Implementation of HMM-Based Human Activity Recognition Using Single Triaxial Accelerometer

Chang Woo HAN\(^{(a)}\), Shin Jae KANG\(^{(b)}\), Nonmembers, and Nam Soo KIM\(^{(c)}\), Member

SUMMARY In this letter, we propose a novel approach to human activity recognition. We present a class of features that are robust to the tilt of the attached sensor module and a state transition model suitable for HMM-based activity recognition. In addition, postprocessing techniques are applied to stabilize the recognition results. The proposed approach shows significant improvements in recognition experiments over a variety of human activity DB.

key words: human activity recognition, triaxial accelerometer, HMM, state transition model

1. Introduction

Monitoring and automatically detecting human activities have many useful applications. Due to the recent advances in sensor technologies, the sensor devices have become compact and portable enough to be attached to the human body without any difficulty. There have been some studies on tracking human movements using multiple accelerometers fixed to the specific places of a human body, such as waist, thigh, wrist, ankle and upper arm [1]–[5]. Our daily activities are categorized into stationary states like standing and lying, active movements such as walking, walking upstairs, moving downward, running and bicycling, and short-time motions like jumping and falling which is important to detect a state of emergency [1],[2],[6].

In order to extract features from the incoming accelerometer signals, a finite length window is applied, which overlaps by 50% with the adjacent windows [3]. To recognize human activities, time, frequency and time-frequency domain feature parameters are usually extracted [1],[6]. It has been reported that the autoregressive (AR) coefficients show better performance than the original time and frequency domain feature parameters [6]. Traditionally, as the time domain features, the mean (DC), standard deviation, energy and correlations of the accelerometer signal in each axis are computed. The accelerometer signal in each direction is also converted to the corresponding frequency domain features by applying the Fourier transform. Recently, time-frequency features have been proposed based on the wavelet analysis which compromises the spectral and temporal resolutions.

In order to recognize human activities, there have been developed a variety of classifiers such as the support vector machine, decision tree, naive Bayes classifier, dynamic time warping and k-nearest neighbor techniques [1],[3]–[7]. Among a variety of pattern recognition techniques, hidden Markov model (HMM) is considered suitable for gesture recognition [8]. One of the major advantages of the HMM-based approaches is that they can handle feature vector sequences of varying length and thereby can track the continuously evolving human activities.

Most of the previous methods are considered to be sensitive to the rotation of the sensor device. When a user attaches the sensor module in a different way from the configuration with which the training DB was collected, the performance of the recognition system will degrade dramatically. Therefore, it is necessary to devise a method to compensate the rotation and misplacement of the sensor module.

In this letter, we propose a new human activity recognition technique using single triaxial accelerometer. A set of composite feature vectors are derived, which are robust to the rotation of the accelerometer. HMMs are applied to classify the feature vector stream extracted from the accelerometer signals. Furthermore, postprocessing algorithms are employed to stabilize the recognition results. From a number of activity recognition experiments, it has been observed that the proposed methods are effective to improve the performance.

2. Robust Feature Extraction

The triaxial accelerometer used in this study senses the acceleration in three dimensional space with a dynamic range of \(\pm 3\ g\) where \(g\) denotes the gravitational acceleration. The sampling rate is 100 Hz with 8 bit resolution per each axis and the data are transmitted to a computing device via Bluetooth connection.

Six basic daily activities, e.g., standing, walking, running, falling, lying and jumping, are defined as the target human activities that should be recognized. We categorize the representative motions based on the analysis of the nonactives, periodic activities and instantaneous activities. For convenience, we define the left-and-right, front-and-rear and up-and-down directions as \(x\), \(y\) and \(z\) axes, respectively. Standing and lying have a little movement but they can be distinguished from each other by the tilt of the sensor. Especially, the bias values in the \(y\) and \(z\) axes are useful to discriminate these two states. Walking and running are
quasiperiodic activities. In the case of running, signal
variation in the $z$ axis is relatively bigger than that obtained from
walking. Falling and jumping are instantaneous activities.
Falling is an action that gives rise to a big acceleration in the
$y$ and $z$ axes and particularly the activity state changes from
standing to lying. Jumping also shows a great acceleration
in the $z$ axis. However, it has additional run-up motion when
compared to falling.

Let $\mathbf{a}_t(t) = [a_{xt}(t), a_{yt}(t), a_{zt}(t)]'$ be a three di-
mensional vector obtained from the triaxial accelerometer at
time $t$ with the prime denoting matrix or vector transpose.
$\mathbf{a}_t(t)$ does not show an actual acceleration data but biased
one due to the gravitational acceleration. The tilt of the
sensor module can be estimated from a long-term aver-
age of the three dimensional acceleration data. Let $\mathbf{b}_k =$
$[b_{x,z}, b_{y,z}, b_{z,z}]'$ be the estimated bias vector. Then, it can be
computed as follows:

$$\mathbf{b}_k = \frac{1}{M} \sum_{t=1}^{M} \mathbf{a}_t(t)$$

with $M$ denoting the number of samples used in the long-
term average.

Since waist is considered less affected by the move-
ments of arms or legs, we assume that the user attaches the
sensor module to the waist belt tightly. Even though every
user attaches the sensor module in a similar way, the degree
of tilt of the sensor module may be different. For that rea-
son, tilt compensation should be taken into consideration.
Let $\mathbf{\tilde{a}}_t(t) = [a_{xt}(t), a_{yt}(t), a_{zt}(t)]'$ be the tilt com-
penated vector which sets the direction of the gravitation parallel
to the $z$ axis, and $\mathbf{b}_r$ be the estimated bias vector of $\mathbf{\tilde{a}}_t(t)$. Then,
$\mathbf{\tilde{a}}_t(t)$ is calculated from the estimated bias as follows:

$$\mathbf{\tilde{a}}_t(t) = \begin{bmatrix}
\cos \theta_2 & (-\sin \theta_1 \sin \theta_2) & (-\cos \theta_1 \sin \theta_2) \\
\sin \theta_1 \cos \theta_2 & \cos \theta_1 & -\sin \theta_1 \\
\sin \theta_2 & \sin \theta_1 \cos \theta_2 & \cos \theta_1 \cos \theta_2
\end{bmatrix} \mathbf{a}_t(t)$$

with

$$\theta_1 = \arctan \left( \frac{b_{x,z}}{b_{z,z}} \right), \quad \theta_2 = \arctan \left( \frac{b_{y,z} \sin \theta_1 + b_{z,z} \cos \theta_1}{b_{z,z}} \right)$$

$$\mathbf{\tilde{b}}_r = \frac{1}{M} \sum_{t=1}^{M} \mathbf{a}_t(t) \approx [0, 0, g]^T$$

where $\theta_1$ is the angle between the projection of $\mathbf{\tilde{b}}_r$ onto $y$-$z$
plane and $z$ axis, and $\theta_2$ is the angle between the rotated bias
vector on $x$-$z$ plane and $z$ axis as shown in Fig. 1. In (2),
$\mathbf{\tilde{a}}_t(t)$ is a vector obtained by rotating $\mathbf{\tilde{a}}_t(t)$ by $\theta_1$
around $x$ axis then by $-\theta_2$ around $y$ axis. Let $\mathbf{\tilde{a}}_t(t) =$
$[a_{xt}(t), a_{yt}(t), a_{zt}(t)]'$ be the bias-removed and tilt com-
penated estimate for the acceleration vector. Then, $\mathbf{\tilde{a}}_t(t)$ is obtained by

$$\mathbf{\tilde{a}}_t(t) = \mathbf{\tilde{a}}_t(t) - \mathbf{\tilde{b}}_r.$$  

The feature parameters for human activity detection are
extracted from $\mathbf{\tilde{a}}_t(t)$. It is not difficult to understand that $a_{zt}(t)$
is responsible for the up-and-down movement. On the other
hand, it is not desired to treat $a_{xt}(t)$ and $a_{yt}(t)$ separately
because a slight misplacement of the sensor will lead to quite
different sensed values. For this reason, it is beneficial to
describe $a_{xt}(t)$ and $a_{yt}(t)$ jointly as a single complex number
and extract rotation invariant feature vectors from that. Let $a_{xy}(t) = a_{xt}(t) + ja_{yt}(t)$ be the acceleration data described in
the $x$-$y$ Gauss plane with $j = \sqrt{-1}$. The feature vectors are
extracted based on $a_{x}(t)$ and $a_{y}(t)$ in order to implement the
rotation invariance property.

The proposed feature vector is composed of short time
means, short time variances, short time Fourier transform
(STFT) and complex AR coefficients, computed over the input
acceleration data. Let a rectangular window be given by
the sequence $u(t) = 1$ for $0 \leq t \leq N - 1$ and zero elsewhere,
with $N$ being the window size, $S$ be the window shift, and
$P$ be the order of AR analysis. Then, the proposed feature
parameters are listed in Table 1. Since the short time means,
$\mu_x(n)$ and $\mu_y(n)$, depend on the degree of tilt from the
initial state, they are useful to distinguish between the states of
standing and lying. The short time variances, $\sigma^2_x(n)$ and
$\sigma^2_y(n)$, are useful to distinguish among the states of move-
ments such as walking and running because the variances
measure the variability of the acceleration signal. The STFT
coefficients, $A_x(n, k)$ and $A_y(n, k)$, are easily computed by
applying the $N$-point fast Fourier transform (FFT) algorithm
[6]. The AR coefficients, $b_x(n, p)$ and $b_y(n, p)$, are obtained by
using the Levinson-Durbin algorithm [6], where the AR
model is represented as follows:

$$a_{xt}(t) = -\sum_{p=1}^{P} b_{x}(n, p) \cdot a_{xt}(t - p) + u(t)$$

$$a_{yt}(t) = -\sum_{p=1}^{P} b_{y}(n, p) \cdot a_{yt}(t - p) + u(t)$$

in which $u(t)$ is a white Gaussian noise and $n$ denotes the
frame index. Since $\mu_x(n)$, $A_x(n, k)$, $A_y(n, k)$ and $b_y(n, p)$

\begin{table}[h]
\centering
\caption{Basic feature parameters for activity recognition ($0 \leq k \leq N - 1$, $1 \leq p \leq P$).}
\begin{tabular}{|c|c|c|}
\hline
 & Windowed data for $n$-th frame & \\
\hline
 & $a_{xt}(t)u(t)$ & $a_{yt}(t)u(t)$ \\
\hline
Short time mean & $\mu_x(n)$ & $\mu_y(n)$ \\
\hline
Short time variance & $\sigma^2_x(n)$ & $\sigma^2_y(n)$ \\
\hline
STFT coefficients & $A_x(n, k)$ & $A_y(n, k)$ \\
\hline
AR coefficients & $b_x(n, p)$ & $b_y(n, p)$ \\
\hline
\end{tabular}
\end{table}
are complex valued, their magnitudes are used as feature parameters for the purpose of keeping the rotation invariance property.

The four kinds of feature parameters are applied to the recognition system. The window size $N$ and the window shift $S$ are set to 32 and 16, respectively, which showed a good performance in our experiments. The number of sample vectors for bias estimation, $M$, is set to 200, which corresponds to two seconds. Among the 32-point FFT results, the magnitudes of the first 11 coefficients excluding the DC component are used. The AR order $P$ is set to 4. Summarizing, a 34-dimensional feature vector $\vec{f}(n)$ is formed for each analysis frame in which

$$\vec{f}(n) = [\mu_x(n), \sigma_x^2(n), |A_x(n, 1)|, \cdots, |A_x(n, 1)|, b_x(n, 1), \cdots, b_x(n, 4),$$

$$|A_{xy}(n)|, \sigma_{xy}^2(n), |A_{xy}(n, 1)|, \cdots, |A_{xy}(n, 1)|, |b_{xy}(n, 1)|, \cdots, |b_{xy}(n, 4)]]. \quad (6)$$

3. HMM-Based Human Activity Recognition

HMM is one of the most successful techniques for sequential or temporal data classification with varying length such as speech. Because each human activity possesses an idiosyncratic pattern, recognition of human activity can be viewed as a typical pattern matching problem. Human activity recognition can be established in a similar way to the conventional speech recognition technique which is based on the HMM. Furthermore, we are motivated by the successful application of HMM to vision-based activity recognition [8], [9].

There are two major differences between human activity recognition and conventional speech or gesture recognition. First, the human activities have to be distinguished from speech or gestures which consist of several steps including start and end [8]. There are not any well-defined subunits in the acceleration data because the human activities are continuous and some of them are possible to last for a long time. From this difference, we assume that there exists only one activity during a prespecified short time interval $T$, which is set to two seconds in our system. This means that a human activity pattern is identified during every interval of length $T$. Since the sampling rate is 100 Hz and the window shift is 16, the most recent 12 feature vectors are input into the recognizer. This indicates that at each time a sequence of feature vectors of a fixed length, i.e., 12 frames which we call a package, is applied to the HMM-based recognizer. Second, the topology of the HMM applied to human activity recognition should be different from that of the conventional speech recognition. In contrast to the simple left-to-right structure shown in Fig. 2(a), we apply the HMM structure in Fig. 2(b), which allows both the backward and skip transitions. This modification is motivated by the fact that in human activity the state transition can occur in a more complicated manner and the starting and ending states may be arbitrary.

The target objects of recognition consist of six activities: standing, walking, running, falling, lying and jumping. Recognition is performed for each package with one frame shift, which outputs the recognition result for every 0.16 seconds as shown in Fig. 3. Let $\tilde{I}_n(n)$ be a six-dimensional vector of likelihoods corresponding to standing, walking, running, falling, lying and jumping states computed from the HMM recognizer with the $n$-th package. Since the six activity states are modeled by separate HMM’s, it is easy to obtain $\tilde{I}_n(n)$ from the given package.

In a number of preliminary experiments, we could observe that the recognized results were unstable especially when the transition in human activity occurred frequently. There happened spurious activity states to occur from time to time. To alleviate this problem, we employ two post-processing methods. First, the likelihood for each target object is smoothed with a forgetting factor $\lambda = 0.6$. Let $I_l(n)$ be the smoothed likelihood vector of $\tilde{I}_n(n)$. Then,

$$I_l(n) = (1 - \lambda) \times I_l(n - 1) + \lambda \times \tilde{I}_n(n). \quad (7)$$

Secondly, transitions to a new activity state are allowed only when the likelihood for this state is maximum for at least eight consecutive packages. This post-control technique enables a smooth transition from one activity state to another while preventing spurious recognition results.

4. Experimental Results

Performance of the proposed human activity recognition algorithm was evaluated on a DB of acceleration signals. The DB was constructed by collecting the signal picked up by a triaxial accelerometer that was attached to the waist belt. The DB enclosed all possible combinations of the aforementioned activities with a total amount of about five and half hours. The DB was divided into two groups: group A and group B. The group A was collected without a sensor rotation and applied to evaluate the performance in a 10-fold cross-validation manner. The group B was obtained with
various tilts of the sensor module and utilized in the tests to validate robustness of the proposed algorithms.

First, we compared the triaxial accelerometer data obtained after tilt compensation with the raw data when the tilt of the sensor module was 0°, 45° and 90°. Figure 4 plots the data obtained when a user was walking. We can see that the original accelerometer data depends much on the tilt of the sensor. In contrast, the tilt compensated data shows a similar pattern regardless of the degree of tilt. From the results, it can be said that the proposed feature extraction technique is robust to the tilt of the sensor module.

We evaluated the performance of the five different HMM-based activity recognition systems: Baseline, Prop, Prop_noPre, PropLTR and Prop_noPost. Baseline is a system built with the left-to-right state transition model in Fig. 2(a) without any tilt compensation or postprocessing. Prop is a system based on the proposed HMM topology given in Fig. 2(b) with both tilt compensation and postprocessing. Prop_noPre, PropLTR and Prop_noPost were similar to Prop but applied without the tilt compensation, with the left-to-right state transition model and without the postprocessing algorithm, respectively.

The performances of the five HMM-based systems were compared in terms of the package classification accuracy and the activity recognition accuracy. The package classification accuracy indicates the ratio of the number of packages that are correctly identified by the HMM recognizer, and it was measured over about 120,000 packages. In contrast, the activity recognition accuracy means the ratio of the number of correctly classified activity states considering substitution, deletion and insertion errors, which were computed over 1,200 activities. Table 2 shows the accuracies of package classification and activity recognition with various tilts of the sensor module. Note that some accuracies of activity recognition are negative because there existed a lot of insertion errors. From the results, we can see that 1) the tilt compensation technique is robust to the rotation of the sensor module, 2) application of the proposed state transition model achieved higher accuracies than that of the left-to-right model, and 3) the postprocessing technique is efficient in terms of the activity recognition accuracy due to the significant decrease in insertion errors. Based on these results, it can be concluded that the proposed three approaches, i.e. tilt compensation, extended HMM topology and postprocessing algorithm, produced much better recognition performance compared with the conventional approaches.

5. Conclusions

In this letter, we have presented a new human activity recognition technique. First, we have derived a set of feature vectors robust to the tilt and misplacement of the attached sensor module. The three dimensional raw data are rotated so that it can match the training data. Next, we have proposed a state transition model suitable for HMM-based activity recognition instead of the simple left-to-right topology. Finally, we have introduced postprocessing technique to stabilize the activity recognition results. In a number of experiments, the proposed approaches have shown good performance in human activity recognition.

Acknowledgments

This work was supported in part by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No. 2009-0083044) and by the Seoul R&B Program (10544).

Table 2 Accuracies of package classification and activity recognition (%).

<table>
<thead>
<tr>
<th>Database group (with tilted sensor module)</th>
<th>Recognizer</th>
<th>Package classification</th>
<th>Activity recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Baseline</td>
<td>90.23</td>
<td>−48.25</td>
</tr>
<tr>
<td></td>
<td>Prop</td>
<td>93.05</td>
<td>78.83</td>
</tr>
<tr>
<td></td>
<td>Prop_noPre</td>
<td>92.95</td>
<td>80.17</td>
</tr>
<tr>
<td></td>
<td>PropLTR</td>
<td>91.61</td>
<td>72.33</td>
</tr>
<tr>
<td></td>
<td>Prop_noPost</td>
<td>91.98</td>
<td>14.42</td>
</tr>
<tr>
<td>B</td>
<td>Baseline</td>
<td>68.32</td>
<td>−306.87</td>
</tr>
<tr>
<td></td>
<td>Prop</td>
<td>92.49</td>
<td>79.37</td>
</tr>
<tr>
<td></td>
<td>Prop_noPre</td>
<td>74.10</td>
<td>53.17</td>
</tr>
<tr>
<td></td>
<td>PropLTR</td>
<td>79.23</td>
<td>47.50</td>
</tr>
<tr>
<td></td>
<td>Prop_noPost</td>
<td>90.96</td>
<td>−8.21</td>
</tr>
</tbody>
</table>

References