Human Activity Recognition with User-Free Accelerometers in the Sensor Networks

Shuangquan Wang
Jie Yang
Institute of Image Processing & Pattern Recognition
Shanghai Jiaotong University
Shanghai, China, 200030
E-mail: {wangshuangquan, jieyang}@sjtu.edu.cn

Ningjiang Chen
Xin Chen
Qinfeng Zhang
Philips Research East Asia
Shanghai, China, 200070
E-mail: {james.nj.chen, mylan.chen, qinfeng.zhang}@philips.com

***Abstract*—Many applications using wireless sensor networks (WSNs) aim at providing friendly and intelligent services based on the recognition of human's activities. Although the research result on wearable computing has been fruitful, our experience indicates that a user-free sensor deployment is more natural and acceptable to users. In our system, activities were recognized through matching the movement patterns of the objects, which tri-axial accelerometers had been attached to. Several representative features, including accelerations and their fusion, were calculated and three classifiers were tested on these features. Compared with Decision Tree (DT) C4.5 and Multiple-Layer Perception (MLP), Support Vector Machine (SVM) performs relatively well across different tests. Additionally, feature selection will be discussed for better system performance for WSNs.**

I. INTRODUCTION

The technological advances in wireless sensor networks (WSNs) enables the development of various applications, such as smart house, intelligent office and health care [1,2], which intend to provide appropriate and friendly services based on the recognition of human's activities. WSNs use many kinds of sensors for different applications, including sound, infrared, light, temperature, accelerometer, magnetometer, etc. Because of their inherent property, accelerometers are most often employed in human activity recognition to measure the speed and orientation changes of human bodies and objects.

Since the movements of human bodies are the main character of their activities, many research groups [3, 4, 5] focus their research on the wearable system. Ambulation, typing, talking and some other activities were distinguished in [6] with five small bi-axial accelerometers. In [7] and [8], similar daily activities, such as standing, walking, climbing up/down stairs and brushing teeth, were analyzed with several classifiers. Paul Lukowicz et al. [9] used body worn sensors to automatically track the progress of maintenance or assembly tasks in a wood shop. [10] measured tri-axes accelerations in freestyle swimming on Japanese top-level college swimmers to analyze and evaluate swimmers' stroke technique.

Although the research result on wearable computing has been fruitful, during our investigation, the users of these systems don't always feel comfortable wearing sensors. We need a more natural way to recognize human's activities and in a more friendly manner. In addition, human's activities are usually complicated and ambiguous, different with each other. The more natural and unbending the human's activities are, the more difficult to recognize.

Human's activities can be represented from three aspects. 1) The most direct method is identifying the movements of human bodies. The wearable systems we introduced above are a typical representation. These wearable sensors mounted on special places of the bodies can be used to recognize even very small activities. 2) A person's activities can also be detected from the movements of the objects associated with these activities. For example, we can infer whether somebody is drinking according to the movements of a cup. No movement, no drinking. But at the same time, deceptive movements must to be excluded through machine learning of the classifier. 3) Person-object interaction may provide clues of actions from a person's approaching to, touching, or moving an object. The recognition results are not as accurate as the first two methods. It does not extract and match the activity patterns as the second method does.

Wearable sensors are often used in health care or medical surveillance, for special applications. While for daily activity recognition, as discussed above, wearing these sensors is uncomfortable for users. The person-object interaction often plays an assistant role to other information. Therefore, we focused on recognizing human's activities from the movements of the objects which keeps the users free of wearing any sensors. Additionally, the analysis on the movement patterns of the object excludes the affect of human's ambiguous activities.

Compared with other methods, which classify a time series based on the values of each time point, the feature-based classification method [11] is verified to be efficient. Representative features are statistical quantities of the digital signal itself. The physical and inherent attributes of accelerometers, accelerations, are often ignored. These attributes enable the recognition of movement or incline of the accelerometer when it is moving or not. In this paper, the acceleration related features are added into the training and testing sets to embody the human's activities. Additionally, feature selection is also considered for better system performance.

Activity recognition is formulated as a classification problem. That is to say, the recognition capability largely relies on the classification algorithm. On the one hand, the algorithm should classify the similar actions it learned from the training. On the other hand, it needs to identify new unknown actions as accurately as possible. The cross-validated test and independent data test were employed with three commonly used classifiers (Decision Tree (DT), Multiple-Layer Perception (MLP), and Support Vector Machine (SVM)) for these two above-mentioned aspects. The results indicated the SVM was the best one for recognizing activities in our scenario.

The rest of the paper is organized as follows. In Section 2, we introduce the accelerometer and its recognition rationale. Feature extraction is presented in Section 3. We describe the classification and feature selection in Section 4. Experiments and discussion on feature selection will be illustrated in Section 5 and Section 6. Section 7 concludes the paper.

II. ACCELEROMETERS

Accelerometers are used to detect the acceleration using the piezoelectric technology. They can be deployed to monitor a person's motion and allow the system give correct feedback.

In our experimental environment, an office room, typical activities include writing, phoning, drinking, walking, working on the computer, and talking with others. Of all these activities, walking can be detected by location information, while working on computer and talking can be discovered through other kinds of sensors or their fusion. Accelerometers are used to recognize writing, phoning and drinking with their strong capability of detecting the changes of speed and orientation. In our recognition framework, each activity was decomposed into several element activities [12]. For instance, during the phoning process, three element activities are: someone is near the phone, sound is detected by the microphone and the telephone receiver is picked up. Although all of these element activities happening simultaneously means a high probability of phoning, the first two has no direct and inevitable evidence to prove that a phone call is being made. The most obvious proof is the position and orientation change of the accelerometer mounted on the receiver. This is the same for drinking and writing. Of course, one may ask: when I move the receiver or pick it up, what will happen? Yes, we will assure that you are really phoning if its direction and movement are similar enough with your phoning activity in practice. Otherwise, the classifier should distinguish this deceptive action.

The tri-axial accelerometer we used detects and transforms changes in capacitance into an analog output voltage, which is proportional to acceleration. This voltage is digitized by an on-board A/D converter and is accessed via a Serial Peripheral Interface (SPI) [13].

A linear model can be used to describe the mapping from electrical data to acceleration [14]:

$$\begin{cases} a_x = k_x v_x + b_x \\ a_y = k_y v_y + b_y \\ a_z = k_z v_z + b_z \end{cases} \quad (1)$$

where v_x, v_y, v_z are the output voltages along three axes, and a_x, a_y, a_z stand for the corresponding accelerations the real world. k_x, k_y, k_z denote the sensitivity parameters, which describe the gain of the sensor, and b_x, b_y, b_z are the zero-g level (offset), which describe the deviation of an actual output signal from the true output signal if there is no acceleration present.

These parameters can be determined by a simple method. Position each of the three axes of the sensor in the direction of gravity and in the opposite direction. Then for each case, note down the sensor outputs and calculate the related parameters directly. For example, when the x-axis is along with and contrary to gravity direction, the sensor outputs are v_{x1} and v_{x2} respectively. The sensitivity and offset parameters are:

$$\begin{cases} k_x = \dfrac{2g}{v_{x1} - v_{x2}} \\ b_x = -\dfrac{k_x (v_{x1} + v_{x2})}{2} \end{cases} \quad (2)$$

where g is the gravitational acceleration. We set $g = 9.8$ in our experiments.

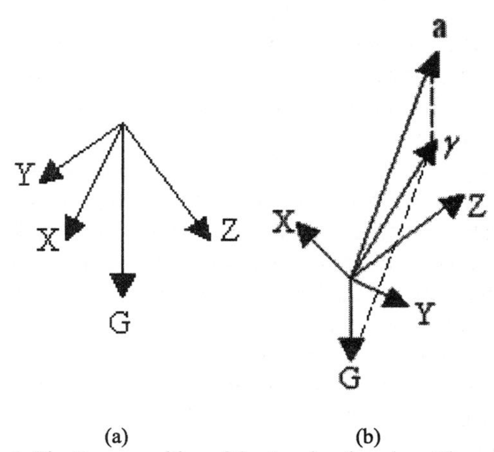

Figure 1. The Decomposition of the Acceleration along Three Axes

As we know, when the sensor is static, the total acceleration of the sensor is due to the gravitational acceleration, which means [14]:

$$a_x^2 + a_y^2 + a_z^2 = g^2 \quad (3)$$

The gravitational acceleration orthogonal is decomposed along three axes (see Fig.1 (a)). If the sensor is dynamic, the accelerations are decompositions of the synthesization of the real acceleration and the gravitational acceleration (see Fig.1 (b)). γ is the synthesization of G and a. X, Y and Z are decompositions of along three axes.

III. FEATURE EXTRACTION

Features were extracted from the raw accelerometer data using a window size of 64 with 32 samples overlapping between consecutive windows. This window size enables fast computation of FFTs for calculating some features. Feature extraction on a sliding window with 50% overlap has demonstrated to be successful in previous work [15]. At a sampling frequency of 32Hz, each window represents 2 seconds. According to our experience, a much shorter window can't seize the activity pattern. However, if it is too long, there will be a delayed response.

Six features were extracted from each of the three axes of the accelerometer. The features were: acceleration, mean, standard deviation, energy, frequency-domain entropy and correlation between axes. In addition, there was a fusion feature, Δ, of three directions representing vibration of the sensor, giving a total of nineteen attributes.

The accelerations, as we mentioned above, are the decomposition along three axes. In addition, the heavier vibration of the accelerometer is, the farther the value of $a_x^2 + a_y^2 + a_z^2$ deviates from g^2. We use $\Delta = \left| a_x^2 + a_y^2 + a_z^2 - g^2 \right|$ as a feature to represent the vibration of the sensor.

The DC feature is the mean acceleration value of the sliding window [6]. Standard deviation was used to characterize the stability of the signal. The energy feature, which can capture the data periodicity [8], was calculated as the sum of the squared discrete FFT component magnitudes of the signal.

Frequency-domain entropy is calculated as the normalized information entropy of the discrete FFT component magnitudes of the signal. This may help to discriminate activities with similar energy values. The correlation feature aims to find out the relationship between axes in three directions [6].

It is obvious that the ranges of these features are different. The correlation ranges from 0 to 1, while energy is often much larger than 1. We normalized all features into $[-1,1]$ for later classification.

IV. CLASSIFICATION AND FEATURE SELECTION

A. Classification

Activity recognition was performed using Decision Tree (DT) C4.5, Multiple-Layer Perception (MLP) neural networks, and Support Vector Machine (SVM). These classifiers have been used in many activity recognition applications [8, 11].

These three classifiers were trained and tested in three ways for each activity. In the self-consistency test, we combined the positive samples and the negative samples of each activity to form the training set. The classifiers were then trained and tested with the same data set.

In the cross-validated test, the above data set was randomly partitioned into n mutually exclusive and exhaustive ones. During this test, each in the n sets was in turn taken out and all the rule parameters were calculated based on the remaining sets. In other words, the actions were predicted using the rule parameters derived from all the other actions except those being identified.

Usually the training set is limited. The generalization capability of the classifier is more important for recognition. We use the leave-one-subject-out validation test to evaluate the classifier' ability to recognize unacquainted actions. Classifiers were trained on activity data for all subjects except one. The classifiers were then tested on the data for only the subject left out of the training data set. This process was repeated for all subjects.

B. Feature Selection

In almost all recognition applications mentioned above, the data was processed in a centralized way. In WSNs, however, only relevant data, which was preprocessed at the nodes, was downloaded to a centralized location. Dimensionality reduction provides a scalable method for this question where careful attention is paid to computing, communication, storage and human factors in order to use them in a near optimal fashion.

In general, reduction methods can be divided into two main categories. Feature selection techniques attempt to reduce dimensionality by discarding some of the original features, whereas feature transform methods attempt to map the original features into a lower dimensional subspace [16]. For the latter, the principal component analysis (PCA) and independent component analysis (ICA) are applied on the 24 dimensional feature vector in [7]. The SVM based feature selection method [17] will be used in our experiments.

V. EXPERIMENTS AND DISCUSSION

A. System Setup

The acceleration data was collected using the KXP74 accelerometers [13]. Its sensitivity is programmed from

$-2.0g$ to $+2.0g$. This can fully meet our demands, for previous work has shown promising activity recognition results from $\pm 2.0g$ acceleration data [18].

The sensor was mounted onto the sensor board and sealed hermetically. The sensor node attached with the sensor board wirelessly transmitted the data via RF signal to the base station, which is connected to the serial port of a laptop through an interface board and a serial cable.

For convenience, the node was fixed to the rear of the telephone receiver for recognizing the phoning activity. The other two were attached to the base of the cup and on the top of the pen (Fig. 2), for recognizing the other two activities.

Figure 2. The Deployment of Accelerometers in Drinking, Phoning and Writing

B. Data Collection

The sensor data of the accelerometer has the following attributes: time, acceleration along x-axis, acceleration along y-axis and acceleration along z-axis. It was collected with a sampling frequency of 32 Hz and no noise filtering was carried out.

Each of these three activities was performed by four subjects. For each activity, everyone executed two different groups of experiments: positive actions and negative actions in accordance to whether this activity takes place or not, each lasted 5 minutes. Any action that represented this activity happening was included in the positive samples. Otherwise, they would be classed into the negative samples. During writing, for example, the subject can write on the table or on the blackboard, slowly or fast, and any gesture he wants. These actions were considered as writing. On the contrary, if the subject rotated the pen with his fingers, played with the pen randomly, all these actions were regarded as not writing. And the same case with drinking and phoning.

When the sensor is static, the gravitational acceleration orthogonal decomposes along three axes, which means that if we attach the sensor on the pen laterally, when rotating the pen along its long axes, the decomposition changes accordingly. The same action may present different acceleration values along three axes when the angle between the real acceleration and the gravitational acceleration changed. The subjects were, therefore, encouraged to do positive actions with diverse rotation angles of the pen to reduce the gravity effect. The same impact will be taken into consideration in drinking and phoning data collection.

The subjects were asked to continuously perform each activity, positive and negative, respectively. Acceleration data collected between the start and stop times were labeled with the name of that activity. The interval between two actions was not excluded despite the fact that this activity label around these times may not correspond to the actual activity.

C. Experimental Results

In the self-consistency test, the three classifiers all perform well with an overall accuracy rate of more than 95%. That is to say, these three classifiers are capable of recognizing the same data that they have learned.

In the cross-validated test, the fold number was set 10. We average the accuracy of 10 tests, and the results are listed in Table I.

TABLE I

THE ACCURACY OF THE CROSS-VALIDATED TEST

Classifier	Accuracy (%)		
	Drinking	Phoning	Writing
DT C4.5	98.00	99.05	97.66
MLP	86.95	93.00	88.41
SVM	89.48	94.32	88.42

As you can see from Table I, recognition accuracy is highest for the decision tree classifier, which got an average accuracy of about 98%. SVM is the second most accurate algorithm although there is only a little difference between it and the MLP. For activities, the phoning recognition is better than drinking and writing recognition. This result is reasonable, because the phoning activity is comparatively simple and the gravitational acceleration only has a little influence under the constraint of the receiver orientations and positions. For the actions learned by the classifiers, in general, similar ones can also be classified correctly, by and large. The DT algorithm displayed strongest capability.

In the leave-one-subject-out validation test, the mean and standard deviation for classification accuracy are summarized in Table II.

TABLE II

THE ACCURACY (MEAN $\pm$ STANDARD DEVIATION) OF THE LEAVE-ONE-SUBJECT-OUT TEST

Classifier	Accuracy (%)		
	Drinking	Phoning	Writing
DT C4.5	67.47 ± 7.37	81.71 ± 8.61	74.97 ± 3.82
MLP	72.50 ± 6.45	82.25 ± 3.56	80.34 ± 4.62
SVM	72.17 ± 6.21	84.28 ± 3.39	80.35 ± 5.89

Table II shows that, in the leave-one-subject-out test, the performance of DT algorithm decreased significantly. The accuracy of MLP and SVM were about 5% higher than that of DT for drinking and writing, and SVM outperformed the other two in phoning. SVM and MLP had nearly equivalent stability, a little better than DT.

In order to examine the performance of SVM and MLP more carefully, confusion matrices based on average accuracy and error rate of three activities are presented in Table III and Table IV.

TABLE III

CONFUSION MATRIX OF MLP RESULTS FOR THREE ACTIVITIES

	Positive samples	Negative samples
Positive samples	79.42%	20.58%
Negative samples	23.11%	76.89%

TABLE IV

CONFUSION MATRIX OF SVM RESULTS FOR THREE ACTIVITIES

	Positive samples	Negative samples
Positive samples	81.80%	18.20%
Negative samples	22.97%	77.03%

From the confusion matrix, we can see that SVM has a low false-classification rate. The false alarm ratio of MLP was a little higher than that of SVM, and the false negation ratio was over 2% more than that of SVM.

As a whole, in the three tests, SVM performed relatively well compared to other two classifiers in stability and generalization capability.

VI. DISCUSSION ON FEATURE SELECTION

In feature selection, using the SVM based feature selection method in [17], we sorted the nineteen attributes according to their emergence order. Beginning with the most important one, the attribute sequence of drinking was:

$P_z > P_x > P_y > C_{yz} > A_z > A_y > C_{xy} > S_y > C_{xz} > S_z > \Delta > G_z > A_x > G_y > S_x > G_x > E_z > E_y > E_x$,

where A is the acceleration; E is the mean; S is the standard deviation; G is the energy; P is the entropy, the subscript x, y, z stand for three axes. C_{xy}, C_{xz}, C_{yz} are correlations of two corresponding axes. Δ is the vibration feature. The sequence showed that, for drinking, the entropy feature, correlation feature and acceleration feature are more important than the energy feature, standard deviation feature and mean feature.

In the same way, the sequences of phoning and writing were:

$A_x > A_y > G_y > P_z > P_y > P_x > G_x > E_x > A_z > G_z > C_{yz} > C_{xz} > C_{xy} > S_x > \Delta > S_y > S_z > E_z > E_y$

and

$P_z > P_x > P_y > \Delta > A_x > G_x > G_y > S_x > S_y > S_z > A_y > A_z > G_z > E_x > C_{yz} > C_{xz} > E_z > E_y > C_{xy}$.

Starting from scratch, we added one attribute into the SVM classifier at a time following the sequence order. Then we can draw the accuracy curves as follows (Fig.3):

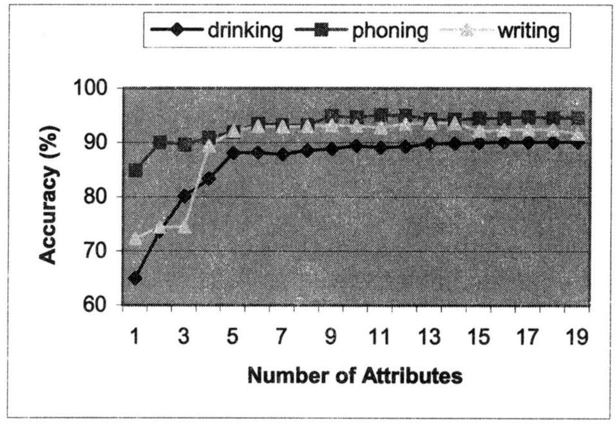

Figure 3: The Accuracy Curve of Drinking, Phoning and Writing

From the diamond line in Fig.3 we can see that when using five attributes, the accuracy of drinking has reached 88%, only 2% lower than that when all attributes were included.

The rectangle line in Fig.3 indicates that the accuracy of phoning could reach 90% using only two attributes, despite the highest value being about 95%. From the attribute sequence we know the first two attributes are accelerations in x-axis and y-axis. When phoning, in most cases, the sensor on the telephone receiver was nearly static. As we analyzed in Section 2, the acceleration of the three axes are a decomposition of gravitational acceleration. So, the orientation and position, which determined the decomposition, are the two most important characters in recognizing the phoning activity.

For writing recognition, as the triangle line shows, when the fifth attribute was included, the accuracy was high enough. From the sixth attribute, the accuracy stopped rising, only with small fluctuations. Also there was a sharp boost between point 3 and point 4.

For these three activities in an intelligent office, generally speaking, five attributes extracted from sensor data streams contained most of the information. Therefore, each node in sensor networks needs to calculate only five attributes and transmit the results to the center. This greatly reduced the amount of computation and communication needed.

At the same time, we found that, the acceleration related attributes contribute much to the accuracy. A_z was included in the first five attributes of the drinking attribute sequence. The phoning attribute sequence contained A_x and A_y, and, Δ and A_x play an active role in writing recognition.

It should be pointed out that the first several features selected were not the best combination. In the accuracy curve of writing, the second and third attributes (P_x and P_y) didn't improve the accuracy too much, but the fourth attribute (Δ) did. Compared with Δ individually, P_x and

P_y contributed more to the classifier if just one attribute is used. Unfortunately, the combination of P_x, P_y and P_z contributed little to the performance. The experiment showed that the accuracy can also reach 88% using only P_x and Δ. That is to say, the second and third attributes are redundant. As stated in [16], the main problem with feature selection techniques is that they are unable to find features that jointly maximize a predefined criterion. It proposed a method to select the most context features based on calculating characteristics of the different features and calculating similarity values for feature pairs using Gaussian kernels. On the contrary, [8] preserved valuable features by removing the ones that were less important than others. For us, selecting efficient features for activity recognition in sensor networks via analysis of relevance and redundancy is our future work.

VII. CONCLUSION

We used the objects attached with sensors for users during activity recognition using WSNs. It eliminates the uncomfortable sense of wearable computing. Acceleration features were analyzed in detail and tested with several other features by three classifiers. SVM performed the best when classifying the learned actions and identifying the new unknown actions with a strong stability and generalization capability. Feature selection was applied to select several most important attributes. The experiment results showed that the accuracy of activity recognition using only five attributes is high enough for these daily activities. The first five attributes of each sequence contained several acceleration attributes proved the validity of the acceleration feature in activity recognition. While we also analyzed the limitation of the used feature selection method, the analysis result can guide our further work.

REFERENCES

[1] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless Sensor Networks: A Survey. *Computer Networks*, vol. 38, no. 4, pp. 393-422, 2002.

[2] T. Hori , Y. Nishida et al. Sensor network for supporting elderly care home. Proceedings of IEEE. Sensors, 2004(2), pp. 575-578.

[3] G.D. Abowd, A.K. Dey, R. Orr, J.A. Brotherton. Context-awareness in wearable and ubiquitous computing. Virtual Reality, (3), pp.200--211, 1998.

[4] T. Starner, B. Schiele, A. Pentland. Visual contextual awareness in wearable computing. In IEEE International Symposium on Wearable Computers, Pittsburgh, October 19-20, 1998, pp. 50-57.

[5] C. Randell and H. Muller. Context awareness by analyzing accelerometer data. In Digest of Papers. Fourth International Symposium on Wearable Computers, pp. 175-176, 2000.

[6] L. Bao, and S.S. Intille. Activity recognition from user-annotated acceleration data. In Proceedings of PERVASIVE 2004. Springer, April, 2004.

[7] J. Mantyla, J. Himberg, T. Seppanen. Recognizing Human Motion with Multiple Acceleration Sensors. In IEEE International Conference on Systems, Man and Cybernetics, Tucson, USA, vol. (3494), pp. 747–752, 2001.

[8] N. Ravi, N. Dandekar, P. Mysore, M. Littman. Activity Recognition from Accelerometer Data. In Proceedings of the 17th Annual Conference on Innovative Applications of Artificial Intelligence, IAAI 2005.

[9] P. Lukowicz, J.A. Ward et al. Recognizing Workshop Activity Using Body Worn Microphones and Accelerometers. Pervasive 2004. LNCS 3001, pp. 18-32, 2004.

[10] Y. Ohgi, M. Yasumura, H. Ichikawa, C. Miyaji. Analysis of stroke technique using acceleration sensor IC in freestyle swimming. The Engineering of SPORT, pp.503-511, 2000.

[11] A. Nanopoulos , R. Alcock , Y. Manolopoulos. Feature-based classification of time-series data. Information processing and technology. Nova Science Publishers, Inc., Commack, NY, 2001.

[12] N.J. Chen, W. Fontijn, X. Chen and Q.F. Zhang. A Framework for Ambient Applications. International Conference on Sensor Networks. Montreal, Canada, August 14, 2005.

[13] http://www.kionix.com/Product%20Sheets/KXP74%20Series.pdf

[14] J.W. Tu, F. Liu. A Motion-Sensor-Based Interactive System. Global Signal Processing Expo and Conference. Santa Clara, USA, September 27, 2004.

[15] R.W. Devaul and S. Dunn. Real-Time Motion Classification for Wearable Computing Applications. Technical report, MIT Media Laboratory, 2001.

[16] P. Nurmi, P. Floréen. Online feature selection for contextual time series data (extended abstract). Presented in the PASCAL Subspace, Latent Structure and Feature Selection Workshop, Bohinj, Slovenia, 2005.

[17] G.Z. Li, J. Yang, G.P. Liu, L. Xue. Feature selection for multi-class problems using support vector machines, In Proceedings of 8th Pacific Rim International Conference on Artificial Intelligence(PRICAI-04), Sheraton, Auckland, August 9 to August 13, LNCS3157, pp. 292-300, 2004.

[18] N.Kern, B. Schiele, and A. Schmidt. Multi-sensor activity context detection for wearable computing. In European Symposium on Wearable Computers, pp. 88-97. IEEE Press, 2003.