Anomaly Detection and Life Pattern Estimation for the Elderly Based on Categorization of Accumulated Data

Taketoshi Mori*, Takahito Ishino†, Hiroshi Noguchi**, Masamichi Shimosaka‡ and Tomomasa Sato§

*tmori@ics.t.u-tokyo.ac.jp
†ishino@ics.t.u-tokyo.ac.jp
**noguchi@ics.t.u-tokyo.ac.jp
‡simosaka@ics.t.u-tokyo.ac.jp
§tomomasasato@jcom.home.ne.jp

Abstract. We propose a life pattern estimation method and an anomaly detection method for elderly people living alone. In our observation system for such people, we deploy some pyroelectric sensors into the house and measure the person's activities all the time in order to grasp the person's life pattern. The data are transferred successively to the operation center and displayed to the nurses in the center in a precise way. Then, the nurses decide whether the data is the anomaly or not. In the system, the people whose features in their life resemble each other are categorized as the same group. Anomalies occurred in the past are shared in the group and utilized in the anomaly detection algorithm. This algorithm is based on "anomaly score." The "anomaly score" is figured out by utilizing the activeness of the person. This activeness is approximately proportional to the frequency of the sensor response in a minute. The "anomaly score" is calculated from the difference between the activeness in the present and the past one averaged in the long term. Thus, the score is positive if the activeness in the present is higher than the average in the past, and the score is negative if the value in the present is lower than the average. If the score exceeds a certain threshold, it means that an anomaly event occurs. Moreover, we developed an activity estimation algorithm. This algorithm estimates the residents' basic activities such as uprising, outing, and so on. The estimation is shown to the nurses with the "anomaly score" of the residents. The nurses can understand the residents' health conditions by combining these two information.

Keywords: Anomaly detection, Monitoring the elderly, Pyroelectric sensors
PACS: 89.20.Ff

INTRODUCTION

Recently, the number of elder people living alone is increasing because of population aging. Therefore, systems that monitor the lives of the elderly and detect anomaly events become an active research area [1–7].

In our research, we assume the system using some pyroelectric sensors that can detect the person's motion. In the system, as shown in figure 1, we deploy a few pyroelectric sensors into the elder people's house. The sensor data is successively obtained all the time. The anomalies we want to detect are elder people's depression, dementia, illness, and so on. The data of each sensor are gathered wirelessly in a data transmitter and are sent to the call center through the Internet or telephone lines. In the center there are some nurses and they receive the result of the anomaly detection algorithms. When anomaly candidate data is discovered, it is informed to one of the nurses with the reason why the data is regarded as the anomaly. If the nurse determines that the candidate is truly an anomaly data, she/he recommend that the elderly should be hospitalized, for example. As for the system, the researches of anomaly detection method to detect rare behaviors or change of life pattern per person have been conducted[8][9]. Our method extends a probabilistic concept of them. The method accumulates the sensor data during long span and analyzes the data to categorize the people whose features in the data resemble each other into the same group. Anomalies occurred in the past are shared in the group and utilized in our anomaly detection algorithm. This is because the anomalies hardly occur and are difficult to discover in one person's data, though it is desirable that the system can adapt to each person's unique life pattern. In addition, we estimate activities which intuitively describe life patterns of people such as uprising, outing, and going to bed. This estimation is displayed with the result of the anomaly detection together, which provides the nurses with precise understanding of the change in the life pattern of the elderly.

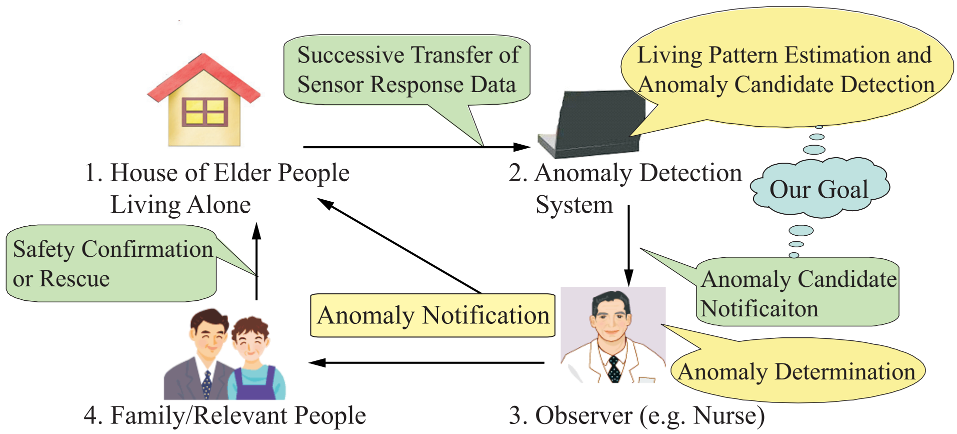

FIGURE 1. The elderly observation system we assume.

The algorithms which cluster and visualize the patterns in human life by applying HMMs or GMM to massive daily sensor log analysis[10] is one of the related researches. This research tried to interpret human behavior modeling result intuitively, and this research resembles our research in this way. The research extracted human behavioral patterns by data-driven approaches based on statistical methods. However, in actual operation of the health monitoring system, the use of common knowledge about general human lives is also an effective approach. For example, in the data of human daily life, we can easily anticipate that there are some regularities attributed to the factor such as season or week. Concretely, the life on weekdays and the life on the weekend should be different in most people. In this research, we utilize these kinds of intuitive knowledge.

GROUPING THE SIMILAR LIFE PATTERNS

It is important to categorize their life pattern with some kind of criterion. Ideally, the system should adapt each of them individually, because the lifestyle varies depending on a person and family. However, it is difficult when the system is actually in operation, as previously mentioned. This is why we adopt the method grouping the similar life patterns. Hereinafter, we discuss some feature values as useful criteria to grouping the residents.

Generally, human lives display some particular regularities, influenced by seasons, weeks, customs, and so on. For example, a person who takes culture lesson on every Tuesday afternoon are usually not at home in Tuesday afternoon. We assume that people who resemble in their regularities are likely to live the similar lives. In this research, we propose using four feature values named "mean total response per day," "week regularity," "season cycle regularity," and "room using pattern," which are thought to illustrate residents' life pattern vividly, and categorize each of them respectively(figure 2). The residents with the same result in each categorization are regarded as the members of the same group whose life pattern resemble each other. "Mean total response per day," "week regularity" and "season cycle regularity" are categorized in two patterns, and "room using pattern" is three. As the result, each resident belongs to one of the $24 (= 2^3 * 3)$ life patterns.

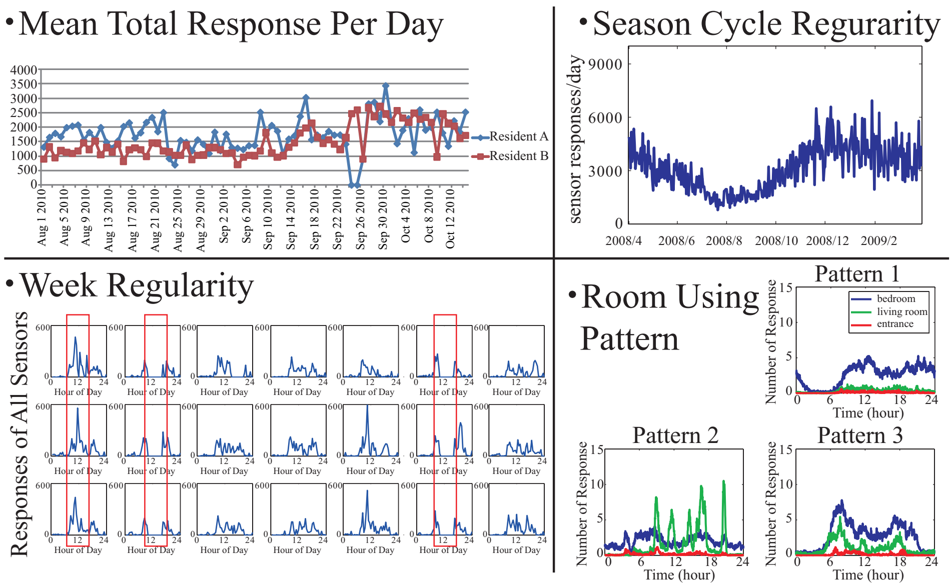

FIGURE 2. The four feature values we propose.

The pyroelectric sensor responds when a heat sources such as humans pass through its field of view. The count increases for each passing of the heat source. The sensors in this research are designed to return an integer value from 0 to 15 per minute for saving energy. We analyze such data from a few pyroelectric sensors per one house. For example, there are three sensors in one's house. The three sensors are located at bedroom, living room and entrance.

"Mean total response per day" represents general briskness of the resident. This value is obtained by summing the total responses of the sensors attached in one house and average it for long days. "Week regularity" is dependency on a weekly activity calculated from the variability of the sensor response among a day of week. This concept is based on a presumption that residents with similar lifestyle display similar characteristic of the sensor response like figure 3. Concretely, this feature value is a standard deviation of the average number of sensor responses in the same day of week. This value is normalized by "mean total response per day" of each resident. According to the presumption above, the smaller this value is, the more regular life the resident lives.

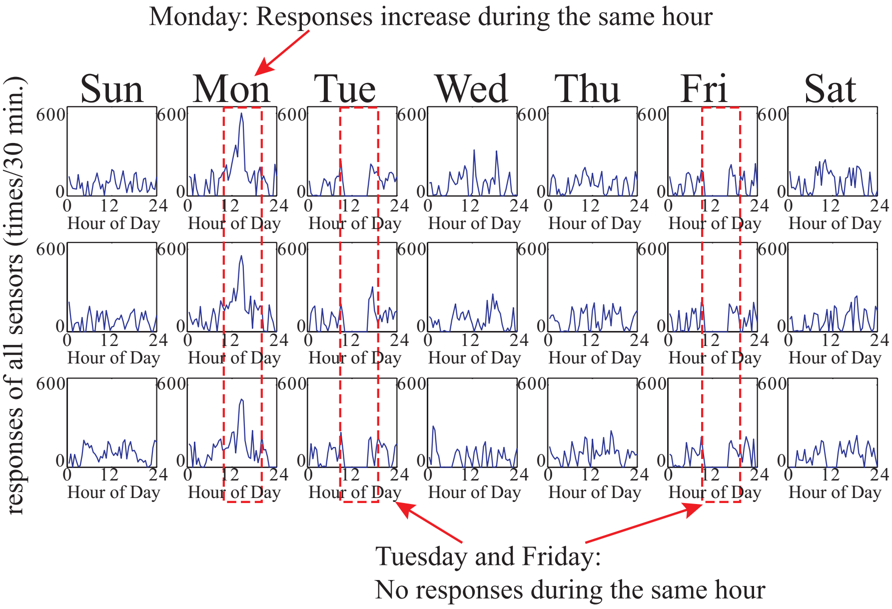

FIGURE 3. Example of "week regularity."

In most of the residents we examined, there was a bent that the number of the sensor response per a day indicates a minimal value in summer, and indicates a maximal value in winter, as depicted in figure 4. "Season cycle regularity" measures this characteristic by calculating standard deviation of the average number of sensor responses for each month. This value is normalized by "mean total response per day" of each resident. These three feature values are classified by k-means clustering algorithm. For each feature, the number of the cluster is 2.

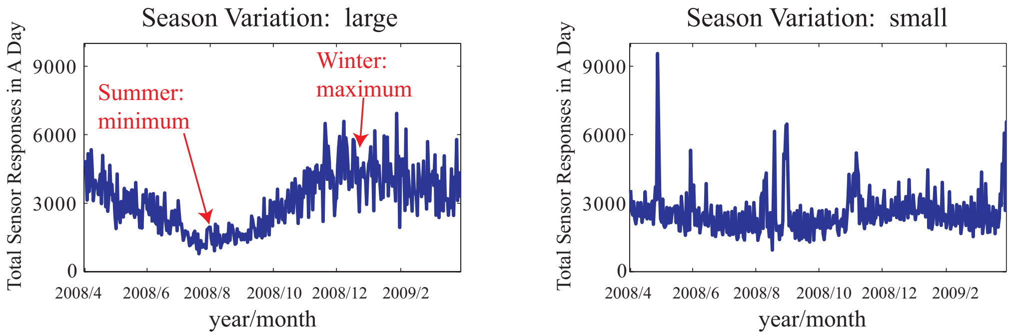

FIGURE 4. Example of "season cycle regularity."

"Room using pattern" is a pattern that is associated with the room layout of each house. This classification is based on the ratio of the sensor response for each sensor attached in a house, and "room using pattern" is divided in three patterns according to this ratio. The three patterns are defined as follows:

- Pattern 1: total sensor responses in a day are occupied by one particular sensor.
- Pattern 2: there are two dominant sensors which respond alternately depending on hours in a day.
- Pattern 3: two dominant sensors respond similarly in most part of the day.

The residents with the same categorization for each feature value are grouped in the same life pattern. Based on this method, we conducted an experiment to categorize several residents into 24 patterns. We used the data of 15 houses, in which there are three sensors in the bedroom, living room and entrance. The residents of all houses are elder people living alone. Figure 5 displays the sensor responses of 6 residents of all 15 residents. In the figure, the number of the sensor response per minute of each sensor averaged for 365 days are displayed. The 6 residents are randomly selected in all residents. Before the experiment, we manually categorized the residents' pattern by looking the shape of the graphs such as figure 5. Then, we applied the categorization algorithm to the data and compared the result with the label we manually categorized. The experiment is conducted with four different learning (averaging the data) span, 3 months, 6 months, 9 months, and one year. The algorithm is originally designed for averaging annual data, especially in "season cycle regularity." We experienced the data of shorter span because we thought that the feature value may represent the difference even in such short span. Besides, shortening the learning span shortens the computation time.

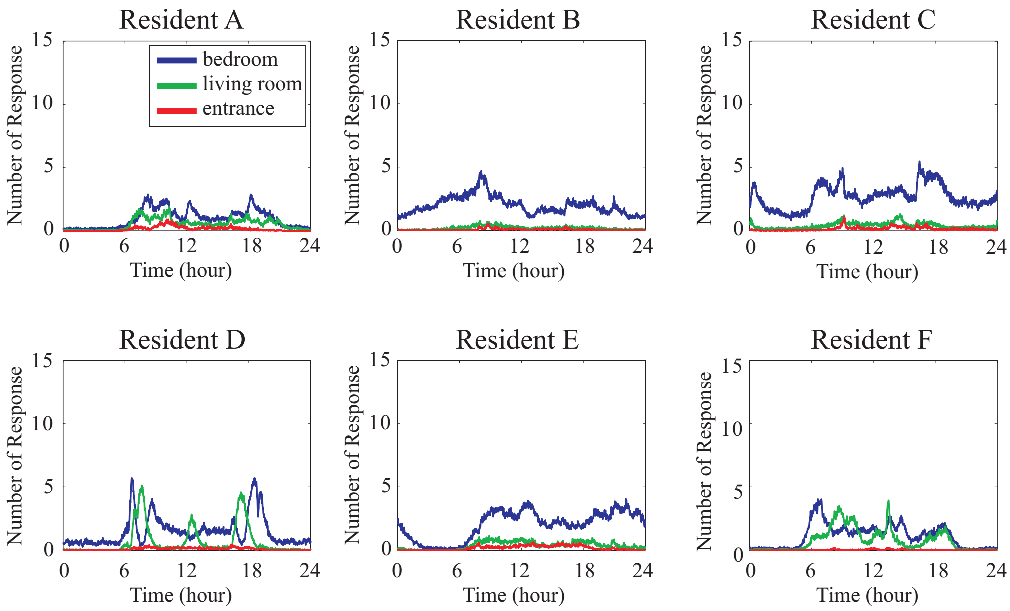

FIGURE 5. Example of the graph of sensor response.

As the result, if the residents are categorized in the same group, there are similar characteristics in the shape of their graph. For example, resident D and F are categorized in the same group. In addition, the result when the span is set to one year was the most valid categorization compared with our intuitive categorization. The reason for this may be that, in four feature values, "mean total response per day" and "season cycle regularity" may be the value varying on annual cycle, and the independence of these two features are high when the learning span is 365 days. We propose these categorization algorithms as one of the possible criteria to describe the residents' life pattern, and decided to use this algorithms in the anomaly detection algorithms in the following.

ACTIVITY ESTIMATION METHOD

The estimation of the simple activities in the residents' life is efficient help of understanding their life patterns. Concretely, we implemented an algorithm that estimates the activity of the resident into three states called "active," "inactive at home," and "outing." This algorithm also detects uprising and going to bed because these activities are important to grasp the resident's life pattern. In addition, outing routines of residents are estimated because some routines such as "outing to a further education school on Tuesday afternoon" can characterize the resident's lifestyle. We label this kind of outing as "habitual outing." An explanation of this algorithm is in the following.

Basically, the state that any sensor in the house responds is called "active." If none of the sensor responds for certain minutes (e.g. 10 minutes), the algorithm counts the response of the sensor at entrance for certain minutes (e.g. 10 minutes) before the time when sensor responses has paused. If the number of responses exceeds a certain threshold value (e.g. 5 times), this sequence of the response are regarded as preparation of outing such as putting on shoes. We determine that this pause of the response is caused by "outing." If not, we determine that the resident is "inactive at home" for some reason such as sleeping.

We conduct accuracy estimation experiments for each algorithm. We compared the result of the algorithm with the correct data obtained with an another method. The experiments resulted in 90.5 percent accuracy in uprising detection, 95.5 percent precision in "outing" detection, and 97.1 percent precision in "habitual outing" detection, which is high enough to grasp the residents'life patterns.

ANOMALY DETECTION METHOD

We define an anomaly as a situation in which the number of the sensor response greatly increase or decrease compared with normal time. To determine whether a situation is anomaly or not, we calculate the increase and decrease of the number of response from the usual time for each time of day. Therefore, if the increase or decrease exceeds a certain threshold, an anomaly occurred at the time. The threshold is calculated from the increase or decrease on the day some kind of anomaly occurred in the past are calculated. The days in which the anomaly occurred are registered in the anomaly database in advance.

Figure 6 displays the concept of anomaly detection. As the figure illustrates, the anomaly which the response increases is defined as "plus anomaly," and the one the response decreases is defined as "minus anomaly." The example of the "plus anomaly" is a case in which the response in the night is active, though the resident is usually in bed in this hour. On the other hand, the example of the "minus anomaly" is a sudden outing.

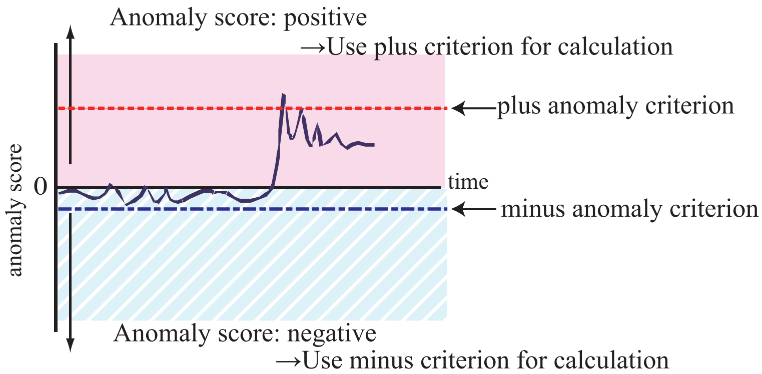

FIGURE 6. The outline of calculating anomaly.

A concrete explanation of this method is as follows:

1. For the day of examination, calculate the value named "thirty-minute difference of response," the number of sensor response for the past 30 minutes from the present time for each minute.
2. Calculate the same value for the past 15 weeks (approximately 3 or 4 months) and average the value for each day of week.
3. Figure out "one-minute anomaly score" by comparing the two values mentioned above by using the equation below.

$$e := \frac{D - D_M}{\alpha + D_M}$$

   Where D is a "thirty-minute difference of response" of each minute and D_M is a "thirty-minute difference of response" averaged for each day of week, for past 15 weeks. α is a constant number and set to 1.0 to prevent e from diverging to infinity. In addition, the one-minute scores are accumulated for the past 10 minutes into "anomaly score."
4. Calculate "anomaly criterion" as the criterion whether the "anomaly score" is an anomaly or not. The criterion is derived from the "anomaly score" on the day an actual anomaly occurred, based on the anomaly data pre-registered in the anomaly database. The criterion is set to the maximum "anomaly score" in the day, if the anomaly occurred in the day is "plus anomaly," and is set to the minimum if the anomaly is "minus anomaly."
5. Calculate the proportion of the "anomaly score" to each "anomaly criterion." The proportion is called "anomaly ratio." If the ratio exceeds a certain threshold, we determine that an anomaly of some type corresponding the "anomaly criterion."

The parameter for calculating D_M was set to 15 weeks because the span was close to the length of a season. The rest of the parameters in this algorithm were determined empirically by inspecting a lot of data.

We conducted an experiment of this algorithm. We used a certain resident's data for the experiment. The data includes more than a year. The last 71 days are used for the anomaly detection and the rest are used for the learning. In the data, there is one drastic increase of the response in the 11th day and there is long absence of the response between the 45th day and the 69th day. The two data are regarded as the "plus anomaly" and the "minus anomaly." We examine whether these two anomalies were detected as anomalies or not.

Figure 7 is the result of applying our anomaly detection algorithm to the living data of a resident. Figure 7(a) and 7(b) represent "plus anomaly" and "minus anomaly" of the same span, respectively, and the threshold to each "anomaly ratio" is 0.75 for "plus anomaly" and 0.98 for "minus anomaly." As the "anomaly ratio," we used the summed ratio of the two sensors which mainly responded in the house. In figure 7(a), an anomaly is accurately detected in the 11th day, in which the number of sensor response drastically increased, and similarly in figure 7(b) anomalies are detected mainly during the span between the 45th day and the 69th day, in which there was no response because the resident was not at home during the span. However, in the detection of "minus anomaly," a lot of other anomalies were detected if we lower the threshold from 0.98 to 0.95, for example. This means that adjustment of the threshold of "minus anomaly" should be carefully done. Besides, in this experiment, the contents of the anomalies are restricted to the cases when the number of response dramatically increases or decreases. This means that the concrete reason of anomaly cannot be figured out. Ideally, the reasons of the detected anomalies should be informed concretely. For example, "the resident is so feeble that she/he needs to be hospitalized." For the future work, we should discover a new anomaly criteria which can describe the reason of the detection more concretely.

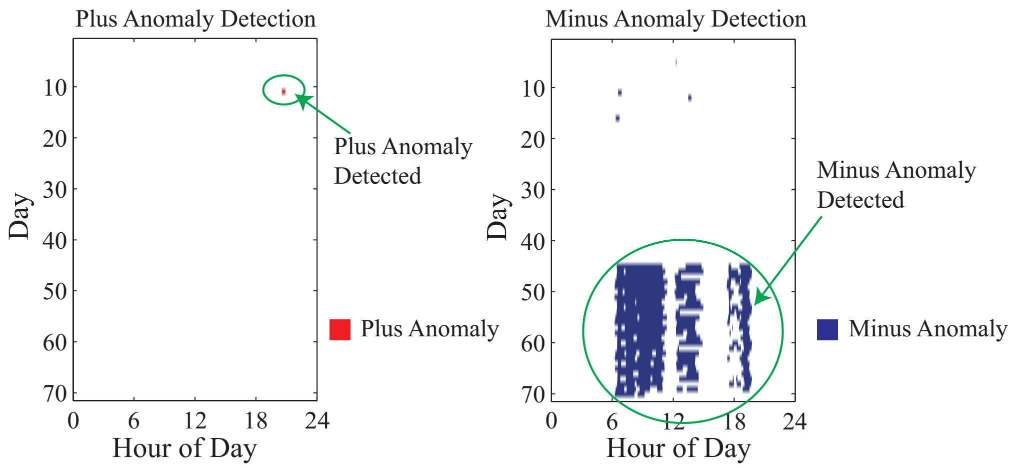

FIGURE 7. Anomaly detection experiment.

BRISKNESS ESTIMATION BY APPLYING ACTIVITY ESTIMATION METHOD

We demonstrate an example of actual observation of the elder people's life by applying the algorithms described above to their living data. Here is the observation on the elder people who started to use Demand Responsive Transport (DRT) service from a certain day in the experimental period. DRT is a kind of transporting system running with flexible routing and scheduling to fulfill the demand of passengers. The aim of the experiment is to observe the influence of encouraging the elder people to go out by using this transport service on their lives. Since four or five pyroelectric sensors are attached in their houses, it is possible to estimate the time of uprising, outing and going to bed by applying the activity estimation algorithm to the data discussed above. On this experiment, a day in which an outing over one hour occurs is called "outing day" and a day with no outings is called "non-outing day." We compared these two types of the day with the value called "briskness," which is calculated by dividing the number of total sensor response in the resident's "active hour" (the time of day except for outing time and sleeping time at night) by the length of "active hour" of the resident as normalization. The result is showed in figure 8.

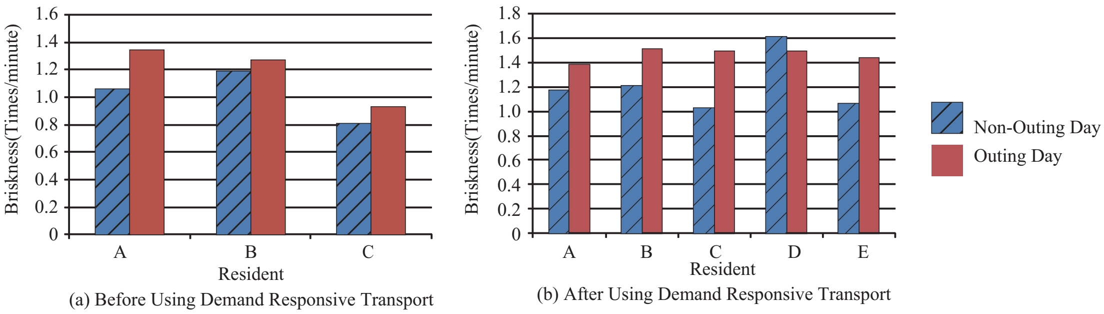

FIGURE 8. "Briskness" of the elderly using Demand Responsive Transports.

Figure 8 illustrates the "briskness" on "outing days" and "non-outing days" for each house. As shown in this figure, the "briskness" on "outing days" exceeds the one on "non-outing days" in all residents except for house D (The resident A and C live alone and the rest are families of husband and wife. Therefore, outing detection algorithm might have not work well on D). The result may be caused by two factors. First, preparation for the outing causes the increase of sensor response. Second, residents feel more excited than usual because of particular activities like outing.

In addition, comparing figure 8(a), which represents the residents' "briskness" before starting to use DRTs with figure 8(b) as their "briskness" after starting to use DRTs, "briskness" of almost all the residents increases by starting to use DRTs. Thus, the "briskness" is one of the criteria for precise understanding the lives of the elder people.

CONCLUSIONS AND FUTURE WORK

In this research, the anomaly detection and behavior pattern estimation method for the elderly living alone was proposed. First, the examined residents are categorized according to the sensor data. Then, the records of the anomaly which happened to the residents are shared in the same group. The categorization result is referred in the anomaly detection algorithms which detect drastic increases and decreases of the sensor response. In addition, basic activities in life such as outing are detected by using the activity estimation algorithm. This estimation and the result of the anomaly detection algorithm are displayed together to the nurses in the elderly observing system. The combination of two information enables the nurses to understand the elder people's lives precisely. In the experiment of monitoring elder people's lives using our algorithms, the fact that promoting outing to the elderly can activate their lives even in the day they did not go out was discovered. This experiment demonstrated that sensor monitoring systems can be utilized not only for the activity estimation and anomaly detection, but also for quantitative understanding tool of the residents' living situation. Therefore, continuous monitoring of the resident's daily life is very important for the sensor monitoring system. Our future work will focus on the study of the life pattern categorizing method. In this research, we manually determine the criteria of the categorization by observing the inclination of the data. However, in the future the criteria should be determined automatically, based on the properties of the data themselves.

REFERENCES

1. Kidd Cory D., R. J. Orr, G. D. Abowd, C. G. Atkeson, I. A. Essa, B. MacIntyre, E. Mynatt, T. E. Starner, and W. Newstetter, "The Aware Home: A Living Laboratory for Ubiquitous Computing Research," in *The Second International Workshop on Cooperative Buildings*, 1999, pp. 191–198.
2. T. Yamazaki, "Ubiquitous Home: Real-life Testbed for Home Context-Aware Service," in *Tridentcom2005 (First International Conference on Testbeds and Reserch Infrastructures for the DEvelopment of NeTworks and COMmunities)*, 2005, pp. 54–59.
3. K. L. Stephen S. Intille, E. M. Tapia, J. S. Beaudin, P. Kaushik, J. Nawyn, and R. Rockinson, "Using a Live-In Laboratory for Ubiquitous Computing Research," in *Parvasive 2006*, 2006, pp. 349–365, lNCS 3968.
4. K. Matsouoka, "Smart House Understanding Human Behaviors: Who Did What, Where, and When," in *Proceedings of the 8th World Multi-Conference on Systems, Cybernetics, and Informatics*, 2004, vol. 3, pp. 181–185.
5. N. Noury, T. Herve, V. Riallé, G. Virone, E. Mercier, G. Morey, A. Moro, and T. Porcheron, "Monitoring Behavior in Home Using a Smart Fall Sensor and Position Sensors," in *Microtechnologies in Medicine and Biology, 1st Annual International Conference on*, Lyon, France, 2000, pp. 607–610.
6. C.-H. Lu, and L.-C. Fu, "Robust Location-Aware Activity Recognition Using Wireless Sensor Network in an Attentive Home," in *Automation Science and Engineering, IEEE Transactions on*, 2009, vol. 6, pp. 598–609.
7. A. Sixsmith, and N. Johnson, "A Smart Sensor to Detect the Falls of the Elderly," in *IEEE Pervasive Computing* pp. 42–47 (2004).
8. T. Mori, R. Urushibata, M. Shimosaka, H. Noguchi, and T. Sato, "Anomaly Detection Algorithm Based on Life Pattern Extraction from Accumulated Pyroelectric Sensor Data," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2008, pp. 2545–2552.
9. K. Oshima, R. Urushibata, A. Fujii, H. Noguchi, M. Shimosaka, T. Sato, and T. Mori, "Behavior Labeling Algorithms from Accumulated Sensor Data Matched to Usage of Livelihood Support Application," in *The 18th IEEE International Symposium on Robot and Human Interactive Communication*, 2009, pp. 822–828.
10. M. Shimosaka, T. Mori, A. Fujii, and T. Sato, "Discriminative Data Visualization for Daily Behavior Modeling," in *Advanced Robotics* **23**, 429–441 (2009).