International Journal of Multimedia and Ubiquitous Engineering
Vol. 10, No. 2 (2015), pp. 129-136
http://dx.doi.org/10.14257/ijmue.2015.10.2.12

Multi-Event Identification with the Single Sensor

Sung-Gil Kim and Juhye Yook[1]

Department of Civil and Environmental Engineering, Kongju National University
[1] Department of Rehabilitation Technology, Korea Nazarene University
{sgkim@kongju.ac.kr}, [1]{jhyook@kornu.ac.kr}

Abstract

Sensors perform various activities to detect in a recent environment of Internet of things and ubiquitous sensor network. A single sensor detects multi-events in some cases while multi-sensors detect a single event and report in some other cases. The study presented a way of data clustering to classify events detected by a single sensor in circumstances where information is obtained by a form of data stream. Data was grouped in three types of events when clustering was applied for sensor data detected and reported. The following results also showed that clustering distinguished the events in the consequent time timeslots. The clusters displayed changing patterns with time. This result would contribute to the study field of context inference based on data stream.

Keywords: *Data Stream Mining, Internet of Things, Data Clustering*

1. Introduction

Many researchers address the themes on Internet of things nowadays. Internet of things is to combine various networks. It uses remote mobile communication network and local wireless network. Internet of things is to attach sensors on numerous objects, to communicate among them, and to send information acquired by them. Many sensors often sense targets for detection if we look at the circumstances of Internet of things closely. That is, many sensors detect a same event and report. This type of detecting activities is common in wireless sensor network. On the other hand, we could consider the condition that a single sensor detects many events. Various types and ways of sensors could be distributed when a sensor network is composed. Many events could occur around a particular sensor, or there would be a case that many sensors could not be attached. ETR and EMR are assistive mobility equipment for individuals with visual disabilities. Those persons prefer ETR that has the same weight, size, and thickness as an existing white cane. They rarely use a smart white cane or other types of assistive mobility equipment different in shapes from, heavier than, or larger than a traditional white cane. Therefore, sensors and materials should be minimized to attach in a smart white cane. Also, we should consider the cases that these kinds of lightweight devices with cheap sensors and simple or poor equipment to detect complex circumstances and acquire information. Each event needs to be distinguished among the acquired data when a single sensor detects multi-events, which is different from when multi-sensors detect a same event. Each event is to distinguish and analyze.

A way of distinguishing each event is necessary when multi-events are detected and reported at the same time by the least use of a sensor in poor condition. The study suggests a way of identifying each event when many events are detected continually by a single sensor.

[1] Corresponding Author

ISSN: 1975-0080 IJMUE

Distinguishing events mixed in sensor data is valuable to recognize or infer circumstances. To do so, data clustering is used. This is conducted for rapid identification of data detected by a single sensor in a period of time with limited resources and without prior information.

The study is consisted as the following. Related research is reviewed in chapter 2, and a way of distinguishing multi-events by a single sensor is suggested in chapter 3. An experiment for the suggested theory is conducted, and the results are evaluated in chapter 4. A conclusion of the study is drawn in chapter 5.

2. Related Research

Studies K-means clustering is used the most among the ways of clustering. K-means algorithm is to select the k number of central point randomly and to assign items closest to the points to clusters. Then, the central points are moved to average positions of all assigned nodes, and they are reassigned. Selecting proper input value of K at the beginning is essential while the fast performance is advantageous [1, 2, 5, 6].

The basic algorithm of K-means clustering is as follows [3, 6, 7].

Step 1. Decide the k number of clusters, and assign one initial value or central point of clusters to each cluster.

1. $(S_1, S_2, \dots, S_K) \leftarrow Select\ Random\ Seeds\ \ (\{x_1, \dots, x_N\}, K)$
2. $for \quad k \leftarrow 1\ to\ K$
3. $do \quad \mu_k \leftarrow S_k$

Step 2. Assign all data to the nearest central points of clusters, using Euclid distance.

4. *While stopping criterion has not been met*
5. $for \quad k \leftarrow 1\ to\ K$
6. $do \quad \omega_k \leftarrow \{\ \}$
7. $for \quad n \leftarrow 1\ to\ N$
8. $do \quad j \leftarrow \arg min_i\ |\mu_i - x_n|$
9. $\omega_i \leftarrow \omega_i \cup \{x_n\}$

Step 3. Calculate new central points of clusters to minimize the distance from assigned data in each cluster.

10. $for \quad k \leftarrow 1\ to\ K$
11. $do\ \vec{\mu}_k \leftarrow \frac{1}{|\omega_k|} \sum_{\vec{x} \in \omega_k} \vec{x}$

Step 4. Repeat Step 2 and 3 until there is little change in the centers of clusters.

12. return $\{\mu_1, \dots, \mu_k\}$

K-means clustering is finished under the following conditions.

(1) Repeat the number of times defined in advance. This limits the performance time of clustering algorithm, but the quality of clustering could decrease for lacking the number of repetition.

(2) Repeat until clusters with vectors do not change. The quality of clustering is high in this condition unless it goes into local minimum. However, its weakness is long performance time.

(3) Repeat until the centers do not change.

(4)Repeat until RSS reaches a critical value or less. The quality of clustering is very good with the completion by this criterion. Actually, this method is combined to finish the performance with setting the number of repetition times.

The advantages of the K-means clustering are easy to perform, linear in time complexity, and able to classify in detail when clusters form a ball.

The problems of the K-means are as the following.

(1) It is sensitive to numerical outliers because the values that are extremely big or small could distort data distribution.

(2) It could get entirely different results by early conditions.

(3)Data discrimination is weak because all information of data is calculated with the same weighted values.

(4)There could be data that do not belong to each circle because the clusters form shapes of circles by using distance values for clustering units. The results would be wrong to measure distances between individual points only with Euclid distance when the shapes of clusters are oval.

(5)The results would be bad if the k number of clusters is inappropriate.

(6) It is sensitive for an outlier and noise.

(7)A lot of resources are consumed to calculate distance matrix equations if the number of events increases.

(8) It is used only with numerical data.

(9) It does not classify clusters well with different size and density.

3. Distinguishing Multi-events from a Single Sensor

There are many constraint conditions to use various or powerful sensors for a smart white cane for people with visual disabilities. They usually expect the same weight and volume of a smart white cane as a conventional white cane. The study is necessary to infer circumstances where the sensors with low capacity and cost are used only [3, 4, 7, 8].

A consideration to use sensors in this condition is that information of many events by a single sensor is mixed. Each event should be discriminated when event information detected by a single sensor includes many events. Then, it is analyzed how each event changes with time. A clue to infer circumstances should be acquired even with poor sensor equipment and analysis tools by distinguishing events mixed in data from a sensor when there are sensor data continually obtained but not with enough resources to analyze. Data clustering is used to solve this matter. Data clustering is valuable for the study because it is appropriate to use in circumstances depending on data analysis without prior information. It is also proper to distinguish many multi-events in single sensor data for dividing into many subsets from a set of data.

A way of distinguishing multi-events from a single sensor is presented next.

A case that data stream $DS = \langle \mathrm{D}_1, \mathrm{D}_2, \mathrm{D}_3, \ldots \mathrm{D}_\mathrm{t} \rangle$is coming by time $TS = \langle \mathrm{T}_1, \mathrm{T}_2, \mathrm{T}_3, \ldots \mathrm{T}_\mathrm{t} \rangle$ is hypothesized. Clusters $\mathrm{C}_1, \mathrm{C}_2, \ldots \mathrm{C}_\mathrm{k}$ are searched by K-means clustering. The process is as the following.

1. An arbitrary sample of the k number obtained at the first *l* seconds in each time interval is selected, and $\mu_1, \mu_2, \ldots \mu_k$ are the centers of early *k* clusters.

1-1. The first *l* seconds are determined by the following formula. Suppose that the size of T_i is *m* seconds in a timeslot, the number of data obtained per second is *n*, and then the number of the clusters is *k*. This is the case that data with the k number of types was obtained randomly. To determine the case, *l* seconds selected as a sample are enough with the minimum value by the following formula.

$$\frac{(nl-k)!k!}{(\mathrm{nl}-1)!} \geq 5$$

l seconds are minimum value by the formula presented above when a sample rate $k/\binom{nl}{k}$, a rate to select the appropriate k number, is considered for selecting the *nl* number of a sample. Also, if the range of the k number value selected is $[\mathrm{a}, \mathrm{b}]$, a center with minimum

error could be selected after a value in the range selected randomly is defined as the center of clusters.

2. The least $\mu_j(1 \le j \le k)$ is classified as *j*th cluster after calculating $min(\|a-\mu\|^2)$ for data $\alpha \in D_i$ entered at T_i $(i \in [1,t])$.

3. Next, μ_j is renewed with the average $\frac{1}{n(C_j)}\sum_{\alpha \in C_i}\alpha$ of a sample assigned into *j*th cluster.

4. Repeat until an average $\mu_1, \mu_2, \ldots \mu_k$ minimizing, $\frac{1}{n(C_j)}\sum_{\alpha \in C_i}\|\mu_i-\alpha\|^2$ a variation of , C_i is resulted.

5. Repeat the process until it is the same as the value μ_j of a prior loop. Repeat 10 times to the maximum.

6. Then, calculate the following. $\frac{1}{n(D_i)}\sum_{\alpha \in D_i}\|\mu_i-\alpha\|^2$ is the total number of variations from each cluster. Data stream is simplified by understanding the trends and tendencies of the total numbers of the variations. Data stream is also understood by getting the value of $\sum_{i=1}^{k}\|\bar{\mu_i}-\alpha\|^2$ with $\bar{\mu}=\frac{1}{k}\sum_{i=1}^{k}\mu_i$.

The study applied K-means clustering used the most out of data clustering to distinguish multi-events mixed in a set of data. Data clustering was processed for data acquired in time intervals at each timeslot because data to analyze is obtained continually. An experiment and evaluation for the theory are performed in the next chapter.

4. An Experiment and Evaluation

Experiment Process

Continual sounds from three types of sound sources were to detect and report to a host. A time interval to detect was 0.1 second, and a timeslot of time series data was 30 seconds. Data in each timeslot included three types of sound events. The results were as follows when they were distinguished using data clustering.

Results

Figure 1 shows the distribution of sound data acquired through a sensor.

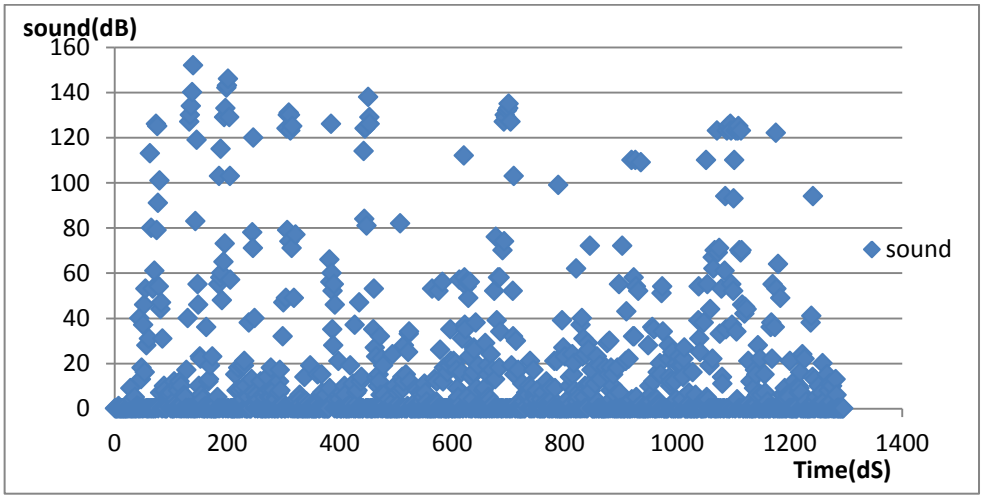

Figure 1. Sound Data Through a Sensor

The horizontal axis represents time, and the vertical axis does sound intensity. Sound data acquired by a sensor had event information with three types of mixed events. Data clustering was conducted for the data divided by 30 seconds of each timeslot. The results are showing from Figure 2 to Figure 6.

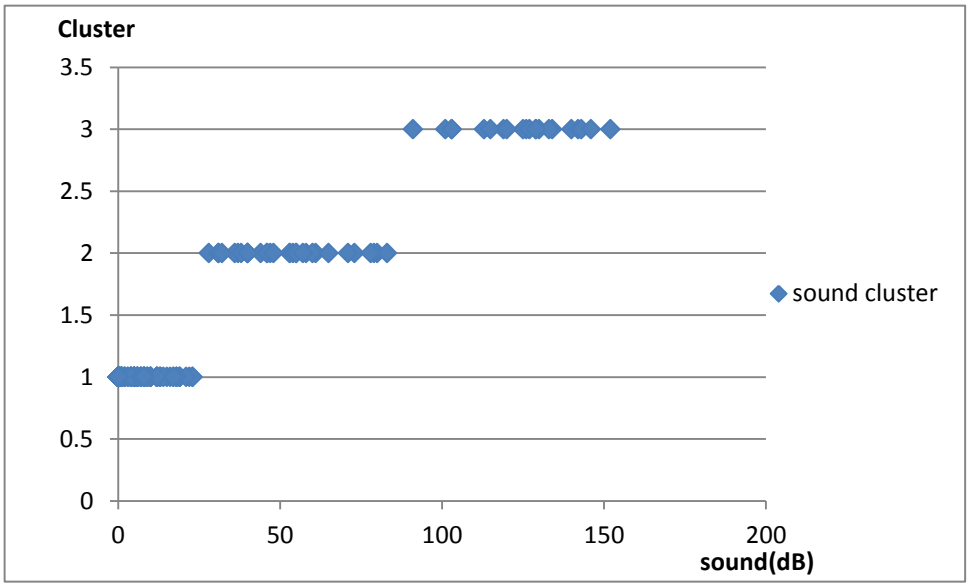

Figure 2. The Result of Clustering for Sound Data from 0.1 to 30 Seconds

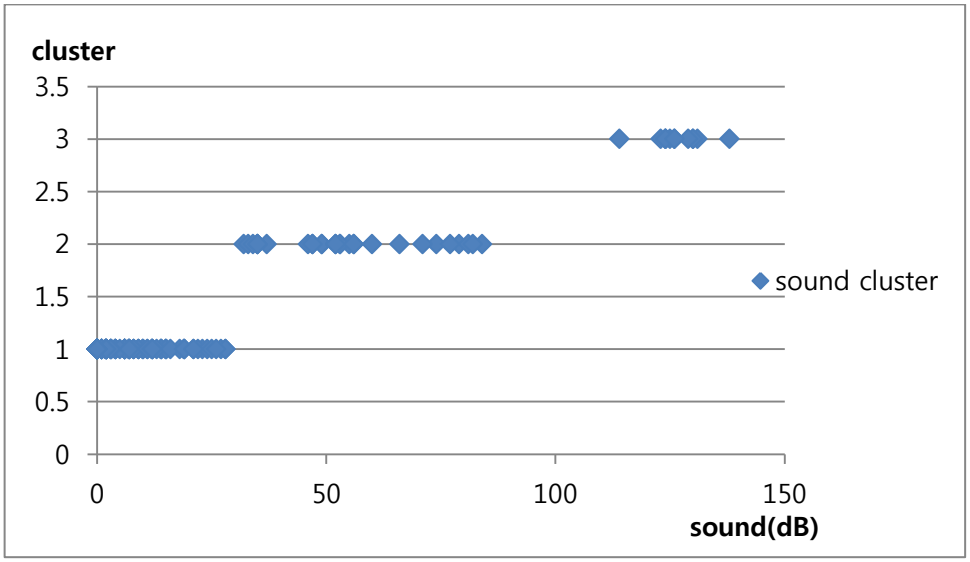

Figure 3. The Result of Clustering for Sound Data between 30.1 and 60 Seconds

The data were distinguished as cluster 1, 2, and 3 as shown in Figure 2 when clustering was carried out for sound data obtained in the first timeslot between 0.1 and 30 seconds.

The data were definitely distinguished from the result of clustering for sound data for the second timeslot between 30.1 and 60 seconds. Only, the pattern of each cluster changes with time.

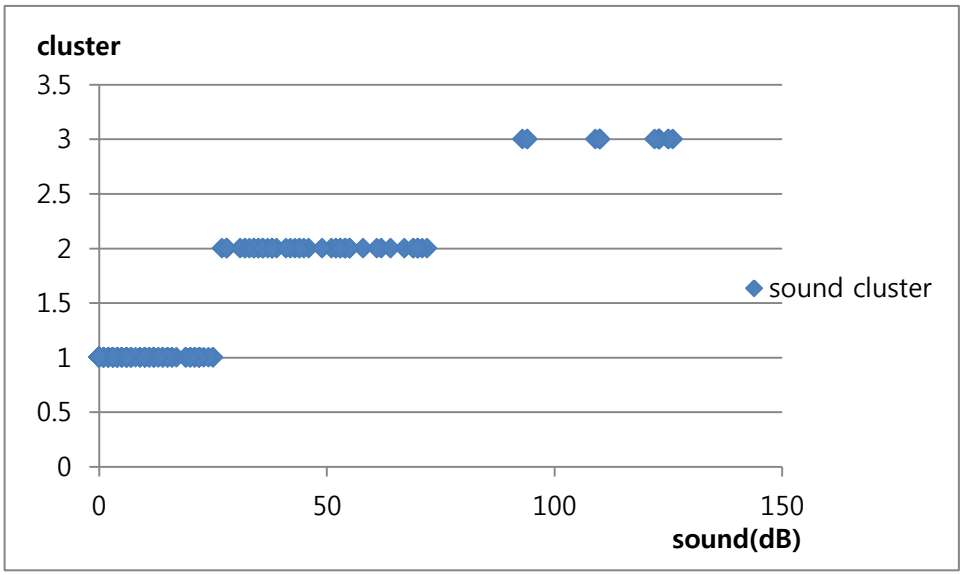

Figure 4. The Result of Clustering for Sound Data between 60.1 and 90 Seconds

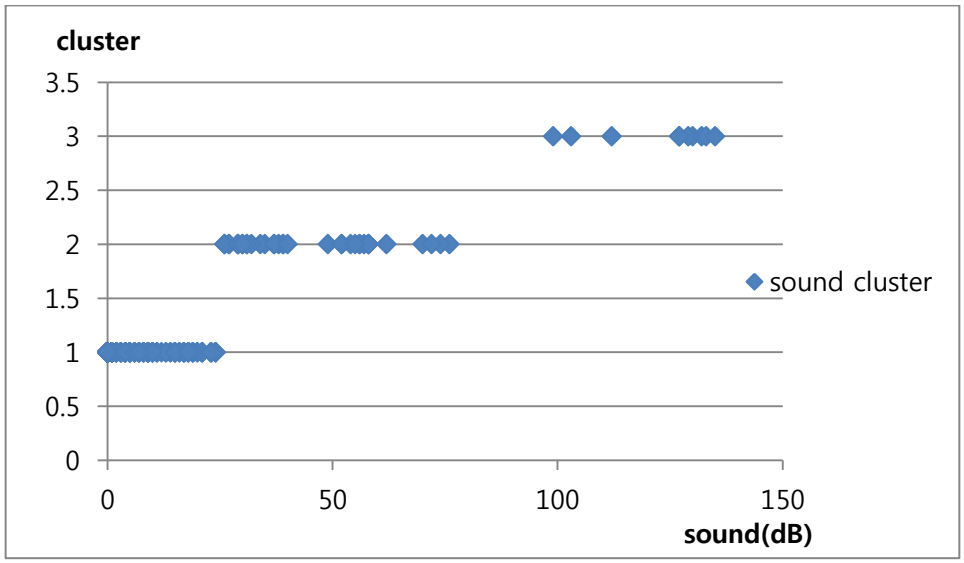

Figure 5. The Result of Clustering for Sound Data between 90.1 and 127.7 Seconds

Each cluster was grouped clearly in both Figure 4 and 5, and the sizes and forms of the clusters change with time. Each event in data could be distinguished for continual measured values of the forms of data stream by data clustering. Event clusters by the result of clustering in each timeslot presented changes with time. The purpose was to differentiate event data mixed in data. Circumstances detected by a sensor could be inferred or recognized through

continual analysis for identifying the changing patterns of each cluster. Distinguishing the multi-events is an important data process prior to employing different sensors and data fusion process with the detected data resulting in more precise data acquisition.

Evaluation and discussion

A single sensor detects multi-events in some cases while multi-sensors detect and report a single event in other cases. It is necessary to distinguish each event from a single sensor. Data clustering is a way of data mining for database gathered and stored, and the study used it for data stream to discriminate data. The study shows that this way is effective to distinguish each different event mixed in the obtained data. This result would be based to apply to and to develop for a smart white cane for individuals with visual disabilities. This would be beneficial to detect and analyze multi-events by a simple sensor function in circumstances with many limits in sizes, weights, and so on.

5. Conclusion

Sensors perform various activities to detect in a recent environment of Internet of things and ubiquitous sensor network. A single sensor detects multi-events in some cases while multi-sensors detect a single event and report in some other cases. The study presented a way of data clustering to distinguish events detected by a single sensor in circumstances where information is obtained by a form of data stream. Data was grouped in three types of events when clustering was applied for sensor data detected and reported in every 0.1 second for each data acquisition timeslot of 30 seconds. The following results also showed that clustering distinguished the events in the consequent time timeslots. The clusters displayed changing patterns with time. This result would contribute to the study field of context inference based on data stream.

Acknowledgements

This research was partially supported by the Korea Nazarene University Research Grants 2014.

References

[1] T.Kanungo, N. S. Netanyahu, C. D. Piatko, R. Silverman and A. Y. Wu, ”An Efficient k-Means Clustering Algorithm, Analysis and Implementation”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, **(2002)** July, pp. 881-892.
[2] M. Dipti and T. Patel, “K-means based data stream clustering algorithm extended with no. of cluster estimation method”, International Journal of Advance Engineering and Research Development (IJAERD), vol. 1, Issue 6, **(2014)** June.
[3] S. Vijayarani and P. Jothi, ”An Efficient Clustering Algorithm for Outlier Detection in Data Streams”, International Journal of Advanced Research in Computer and Communication Engineering, vol. 2, Issue 9, **(2013)** September, pp. 3657-3665.
[4] H. M. Koupaie, K. Lumpur, S. Ibrahim, K. Lumpur and J. Hosseinkhani, “Outlier Detection in Stream Data by Clustering Method”, International Journal of Advanced Computer Science and Information Technology (IJACSIT), vol. 2, no. 3, **(2013),** pp. 25-34.
[5] A. K. Jain, “Data Clustering, 50 Years Beyond K-Means”, To appear in Pattern Recognition Letters, vol. 31, Issue 8, **(2010),** pp. 651-666.
[6] M. Khalilian, N. Mustapha, N. Suliman and A. Mamat, “A Novel K-Means Based Clustering Algorithm for High Dimensional Data Sets”, Proceedings of the International Multi-Conference of Engineers and Computer Scientists, vol. 1, **(2010),** pp. 503-507.
[7] M. R. Ackermann, M. Märtens, C. Raupach, K. Swierkot, C. Lammersen and C. Sohler, “StreamKM++: A Clustering Algorithm for Data Streams”, ACM Journal of Experimental Algorithmics, vol. 17, **(2012)** January.

[8] F. N. Eduardo, A. F. Antonio and C. F. Alejandro, "Information fusion for Wireless Sensor Networks: Methods, Models, and Classifications", ACM Computing Surveys, vol. 39, no. 3, **(2007)** August.
[9] X. Zhu, C. Zhou, W. Guo, D. Chen and K. Liu, "An Optimization Technique for Spatial Compound Joins Based on a Topological Relationship Query and Buffering Analysis in DSDBs with Partitioning Fragmentation", International Journal of Database Theory and Application, vol. 5, no. 4, **(2012)** December, pp. 45-60.
[10] M. Prabukumar and J. C. Clement, "Compressed Domain Contrast and Brightness Improvement Algorithm for Colour Image through Contrast Measuring and Mapping of DWT Coefficients", International Journal of Multimedia and Ubiquitous Engineering, vol. 8, no. 1, **(2013)** January, pp. 55-70.
[11] V. Budyal and S. S. Manvi, "Intelligent Agent Based Delay Aware QoS Unicast Routing in Mobile Ad hoc Networks", International Journal of Multimedia and Ubiquitous Engineering, vol. 8, no. 1, **(2013)** January, pp. 11-28.
[12] J. Kim and Y. Y. Cho, "Efficient Character Segmentation using Adaptive Binarization and Connected Components Analysis in Ubiquitous Computing Environments", International Journal of Multimedia and Ubiquitous Engineering, vol. 8, no. 2, **(2013)** March, pp. 89-100.

Authors

Sung-Gil Kim, he is a professor at Kongju National University. He received the B.S. degree in Architectural Engineering from Yeonsei University in 1988, and the M.S. in Urban Planning from Yeonsei University in 1990. He received Dr.-Ing in Housing and Transportation from Hamburg Technical University in Germany in 2013. His research interests are included in data fusion in Ubiquitous City.

Juhye Yook, she received the M.Ed. in 1995 and the Ph.D. in 2000 in Special Education from the University of South Carolina in the U.S. She worked for the Korea Employment Promotion Agency for the Disabled (2000~2004) before working at the Korea Nazarene University up until now. Her research interests are computer/information access and assistive technology.

