Anomaly Detection Algorithm Based on Life Pattern Extraction from Accumulated Pyroelectric Sensor Data

Taketoshi MORI, Ryo URUSHIBATA, Masamichi SHIMOSAKA, Hiroshi NOGUCHI, Tomomasa SATO.
The University of Tokyo

ABSTRACT

This paper describes an algorithm of behavior labeling and anomaly detection for elder people living alone. In order to grasp the person's life pattern, we set some pyroelectric sensors in the house and measure the person's movement data all the time. From those sequential data, we extract two kinds of information, time and duration, and calculate two-dimensional probabilistic density function of them. Using this function, we try to classify behavior labels and detect anomaly. Here, we assume two kinds of anomaly, "the rare behaviors" and "the changes of life pattern". The algorithm is confirmed to work on real behavior data through the experiment on about 400 days data.

I. INTRODUCTION

Recently, the number of elder people living alone is increasing sharply (especialy, in Japan and some developed countries). Therefore, systems of monitoring their life and detecting their symptom as soon as possible become an active research area[1][2]. In order to realize such system in the near future, we have been using a sensor system that is comparatively cheap[3]. Concretely, we decide to use a few pyroelectric sensors per one house. For example, a person's house is set totally three sensors at bedroom, living room and entrance.

We assume elder people's depression, dementia, illness, and so on, as the target of this system. But it is difficult to define a sensor data's pattern of those symptoms absolutely and generally, because such patterns will change per person. For example, if there is no sensor reaction for several hours, we are apt to think that he fell down because of some illness. However, if the person often goes trip, such sensor reaction is not so rare and we should regard these reactions as normal pattern. Considering such individual difference, we decide to detect rare behaviors or change of life pattern per person. As for anomaly detection, we usually consider the former, but changes of life patterns such as shortening of sleep time or disappearance of going out is to be detected. Using such probabilistic concept has the merit that the system can adapt to each person's unique life pattern.

The data given by our pyroelectric sensors are the frequency of reaction per one minute. Its time span, one minute, is very severe to grasp person's life pattern, but for realizing such cheap cost system, this span will be inevitable.

Therefore, the purpose of our research is to construct an algorithm to detect rare behaviors and changes of life pattern of elder people living alone by analysis of sensor data that are the frequency of reaction.

II. OUTLINE

Here, we describe the outline of our system. This is expressed in Figure1.

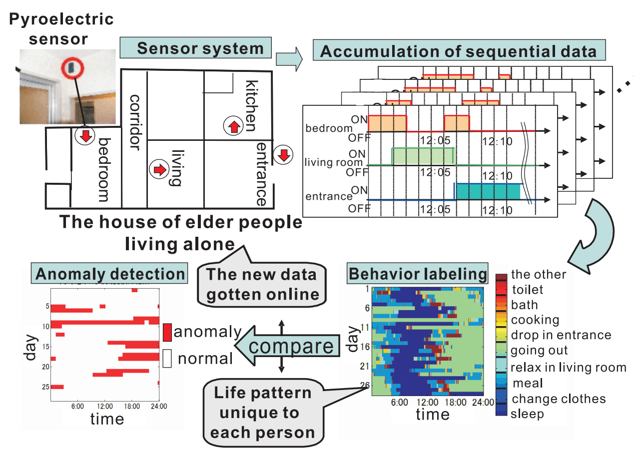

Fig. 1. System Outline

In order to detect rare behaviors or changes of life pattern, we need to know the person's life pattern. For example, a person typically wakes up about 8 a.m, goes toilet, then eats breakfast and relaxes at living room, eats lunch about 1 p.m, goes out to visit her husband's grave, cooks and eats dinner watching TV during 7 p.m to 9 p.m and goes to bed about 10 p.m, and so on. To know such life pattern, we have to assign the same label to the same behavior. After here, we call this labeling "behavior labeling".

To label behavior, the system has to extract some characteristics of behaviors from sensor data and classify them. In this paper, we select information of space, time and duration as the characteristics of behaviors. To abstract the frequency of sensor reaction, the system converts each sensor data into "on-off" data. Here, "on" means "The person stays here at that time". The person is thought to be performing a certain behavior during "on" time. Therefore, the system labels such "on" segments. For example, in the right top figure of Figure1, an "on" segment at living room may express lunch, while an "on" segment at entrance may express preparation of going out.

Our system detects rare behaviors and changes of life pattern in the process as follows:

1) Converts pyroelectric sensor data into ”on-off” room data.
2) Classifies ”on” segments into behavior labels and grasp one’s life patterns from space, time, and duration point of view.
3) Compares new sensor data with the life patterns and judges whether the new data contains anomaly or not.

This paper is composed as follows: Section III, IV and V present the algorithm of pyroelectric sensor data clustering, behavior labeling, and anomaly detection respectively. Section VI provides some experimental results of the anomaly detection. Section VII concludes this paper.

III. Extraction of Behavior Cluster from Pyroelectric Sensor Data

First, the system extracts behavior cluster from sensor data of reaction frequency by converting them into ”on-off” data. Figure2 expresses this outline. As described already, state of ”on” means ”from the start of a behavior to the end of it”, that is, each ”on” segment corresponds to one behavior in the room. For this reason, one usual behavior will occur at one place, not at multiple places. For example, sleep occurs only at bedroom, and eating lunch occurs only at living room.

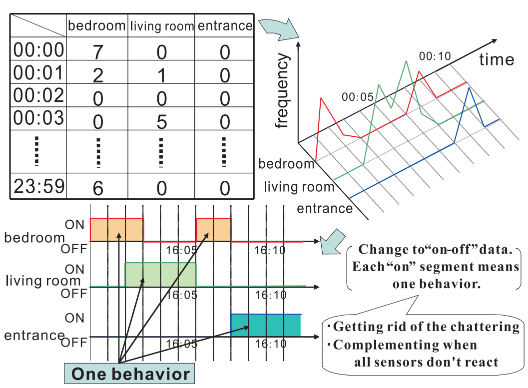

Fig. 2. Exraction of Behavior Cluster

When labeling multi-dimensional sequential data, we are often bothered by the segmentation problem, that is, ”where we should cut the data”. In this case, it means ”where a behavior should be regarded to change to another behavior”, and it is usually thought as a difficult problem. But we can solute such problem by regarding one ”on” segment as one behavior.

Basically, when the system converts the sensor data into ”on-off” data, if reaction frequency is more than 0, the ”on-off” at the time is regarded as ”on”, and if 0, as ”off”. However, the system needs to adjust this rule because it has the most important hypothesis that one ”on” segment means one behavior. Concretely, the system adds two processes as follows:

- Getting rid of chattering
- Complementing when all sensors don’t react.

The former means ignoring short reaction or short non-reaction. For example, let’s think the case when a person changed his clothes at bedroom, passed through kitchen, and went out. We want to label this sequence of behavior pattern as ”changing clothes” $\rightarrow$ ”going out”. So even though he passed through kitchen, that is, kitchen sensor reacted, we want to ignore this reaction. Thus, in order not to label such short sensor reaction of only passing as one behavior, short reactions should be ignored. As such, short non-reaction should be also ignored.

The latter means that a certain label should be assigned every time. It mainly intends for sleeping and going out. Our pyroelectric sensor basically reacts only when something moves, so it doesn’t react when sleeping. And, of course, it doesn’t react when going out. But in order to label ”Sleep” or ”Going out”, such time should be regarded as ”on”, that is, bedroom sensor needs to be ”on” when he is sleeping, and entrance sensor needs to be expressed as ”on” when he is going out. Basically, when a person enters the place or goes out of the place, the place’s sensor will react. This means that the sensor will be ”on” at the start and end of a behavior. So when all sensors don’t react, the place that should be ”on” is decided from the reaction before and after the time.

IV. Behavior Labeling Algorithm

Next, the system assigns those ”on” segments some behavior labels. Here, the system doesn’t need to know what behavior the label means. Such identification wouldn’t be universal. So the system only assigns the same kind of behavior the same label.

First, assuming that the same behavior almost occurs at the same place, the system treats each sensor’s ”on” segments independently. So bedroom’s ”on” and kitchen’s ”on” are classified as the different behaviors. In this process, we can classify ”on” segments into the sensor-number’s labels at least. These labels are equivalent to room labels such as ”bedroom”, ”living room”, ”kitchen”, and so on. However, we premised to classify not by the place but by the behavior, which is deeper level than place. We need to think an algorithm of classification of each place’s ”on” segments into several behaviors occurring at the same place.

Now, we pay attention to the two kinds of information; time and duration. Here, ’time’ means the time of the segment’s middle point. For example, if bedroom’s ”on” keeps for 8 hours at midnight, we think the segment must be ”Sleep”. But if bedroom’s ”on” keeps only for 10 minutes at morning, we think it must be ”Changing clothes or something except sleep”. Thus, we assume that same behavior should have almost the same time and duration. Then, the system classifies the behaviors occurring at the same place by measuring the differences of time and duration. After all, the system extracts the two-dimensional vector of attributes, time and duration, from each ”on” segment, and classifies the vectors.

Some basic clustering algorithms such as the k-means clustering or the hierarchical clustering are known widely[4][5][6]. However, they often give mainly two problems as follows:

- There is no absolutely stable way to know the best cluster number.
- Results will change with the initial parameters.

In this research, then, we suggest the original algorithm that makes it possible to decide the cluster number and the initial parameters automatically by examining the density difference of the plot figure. We describe its concrete algorithm as follows.

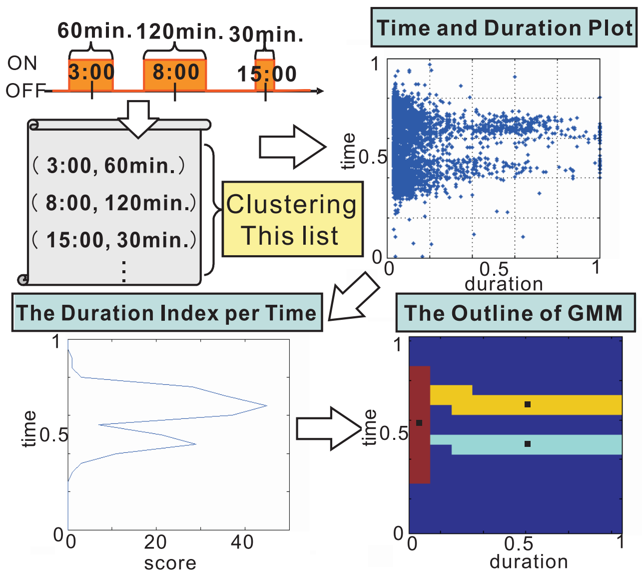

Fig. 3. Estimation of Mixture Number and Initial Parameter of GMM

First, the algorithm plots the two-dimensional data on the plane which axes are time and duration. Here, both dimensional ranges are abridged into [0, 1] in order to make their contribution equal. In case of time, the range from 00:00 to 23:59 is abridged into [0, 1]. In case of duration, the range from minimum to maximum is abridged into [0, 1], but considering overwhelmingly large value, the algorithm gets rid of top 1 percent data at this abridgement.

Now, we must classify these plot data into some classes. For the purpose, we tried estimating the probabilistic model which is thought to generate the plots. As mentioned above, we assume that same behaviors have similar time values and similar duration values, so the probabilities of time and duration of each behavior's center at the model should be high. Considering this condition, we decide that the probabilistic model should be the Gaussian mixture model (called GMM from here), that is composed of each Gaussian distribution expressing each behavior. Thus, when a new plot is given, the algorithm calculates relative likelihoods toward each Gaussian distribution and classifies the plot into the behavior label which is corresponded to the Gaussian distribution that has the maximum relative likelihood.

Here, we explain this algorithm in detail. First, the algorithm estimates the generous GMM's outline, that is, the number of mixture and the initial parameters of each Gaussian model (mean, covariance, and weight) from density of the time and duration plot figure. This process is shown

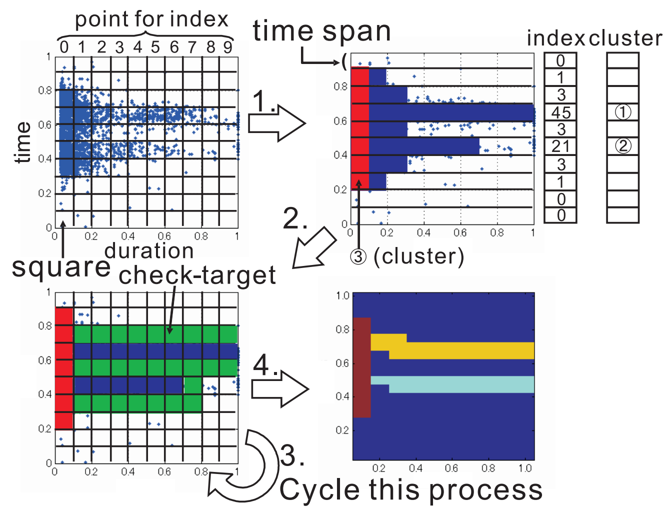

Fig. 4. Calculation Process of Initial GMM

at Figure4. Considered from the previous paragraph, a high density part of the plot figure means one behavior. By observing some plot figures of sensors like the right upper figure of Figure3, we noticed that there are mainly two ways of spreading that plots of each behavior have as follows:

1) The cluster that the duration mean value is small (almost zero), and the time dispersion is large while the duration dispersion is small. (called ”the long-length cluster”)
2) The cluster that the duration mean value is large to some extent, and the time dispersion is small while the duration dispersion is large. (called ”the long-side cluster”)

It is easy to find the long-length clusters because they have many plots and are known to have small duration mean in advance. Conversely, it is difficult to find the long-side clusters because they usually have relatively small numbers of plots so that the results will be different according to definition easily. Thus we suggest the process of discovery of the long-side cluster as follows. This process is shown in Figure4.

First, the process determines the duration index per each time span as follows. Here, the time span means the sequence of time like ”from 8:00 to 9:00” (shown in Figure4). The process divides the abridged duration range [0, 1] into ten parts, and assign the point 0,1,2,$\cdots$,9 to each part from the short duration sequentially. Then, when the process sees a rectangle of a time span which cut up from the plot figure, if the plot density of a square is higher than a threshold (call this state ”On”), the process adds the point corresponding to the duration of the square to the index of the time span. Here, the word ”square” means the square or rectangle part of the plot figure like ”the rectangle of time range from 8:00 to 9:00 and duration range from 30minutes to 1 hour” (shown in Figure4). After all, the index of each time span, $P(t)$, is defined as follows:

$$P(t) = \sum_{i=0}^{9} (the\ point\ of\ square\ which\ state\ ''On'')$$

We defined this index based on the two reasons as follows:

- The larger the duration dispersion is, the larger the index becomes.
- The larger the duration value is, the larger the index becomes.

The left bottom figure of Figure3 is the calculation result of this index toward the right upper figure's plots. We can confirm that the index of the time span where the long-side clusters are guessed to exist from right upper figure's plot is properly higher than another time span.

Then, using this index, the system creates the clusters consisted of square as the following process:

1) The process sets a threshold of the difference of index, and if the index at a time span is higher than the one at the next time span by the threshold, it decides that a cluster should exist at the time span.
2) The process checks the squares at the time span given a cluster one by one, and if the density of the square is larger than a, it regards the square as a part of the cluster. Here, a is a fixed number (usually set to make the cluster number not exceed 3 or 4).
3) The process checks the squares near the squares which is regarded as the part of some cluster at 2) (8 neighbor squares per one square) one by one, and if the square's density is larger than a, it regards the square as the part of the same cluster.
4) The process repeats the same check toward the squares generated at 3) until the range of the cluster becomes independent.

We show the result of this process toward the plot figure of Figure3 as the right bottom figure. Comparing this figure with the right top one, we can confirm that the square which has high density of plots is actually assigned a cluster.

As the result, the algorithm can calculate the mixture number and the initial parameters of GMM automatically, i.e., the mixture number is equal to the cluster number, and the initial parameters of the GMM, means, dispersions, and weight of each dimension, are given as follows:

- mean $\rightarrow$ the center of each cluster
- covariance $\rightarrow$ the covariance culcurated by the plots in each cluster
- weight $\rightarrow$ the relative number of plots in each cluster

Though the system can estimate the outline of GMM, this isn't the correct model, so it needs to calculate the correct model of GMM by using this outline and the plot data. For this, we apply the EM-method (Estimate and Maximization); every time when a new datum is given, the model updates its parameters in order to maximize the likelihood of all data that contain the new datum, too.

In addition, these data are the sequential data that are accumulated during long span and often change the tendency. In order to adapt to such changes as soon as possible, the system needs to change the weight of the past data; as the data become old, the weight of the data should be lightened. In this system, we introduce the discounting parameter $r(0 < r < 1)$, and when a new datum is given, the old data are multiplied by $(1-r)$ in order to make the weight of the past data lighter slowly.

Now, we explain the SDEM (Sequential Discounting Expectation and Maximization) algorithm which calculates the likelihood of GMM, $p(y|\theta)$.

$$p(y|\theta) = \sum_{i=1}^{k} c_i p(y|\mu_i, \lambda_i)$$

Here, k is the number of the Gaussian distribution, each $p(y|\mu_i, \lambda_i)$ is a two-dimensional Gaussian distribution with density specified by mean μ_i and covariance matrix λ_i. And θ is the vector of parameter:

$$\theta = (c_1, \mu_1, \lambda_1, \cdots, c_k, \mu_k, \lambda_k)$$

Then, we have to calculate θ which makes $\sum_{j=1}^{t} logp(y_j|\theta)$ maximization (Maximum likelihood estimator). This algorithm is shown in Table I. In this algorithm, a parameter α is introduced in order to make the estimated value of c_i stable, and this value is usually set from 1.0 to 2.0.

The initial parameter $c_i^{(0)}$, $\mu_i^{(0)}$, $\lambda_i^{(0)}$ was defined according to [7] that "k is given, $c_i^{(0)} = 1/k$, and $\mu_i^{(0)}$ are set so that they are uniformly distributed over the data space, and $\lambda_i(0)$ is set rightly to represent the distribution of the data". However, as mentioned above, the system already calculated these initial parameters. In this sense, we suggest the improved algorithm of SDEM.

By repeating the part $Repeat$, the GMM is adapted to the distribution of the plot, shown as the left figure of Figure5. Though the model is cut at the density value up to 3 in this figure, the maximum of the Gaussian distribution of the short duration is much higher than 3. This means that the large part of the plot data has short duration.

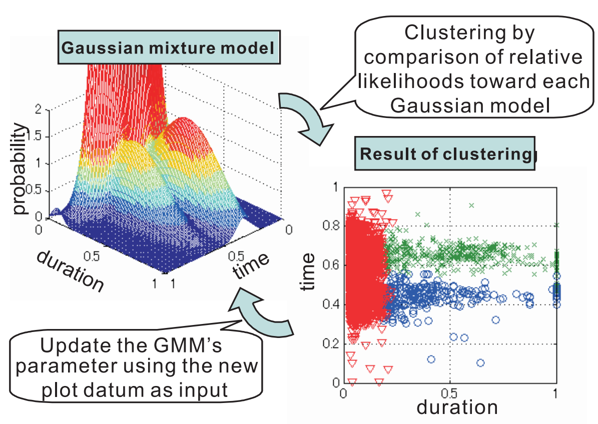

Fig. 5. Cycle of Learning of GMM and Clustering

Then, when a new datum y_t is given, its relative likelihoods toward each Gaussian distribution L_i are given as follows:

TABLE I

IMPROVED SDEM ALGORITHM

$t := 0$

$(0 < r < 1 : discounting\ parameter, \alpha > 0, k : given)$

Initialize: $c_i^{(0)}, \quad \mu_i^{(0)}, \quad \lambda_i^{(0)} (Calculated\ above.)$

$Repeat:$

Input: y_t

For $\quad i = 1, \cdots, k$

$$\gamma_i^{(t)} = (1 - \alpha r) \frac{c_i^{(t-1)} p(y_t | \mu_i^{(t-1)}, \lambda_i^{(t-1)})}{\sum_{i=1}^{k} c_i^{(t-1)} p(y_t | \mu_i^{(t-1)}, \lambda_i^{(t-1)})} + \frac{\alpha r}{k}$$

$(calculating\ posterior\ probabilities)$

$$\begin{aligned} c_i^{(t)} &= (1-r) c_i^{(t-1)} + r \gamma_i^{(t)} \\ \bar{\mu}_i^{(t)} &= (1-r) \bar{\mu}_i^{(t-1)} + r \gamma_i^{(t)} y_t \\ \mu_i^{(t)} &= \frac{\bar{\mu}_i^{(t)}}{c_i^{(t)}} \\ \bar{\lambda}_i^{(t)} &= (1-r) \bar{\lambda}_i^{(t-1)} + r \gamma_i^{(t)} y_t y_t^T \\ \lambda_i^{(t)} &= \frac{\bar{\lambda}_i^t}{c_i^{(t)}} - \mu_i^{(t)} \mu_i^{(t)T} \end{aligned}$$

$(updating\ parameters)$

$t := t + 1$

$$L_i = \frac{c_i^{(t-1)} p(y_t | \mu_i^{(t-1)}, \lambda_i^{(t-1)})}{\sum_{i=1}^{k} c_i^{(t-1)} p(y_t | \mu_i^{(t-1)}, \lambda_i^{(t-1)})}$$

And the new datum is classified into the cluster corresponding to the Gaussian distribution which has the maximum relative likelihood (, max $L_i (i = 1, \cdots, k)$). The result of the clustering is shown at the right figure of Figure5. We can confirm that the result of clustering will be similar to the result that we human try to classify, so we can conclude that the algorithm is probably appropriate.

After all, as shown at Figure5, the system repeats the two calculations every time when a new datum are given:

1) Classifies the new datum by comparing relative likelihoods toward each Gaussian distribution.
2) Updates the GMM by the SDEM algorithm, treating the new datum as input.

This process is the labeling per one place. Then, by gathering the classification results of all sensors, the system becomes able to label the multi-dimensional sequential data the behavior labels which have more kinds than the number of sensors. We show the result in Figure6. The horizontal axis means time and the vertical axis means day, and the same label is shown by the same color. The legends mean the relation between colors and behavior labels. Here, a behavior label is named as ”*the room name + number*” like bedroom2, entrance3, and so on.

From Figure6, we can confirm that the same behaviors are probably classified into the same labels. For example, the behaviors occurring at midnight are classified into the same label ”bedroom1”, which probably show ”Sleep”. Also, we can grasp the person’s two kind of pattern of going out that colors are yellow and orange, and this proves that the system can classify the same behavior into multiple patterns (as long as these patterns have the clearly different time or duration patterns). In addition to this analogy, we actually have confirmed that the same behavior is assigned the same label in the short term preliminary experiment with ground truth behavior label. After all, we conclude that the system can label the sensor data the behavior labels, which is adapted to the person’s unique behavior patterns.

In addition, as the purpose of the labeling, the result shows the person’s life pattern; for example, the examinee of Figure6 wakes up about 8 o’clock, and stays at living room or goes out in the morning, and maybe eats lunch at noon at living room, and sometimes naps at bedroom after lunch, and goes out afternoon, then stays at living room from 6 p.m to 8 p.m, and goes to bed about 8 p.m. By grasping the person’s life pattern like above, we may be able to detect anomaly behaviors or changes of life patterns. So, we suggest the algorithm of anomaly detection at the next section.

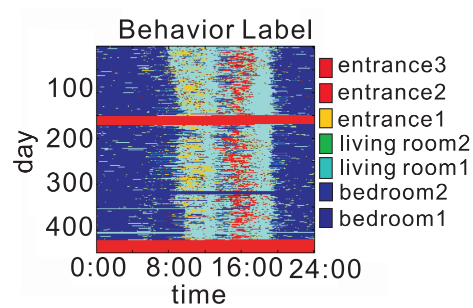

Fig. 6. Example Result of Behavior Labeling

V. ANOMALY DETECTION ALGORITHM

Next, we explain the algorithm of anomaly detection. As mentioned above, we assume rare behaviors and changes of life pattern as anomaly. We call the rare behaviors ”the local anomaly” and the changes of life pattern ”the global anomaly” from here.

First, we mention the way to detect the local anomaly. When thinking about the novelty factor of a behavior, we picked up time and duration as the main factor again. For example, sleeping longer than usual may be the coma, or too long absence may imply the traffic accident. Such behaviors belong to ”the duration anomaly”. As other examples, going out at midnight may be loitering because of dementia, or sleeping in the daytime may express sudden down by cerebral infarction. Such behaviors belong to ”the time anomaly”.

In order to judge a behavior as usual or as anomaly, the system measures the likelihood of the behavior calculating the joint probability of time and duration. Then, if the likelihood is extremely low, the behavior should be judged as anomaly. To calculate the joint probability of time and duration, the system computes the probability of the fact that the behavior occurred at the time ($P(x)$) at first. After that, the system computes the conditional probability of the fact that the behavior kept for the duration on the condition that the behavior occurred at the time ($P(y|x)$). Then, the joint probability is given as the multiplication of these two probabilities ($P(x, y) = P(x) \times P(y|x)$). The Figure7 expresses this process.

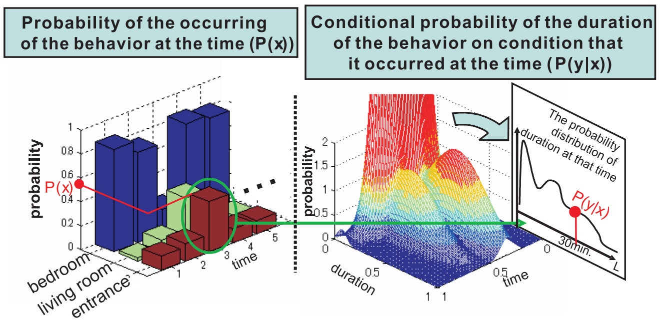

Fig. 7. Calculation of the Joint Probability of Time and Duration

To calculate the former, we use the SDLE (Sequential Discounting Laplace Estimation) algorithm [7]. This algorithm is basically the Laplace Estimation, but specified by the discounting parameter. The reason to add this parameter is already mentioned in the SDEM algorithm. Its algorithm is shown in TableII. Here, M means the cell number (in this case, it is equal to 2; ”on” or ”off”), and $A_1, \cdots, A_M$ mean partitioning of the domain (in this case, A_1 means state ”on” and A_2 means state ”off”).

The latter, ”The probability of the behavior's duration in condition that it occurred at the time”, was actually calculated already, because ,as shown at Figure7, the probability distribution of duration at the time is equal to the time's cross section of the GMM that was estimated at the behavior labeling process. Therefore, this conditional probability corresponds to the probability density of the GMM at that time and duration.

Then, by multiplying these two probabilities, the system can calculate the joint probability of time and duration. In this paper, we define the first anomaly as the behaviors that have the joint probability of minimum 0.1 percent and the second anomaly as that of minimum 1 percent.

Next, we mention the way to detect the global anomaly. Here, the global anomaly means the changes of the person's life pattern. We assume that the changes of life pattern mean the extinction of the behavior that occurred regularly before, or the clear change of frequency, time, or duration of such behaviors.

In order to detect such changes, the system checks the change of the probabilistic model. For example, if we want to know the change of time and duration distribution of a behavior, we should examine the change amount of the

TABLE II

SDLE ALGORITHM

$t := 0$

The cell group:$A_1, \cdots, A_M$

Initialize: $T_i^{(0)} = 0 \quad (0 < r < 1, \beta > 0 : given)$

$(r : discounting\ parameter)$

Repeat :

Input: x_t

For the cell number i ($i = 1, \cdots, M$)

$$T_i^{(t)} = (1 - r)T_i(t - 1) + \delta_i x_t$$

$(calculating\ the\ statistic\ of\ each\ cell)$

$$\delta_i(x_t) = \begin{cases} 1 & (\text{if } x_t \in No.\ i\ cell), \\ 0 & (\text{otherwise}) \end{cases}$$

$$q_i(x) = \frac{T_i + \beta}{\sum_{k=1}^{t}(1 - r)^{(t-k)} + M\beta}$$

$(Laplace\ Estimating)$

$$p(x) = \frac{q_i(x)}{|A_i|}$$

$(calculating\ the\ probability\ of\ each\ symbol)$

$t := t + 1$

GMM of time and duration of the behavior, which is equal to the volume of the changing part (increasing part or decreasing part) that is calculated when compared with the past model, as shown at Figure8.

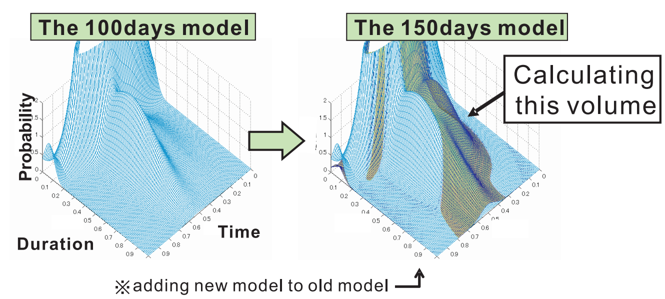

Fig. 8. Example of Model Change

Basically, the system regards the model which changes extremely as anomaly, but if it measures the difference between the present model and the model some span before, it can't get rid of the factor of periodic changes such as season or week, and can't cease to regard the change of return to normal model from abnormal model as anomaly. The system solutes such problems by calculating the differences between the present model and all the past day's models and regarding the change amount of the day as the mean of those differences. Here, we add the function of discounting for the same reason as the SDEM algorithm and SDLE algorithm; in order to make much of newer data.

In such way, the system calculates the changes of all models every day and gets many one-dimensional sequential data. Then, the system regards a day as anomaly if the degree of leaning of each data during recent one month is plus and larger than the value which multiplies some fixed number with the degree of leaning from four month ago till one month ago. In other words, the system regards the global anomaly as when the change amount rapidly increases, because a model doesn't change so much in a day.

VI. Result of Examination

Here, we show some results of examination. The data used in these examinations are the data of the real elder people living alone. To get the reference record, we call them every other week and ask their health condition.

First, we examine the local anomaly detection. The health condition of this examinee was actually getting worse, and her family took her to their house in order to take care of her about the 400th day from the beginning of the sensor data accumulation. Using this valuable data, we confirmed that the system really can detect the anomaly of time and duration, and that we can get a certain result which shows the foretaste of the fact that her condition was actually getting worse.

The result is shown at Figure9. The top figure shows the clustering result, in which the horizontal axis means day and the vertical axis means time. The middle figure shows the anomaly detection result, in which red parts mean the first anomaly (minimum 0.1 percent) and yellow parts mean the second anomaly (minimum 1 percent). The bottom figure shows the frequency of anomaly per month, that is, the frequency plots of every other month. And the day her family took her is shown by the green line.

From the top and middle figure, we can confirm that the system detected the long absence at about the 150th day and the long stay at bedroom at about the 300th day; the duration anomaly can be detected by this system. This means that the system may probably detect the sudden down by some illness or some accident out of the house. Also, the system detected some behaviors except sleep occurred in midnight; the time anomaly can be detected. This means that the system may be able to detect the insomnia or the loitering by dementia.

In addition, from the bottom figure we can see the increase tendency of the anomaly frequency, even though the tendency is gradual. Usually, the anomaly frequency decreases because the anomaly threshold keeps getting high; if two behaviors at the 100th day and the 300th day have the same duration which is much longer than usual, the behavior at the 300th day can't be regarded as anomaly because the same duration's behavior already occurred at the 100th day, so the probability of the duration is not so low, even though the same behavior at the 100th day is regarded as anomaly. Thus, logically, the increase tendency of the anomaly frequency is not usual, and we thought the tendency shows the aggravation of the examinee's health condition. Therefore, we conclude that the system is able to omen the aggravation of health condition.

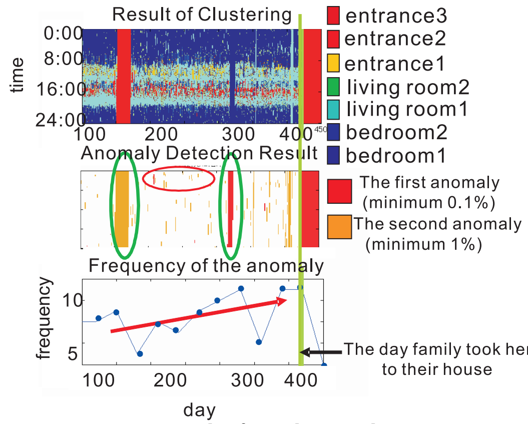

Fig. 9. Result of Local Anomaly Detection

Next, we examined the global anomaly detection algorithm. The examinee of Figure11 has the change of life pattern; she slept at bedroom until the 110th day, but after then, she changed the place of sleep from bedroom to living, and at about the 200th day she returned the place to bedroom again. So we examined whether the system can detect this change of life pattern or not.

At first, we show the change of the GMM of living sensor at Figure10. The left figure shows the time and duration plot at living until the 140th day. In this figure, the plot group at right bottom (means long duration and at midnight) means sleep and we observed that this plot group generated rapidly from the 110th day. The right figure shows the GMM at living room. The blue model shows the model at the 140th day and the white one shows the model at the 100th day. Comparing these two models, we can easily see that the volume of their difference is extremely large. This means that the system may estimate that there was some life pattern change concerning a behavior occurring at living room. Then, we confirm this next.

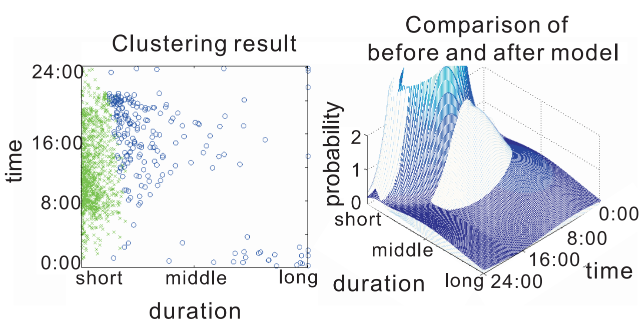

Fig. 10. Model Change of GMM at Living room

We show the result of this examination at Figure11. The top figure shows the clustering result. The second figure shows the total change of the probability of "on" per each behavior at each time. The third figure shows the volume change of the GMM; the red line shows that of living room and the green one shows that of bedroom. The bottom figure shows the result of anomaly detection using the results of the second and third figure; three rows show the result of anomaly detection of total ”on” probability, of the GMM at bedroom, and of the GMM at living room from the top respectively. In this figure, a red part means anomaly.

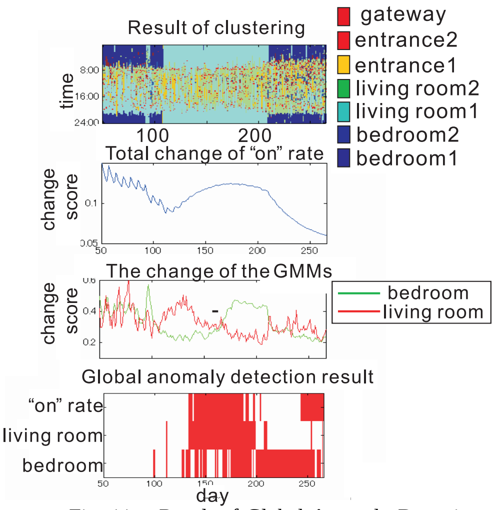

Fig. 11. Result of Global Anomaly Detection

From the top figure, we can easily confirm the change of sleeping place from bedroom to living room. But we want to detect this change automatically and as soon as possible. To see the second figure, the total ”on” probability changes clearly at this change point; the tendency of the change score was decreasing until the 110th days, but after about the 115th days, the tendency changed into increase. In addition, we can confirm that at the 210th days, the day when the pattern of sleep place returned to the bedroom, the score didn't change into increase; by taking the difference toward all past data, we can cease to detect the model change by the return to the normal pattern again. From the third figure, we can confirm that the living room model changed relatively soon after the change point and the bedroom model changed relatively slowly. We recognize that the difference generates because living model is updated by the new data of sleep, while bedroom model is updated by the custom data except sleep. It means that the system can estimate the details of the change from the adapting speed of each GMM; in this case, the adapting speed of the GMM of living room is faster than that of the GMM of bedroom, so the system may be able to guess that a behavior at bedroom disappeared and a behavior at living room appeared in place of the bedroom one. From the bottom figure, that is the final results of the global anomaly detection, we can confirm that the system actually can detect the change of sleeping place after a few week from the change point. When we human judge the pattern change, we may conclude this after we are convinced the fact that the changed pattern keeps occurring for a degree of length. This system also judges the change by the same way, so the detection naturally delays a few weeks after the change. From these, we conclude that the global anomaly detection system can detect the changes of life pattern of elder people living alone.

VII. CONCLUSION

In this paper, we presented the automatic monitoring system for elder people living alone composed of only a few pyroelectric sensors so that the system can be used practically in the near future. The system can grasp the person's unique life pattern by accumulating his long-range life data, and adding the discounting function makes it possible to adapt to newer life pattern. Concretely, the system classifies the behaviors at the same place by time and duration. In this process, it can estimate the best number of behaviors automatically by investigating the plot figure of these two kinds of information. From such life pattern, the system is able to detect anomaly of broad range; the rare behaviors as the local anomaly and the changes of life pattern as the global anomaly. We confirmed by some examinations that this system could actually detect such anomaly. In addition, we were able to get the result that the system could omen the aggravation of physical condition actually.

REFERENCES

[1] Kidd Cory D. and Robert J. Orr and Gregory D. Abowd and Christopher G. Atkeson and Irfan A. Essa and Blair MacIntyre and Elizabeth Mynatt and Thad E. Starner and Wendy Newstetter. The Aware Home: A Living Laboratory for Ubiquitous Computing Research. In Proceedings of the Second International Workshop on Cooperative Buildings, pp. 191–198, 1999.

[2] Taketoshi Mori and Hiroshi Noguchi and Aritoki Takada and Tomomasa Sato. Sensing Room: Distributed Sensor Environment for Measurement of Human Daily Behavior. In 1st International Workshop on Networked Sensing Systems(INSS), pp. 40–43, 2004.

[3] Taketoshi Mori and Akinori Fujii and Masamichi Shimosaka and Hiroshi Noguchi and Tomomasa Sato. Typical Behavior Patterns Extraction and Anomaly Detection Algorithm Based on Accumulated Home Sensor Data. In FGCN2007Workshop: International Symposium on Smart Home (SH07), pp. 12-18, 2007.

[4] K. Matsuoka and H. Yamamoto and S. Inoue and S. Tanaka and K. Ohta and Y. Matsumura and T. Kometani and K. Fukushima.Smart House monitoring living behaviors for finding un-usual conditions.In Proceedings of ROBOMEC'04,2P1-L2-34, pp.1–4, 2004.

[5] Taketoshi Mori and Aritoki Takada and Yasuhiko Iwamura and Tomomasa Sato. Automatic Human Life Summarization System in Sensory Living Space. In Proceedings of IEEE International Conference on Systems, Man and Cybernetics(SMC), pp. 1583–1588, 2004.

[6] Padhraic Smyth. Clustering Sequences with Hidden Markov Models. In Advances in Neural Information Processing Systems, pp. 648-655, 1997.

[7] Kenji Yamanishi, Junichi Takeuchi, Graham J. Williams, and Peter Milne. On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms. In Knowledge Discovery and Data Mining, pp. 320-324, 2000.