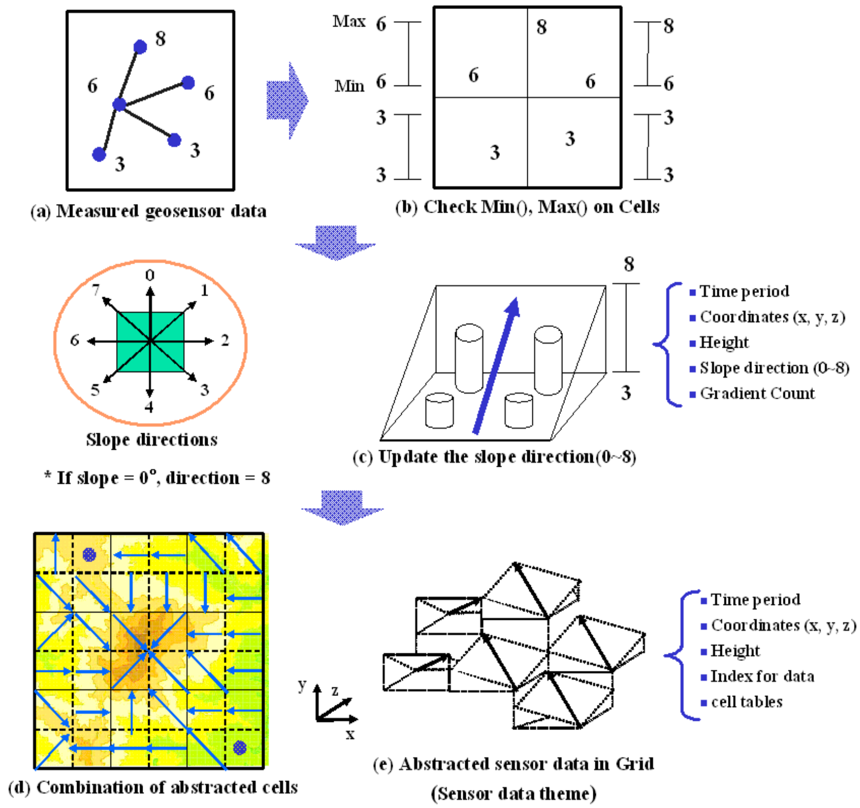


Fig. 1. The sensor data abstraction

We propose sensor data abstraction model, in which sensor data is represented in individual grids. The grid for each type serves as input for spatio-temporal sensor queries, which are sophisticated data interpretation in most cases. In our approach, we use the data abstraction model as a preprocessing step for interpreting the conditions at the remote place, thus, as a basic data representation model for an environmental monitoring application [11]. It focuses only on current data. This abstract data is used for making useful information by combining other abstracted data types.

The proposed sensor data abstraction is shown in Figure 1. The data is summarized on each cell after receiving the sensor data transmitted from geosensor network at (a).

We choose the minimum and the maximum value for representing values within a cell at Figure 2(b). The other values of the cell are included in the value boundary, which consists of min() and max(). The size of the cell is defined depending on some conditions such as the number of sensors in a cell, data feature, and application function. In a cell, 1~8 sensors are useful to present the trend of data, because a cell includes 4 subcells which present a min() and a max() with two data values within each subcell. It can include more than 8 observed values but the data representation is not different. The size is also defined by a data feature. For example, it does not require many sensors for observing wind direction and speed, because both phenomena don't frequently changed over a larger area. In this paper, we consider only a static cell size since changing a cell area also requires the time to process it. It would be an obstacle for the rapid data processing.

The coordinates (x_1, y_1, z_1, x_2, y_2, z_2) indicate the locations of cells in a grid. The time period (t_1, t_2) of a cell includes all of the observation time of the sensor data at this location. It contains the latest measurement time, because it is updated whenever the observed data is updated. At (c), the height is a difference between min() and max() in the cell. If max() is 8 and min() is 3, the height is 5 and z coordinate is 3, which starts at the min(). The slope direction is derived from the relative positions of two subcells as shown in Figure 2. One is the "M" subcell, which contains the maximum value, and the other is the "m" subcell which has a minimum value.

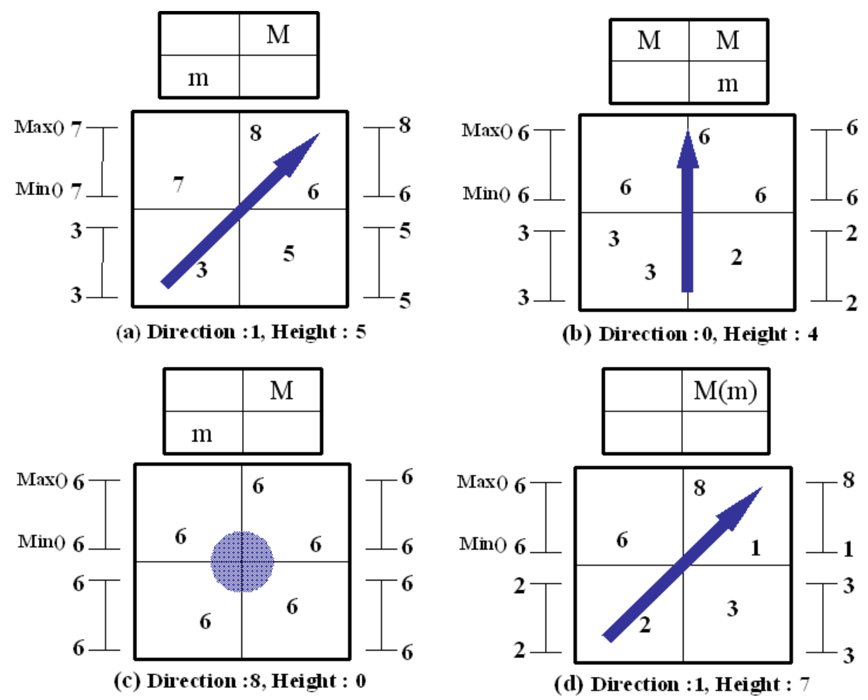


Fig. 2. The derived slope direction

The direction vector is a vector pointing from m subcell to M subcell. Namely, it presents the direction from the min() to the max() in a cell (Figure 2). In a small grid above each cell, M indicates the M subcell and m the m subcell. If the slope angle = 0°, the direction = 8. It means that all the sensor data has the same value and the slope is flat. It is presented a dot in the grid at (c) in Figure 2. If a subcell includes a maximum and a minimum values like (d), the direction points towards the M subcell. Namely, the direction always points the M subcell. When it present the data in a cell, this slope is better than a horizontal plane, because the slope present the data trend. For example, if there are two horizontal planes, which have same height, they are only equal. However, if there are two slopes with same height, they can present various kinds of data motion even though they are same height.

4 Updating Data to the Abstract Model

It is essential to support frequent data update in sensor data monitoring applications, especially for real-time monitoring and warning systems. Such a system has to detect and cope with the observed emergency situation such as air pollution, forest fire, battlefield analysis, etc. Sensor data likely changes dynamically over time. The abstracted data representation model employs a mapping table for rapid data update.

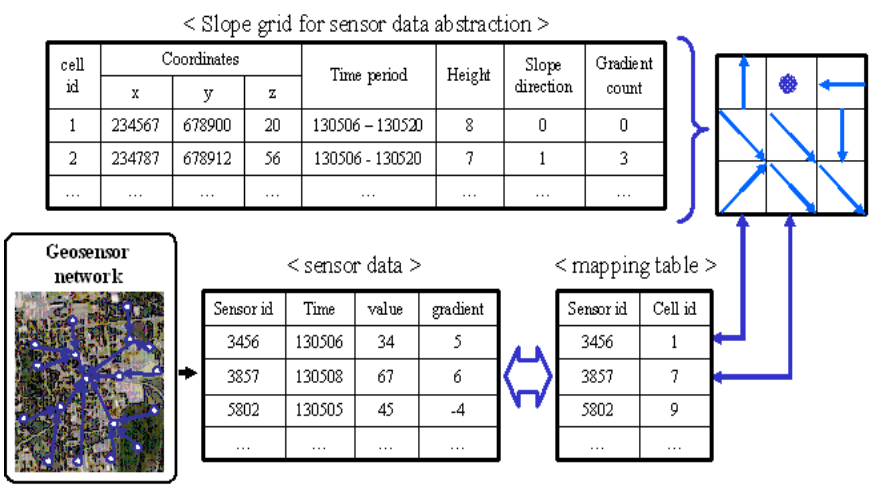


Fig. 3. The sensor data update structure in the SGSA

The update process of the abstract data model is shown in Figure 3. The transmitted data is stored in an incoming sensor data table or data stream. The sensor data table stores the sensor id, the measurement time stamp, the sensed value, and the gradient. The gradient shows data variation derived from past value. The mapping table shows the cell id and included sensor id. Whenever the locations of sensors are changed, it finds the cells, which include the locations. This data is summarized in the abstracted data table for each cell. Here, the slope direction and the gradient count are also updated by changing the minimum and maximum values. However, if there is already a lower minimum value or higher maximum value, the values are not changed. So, it is designed to update the values only whenever the time period or the slope direction is changed. The updatable area in the SGSA can be limited by defining the boundary of a specific area depending on applications.

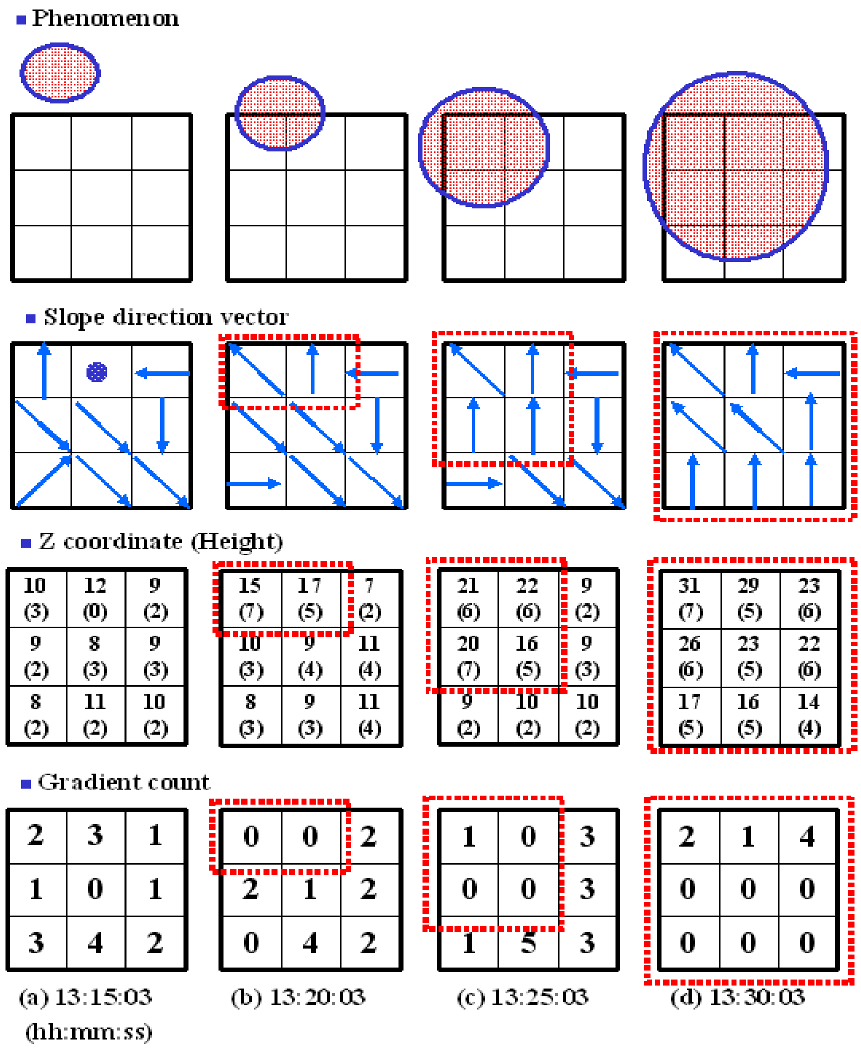


Fig. 4. The representation update of the SGSA

Figure 4 shows the changed representation of the SGSA depending on an observed phenomenon. For example, we can imagine a temperature change on a fire area. Users of environmental monitoring applications can understand and cope with the event by analyzing these changed values. When a geosensor network detects a phenomenon, it changes the attributes in SGSA such as slope direction, z coordinate, height, and gradient count. First, the z coordinate and the height are changed by updating the sensed data. Next, the slope direction is derived from the m and M subcells in each cell. Users can recognize the trend change of the observed data by searching the direction and the height, because the direction points are higher. Finally, the gradient count is updated depending on the direction change. The gradient count shows how long the value of the direction has been kept. Whenever the sensor data is updated, if the direction is not changed, it increases the count. In other words, a high count means the stable condition in the cell. If the direction is changed, it sets the count to 0. It means that the condition in the cell is changed. If the count is almost 0 in a long time, it indicates the fact that the condition of the cell, or the area, is frequently changed. This count is a useful factor for evaluating the dynamically changing condition of an area. Users can find some areas (cells), which shows recently changed data motion by searching low gradient count. These areas are utilized for finding a boundary of a phenomenon or recognizing areas which show frequently changed data motion. These attributes are used for tracking and analyzing a phenomenon.

5 Utilization in Sensor Data Query Processing

In sensor data monitoring applications, users are interested in posing higher-level queries to understand the information collected via sensors in a situation in a specific area. For example, a habitat monitoring system analyzes different conditions among groups of animals or an environmental monitoring application tracks the interesting phenomena such as red tide, a forest fire, and air pollution.

The abstract data types are used as the basis for answering a query, because of its summarized data representation on spatial area by describing a data gradient, min(), max(), and gradient count. An example of query processing on the abstracted data is shown in Figure 5. Assume, a user asks the query "find all wildfire fires in the area" at (a). Once the query is parsed, the monitoring system registers it as a continuous query. In order to answer the query, the system creates subqueries to analyze the related abstracted data within the observation area at (b). For example, two kinds of subqueries are issued to the temperature theme (abstract data) such as "find the area(s) detecting high temperature" and "find the area(s) detecting a high slope." The result of the first subquery can be a factor for finding a currently burning area or a burned area, because it is already hot. The result of the second subquery can be an indicator for finding areas, which could be burnt in near future by checking the high slope and the gradient count. The high slope and low gradient count mean that the data on the cell is rapidly changed. The cell detecting a high slop or a low gradient count could be a boundary of fire area. During this query process, if it requires a detail data on a cell, it can also search the sensor data by using the mapping table in Figure 3. At (c), the results can be combined after these two kinds of results are returned from several abstract data layers. Some information is extracted from the combined area depending on the predefined information about a fire area. It then can provide the users with a probability of an event at (d).

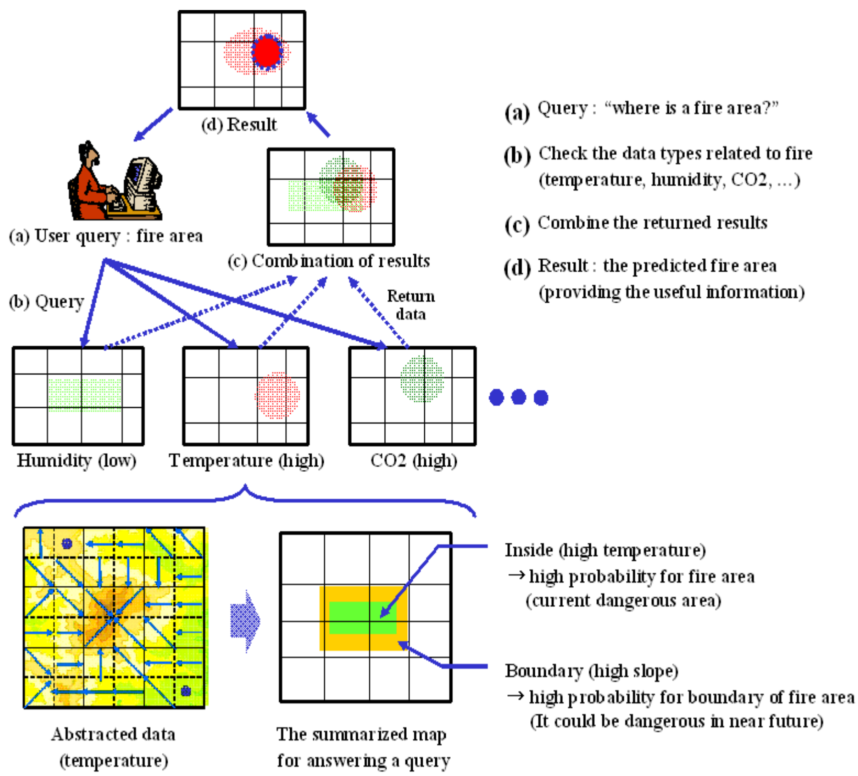


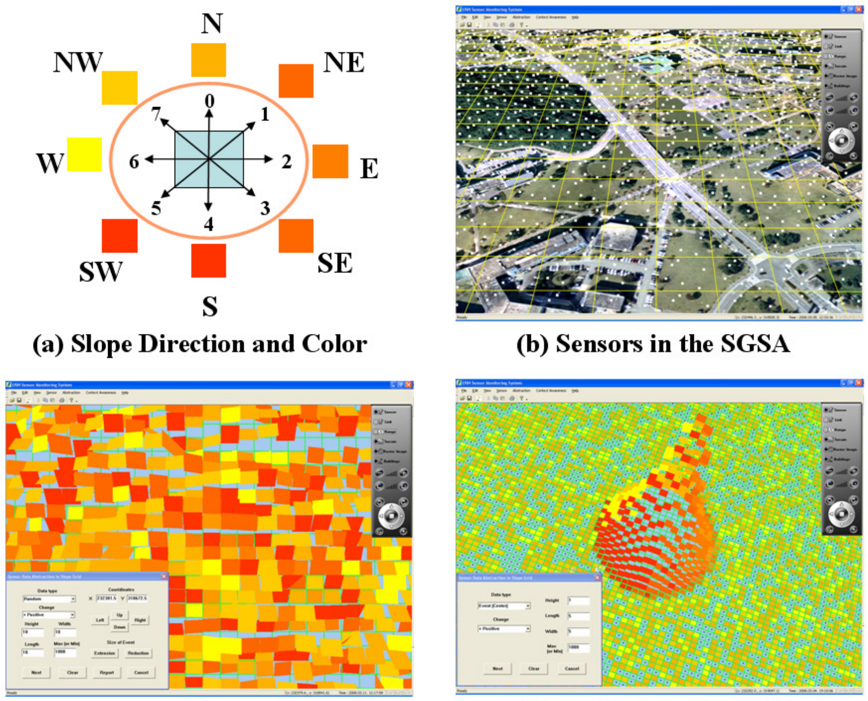
Fig. 5. The update of the abstracted data in the Grid

The results of the query will be continuously updated, since most of queries are continuous queries. Therefore, whenever sensor data is updated, the query result needs to be updated. These steps show the example of query processing with the abstracted model. It should predefine the relation among phenomenon (e.g. fire) and data types before the query processing. It could include some kinds of rules for combining several data types.

6 Implementation

The designed slope grid is implemented with 60,000 simulated static sensors and

10,000 cells as shown in (b) of Figure 6. It assumes that all of the observed data is transmitted to the central processing server. A data generator is used for creating simulated sensor data. Whenever the sensor data is updated by the generator, it updates the attributes of the cells in the grid such as max(), min(), and the slope direction. The slopes and gradient counts of cells are derived from the attribute values in each cell. Finally, it makes a summarized 3 dimensional map with SGSA for presenting the conditions of a remote place. This map is changed over time depending on the observed data change.



(a) Slope Direction and Color (b) Sensors in the SGSA

(c) Random Data in the SGSA (d) Event Data in the SGSA

Fig. 6. Data representation in SGSA

Two kinds of data types are used for testing the abstraction model such as random data and event data. When it generates all of random sensor data over time, the slopes are continuously changed like a wave at (c). Event data is used to detect an accident such as a fire, pollution, and red tide areas. The simulated event data is generated according to the parameters of sensor data generator such as height, width, and maximum (or minimum) value. The attributes of the cells are also updated according to the abstraction model. There are various kinds of data representation depending on the data change such as a variety of natural geographical features. In (d) event data of Figure 6 shows rapidly changed slopes in a specific area like a fire area. The maximum value and boundary of the event is used to process user queries as shown in Figure 5. In order to get the useful data in SGSA, some rules are used for extracting the specific cells having specific conditions. For example, to get the cells having the high value over mean(), it uses a rule that is min() in cells > mean(). To get the boundaries of a phenomenon, it also uses rules that are height > 20 and gradient count < 5, because high height and low gradient count mean rapidly changed condition in a cell.

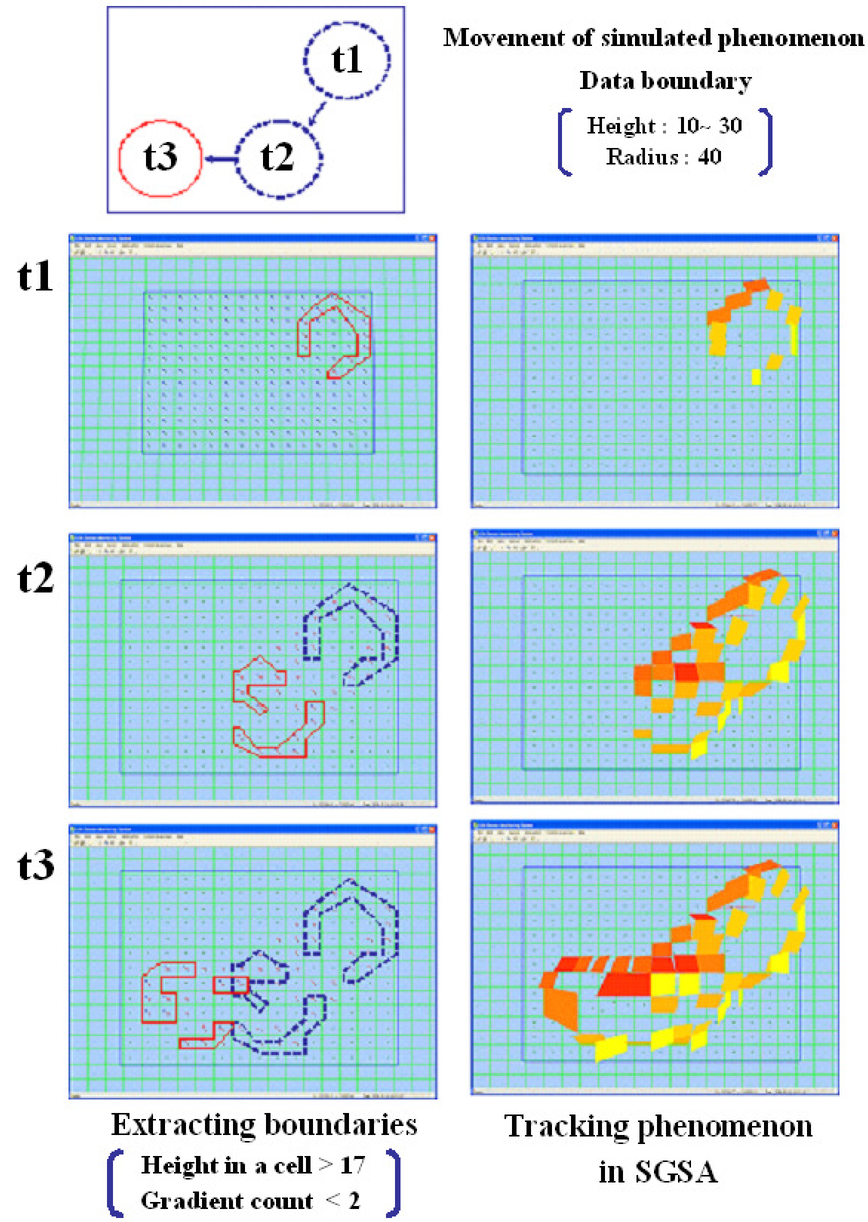


Fig. 7. Tracking the simulated phenomenon with SGSA

Figure 7 shows the tracking phenomenon which is simulated by data generator. The generator makes the simulated data with predefined conditions such as height (10~30), radius (40 m). The created circle having high values is moved over time according to the user defined route. When the circle is detected by sensors at t1, it makes a map of data with SGSA. The map is changed depending on the simulated circle's movement. We can extract the boundaries of the circle by simple rules for finding the cells which have the height over 17 and the gradient count less than 2. The red lines indicate the current detected boundaries on the left side pictures. The dot blue lines presented past boundaries. On the right side pictures, the boundaries are shown by SGSA. From these abstracted data, we can extract the information of the detected phenomenon such as the speed of movement, the boundaries, and the height of data values. It is useful to answer these questions such as "How fast does the phenomenon travel?", "What is the data gradient average in its boundaries?" To support a complex query, it is required to combine the information extracted from the several SGSAs for various data types.

7 Conclusions

Environmental monitoring applications deal with the data sets of different types as well as merge them to enhance our understanding of the circumstance. It is required to abstract the sensor data as a preprocessing step for interpreting the condition on a specific area. We designed the data representation abstraction for large-scale geosensor data in a central monitoring system. The slopes on cells represent the observed data in grid like a summarized contour map. It also stores a gradient count and a slope direction for finding the condition change in a wide area. It could be utilized as a data representation model in the environmental monitoring applications. In the future, this model can be extended by using the combination method among the heterogeneous abstracted data types for a sensor data fusion.

References

1. J. Elson, D. Estrin, "Sensor networks: a bridge to the physical world," Wireless Sensor Networks, pp. 3-20, 2004.
2. S. Nittel, A. Stefanidis, "GeoSensor Networks and Virtual GeoReality," GeoSensors Networks, p. 296, 2005.
3. K. Martinez, J. K. Hart, R. Ong, "Environmental Sensor Networks," IEEE Computer, Vol. 37, No. 8, pp. 50-56, 2004.
4. C. Y. Chong, S. P. Kumar, "Sensor Networks: Evolution, Opportunities, and Challenges," Proceedings of the IEEE, Vol. 91, No. 8, pp. 1247-1256, 2003.
5. Mario Lopez, Joseph K. Berry, "Use Surface Area for Realistic Calculations," GeoWorld, pp. 22-23, 2002.
6. N. Xu, "A Survey of Sensor Network Applications," IEEE Communications Magazine, Vol. 40, No.8, pp. 102-114, 2002.

7. A. R. Ilka, C. Gilberto, A. Renato, M. V. M. Antônio, “Data-Aware Clustering for Geosensor Networks Data Collection,” Anais XIII Simpósio Brasileiro de Sensoriamento Remoto, INPE, pp. 6059-6066, 2007.
8. A. Mainwaring, J. Polastre, R. Szewczyk, D. Culler, J. Anderson, “Wireless Sensor Networks for Habitat Monitoring,” ACM International Workshop on Wireless Sensor Networks and Applications, EUA, pp. 88-97, 2002.
9. H. B. Milburn, A. I. Makamua, F. I. Gonzalez. “Real-Time Tsunami Reporting from the Deep Ocean,” OCEANS’96, MTS/IEEE conference, Vol. 1, pp. 390-394, 1996.
10. ALERT, http://www.alertsystems.org.
11. Y. J. Jung, Y. K. Lee, D. G. Lee, M. Park, K. H. Ryu, H. C. Kim, K. O. Kim, “A Framework of In-situ Sensor Data Processing System for Context Awareness,” ICIC, pp. 124-129, 2006.
12. E. Biagioni, K. Bridges, “The application of remote sensor technology to assist the recovery of rare and endangered species,” In Special issue on Distributed Sensor Networks for the International Journal of High Performance Computing Applications, Vol. 16, No. 3, 2002.
13. J. K. Hart, J. Rose., “Approaches to the study of glacier bed deformation,” Quaternary International, Vol. 86, pp. 45-58, 2001.
14. B. Delaunay: Sur la sphère vide, Izvestia Akademii Nauk SSSR, “Otdelenie Matematicheskikh i Estestvennykh Nauk,” Vol. 7, pp. 793-800, 1934.
15. Generic Mapping tools, http://gmt.soest.hawaii.edu.
16. D. Chu, A. Deshpande, J. M. Hellerstein, W. Hong, “Approximate Data Collection in Sensor Networks using Probabilistic Models,” International Conference on Data Engineering, p. 48, 2006.
17. D. Tulone, S. Madden, “PAQ: Time Series Forecasting For Approximate Query Answering In Sensor Networks,” Lecture Notes in Computer Science 3868, pp. 21-37, 2006.
18. S. Madden, M. J. Franklin, J. M. Hellerstein, W. Hong, “Tag: a tiny aggregation service for ad-hoc sensor networks,” SIGOPS Operating Systems Review, p. 31-46, 2002.
19. D. Goldin, “Faster In-Network Evaluation of Spatial Aggregation in Sensor Networks,” Int’l IEEE Conference On Data Engineering, p.148, 2006.
20. P. Rigaux, M. Scholl, A. Voisard, “Spatial Databases with application to GIS,” Morgan Kaufmann Publishers, 2002.
21. W. Choi, “Adaptive cell-based index for moving objects,” Data & Knowledge Engineering archive, Vol. 48, Issue 1, pp. 75-101, 2004.