

RESEARCH

Open Access

Recognition of human activities with wearable sensors

Weihua He*, Yongcai Guo, Chao Gao and Xinke Li

Abstract

A novel approach for recognizing human activities with wearable sensors is investigated in this article. The key techniques of this approach include the generalized discriminant analysis (GDA) and the relevance vector machines (RVM). The feature vectors extracted from the measured signal are processed by GDA, with its dimension remarkably reduced from 350 to 12 while fully maintaining the most discriminative information. The reduced feature vectors are then classified by the RVM technique according to an extended multiclass model, which shows good convergence characteristic. Experimental results on the Wearable Action Recognition Dataset demonstrate that our approach achieves an encouraging recognition rate of 99.2%, true positive rate of 99.18% and false positive rate of 0.07%. Although in most cases, the support vector machines model has more than 70 support vectors, the number of relevance vectors related to different activities is always not more than 4, which implies a great simplicity in the classifier structure. Our approach is expected to have potential in real-time applications or solving problems with large-scale datasets, due to its perfect recognition performance, strong ability in feature reduction, and simple classifier structure.

Keywords: activity recognition, generalized discriminant analysis, relevance vector machines, wearable sensors.

1 Introduction

Activity recognition has become one of the most active topics in context-aware computing, due to its wide application prospects in industrial, educational, and medical domains [1-3]. For instance, characterization of the activities of assembly-line workers can increase the safe reliability and improve the productivity [1]; as another example, monitoring the human activity of daily life can provide very useful information for medical diagnosis, elderly care, as well as the assessment of individuals' physical and mental conditions [2].

Early studies in activity recognition employed vision-based systems with single or multiple video cameras, which remains the most common means to date [4,5]. In general, such systems may be acceptable and practical in a laboratory or well-controlled environment. However, when the activities take place in the real-home setting or outdoors environment, the accuracy of activity recognition would be affected by variable lighting condition or the clutter disturbance [6]. Wearable-sensor-

based system offers an appropriate alternative to activity recognition [7-11], which is inherently immune to the shadow and occlusion effects. Furthermore, compared with the vision-based systems, this kind of systems would not supply additional privacy information, thus the subjects may act more naturally as in their daily life. Another major advantage in using wearable-sensor-based systems is that the cost of storing and processing data can greatly be reduced. Therefore, in this article we will focus on developing an effective approach for activity recognition with wearable sensors.

Feature dimension reduction (which can be seen as a feature selection operation) is an important and essential procedure before classification [6-8]. A high-dimensional feature would result in the following problems: First, some features are irrelevant or redundant and cannot provide supportive information for classification (in the worst sense, inappropriately emphasizing such features may even hinder the recognition accuracy); second, the training of classifiers in a high-dimensional space would be difficult and time consuming. Therefore, it is desirable to effectively reduce the dimension of the feature set before performing the classification task. Principal

* Correspondence: heweihua@cqu.edu.cn

Key Laboratory of Optoelectronic Technology and Systems of the Ministry of Education of China, Chongqing University, Chongqing 400030, China

component analysis (PCA) and linear discriminant analysis (LDA) are two classical techniques for data dimension reduction [7-9,12,13]. The essence of the PCA is to find the optimal projection directions that maximally preserve the data variance. However, it does not take into account the class label, thus the 'optimal' projection directions may not give the most discriminating feature. While LDA seeks the optimal projection directions that maximize the ratio of between-class scatter and within-class scatter. However, it cannot capture the nonlinear relationships among samples. To overcome the weakness of LDA, the generalized discriminant analysis (GDA) based on the kernel trick has been proposed in [14], which can be viewed as an extension of LDA. Actually, GDA has been proved superior to LDA in many applications such as face recognition, document analysis, and image retrieval [15-17].

As to the existing classification methods, most of them can be divided into two categories, depending on whether they are based on the kernel-based learning or not. The non-kernel-based learning classification methods [for instance, the k -nearest neighbor (k -NN) method] usually give equal or comparative weight to each training sample, which may not be reasonable in some cases. While the kernel-based learning classification methods [for instance, the support vector machines (SVM) method] try to pick out the informative samples for classification, thus these methods usually achieve higher recognition accuracy and at the same time maintain relative sparsity of the support vectors. In the last decade, the Bayesian theory was introduced into the design of classification methodology, of which the relevance vector machines (RVM) method [18] is a representative. As a Bayesian extension of the SVM, the RVM can provide posterior probabilistic output for class memberships. Furthermore, the RVM requires dramatically fewer number of relevance vectors (RVs), which means it is more suitable for real-time applications.

Extensive study has been done on human activity recognition with wearable sensors [7-11]. Fleury et al. [7] recognized seven kinds of human daily activities. Data associated with activities were collected by the infrared sensor, the temperature and hygrometry sensor, and the wearable kinematic sensors. A feature set was extracted from the raw sensor data by PCA, and then SVM method was employed in the classification process. The cross-validation test achieved an overall recognition accuracy of 86%. Khan et al. [8] used LDA to generate feature set and artificial neural networks (ANNs) to recognize human activities with data recorded by the body-worn accelerometers. Altun et al. [9] constructed a hardware system with gyroscope, accelerometer, as well as magnetometer, and did a comparative study on human activity recognition. They considered in total

seven classification methods: Bayesian decision, decision tree, the least-squares method, k -NN, dynamic time warping, SVM, and ANN, where the SVM achieved the best recognition performance in the 'leave-one-out' cross-validation process.

In this article, we put forward a new approach for human activity recognition. We employ GDA to reduce the feature dimension and then construct a multiclass RVM classifier to perform the classification task. To the best of authors' knowledge, both the GDA and RVM techniques are applied in the wearable-sensor-based system for the first time.

The rest of this article is organized as follows. Section 2 provides a briefly description of the Wearable Action Recognition Dataset (WARD) [19]. The detailed information about the feature extraction and the feature dimension reduction process is presented in Section 3. In Section 4, we first provide a review of the RVM classification technique and then introduce the construction procedure for the multiclass RVM classifier. Section 5 reports the experimental results. Conclusions are given in Section 6.

2 Wearable action recognition dataset

In this article, we used a publicly available dataset called WARD <http://www.eecs.berkeley.edu/~yang/software/WAR/> for human activity recognition, which was introduced by Yang et al. [19]. To construct this dataset, 20 volunteers, i.e., 7 females and 13 males with a wide age range (from 15 to 75 years old), were enrolled to collect 13 activities. All the involved 13 kinds of activities are listed in Table 1.

The data were recorded by five custom-built sensor boards, which were attached to different body parts: two on the wrists, one on the waist, and two on the ankles. Each sensor board has been equipped with two sensors: a tri-axial accelerometer with the range of ± 2 g (m/s^2) and a bio-axial gyroscope with the range of $\pm 500^\circ/s$. Since the output data format of each accelerometer and gyroscope would have 3 and 2 dimensions, respectively, the activity signal are totally given in 25 dimensions.

Figure 1 provides instances of measured signal with three different activities related to the same subject. Subfigures in the first row show the three-dimensional data recorded by the accelerometer, while subfigures in the second row show the two-dimensional data recorded by the gyroscope located at the waist. As we expected, there are much differences in the magnitude and period of the measured signal among different activities.

3 Feature extraction and reduction

In this section, we first describe the data preprocessing procedure, and then present the feature extraction and

Table 1 Description of activities performed by each subject

| Activities performed | Description |
|------------------------|--|
| Rest at standing (ST) | The subject stands still for more than 10 s |
| Rest at sitting (SI) | The subject sits still for more than 10 s |
| Rest at lying (LY) | The subject lies still for more than 10 s |
| Walk forward (WF) | The subject walks straight forward for more than 10 s |
| Walk left-circle (WL) | The subject walks counterclockwise for more than 10 s |
| Walk right-circle (WR) | The subject walks clockwise for more than 10 s |
| Turn left (TL) | The subject stays at the same position and turn left for more than 10 s |
| Turn right (TR) | The subject stays at the same position and turn right for more than 10 s |
| Go upstairs (Up) | The subject goes up a flight of stairs |
| Go downstairs (Down) | The subject goes down a flight of stairs |
| Jog (Jog) | The subject jog straight forward for more than 10 s |
| Jump (Jump) | The subject stays at the same position and jump for more than 5 times |
| Push wheelchair (Push) | The subject pushes a wheelchair/walker for more than 10 s |

data normalization process. Finally, the GDA is introduced for feature reduction.

3.1 Data preprocessing

In general, it is unnecessary to analyze the entire recorded data for activity recognition. Therefore, we divided every raw sample data into small windows before feature extraction. In order to sufficiently capture the information of human activity and be convenient for the FFT-based computation of the frequency-domain features, the window length is set to be 2^n . Four different window lengths, 32 samples, 64 samples, 128 samples, and 256 samples, were investigated, and the best recognition performance can be achieved with the window length of 256 samples. It can be explained as

follows. At a sampling rate of 30 Hz, the time duration related to the four different window lengths would approximately be 1, 2, 4, and 8.5 s, respectively. Bouten et al. [20] had pointed out that the frequency of human daily activity mainly ranged from 1 to 18 Hz; therefore, the time duration of 8.5 s would cover at least several cycles of human activities and accordingly the activity recognition performance would be more reliable, i.e., shorter time duration would cover less activity cycles, thus the recognition performance would gain lower credibility. In this study, the window lengths larger than 256 samples were not taken into consideration. There are two reasons behind it: (1) for some activities, the number of samples contained in the raw sample data is less than 512. For example, the raw sample data as to

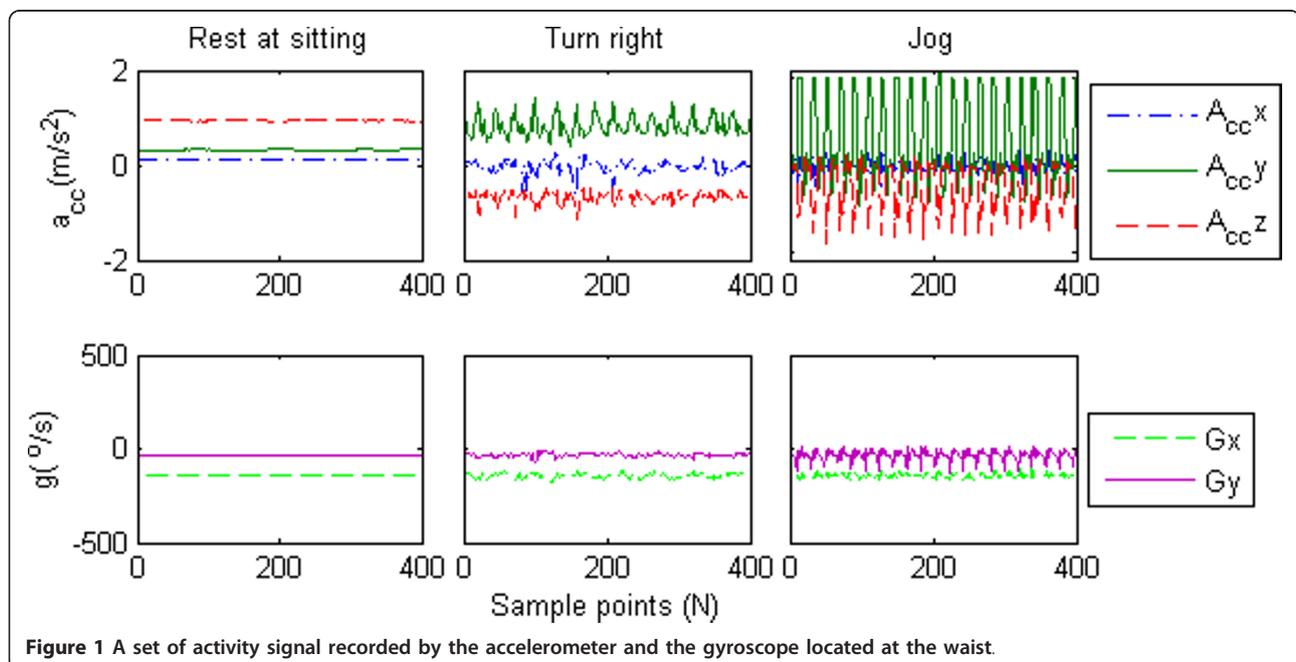


Figure 1 A set of activity signal recorded by the accelerometer and the gyroscope located at the waist.

Table 2 Recognition results with different feature reduction techniques

| Methods | Yang et al. [19] | Huynh [24] | Our approach |
|--------------|------------------|------------|--------------|
| Accuracy (%) | 93.46 | 96.97 | 99.20 |

the “Jog” activity of the test subject 5 only contains 397 samples. (2) As shown below in Table 2, with the window length of 256 a recognition rate as high as 99.2% can be achieved by our method. Increasing the window length may slightly improve the recognition rate, but will result in a longer delay.

As a result, we set the length parameter of the window to be 256 samples in the following experiments. Then, the full data size in a truncated window would be 256×25 , where 25 equals the dimensions of activity signal, i.e., each sensor node can provide five-dimensional activity signal, and there are in total five sensor nodes, and there are 128×25 data overlapping between consecutive windows.

3.2 Feature extraction

The features derived from time domain and transform domain are both used for activity recognition in most previous studies [9,21]. In this study, we simply choose the features most frequently adopted by previous researchers, rather than deliberately select them. Specifically, the selected time domain features include the mean value, variance, skewness, and kurtosis, which can be expressed as follows:

$$\mu_i = E \{s_i\} = \frac{1}{N} \sum_{n=1}^N s_{i,n} \quad (1)$$

$$\sigma_i^2 = E \{(s_i - \mu_i)^2\} = \frac{1}{N} \sum_{n=1}^N (s_{i,n} - \mu_i)^2 \quad (2)$$

$$\text{ske}_i = \frac{E \{(s_i - \mu_i)^3\}}{\sigma_i^3} = \frac{1}{N\sigma_i^3} \sum_{n=1}^N (s_{i,n} - \mu_i)^3 \quad (3)$$

$$\text{kur}_i = \frac{E \{(s_i - \mu_i)^4\}}{\sigma_i^4} = \frac{1}{N\sigma_i^4} \sum_{n=1}^N (s_{i,n} - \mu_i)^4 \quad (4)$$

where $N = 256$, $s_{i,n}$ represents the n th data value in the i th dimension associated with some window. While the selected transform domain features include the magnitudes of the maximum five peaks of the resultant fast Fourier transform (FFT), as well as the magnitudes of the maximum five peaks of the resultant cepstrum coefficients, which can be calculated as follows:

$$X_i(k) = \sum_{n=1}^N s_{i,n} e^{-j2\pi kn/N}, \quad k = 1, 2, \dots, N \quad (5)$$

$$C_i(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log X_i(e^{j\omega}) e^{j\omega n} d\omega \quad (6)$$

The six types of feature presented above can be employed for activity recognition.

Figure 2 gives an instance of signal representations both in the time and the transform domain, related to some windows for two specific activities. Figure 2a, b is the signal along z -axis related to the walking forward and the jump activity, recorded by the accelerometer located at the waist. Figure 2c, d is the resultant FFT of the signal in (a) and (b), respectively (with the maximum five peaks of FFT marked with ‘O’), while Figure 2e, f shows the resultant cepstrum of the signal in (a) and (b), respectively (with the maximum five peaks of cepstral coefficients marked with ‘O’).

According to the above feature extraction procedure, a feature vector with 14 elements can be obtained from each window along every dimension, i.e., 4 elements in the time domain and the remaining 10 elements in the transform domain. Thus, the total dimension of the feature vectors would be 350×1 .

3.3 Data normalization

Commonly, the selected features are heterogeneous. Directly taking the feature data acquired as in Section 3.2 for subsequent classification may lead to problems, especially when the distribution characteristic of the selected features witnesses dramatical discrepancies. Therefore, all the features should be normalized before constructing the classifiers. For simplicity, we do this study for each feature in a linear way as follows:

$$y_j = \frac{x_j - \min \{f_j\}}{\max \{f_j\} - \min \{f_j\}} \quad (7)$$

where x_j denotes the data value of j th feature before normalization, while the symbols ‘ $\max\{f_j\}$ ’ and ‘ $\min\{f_j\}$,’ respectively, represent the maximum and the minimum values related to this feature throughout the whole dataset. Thereafter, all the feature values will fall into the range [0 1].

3.4 Feature dimension reduction by GDA

As described above, the GDA is a nonlinear data dimension reduction method based on kernel function learning technique, which will be used to deal with the normalized feature vectors.

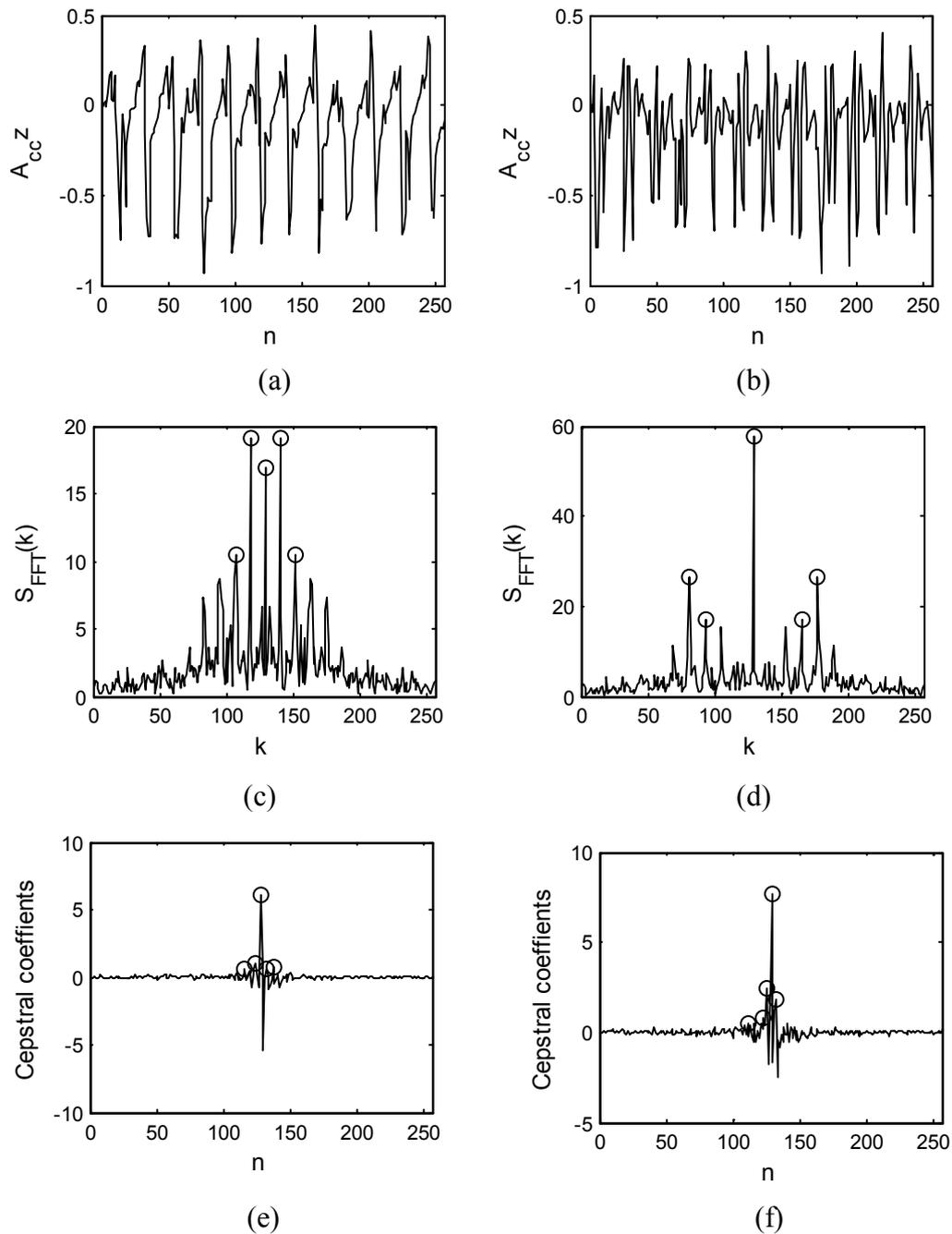


Figure 2 An instance of activity signal representations both in the time domain and in the transform domain: (a, b) The original data in some window respectively related to the walking forward and the Jump activity, recorded by the accelerometer located at the waist along z-axis. (c, d) Resultant FFT of the signal in (a, b), respectively (with the maximum five peaks of FFT marked with 'O'). (e, f) Resultant cepstrum of the signal in (a, b), respectively (with the maximum five peaks of cepstral coefficients marked with 'O').

Given a training dataset $\{\mathbf{x}_i\}_{i=1}^N$ containing L classes, with n_l samples belong to the class l (i.e., $N = \sum_{l=1}^L n_l$),

the GDA operation on it consists of two steps: first, the data \mathbf{x}_i will be transformed from the original feature space R into a new one F via a nonlinear mapping $\varphi: R$

→ F , $\mathbf{x}_i \mapsto \varphi(\mathbf{x}_i)$, and then the classical LDA is carried out in this new feature space.

The between-class scatter and within-class scatter matrix in the feature space F are defined as:

$$B = \frac{1}{N} \sum_{l=1}^L n_l \bar{\varphi}_l (\bar{\varphi}_l)^T \quad (8)$$

$$V = \frac{1}{N} \sum_{l=1}^L \sum_{k=1}^{n_l} \phi(\mathbf{x}_{lk}) \phi(\mathbf{x}_{lk})^T \quad (9)$$

where \mathbf{x}_{lk} denotes the element k of the class c_l and $\bar{\varphi}_l$ represents the mean of the class c_l in space F :

$$\bar{\varphi}_l = \frac{1}{n_l} \sum_{k=1}^{n_l} \phi(\mathbf{x}_{lk}) \quad (10)$$

The GDA method would find the projection matrix ν that maximizes the ratio:

$$V = \frac{\nu^T B \nu}{\nu^T V \nu} = [v_1, v_2, \dots, v_t] \quad (11)$$

Note that explicitly carrying out the mapping φ means a demanding task. Therefore, the skill of reproducing kernels has been adopted when deriving the projection matrix. Since the rank of B is no more than $L-1$, the upper boundary value of t is $L-1$. More details about the GDA are available in [14].

After performing the GDA on the normalized activity feature vectors, the dimensions of them would be reduced to n ($n \leq L-1$). In our case, there are in total $L = 13$ activities, thus the dimension of feature vectors will be reduced to no more than 12. We have tested the performance of GDA with different reduced dimensions ranging from 1 to 12, and finally set the dimension parameter to be 12 as it can provide the best recognition performance. We also used PCA and LDA for comparison. The resultant three-dimensional features with highest weights by each method are picked out and drawn in Figure 3. It can be seen that the GDA can capture the discriminate information better than PCA and LDA, and this would be a good prognosis for the recognition performance.

4 RVM classification techniques

In this section, we will review the binary RVM theory, and then introduce a multiclass RVM model for solving the multi-class problem. Based on that, the multi-dimension activity feature vectors can be classified.

4.1. The binary RVM

RVM is originally designed for the binary classification problem. Given a training dataset of N input target pairs

$\{\mathbf{x}_n, t_n\}_{n=1}^N$, where $\mathbf{x}_n \in R^m$ is the training sample and $t_n \in \{0,1\}$ is the target value of \mathbf{x}_n . Supposing the posterior probability of \mathbf{x}_n with $t_n = 1$ is given by $P(t_n = 1 | \mathbf{x}_n, \mathbf{w}) = \sigma\{y(\mathbf{x}_n; \mathbf{w})\}$, where $\sigma(y)$ is the logistic sigmoid function defined by $\sigma(y) = 1/(1+e^{-y})$, then the posterior probability of \mathbf{x}_n with $t_n = 0$ can be expressed as $P(t_n = 0 | \mathbf{x}_n, \mathbf{w}) = 1 - \sigma\{y(\mathbf{x}_n; \mathbf{w})\}$. On assumption that the input variables \mathbf{x}_n are independent of each other, the likelihood on the entire set of training samples can be calculated using the Bernoulli distribution:

$$P(\mathbf{t} | \mathbf{w}) = \prod_{n=1}^N \sigma\{y(\mathbf{x}_n; \mathbf{w})\}^{t_n} [1 - \sigma\{y(\mathbf{x}_n; \mathbf{w})\}]^{1-t_n} \quad (12)$$

where \mathbf{w} is a weight vector represented by $\mathbf{w} = [w_1, w_2, w_3, \dots, w_N]^T$. Here, we assume that \mathbf{w} can be well described by the zero-mean Gaussian distribution with variance α_n^{-1} , i.e., $p(\mathbf{w} | \alpha) = \prod_{n=1}^N N(w_n; 0, \alpha_n^{-1})$.

In order to find the ‘most probable’ weights \mathbf{w}_{MP} , an iterative procedure is utilized, which is based on a Laplace approximation. With a fixed value of α , the logarithm of the posterior distribution over the weight \mathbf{w} is given by

$$\begin{aligned} \ln P(\mathbf{w} | \mathbf{t}, \alpha) &= \ln P(\mathbf{t} | \mathbf{w}) P(\mathbf{w} | \alpha) - \ln P(\mathbf{t} | \alpha) \\ &= \sum_{n=1}^N \{t_n \ln y(\mathbf{x}_n; \mathbf{w}) + (1 - t_n) (1 - \ln y(\mathbf{x}_n; \mathbf{w}))\} - \frac{1}{2} \mathbf{w}^T \mathbf{A} \mathbf{w} + \text{const} \end{aligned} \quad (13)$$

where $\mathbf{A} = \text{diag}(\alpha_0, \alpha_1, \dots, \alpha_N)$. When maximizing the value of above expression, the mean value of \mathbf{w}_{MP} and its covariance Σ_{MP} should be

$$\mathbf{w}_{MP} = \mathbf{A}^{-1} \Phi^T (\mathbf{t} - \mathbf{y}) \quad (14)$$

$$\Sigma_{MP} = (\Phi^T \mathbf{B} \Phi + \mathbf{A})^{-1} \quad (15)$$

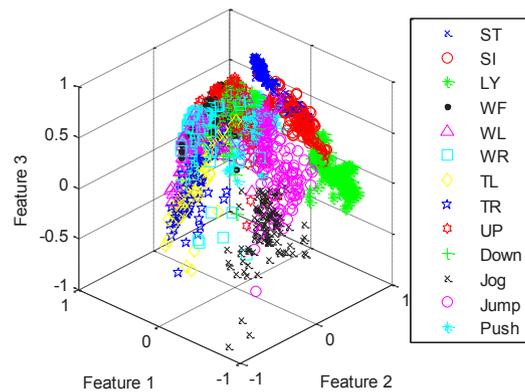
where \mathbf{B} is an $N \times N$ diagonal matrix with elements $b_n = y_n(1-y_n)$, the vector $\mathbf{y} = (y_1, y_2, \dots, y_N)^T$, and Φ is the design matrix with elements $\Phi_{ni} = \varphi_i(\mathbf{x}_n)$. The detail information about the derivation of the above two equations is provided in Appendix.

Consequently, the corresponding marginal likelihood can be expressed as

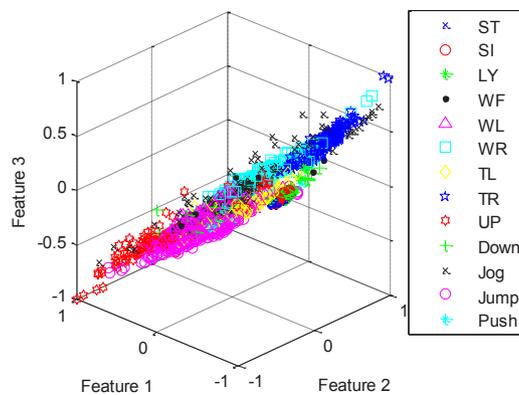
$$\begin{aligned} \log P(\mathbf{t} | \alpha) &= \int \log P(\mathbf{t} | \mathbf{w}) P(\mathbf{w} | \alpha) d\mathbf{w} \\ &= P(\mathbf{t} | \mathbf{w}_{MP}) P(\mathbf{w}_{MP} | \alpha) (2\pi)^{M/2} |\Sigma_{MP}|^{1/2} \end{aligned} \quad (16)$$

To maximize it, the parameter α_n should be updated as follows:

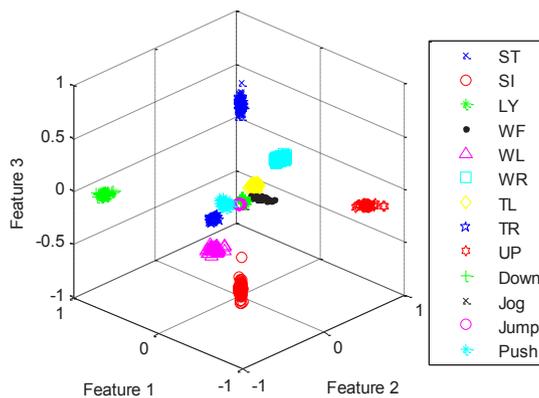
$$\alpha_n^{\text{new}} = \frac{1 - \alpha_n \text{Diag}_n}{w_n^2} \quad (17)$$



(a) 3-D PCA feature space for different activities



(b) 3-D LDA feature space for different activities



(c) 3-D GDA feature space for different activities

Figure 3 Scatter plots of the three most important features respectively picked out by PCA (up), LDA (middle) and GDA (bottom).

in each iteration, where w_n denotes the n th element of the estimated posterior weights \mathbf{w}_{MP} , and Diag_n represents the n th diagonal element of the posterior covariance matrix Σ_{MP} from Equation (15). The above

procedure via Equations (13)-(17) is repeated until when the preset convergence criterion is met. Up to this point, the training stage for the binary classifier is completed.

In the classification stage, the test sample \mathbf{x} will be classified to the class $t \in \{0,1\}$, which maximizes the conditional probability $P(t|\mathbf{x}, \mathbf{w})$.

4.2. Multiclass RVM

Traditional RVM solves the binary classification problem. However, the practical activity recognition task usually involves a multiclass task. For instance, the WARD dataset contains totally 13 kinds of activities. Therefore, the RVM must be extended to the multiclass situation. The first possible scheme is to directly generalize the RVM to the multiclass RVM as in [18]. However, in this case the size of the covariance V would linearly scale with the total number of involved classes, which is a disadvantage from the computational perspective [18]. The second possible solution is to consider the L -class problems as a set of two-class problems. In this case, the simplest way called 'one-versus-all' is to train L individual binary classifiers and integrate them together. The test sample \mathbf{x} will be classified to the class t_i on condition that

$$p(t_i|\mathbf{x}, \mathbf{w}) = \max \{p(t_m|\mathbf{x}, \mathbf{w})\}, m = 1, 2, \dots, L \quad (18)$$

5 Experimental results

In this section, we first examine the convergence characteristic of the constructed RVM classifiers and present the classification results for the proposed approach with threefold cross validation on WARD (Section 5.1). Then, we compare the recognition performances with different feature reduction techniques and different classification techniques, respectively. We also show a comparison between our approach and other existing methods on the same dataset (Section 5.2).

5.1. Recognition performance with the proposed approach

As described in Sections 3 and 4, the feature vectors are extracted from the measured activity signal and compressed into 12 dimensions by GDA, which subsequently are classified with the multiclass RVM technique. To evaluate the recognition performance of our approach, we did threefold cross validation on the entire WARD dataset. Specifically, all the feature vectors were randomly divided into three partitions, of which one partition was retained as the validation set for testing, and the remaining two partitions were used for training. Such cross-validation process was performed for three times, so as to ensure each partition was validated. The whole procedure was repeated for ten times, and the resultant recognition rates were averaged.

During the cross-validation process, we established an RVM classifier for each kind of activity, i.e., there were

in total 13 different RVM classifiers for all kinds of activities. To examine the convergence characteristic of the constructed classifiers, we monitored their marginal likelihood versus the iteration numbers. As shown in Figure 4, the likelihood of 'ST' activity (denoted by the solid line and squares) quickly converged after about 12 iterations, and those of the remained activities also show the similar tendency, which proves the consistency of the classifiers and the convenience to construct them.

Figure 5 shows the recognition results with different numbers of feature types mentioned in Section 3. It can be seen that the recognition accuracy gradually get improved as the number of employed feature types increases, which indicates that the different types of feature considered in this article can provide complementary information. As a result, all the six types of features are used for recognition in the following experiments.

The confusion matrix related to the recognition result (with all six types of feature) is given in Table 3. It can be observed that the confusion occurrences are distributed in an unbalance way. For instance, the three most confused pairs are 'Up' and 'Jog', 'SI' and 'ST', as well as 'ST' and 'SI'. The confusion rates between them have, respectively, reached 2.4, 1.45, and 1.31%, while some pairs such as 'LY' and 'Push', 'TR' and 'WR', as well as 'TL' and 'Jump' are never confused with each other. It probably can be explained as follows. The subjects do not have to move the ankles to perform both 'SI' and 'ST' activities, and at the same time the accompanying movement of waist is always quite small. Thus, the sensor nodes would provide less discriminable information for those two kinds of activities. While 'Up' is sometimes misclassified as 'Jog' mainly because both activities may involve the feet rising action. A possible solution for improving the discrimination between similar activities is to include more sensors located at knees or thighs, or deploying other kinds of sensors. For instance, sensors such as location sensors can be used to keep track of a subject's body movement. Another interesting point is that the confusion rates are not necessarily equivalent between a specific pair. For instance, though the 'Up' activity may be judged as 'Jog' with a rate of 2.4%, the 'Jog' is never mistaken as 'Up'. Such phenomenon may provide cues for further improvement on the selection of features.

5.2. Comparative evaluation

Extensive comparisons have been made to thoroughly examine the performance of the proposed approach. This section reports the comparison results.

First, we compare the recognition performances of our proposed feature reduction with two other feature reduction techniques, i.e., PCA and LDA. The evaluation is performed on the WARD dataset, and the same

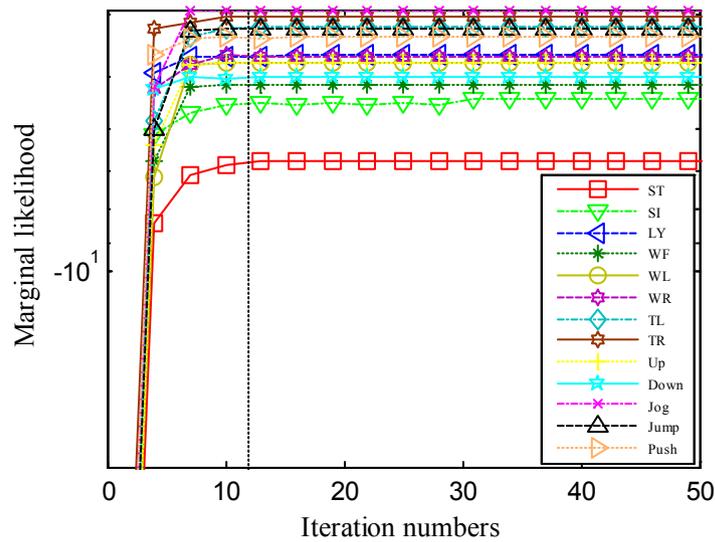


Figure 4 The marginal likelihood versus iteration numbers.

classifier RVM is employed. The comparison results are listed in Table 4. The GDA achieves the recognition rate of 99.2%, which is 22.9% higher compared with the PCA and 58.9% higher compared with the LDA. This consequence is as per our expectation. On one hand, the distribution of wearable data is nonlinear and complex duo to the factors such as measurement noise, outliers, and other variation. On the other hand, both PCA and LDA are linear feature reduction techniques. Thus, it is difficult for them to capture nonlinear relationship with a linear mapping. On the contrary, GDA is nonlinear extension of LDA. It can transform the original data space to a feature space by a nonlinear mapping

through kernel methods, where is more likely to be linearly separable than in the original data space. Therefore, the GDA can provide more reliable and robust solution to activity recognition problem.

We subsequently compare the recognition performances between the RVM classifiers and three other popular classification techniques, i.e., k -NN, Bayesian decision, and SVM. The evaluation is also performed on the WARD dataset, and the same feature reduction technique (GDA) is employed. In general, recognition accuracy is sensitive to the parameters of classifiers. For fairly comparison, it would be desirable to adopt optimal parameters for each kind of classifier, respectively. Specifically, as to the k -NN classifier, we set $k = 3$ since in this article we can achieve the highest recognition accuracy under this condition; While for the RVM classifier, the Gaussian kernel is employed and the optimal bandwidth value is be found by the following simple method: we increased the bandwidth value with a constant step of 0.05 over the range of [0.05 1], and trained the RVM classifier over the whole training set. The bandwidth parameter 0.15, which maximized the classification accuracy, was chosen for the following experiments. As to the construction of the multiclass SVM classifiers, we also adopted the Gaussian kernel and the ‘one-versus-all’ strategy, which are the same as in the multiclass RVM model. Its optimal values of two controlling parameters, i.e., bandwidth and regularizing parameter C , are also fixed with the same searching strategy as in RVM (in this article, the optimal bandwidth and regularizing parameters are set to be 0.25 and 10,

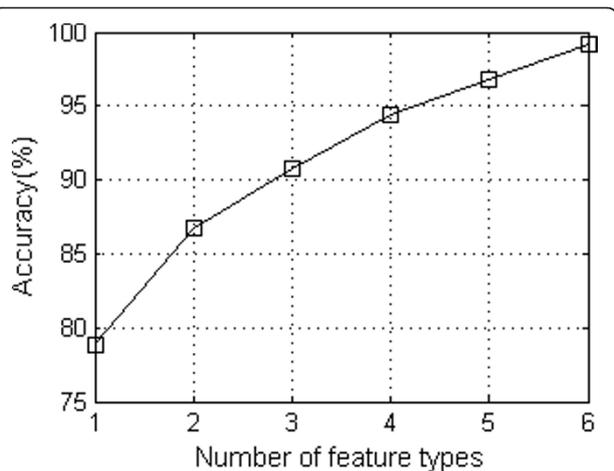


Figure 5 Recognition accuracy of our approach versus number of feature types.

Table 3 Confusion matrix related to the ten times threefold validation by our approach

| True | Classifieds | | | | | | | | | | | | |
|------|-------------|-----|-----|-----|-----|-----|-----|-----|-----|------|-----|------|------|
| | ST | SI | LY | WF | WL | WR | TR | TL | UP | Down | Jog | Jump | Push |
| ST | 349 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| SI | 5 | 339 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| LY | 0 | 0 | 343 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| WF | 0 | 0 | 0 | 203 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| WL | 0 | 0 | 0 | 2 | 328 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| WR | 0 | 0 | 0 | 0 | 0 | 323 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| TR | 0 | 0 | 0 | 0 | 0 | 0 | 304 | 0 | 0 | 0 | 0 | 0 | 0 |
| TL | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 318 | 0 | 0 | 0 | 0 | 0 |
| UP | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 274 | 1 | 4 | 1 | 0 |
| Down | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 1 | 219 | 2 | 0 | 0 |
| Jog | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 159 | 0 | 0 |
| Jump | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 260 | 0 |
| Push | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 236 |

respectively). Table 5 shows the recognition accuracies by these different classification techniques. The RVM gives the recognition rate as high as 99.2%, followed by the SVM and the Bayesian decision, respectively, with recognition rates of 97.7 and 95.9%, while the k -NN performs the worst, of which the recognition rate only reaches 88.3%. The reason behind it may be as follows: k -NN is a linear classifier, which calculates the similarity between the test sample and the training samples. Since it does not take the data distribution into account, k -NN may not be suited for dealing with noise data. As for SVM, it is based on the principle of structural risk minimization. The final classifier obtained by SVM depends only on the “borderline” samples in the training samples, i.e., support vectors (SVs). These SVs are located near the decision boundary of the classifier. This makes the SVM sensitive to noises or outliers and patterns that were wrongly classified lie near the separation hyper-plane. As for RVM, it is a Bayesian extension of the SVM. The final classifier obtained by RVM depends only on fewer samples in the training samples, i.e., RVs. Unlike SVM, these RVs are formed by samples appearing to be more representative of the classes, which are located away from the decision boundary of the classifier. Therefore, the RVM has a better generalization ability and more robust to noises or outliers.

Additionally, Table 5 also gives the N_{SV} and N_{RV} which represent the number of support vectors in the SVM model and the number of relevant vectors in the

RVM model, respectively. They reflect the classifiers’ structural complexity [21]. Although the values of N_{SV} are larger than 70 in most cases, the values of N_{RV} are always not more than 4, i.e., the sparsity property of the RVM is much better than the SVM. Therefore, the results in Table 5 demonstrate that RVM has remarkable advantage both in the recognition accuracy and the sparsity of RVs.

To further evaluate the performance of our approach, we also employed other conventional metrics [22,23] including the precision, the recall rate, the F index and the specificity rate, which can be described as

Table 4 Recognition results with different feature reduction techniques

| Reduction technique | PCA | LDA | GDA |
|---------------------|-------|-------|-------|
| Accuracy (%) | 76.31 | 40.30 | 99.20 |

Table 5 Recognition results with different classification techniques

| Activity | k -NN Acc (%) | Bayesian Acc (%) | SVM | | RVM | |
|--------------|-----------------|------------------|-------------|-------------|-------------|-----------|
| | | | Acc (%) | N_{SV} | Acc (%) | N_{RV} |
| ST | 74 | 91 | 93 | 121 | 99 | 4 |
| SI | 71 | 91 | 96 | 100 | 99 | 2 |
| LY | 74 | 99 | 99 | 100 | 100 | 3 |
| WF | 90 | 94 | 95 | 178 | 99 | 2 |
| WL | 94 | 98 | 99 | 74 | 99 | 2 |
| WR | 94 | 99 | 100 | 64 | 100 | 2 |
| TR | 98 | 100 | 99 | 82 | 100 | 2 |
| TL | 97 | 100 | 100 | 70 | 100 | 2 |
| UP | 91 | 91 | 95 | 145 | 97 | 2 |
| Down | 93 | 84 | 96 | 148 | 98 | 3 |
| Jog | 100 | 100 | 100 | 100 | 100 | 2 |
| Jump | 98 | 99 | 99 | 76 | 100 | 2 |
| Push | 89 | 99 | 99 | 44 | 100 | 2 |
| Total | 88.3 | 95.9 | 97.7 | 1302 | 99.2 | 30 |

$$\begin{aligned} \text{pre} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, & \text{rec} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \\ F &= \frac{2}{1/\text{pre} + 1/\text{rec}}, & \text{spe} &= \frac{\text{TN}}{\text{FP} + \text{TN}} \end{aligned} \quad (19)$$

where TP (true positive) refers to the number of positive samples classified as positive; FP (false positive) refers to the number of positive samples classified as negative; FN (false negative) denotes the number of negative samples classified as positive; and TN (false negative) denotes the number of negative samples classified as negative. The comparison results regarding all these metrics are plotted in Figure 6. It can be seen that the RVM achieves the highest scores of them, which are 0.9918, 0.9903, 0.9910, and 0.9993, respectively.

Figure 7 highlights the relationship between the true positive rate (TPR) and false positive rate (FPR) related to the different classification techniques. It shows that the result of RVM is almost perfect, with the highest TPR of 99.18% and lowest FPR of 0.07%. The SVM and Bayesian decision method appear a little worse, while the *k*-NN method performs the worst, with the lowest TPR value of 89.04% and the highest FPR of 1.08%.

Finally, we also compare the performance of our approach with other existing methods on the same dataset. The quantitative results have been reported in Table 2. It can be seen that our approach outperforms other alternatives in terms of recognition rate. Specifically, Yang et al. [19] employed a distributed sparsity classifier to classify human activities and gave the recognition

accuracy of 93.46% with all five sensor nodes, which is about 6% lower than that of our approach. They modeled the distribution of multiple action classes as a mixture subspace model and represented the test sample with linear approximation of all training samples. However, this linear representation structure may meet limitations in describing the test samples of the complex activities. Huynh [24] combined a generative model (multiple eigenspace) with SVM classifier into their activity recognition framework. Since the multiple eigenspace approach has advantages in the representing the structure of the input data and SVM has good discriminability, they achieved higher recognition accuracy than that of Yang’s method, but still 2.23% lower than that of our approach. We also noted that Yang et al.’s method can be adaptive to the alteration of sensor configuration by constructing a global projecting matrix, thus it has advantages over Huynh’s as well as our approach in dealing the problems if one or more sensor nodes (or sensor) failures. One possible way for us in handling these problems is to individually train the data from individual sensor, and then fuse the classification results with the valid sensors at the decision level.

It is also worth noting that the test data we use are different from that used in the study of [25]. In our test data, the number of sensor nodes, the number of test subjects, and the types of activities are 5, 20, and 13, respectively, while the corresponding numbers and types in [25] are 8, 3, and 12, respectively. Therefore, our results are not compared with those in [25].

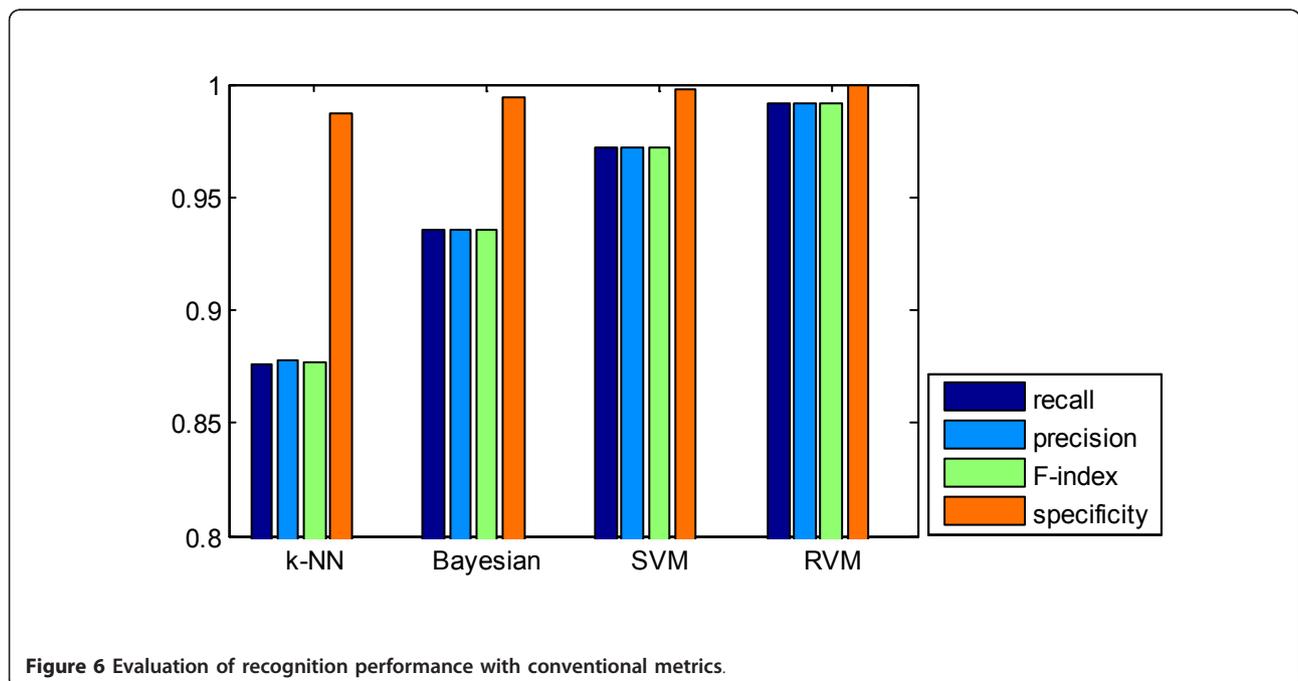


Figure 6 Evaluation of recognition performance with conventional metrics.

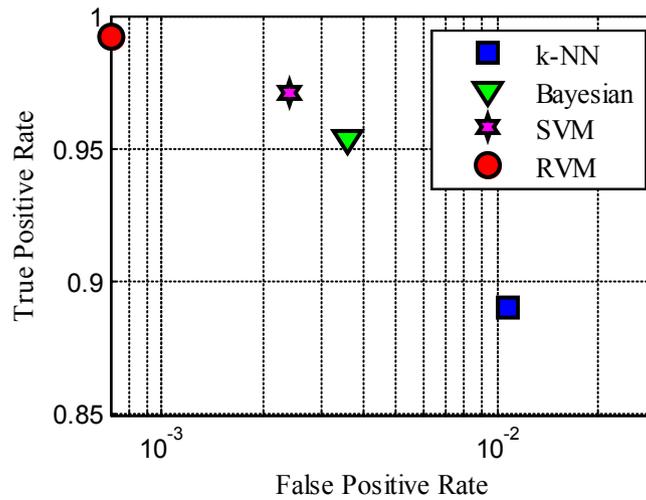


Figure 7 TPRs & FPRs with different classification techniques.

6 Conclusions

We put forward a novel approach for the recognition of human activities with wearable sensors, by combining the GDA and the RVM techniques. To the best of authors' knowledge, both of these techniques are applied to this domain for the first time.

As a powerful data dimension reduction method, the GDA can sharply reduce the dimension of the feature space, while maintaining the most discriminative information among different activities. Specifically, in this article the dimension of the feature vectors has even been reduced from 350 to merely 12, which can greatly speed up the subsequent training process. Meanwhile, the RVM also shows great extension flexibility to the multiclass classification problems. Experimental results on the WARD dataset demonstrated that the RVM technique not only provides the highest recognition rate, the highest TPR, as well as the lowest FPR compared with the conventional classification techniques, but also possesses much simpler classifier structure in contrary to the SVM. In conclusion, our approach would have advantages in real applications or solving problems with large-scale datasets, due to its perfect recognition performance, strong ability in feature reduction, and simple classifier structure.

Appendix: Derivations of Equations (14) and (15)

The RVM model takes the form of a linear combination of basis function transformed by a logistic sigmoid function

$$\mathbf{y}(\mathbf{x}, \mathbf{w}) = \sigma(\mathbf{w}^T \phi(\mathbf{x})) = 1 / \left[1 + e^{-\mathbf{w}^T \phi(\mathbf{x})} \right] \quad (20)$$

The gradient vector of the log posterior distribution, which from Equation (13), is given by

$$\begin{aligned} \nabla_{\mathbf{w}} \ln P(\mathbf{w} | \mathbf{t}, \alpha) &= \sum_{n=1}^N \left[t_n \frac{\phi(\mathbf{x}_n) e^{-\mathbf{w}_n \phi(\mathbf{x}_n)}}{1 + e^{-\mathbf{w}_n \phi(\mathbf{x}_n)}} - (1 - t_n) \frac{\phi(\mathbf{x}_n)}{1 + e^{-\mathbf{w}_n \phi(\mathbf{x}_n)}} \right] - \mathbf{A} \mathbf{w} \\ &= \sum_{n=1}^N [t_n \phi(\mathbf{x}_n) \gamma_n e^{-\mathbf{w}_n \phi(\mathbf{x}_n)} - (1 - t_n) \phi(\mathbf{x}_n) \gamma_n] - \mathbf{A} \mathbf{w} \\ &= \sum_{n=1}^N [t_n \phi(\mathbf{x}_n) \gamma_n (1 + e^{-\mathbf{w}_n \phi(\mathbf{x}_n)}) - \phi(\mathbf{x}_n) \gamma_n] - \mathbf{A} \mathbf{w} \\ &= \sum_{n=1}^N [(t_n - \gamma_n) \phi(\mathbf{x}_n)] - \mathbf{A} \mathbf{w} \\ &= \Phi^T (\mathbf{t} - \mathbf{y}) - \mathbf{A} \mathbf{w} \end{aligned} \quad (21)$$

where the vector $\mathbf{y} = (y_1, y_2, \dots, y_N)^T$, and Φ is the design matrix with elements $\Phi_{ni} = \phi_i(\mathbf{x}_n)$. By setting Equation (21) to zero, the mean \mathbf{w}_{MP} of the Laplace approximation is represented by $\mathbf{w}_{\text{MP}} = \mathbf{A}^{-1} \Phi^T (\mathbf{t} - \mathbf{y})$.

Hessian matrix of the log posterior distribution, which from Equation (13), is given by

$$\begin{aligned} \nabla_{\mathbf{w}} \nabla_{\mathbf{w}} \ln P(\mathbf{w} | \mathbf{t}, \alpha) &= \nabla_{\mathbf{w}} \left\{ \sum_{n=1}^N [\phi(\mathbf{x}_n) (t_n - \gamma_n)] - \mathbf{A} \mathbf{w} \right\} \\ &= \sum_{n=1}^N \left[\frac{e^{-\mathbf{w}_n \phi_n}}{[1 + e^{-\mathbf{w}_n \phi(\mathbf{x}_n)}]^2} \phi(\mathbf{x}_n) \phi(\mathbf{x}_n)^T \right] - \mathbf{A} \\ &= - \sum_{n=1}^N [\gamma_n (1 - \gamma_n) \phi(\mathbf{x}_n) \phi(\mathbf{x}_n)^T] - \mathbf{A} \\ &= -(\Phi^T \mathbf{B} \Phi + \mathbf{A}) \end{aligned} \quad (22)$$

where \mathbf{B} is an $N \times N$ diagonal matrix with elements $b_n = \gamma_n(1 - \gamma_n)$. At convergence of the iterative reweighted least squares algorithm, the negative Hessian represents the inverse covariance matrix for the Gaussian approximation to the posterior distribution [26]. Then, the

covariance matrix Σ_{MP} is represented by $\Sigma_{MP} = (\Phi^T B \Phi + A)^{-1}$.

Abbreviations

ANN: artificial neural networks; FFT: fast Fourier transform; FN: false negative; FP: false positive; FPR: false positive rate; GDA: generalized discriminant analysis; LDA: linear discriminant analysis; k -NN: k -nearest neighbor; PCA: principal component analysis; RVM: relevance vector machines; SVM: support vector machines; TN: false negative; TP: true positive; TPR: true positive rate; TPR: true positive rate; WARD: Wearable Action Recognition Dataset.

Acknowledgements

The authors would like to thank A. Y. Yang et al for providing the WARD activity dataset. Many thanks also go to the PhD candidate Chaoying Tang, in the Nanyang Technological University. The study was supported by the Key Project of Chinese Ministry of Education (Grant No. 108174), the Chongqing Municipal Natural Science Foundation of China (Grant No. CSTC2008BB3169).

Competing interests

The authors declare that they have no competing interests.

Received: 15 July 2011 Accepted: 10 May 2012 Published: 10 May 2012

References

1. JA Ward, P Lukowicz, G Troster, TE Starnier, Activity recognition of assembly tasks using body-worn microphones and accelerometers. *IEEE Trans Pattern Anal Mach Intell.* **28**(10), 1553–1567 (2006)
2. NF Ince, CH Min, A Tewfik, D Vanderpool, Detection of early morning daily activities with static home and wearable wireless sensors. *EURASIP J Adv Signal Process.* **2008**, 273130 (2008)
3. A Bulling, JA Ward, H Gellersen, G Troster, Robust recognition of reading activity in transit using wearable electrooculography. *Pervas Comput.* **5013**, 19–37 (2008)
4. HM Qian, YB Mao, WB Xiang, ZQ Wang, Recognition of human activities using SVM multi-class classifier. *Pattern Recognit Lett.* **31**(2), 100–111 (2010)
5. C Liu, PC Yuen, Human action recognition using boosted EigenActions. *Image Vis Comput.* **28**(5), 825–835 (2010)
6. N Ikizler, P Duygulu, Histogram of oriented rectangles: a new pose descriptor for human action recognition. *Image Vis Comput.* **27**(10), 1515–1526 (2009)
7. A Fleury, M Vacher, N Noury, SVM-based multimodal classification of activities of daily living in health smart homes: sensors, algorithms, and first experimental results. *IEEE Trans Inf Technol Biomed.* **14**(2), 274–283 (2010)
8. AM Khan, YK Lee, SY Lee, TS Kim, A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer. *IEEE Trans Inf Technol Biomed.* **14**(5), 1166–1172 (2010)
9. K Altun, B Barshan, O Tuncel, Comparative study on classifying human activities with miniature inertial and magnetic sensors. *Pattern Recognit.* **43**(10), 3605–3620 (2010)
10. YP Chen, JY Yang, SN Liou, GY Lee, JS Wang, Online classifier construction algorithm for human activity detection using a tri-axial accelerometer. *Appl Math Comput.* **205**(2), 849–860 (2008)
11. M Li, V Rozgic, G Thatte, S Lee, M Annavaram, U Mitra, DS Metz, S Narayanan, Multimodal physical activity recognition by fusing temporal and cepstral information. *IEEE Trans Neural Syst Rehabil Eng.* **18**(4), 369–380 (2010)
12. TK Kim, J Kittler, Locally linear discriminant analysis for multimodally distributed classes for face recognition with a single model image. *IEEE Trans Pattern Anal Mach Intell.* **27**(3), 318–327 (2005)
13. K Altun, B Barshan, Human activity recognition using inertial/magnetic sensor units. *Human Behav Understand.* **6219**, 38–51 (2010)
14. G Baudat, FE Anouar, Generalized discriminant analysis using a kernel approach. *Neural Comput.* **12**(10), 2385–2404 (2000)
15. YS Huang, WC Liu, Face recognition based on complementary matching of single image and sequential images. in *The Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Kyoto, Japan 673–676 (September 2009)

16. L Bai, L Shen, A fast and robust Gabor feature based method for face recognition. in *The IEE International Symposium on Imaging for Crime Detection and Prevention*, London, United Kingdom 95–98 (June 2005)
17. M Wang, HT Xu, GY Hao, XD Zhou, W Wang, Q Zhang, B Shi, PictureBook: a text-and-image summary system for web search result. in *The 24th International Conference on Data Engineering*, Vols 1-3, Cancun, Mexico 1612–1615 (April 2008)
18. ME Tipping, Sparse Bayesian learning and the relevance vector machine. *J Mach Learn Res.* **1**(3), 211–244 (2001)
19. AY Yang, R Jafari, SS Sastry, R Bajcsy, Distributed recognition of human actions using wearable motion sensor networks. *J Ambient Intell Smart Environ.* **1**(2), 103–115 (2009)
20. CVC Bouten, KTM Koekkoek, M Verduin, R Kodde, JD Janssen, A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. *IEEE Trans Biomed Eng.* **44**(3), 136–147 (1997)
21. XD Wang, MY Ye, CJ Duanmu, Classification of data from electronic nose using relevance vector machines. *Sensors Actuators B Chem.* **140**(1), 143–148 (2009)
22. MMN Biccocchi, F Zambonelli, Detecting activities from body-worn accelerometers via instance-based algorithms. *Pervas Mob Comput.* **6**, 482–495 (2010)
23. H Junker, O Amft, P Lukowicz, G Troster, Gesture spotting with body-worn inertial sensors to detect user activities. *Pattern Recognit.* **41**(6), 2010–2024 (2008)
24. DTG Huynh, Human activity recognition with wearable sensors. PhD dissertation, Technische Universität Darmstadt, Darmstadt, Germany (2008)
25. AY Yang, S Iyengar, S Sastry, P Kurloski, R Jafari, Distributed segmentation and classification of human actions using a wearable motion sensor network. in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Anchorage, AK, United States 1–8 (June 2008)
26. DB Rubin, Iteratively reweighted least squares. *Encycl Stat Sci.* **4**, 272–275 (1983)

doi:10.1186/1687-6180-2012-108

Cite this article as: He et al.: Recognition of human activities with wearable sensors. *EURASIP Journal on Advances in Signal Processing* 2012 2012:108.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com