EXTRACTING PERSONAL USER CONTEXT WITH A THREE-AXIS SENSOR MOUNTED ON A FREELY CARRIED CELL PHONE

Toshiki Iso
Network Laboratories, NTT DoCoMo, Inc.
3-5 Hikarino-oka, Yokosuka, Kanagawa, Japan

Kenichi Yamazaki
Network Laboratories, NTT DoCoMo, Inc.
3-5 Hikarino-oka, Yokosuka, Kanagawa, Japan

Keywords: Context extraction, sensors, cell phone, wavelet packet, Kohonen self-organizing map, ubiquitous service, accelerometer.

Abstract: To realize ubiquitous services such as presence services and health care services, we propose an algorithm to extract "personal user context" such as user's behavior; it processes information gathered by a three-axis accelerometer mounted on a cell phone. Our algorithm has two main functions; one is to extract feature vectors by analyzing sensor data in detail by wavelet packet decomposition. The other is to flexibly cluster personal user context by combining a self-organizing algorithm with Bayesian theory. A prototype that implements the algorithm is constructed. Experiments on the prototype show that the algorithm can identify personal user contexts such as walking, running, going up/down stairs, and walking fast with an accuracy of about $88[\%]$.

1 INTRODUCTION

Cell phones are becoming indispensable in daily life and are carried everywhere due to their sophisticated functions such as camera, music player, IrDA, wireless LAN, GPS and IC-chip for electronic wallet (NTT DoCoMo, Inc.,). Since cell phones have become personal assistants, they are well placed to identify user behavior. Mobile service providers are demanding user context such as user behavior because they want to provide appropriate services that are suitable for the situation of the mobile user. Identifying "personal user context" is especially important in advancing health care services and service navigation.

There are many related works on context-aware systems using sensors in mobile environments (Siewiorek et al., 2003), (Krause et al., 2003), (Kern et al., 2004), (K. V. Laerhoven, 2003), (Miao et al., 2003), (DeVaul and Dunn,), (Clarkson et al., 2000), (Healey and Logan,), LBao2004, (M. Unuma, 2004), (Randell and Muller, 2000).

However, these conventional methods have two practical issues; one is that users need to wear some sensors on specific parts of their bodies, the other is that their feature extraction output is suitable only for a limited range of user context. In regard to these issues, most papers adopted the wearable computer approach and sensor data was extracted by FFT-based approaches. Thus, they required several sensors to be fixed to different parts of the user's bodies to achieve a high degree of accuracy. Moreover, it was difficult to analyze the localized wave data present in the sensor data because Fourier transform has lower time-frequency resolution than wavelet transform. Their methods are not realistic because wearing many sensors is very cumbersome. On the other hand, one of the related works (Si et al., 2005) proposed a method of extracting context that is independent of sensor position, however, their method had limited context extraction performance because feature extraction was based on the magnitude of three-axis sensor data.

As a solution to these issues, we proposed the "PerContEx" (Iso et al., 2005) system which could extract personal user context by applying an algorithm to data collected from the user's cell phone. While the system placed no constraints on how the phone was carried, the high computational costs of the algorithm meant that it was not always possible to realize ubiquitous services that require real-time processing. Therefore, as another approach, we propose here a method of extracting personal user context by subjecting the data collected by sensors tailored to the user's activities to wavelet packet decomposition. It can identify high detailed user contexts such as walking at normal speed, running, walking fast, and going up/down stairs. This is because the wavelet packet de-

Iso T. and Yamazaki K. (2006).
EXTRACTING PERSONAL USER CONTEXT WITH A THREE-AXIS SENSOR MOUNTED ON A FREELY CARRIED CELL PHONE.
In *Proceedings of the International Conference on Signal Processing and Multimedia Applications*, pages 224-231
DOI: 10.5220/0001568902240231

composition provides a finer analysis of sensor data than FFT.

Section 2 explains the concept of processing in PerConEx which extract personal user context. Section 3 describes detail algorithm at each stages in PerContEx. We then show some experiments. Finally, we draw several conclusions and describe future works.

2 OVERVIEW OF PROCESSING

Since we use a three-axis sensor to implement our proposals, our system must satisfy the following requirements;

- feature extraction: it should deal with a variety of personal user contexts
- context extraction: it should be independent of how the three-axis sensor is carried.

To satisfy the first requirement, we propose a method to extract features from sensor data by using the Best Basis method with wavelet packet decomposition. To satisfy the second requirement, we propose a context extraction method that consists of a self-organizing algorithm and Bayesians theory; a variety of pseudo-data is used as training data for context clustering.

We show the overall processing flow of the algorithm in Figure 1. The algorithm has three stages: feature extraction, learning, and personal user context extraction.

At first, in order to handle the variety of signal patterns which are possible, we use wavelet packet decomposition (WPD) for wave analysis in the feature extraction stage. This is because WPD provides finer analysis in each frequency range through the use of localized orthogonal basis functions. Moreover, by using information criteria, we can obtain decomposed signals which can efficiently representing the features of signal patterns. We extract feature vectors from the decomposed signals.

Second, in order to avoid the influence of data fluctuation and noise, we use the Kohonen self-organizing map (KSOM) as a non-linear learning method. This is because KSOM can provide not only robustness to noise and signal fluctuation, but also flexibility in terms of the number of clusters. By calculating appearance probabilities of personal user contexts, we can extract the most representative feature vectors and time-series data while avoiding the vague vectors. Moreover, in order to achieve robustness with regard to where the sensors are set, we prepare a variety of KSOMs and train them using pseudo-data generated from the most representative time-series data by rotational transformation.

Third, in order to extract personal user context, we use all feature maps including the appearance possibilities of personal user context trained by both observed and pseudo-data. Finally, we can extract the personal user context from three-axis sensor data by setting adequate criteria. We describe the three stages below.

3 ALGORITHM

3.1 Feature Extraction Stage

Noise is naturally present in time-series data from sensors mounted on mobile devices. Moreover, in the case of cell phones carried freely, the time-series data often contain signals generated by non-targeted behavior. It is difficult to make a comprehensive disturbance model because of the variety of disturbances. Under the assumption that the majority of time-series data is generated by the targeted behaviors, we propose a method to extract features by analyzing the time-series data decomposed by a combination of wavelet packet and information entropy.

3.1.1 Sensor Data Extraction and Preprocessing

At first, we divide the time-series data from a three-axis sensor into frame data. We define t_s as the time range. We assume that the three-axis sensor does not experience twisting in time frame t_s as follows.

$$\{t_s \mid \frac{1}{2}(s-1)\Delta T \leqq t_s < \frac{1}{2}(s+1)\Delta T\} \quad (s = 1, 2, \cdots) \tag{1}$$

where ΔT is frame length. To divide the sensor data, some overlap time must be set to get data stability. At this time, we define an observed three-axis sensor data $\boldsymbol{a}(t)$ as follows;

$$\boldsymbol{a}(t) = (a_x(t), a_y(t), a_z(t)) \tag{2}$$

Feature extraction is preceded by preprocessing operations such as data and time calibration.

3.1.2 Wavelet Packet Decomposition

Time-series data analysis is commonly based on Fourier transform. Actually, it is a very strong tool for feature extraction if we can assume that the input represents regular periodic data that extends over relatively long periods. Unfortunately, this assumption fails in the case of time-series data generated by the user's behavior; the data here is generated by many short-term events. Therefore, wavelet transform should be applied rather than Fourier transform. Wavelet transform is very common in the fields of signal processing applications such as image compression and analysis of electrocardiograms (Ishikawa, 2000).

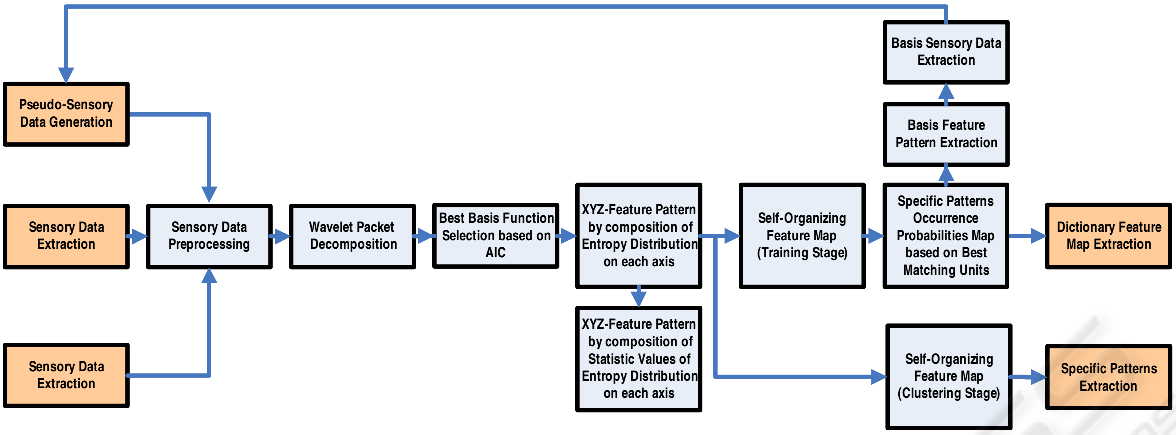


Figure 1: Processing flow.

Wavelet packet analysis provides more detail because it uses a splitting algorithm that downsamples not only the scaling components but also the wavelet components.

In order to realize the flexible analysis of sensor data, we use discrete wavelet packet decomposition (DWPD). By using DWPD, the three-axis sensor data $\boldsymbol{a}(t)$ can be represented at level p $(p = 0, 1, \ldots, p_{max})$ as follows;

$$\boldsymbol{a}(t) = \sum_{q=0}^{2^p-1} \boldsymbol{u}^{(p,q)}(t) \;\; (q = 0, 1, \ldots, 2^{p_{max}} - 1) \quad (3)$$

where $\boldsymbol{u}^{(p,q)}(t)$ is q-th of decomposed data at level p.

3.1.3 Best Basis based on Information Entropy

In order to find the best basis for $\boldsymbol{a}(t)$, we define a cost function of information entropy $L_i(p, q)$ as follows;

$$L_i(p,q) = -\sum_{t_s} |u_i^{(p,q)}(t_s)|^2 \log |u_i^{(p,q)}(t_s)|^2$$

$$(i = x, y, z) \quad (4)$$

We find the pairs of $(p_{B_i}(k), q_{B_i}(k))$ that minimize $L_i(p, q)$.

$$(p_{B_i}(k), q_{B_i}(k)) = \arg \max_{p,q} (L_i(p,q)) \quad (5)$$

where k is kth of the best basis. Then, $a_i(t_s)$ $(i = x, y, z)$ can be represented as follows;

$$a_i(t_s) = \sum_k u_i^{(p_{B_i}(k), q_{B_i}(k))}(t_s) \quad (6)$$

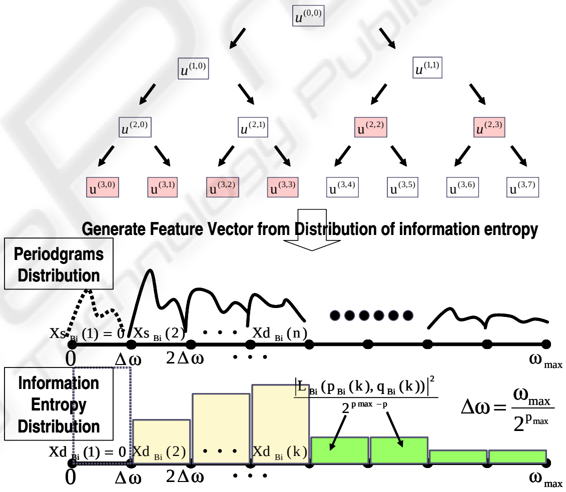


Figure 2: Feature Vectors generated by Wavelet Packet Decomposition.

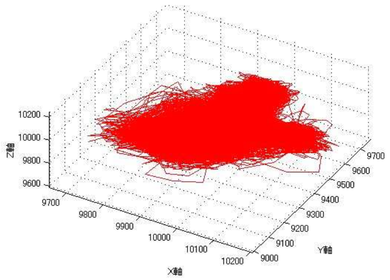
3.1.4 Definition of Feature Vectors

As a wavelet packet is a kind of bandpass filter, each $u_i^{(p_{B_i}(k), q_{B_i}(k))}$ are filtered at their frequency ranges. We calculate both periodgrams and information entropy of each feature each $u_i^{(p_{B_i}(k), q_{B_i}(k))}$ $(i = x, y, z)$ (see Figure 2). Next, we calculate the feature vectors $\boldsymbol{Xs}_{B_i}$ by distributing the periodgrams of the best basis to the maximum level p_{max} and we define the feature vectors $\boldsymbol{Xs}$ as follows;

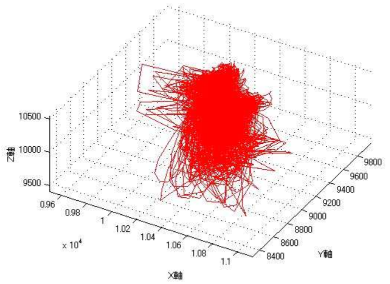
$$\boldsymbol{Xs} = \left(\boldsymbol{Xs}_{B_x} \; \boldsymbol{Xs}_{B_y} \; \boldsymbol{Xs}_{B_z} \right) \quad (7)$$

where

$$\boldsymbol{Xs}_{B_i} = (Xs_{B_i}(1), \ldots, Xs_{B_i}(n), \ldots, Xs_{B_i}(N_{max})) \quad (8)$$



(a) Walking on flat road



(b) Going up stairs

Figure 3: Trajectory of Three-Axis sensor data.

$$Xs_{B_i}(n) = \frac{1}{2\Delta T}\sum_{t_s} |u_i^{(p_{B_i}(k),q_{B_i}(k))}(t_s)exp(-j\omega_n t_s)|^2 \quad (9)$$

and

$$\omega_n = \frac{\omega_{max}}{2^{p_{max}}} n(p,q) \quad (10)$$

$$2^{p_{max}-p}q \leqq n(p,q) \leqq 2^{p_{max}-p}(q+1)-1 \quad (11)$$

where ω_n represents frequency. The above formula does not contain $Xs_{B_i}(0)$, the DC component, because basically it contains gravity of center, moreover, not DC but AC components contain effective features for determining user behavior. Conventional methods have a critical constraint in that sensors must be fixed to specific parts of the user's body. Therefore, in order to extract what context they wanted, they selected specific components of three-axis sensor data or used the magnitude of three-axis sensor data. However, these methods can extract only a limited range of contexts because they treat the sensor data as scalar data, not vector data.

Figure 3 shows trajectories of three- axis acceleration data for the user actions of walking on a flat road and up stairs. Both actions yield characteristic three-dimensional patterns, a kind of chaos attractor. It is difficult to detect features accurately by projecting just one or two axes. Therefore, we use all components of sensor data as vector data which allows us to assess as many kinds of user behavior as possible. We define two kinds of feature vectors, $\boldsymbol{Xs}_{B_i}$ and $\boldsymbol{M}_i$, as vector and statistic of information entropy, respectively. $\boldsymbol{Xs}_{B_i}$ is the feature vector used to detect similar sensor data while $\boldsymbol{M}_i$ is for analyzing the attribution of sensor data.

At first, we define the feature vector $\boldsymbol{Xd}$ which is composed of feature vector $\boldsymbol{Xd}_{B_i}$ by distributing the information entropy of the best basis to the maximum level p_{max} as follows;

$$\boldsymbol{Xd} = \left(\boldsymbol{Xd}_{B_x}\ \boldsymbol{Xd}_{B_y}\ \boldsymbol{Xd}_{B_z} \right) \quad (12)$$

where

$$\boldsymbol{Xd}_{B_i} = (Xd_{B_i}(1), \ldots, Xd_{B_i}(k), \ldots, Xd_{B_i}(2^{p_{max}-1})) \quad (13)$$

$$Xd_{B_i}(k) = \frac{|L_{B_i}(p_{B_i}(k), q_{B_i}(k))|^2}{2^{p_{max}-p}} \quad (14)$$

This allows us to calculate the momenta $\boldsymbol{M}_i$ as follows;

$$\boldsymbol{M}_i = (M_i(1), \ldots, M_i(l), \ldots, M_i(N_l)) \quad (15)$$

where

$$M_i(\boldsymbol{Xd}_{B_i}) = \sum_l (l - M(1))^n Xd_{B_i}(l) \quad (n = 1,2,3,4) \quad (16)$$

and

$$M(1) = \frac{\sum_l lXd_{B_i}(l)}{\sum_l Xd_{B_i}(l)} \quad (17)$$

Afterward, we use the $\boldsymbol{Xs} = (\boldsymbol{Xs}_{B_x}\ \ \boldsymbol{Xs}_{B_y}\ \ \boldsymbol{Xs}_{B_z})$ and $\boldsymbol{M} = (\boldsymbol{M}_x\ \boldsymbol{M}_y\ \boldsymbol{M}_z)$ as feature vectors of frame in three-axis sensor data for extracting personal user context.

3.2 Training Data Learning Stage

Most human action such as walking exhibits fluctuations and noise. Moreover, they are always changing. This suggests that non-linear approaches to clustering them are better than linear ones such as principle components analysis, because the linear approaches are sensitive to the influences of fluctuation and noise. From among non-linear approaches available, we choose the Khonen self-organizing map (KSOM) (Kohonen.,) for clustering personal user context such as human behavior. This is because KSOM offers competitive learning and also has reasonable computational costs; it eliminates the need to set the number of clusters beforehand. However,

Table 1: Dictionary Data Example.

Feature Vector	Personal User Context
$\boldsymbol{Xd}(t_1)$	Walking at normal speed
$\boldsymbol{Xd}(t_2)$	Walking fast
$\cdots$	$\cdots$
$\boldsymbol{Xd}(t_k)$	going up to stairs
$\boldsymbol{Xd}(t_{k+1})$	going down to stairs
$\cdots$	$\cdots$
$\boldsymbol{Xd}(t_{k+N})$	running

KSOM has several problems. One of them is its weakness in deciding cluster borders on the learnt feature map because of the existence of uncertain cells. In order to overcome this problem, we apply a probabilities-based technique after learning the feature map. Under our proposal, we can prepare dictionary data (see Table 1) whose labels are kinds of personal user context. This allows us to calculate the probability. By using the probability information, we can obtain the sensor data most representative of each personal user context.

3.2.1 Competitive Training by KSOM

We use feature vector $\boldsymbol{Xs}$ as main feature vector and feature vector $\boldsymbol{M}$ as auxiliary feature vector, respectively. Because $\boldsymbol{Xs}$ can contain more detail characters of sensor data than $\boldsymbol{M}$. At this stage, we use feature vector $\boldsymbol{Xs}$ as input vectors $\boldsymbol{X}_i$ to train KSOM. They are generated from three-axis sensor data associated with each personal user context. In training, the KSOM measure and update rule are set as follows.

$$\|\boldsymbol{X}_i - \boldsymbol{X}_{n_c}\| = \min_c\{\|\boldsymbol{X}_i - \boldsymbol{X}_c\|\} \tag{18}$$

$$\begin{aligned}\boldsymbol{X}_{n_c}^{Nearest}(t+1) &= \boldsymbol{X}_{n_c}^{Nearest}(t)\\ &+ \delta(t)h_{ci}(r(t))\{\boldsymbol{X}_i(t) - \boldsymbol{X}_{n_c}^{Nearest}(t)\}\end{aligned} \tag{19}$$

where $\boldsymbol{X}_{n_c}(t)$, $\boldsymbol{X}_{n_c}^{Nearest}(t)$, $\boldsymbol{X}_i(t)$ and c are the vectors of the cell nearest to $\boldsymbol{X}_i(t)$, the neighborhood vectors of $\boldsymbol{X}_{nc}(t)$, the input feature vectors at time t in training and a cell in the KSOM, respectively. $\delta(t)$, h_{ci} and $r(t)$ are the learning rate, the neighborhood kernel around the winning unit n_c, and neighborhood radius, respectively. After training, we can obtain a 2D array of cells on the feature map represents the distribution of training data. The cells also hold information about the number of times each personal user context appeared.

3.2.2 Detecting Representative Vectors for Personal User Context

Based on the information about the number of times in each personal user context appears, we make a probability map by calculating the personal user context appearance probabilities in each cell. We define the personal user context appearance probabilities $P(C_{n_c}|\boldsymbol{X}_{n_c})$ as follows;

$$P(C_{n_c}|\boldsymbol{X}_{n_c}) = \frac{P(\boldsymbol{X}_{n_c}|C_{n_c})P(C_{n_c})}{\sum_{k_c} P(\boldsymbol{X}_{n_c}|C_{k_c})P(C_{k_c})} \tag{20}$$

We also define probabilities $P(\boldsymbol{X}_{n_c}|C_{n_c})$ and $P(C_{n_c})$.

$$P(\boldsymbol{X}_{n_c}|C_{n_c}) = \frac{N(\boldsymbol{X}_{n_c}(C_{n_c}))}{\sum_{k_c} N(\boldsymbol{X}_{k_c}(C_{k_c}))} \tag{21}$$

$$P(C_{n_c}) = \frac{N(C_{n_c})}{\sum_{k_c} P(C_{k_c})} \tag{22}$$

where C_{n_c} and $\boldsymbol{X}_{n_c}$ represent the appearance of personal user context n_c and the selected cell on the feature map as nearest to input vector $\boldsymbol{X}$, respectively. $P(\boldsymbol{X}_{n_c}|C_{n_c})$ and $P(C_{n_c})$ are the conditional probability of appearance C_{n_c} given $\boldsymbol{X}_{n_c}$ and the prior probability of C_{n_c}, respectively. $N(\boldsymbol{X}_{k_c}(C_{k_c}))$ represents amount of $\boldsymbol{X}_{k_c}(C_{k_c})$.

We use the probability map of personal user context appearance for two purposes; to find representative feature vectors for personal user context and to output possibilities of the existence of personal user context from three-axis sensor data. To realize the former, we find the representative feature vectors when the probability satisfies the below condition.

$$P(C_{n_c}|\boldsymbol{X}_{n_c}) > Th_p \tag{23}$$

where Th_p is a threshold to judge whether $\boldsymbol{X}_{n_c}$ is representative of C_{n_c}. We can obtain the best representative sensor data corresponding to $\boldsymbol{X}_{n_c}$ which are found as described above.

3.2.3 Training by using Pseudo Sensor Data

In order to achieve robustness with regard to the positions of the sensor, we have already proposed a method based on non-linear simultaneous equations. Unfortunately, its computational cost is excessive and it often fails to provide the real-time processing needed by ubiquitous services. Therefore, the feature map includes the possibilities generated from not only observed training data but also pseudo-data, which are generated from the best representative sensor data by rotational transformation. After obtaining the best representative sensor data (see section 3.2.2), we transform them using the rotation matrix as described below;

$$\begin{aligned}\boldsymbol{a}_p = &\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha_n & -\sin\alpha_n \\ 0 & \sin\alpha_n & \cos\alpha_n \end{pmatrix}\begin{pmatrix} \cos\beta_n & 0 & \sin\beta_n \\ 0 & 1 & 0 \\ -\sin\beta_n & 0 & \cos\beta_n \end{pmatrix}\\ &\times\begin{pmatrix} \cos\gamma_n & -\sin\gamma_n & 0 \\ \sin\gamma_n & \cos\gamma_n & 0 \\ 0 & 1 & 0 \end{pmatrix}\boldsymbol{a}_r\end{aligned} \tag{24}$$

Table 2: Probability Map Example.

Feature	Probabilities [%]			
Vector	WN	WF	$\cdots$	US
$\boldsymbol{X}_1$	82.3	0.0	$\cdots$	17.7
$\boldsymbol{X}_2$	73.7	25.6	$\cdots$	0.7
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$\boldsymbol{X}_k$	28.6	0.0	$\cdots$	71.4

where α_n, β_n and γ_n are the angles around x, y and z-axis, respectively. By increasing these angles in steps of $\Delta\theta$, we can obtain a variety of pseudo sensor data $\boldsymbol{a}_p$. Subsequent training with each pseudo-data in angle sets $(\alpha_n, \beta_n, \gamma_n)$ following the approach given in section 3.2.1 and 3.2.2, yields as many number of feature maps as the angle sets.

3.3 Personal User Context Extraction Stage

In order to recognize personal user context from three-axis sensor data, we use all feature maps with the possibilities trained by both observed and pseudo sensor data. The specific pattern detection stage proceeds as follows;

1. Extract three-axis sensor data
2. Convert them into frame data and calibrate them in terms of scale
3. Decompose them using the best basis approach by wavelet packet decomposition and calculate feature vectors from their spectrogram
4. Calculate residential errors between them and the nearest cells in each feature map
5. Find the most suitable cell $\boldsymbol{X}_{bm}$ for the below conditions;

$$\begin{cases} \boldsymbol{X}_{bm} & = & \arg\min_{\boldsymbol{X}_c}\{\|\boldsymbol{X} - \boldsymbol{X}_c\|\} \\ \|\boldsymbol{X}_{bm} - \boldsymbol{X}_c\| & < & Th_{bm} \\ Corr(\boldsymbol{M}_{bm}, \boldsymbol{M}_c) & > & Th_{corr} \end{cases} \tag{25}$$

where $\boldsymbol{M}_{bm}$ and $\boldsymbol{M}_c$ are the feature vectors of the information entropy distributions of $\boldsymbol{X}_{bm}$ and $\boldsymbol{X}_c$, respectively. $Corr(\boldsymbol{M}_{bm}, \boldsymbol{M}_c)$ means the normalized crosscorrelation between $\boldsymbol{M}_{bm}$ and $\boldsymbol{M}_c$. Th_{bm} and Th_{corr} are thresholds to judge feature vector similarity.

An example of the probability map described in the previous section in given in the below table. Retrieving the most suitable cell $\boldsymbol{X}_{bm}$ from Table 2 yields personal user context.

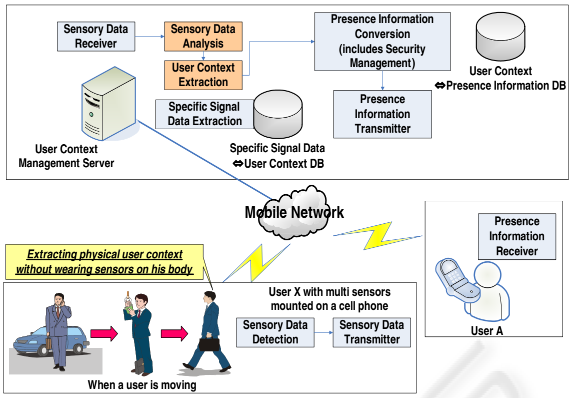


Figure 4: PerContEx System.

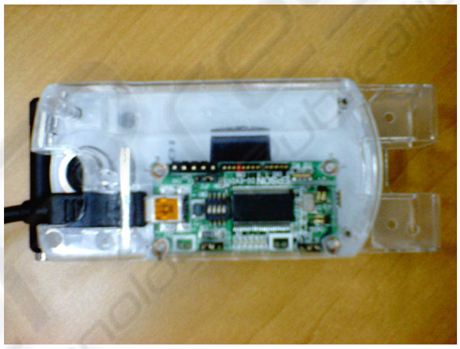


Figure 5: A Three-Axis Accelerometer mounted on A Cell Phone.

4 EXPERIMENTS

We validate of our proposed method using a prototype system. The experiments targeted four user contexts associated with movement: walking, going up/down stairs, walking rapidly, and running. These targets are thought to be relatively hard to discriminate as well as being extremely useful in health advice services such as calculating a person's calorie consumption.

4.1 Experimental Conditions

Our system is composed of a cell phone with three-axis sensor and a PerContEx server. The cell phone transmits the sensor data to the PerContEx server. The PerContEx server extracts the user's context the data and send the results of context extraction to the cell phone or another personal computer (see Figure 4). The experimental conditions of the system are as follows; We show experimental conditions of the system as follows;

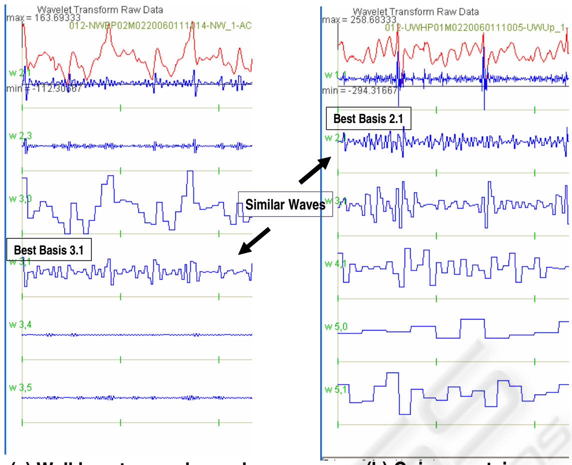
- System specifications

  - Cell phone: IMT-2000 Terminal (NTT DoCoMo P900i, see Figure 5)
  - Sensor: three-axis accelerometer (Hitachi: H48C)
  - PerContEx server: Personal Computer (DELL: Dimension9150)
    - OS:WindowsXP Pro.
    - CPU: PentiumD830 $3[GHz]$
    - MEM:$2[GB]$
- Parameters
  - Mother wavelet: Harr Function
  - Scale level: 5
  - KSOM: $15[cell] \times 10[cell]$
- Data specifications
  - Personal user context: five kinds of gait pattern
    - walking at normal speed (WN),
    - going up stairs (US)
    - going down stairs (DS)
    - walking fast (WF)
    - running (RN)
  - Carrying style: placed in user's breast pocket and hip packet
  - Real data: data sampling: $100[Hz]$
    - training data $30[min]$
    - test data $50[min]$
  - Pseudo-data:
    - step of rotation angle$15[deg]$
    - range of rotation angle (round x, y, z-axis, respectively) from $0[deg]$ to $45[deg]$
  - frame length: $3[sec]$ (overlap time: $1.5[sec]$)

4.2 Results and Estimation

Figure 6 shows the results of identifying "walking at normal speed" and "going up stairs". As both movements contain the behavior of "walking", we can see similar wave patterns at the best basis level 2 and 3. Thus, the other best bases are useful for discriminating between "walking at normal speed" and "going up stairs" and noise. After feature extraction, we made a feature map with personal user context appearance probabilities by using real data and pseudo sensor data as described above. Figure 7 shows the resulting feature map. We can see that each personal user context has its own feature map. Table 3 Lists the accuracy of personal user context extraction. In the experiment, we identified the context which has the maximum probabilities in probabilities of the other contexts.

According to the above Table 3, our proposed algorithm can extract and discriminate the above gaits as personal user context independent of cell phone position. Other tests showed that the proposed algorithm can extract not only walking (WN) and running (RN) but also going up/down stairs (US/UD) and walking fast (WF). The most difficult gait to identify was DS ("going down stairs"). We think that DS has weaker constraints on body movement than the other gaits such as US ("going up stairs") and RN ("running"). Many DS events were misidentified as "walking at normal speed" or "walking fast". We need to enhance the algorithm to provide a more detail analysis of the sensor data.



(a) Walking at normal speed (b) Going up stairs

Figure 6: Sensor Data Decomposited by Wavelet Packet.

Table 3: the accuracy of personal user context extraction.

Cell Phone Position	Probabilities $[\%]$				
	WN	US	DS	WF	RN
Breast pocket	83.2	79.3	76.6	88.4	94.2
Hip pocket	85.1	88.5	72.1	94.9	92.0
Average	84.2	84.1	74.4	91.6	93.1

5 CONCLUSION

We proposed an algorithm to extract personal user context; it uses feature extraction based on wavelet packet decomposition and KSOM with the context appearance probabilities. Experiments on a prototype system showed that in case that the cell phone with the sensor was not fixed in user's pockets, it could extract personal user context at the accuracy of about $88[\%]$. Moreover, we showed that, with only a single three-axis accelerometer, it could extract personal user contexts such as walking, running, walking fast, and going up/down stairs at the accuracy of about $80[\%]$.

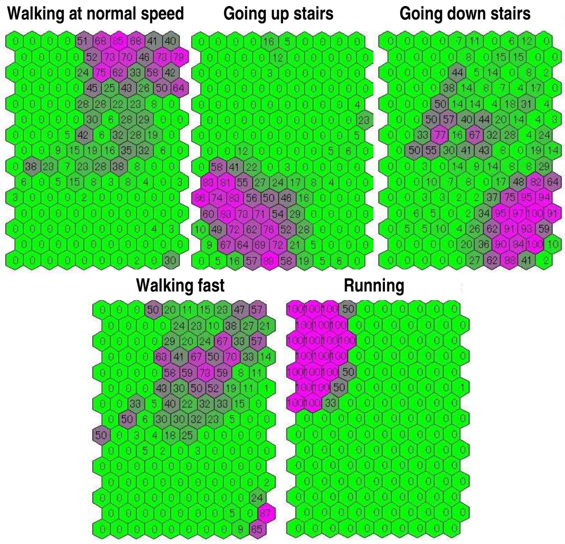


Figure 7: Learnt Feature Map with Probability Information.

The results prove that the algorithm can not only reduce the constraints placed on users but also be applied for ubiquitous services which require personal user context.

Future works is to realize a more comprehensive analysis of sensor data in order to identify personal user context in more complicated situations such as carrying the cell phone by hand. We will then apply the algorithm to ubiquitous services such as presence and health care services.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. K. Imai, NTT DoCoMo Network Labs for his encouragement. We also thank Mr. M. Kinoshita, NTT-IT Corporation and Mr. H. Miyoshi, Seiko Epson Corporation for their technical support in the experiments.

REFERENCES

Clarkson, B., Mase, K., and Pentland, A. (2000). Recognizing user's context from wearable sensor's: Baseline system. Technical Report 519.

DeVaul, R. W. and Dunn, S. *Real-Time Motion Classification for Wearable Computing Applications*. http://www.media.mit.edu/wearables/mithril/realtime.pdf.

Healey, J. and Logan, B. *Wearable Wellness Monitoring Using ECG and Accelerometer Data*. http://www.hpl.hp.com/techreports/2005/HPL-2005-134.pdf.

Ishikawa, Y. (2000). *Wavelet Analysis for Clinical Medicine*. IGAKU-SHUPPAN.

Iso, T., Kawasaki, N., and Kurakake, S. (2005). Personal context extractor with multiple sensors on a cell phone. In *7th IFIP International Conference on Mobile and Wireless Communications Networks, 2005*.

K. V. Laerhoven, A. Schmidt, H. W. G. (2003). Limitations of multi-sensor context-aware clothing. 7(3).

Kern, N., Antifakos, S., Schiele, B., and Schwaninger, A. (2004). A model for human interruptability: experimental evaluation and automatic estimation from wearable sensors. In *8th IEEE International Symposium on Wearable Computers*.

Kohonen., T. *SOM Toolbox 2.0*. The Laboratory of Information and Computer Science in the Helsinki University of Technology, http://www.cis.hut.fi/projects/somtoolbox/.

Krause, A., Siewiorek, D. P., Smailagic, A., and Farringdon, J. (2003). Unsupervised, dynamic identification of physiological and activity context in wearable computing. In *7th IEEE International Symposium on Wearable Computers*, pages 8–17.

M. Unuma, K. Kurata, A. T. T. H. (2004). Autonomy position detection by using recognition of human walking motion. J87-A(1):78–86.

Miao, T., Shimizu, T., Miyake, S., Hashimoto, M., and Shimoyama, O. (2003). Alteractions of complexity measures during subsidiary task in multi-attribute operations. In *Human Interface Symposium*, pages 791–792.

NTT DoCoMo, Inc. http://www.nttdocomo.com/home.html.

Randell, C. and Muller, H. (2000). Context awareness by analysing accelerometer data. Technical Report CSTR-00-009.

Si, H., Kawahara, Y., Kurasawa, H., Morikawa, H., and Aoyama, T. (2005). A context-aware collaborative filtering algorithm for real world oriented content delivery service. In *Ubicom2005 Metapolis and Urban Life Workshop*.

Siewiorek, D., Smailagic, A., Furukawa, J., Krause, A., Moraveji, N., Reiger, K., Shaffer, J., and Wong, F. L. (2003). Sensay: A context-aware mobile phone. In *7th IEEE International Symposium on Wearable Computers*, pages 248–257.