Identifying Physical Traits with Body-worn Accelerometers: A Preliminary Investigation

Akira Masuda†
†School of Engineering
Osaka University, Japan

Takuya Maekawa‡
‡Graduate School of Information Science and Technology
Osaka University, Japan

masuda.akira@ise.eng.osaka-u.ac.jp, maekawa@ist.osaka-u.ac.jp

Abstract
This paper describes our preliminary investigation of the end user physical characteristics (e.g., gender, dominant hand, and skill at sport) that can be successfully estimated solely from body-worn accelerometers. For this purpose we use the huge quantities of data, which include 14880 labeled activities (e.g., walking and running) obtained from 61 subjects. We try to estimate various kinds of characteristics based on our simple idea ‘When the activity sensor data of two users are similar, the physical characteristics of the two users may also be similar.’ We consider that estimating the end user’s physical characteristics will enable us to realize new kinds of applications that automatically recommend information/services to an end user according to her estimated physical characteristics such as gender and weight.

Keywords: Body-worn accelerometer, Physical characteristics, Daily activity

1 Introduction
Advances in sensing and wireless communication technologies have led to the low cost production of small wireless body-worn sensor devices. In the near future, end users will wear small sensor-embedded devices such as wrist-watches, cellphones, and shoes, and their daily lives will be continuously recorded and recognized [1] by these sensors. Also, recently, several amusement parks (e.g., Disney World) have employed RFID sensor bracelets to provide personalized service for visitors. In this paper, we assume such a near-future sensor environment and attempt to estimate information about an end user’s physical traits such as height, weight, gender, dominant hand, and age solely by employing acceleration data obtained from body-worn accelerometers. We believe that the automatic estimation of physical characteristics with body-worn sensors will prove very useful in many application domains. We show some examples.

(1) We can construct a personalized recommender system that provides user-specific advertisement/information according to the user’s estimated characteristics. For example, when the system wishes to provide an advertisement about clothes, it can automatically recommend clothes that match the user’s height and weight. The system can also automatically recommend sports products according to the user’s dominant hand. Furthermore, the system can effectively recommend health-enhancing products and provide the end user with tips about maintaining a healthy life according to her estimated health indicators, e.g., weight and BMI.

(2) While the investigation described below mainly focuses on known (apparent) physical characteristics such as height, dominant hand, and sports experience, we consider our method applicable to the automatic estimation of hidden physical characteristics such as diseases, health indicators, and sporting ability. In particular, the automatic diagnosis of diseases that affect body movements by using always-on accelerometers could be a significant application of wearable sensor systems. In fact, several physiology studies have determined the difference between normal subjects’ gait acceleration signals and those of subjects with a disease, e.g., between healthy subjects and subjects with diabetes [2].

2 Estimation method

2.1 Outline of method
Our method consists of main two procedures; Preparation and Physical characteristic estimation as summarized in Fig. 1. In the Prepa-
We obtain labeled acceleration data including various kinds of activities from many training users in advance. Fig. 1(a) shows the outline of this procedure. In this procedure, (1) we extract features from the sensor data, (2) we compute similarities between the activities of training users according to each activity class by using the extracted features and construct rankings of each training user that include other training users in ascending order of similarity to the training user, (3) we compute attribute sets from the rankings to construct a model that estimates a certain kind of physical characteristic, e.g., gender or age, and (4) we then train the model by using the computed attributes and answers. An answer corresponds to the training user’s physical characteristic value that we want to estimate, e.g., gender or age. We compute the attributes from the physical characteristics of other similar training users based on our idea that when the activity sensor data of two users are similar, the physical characteristics of the two users may also be similar. When we want to estimate a height, for example, we use the heights of similar training users to compute the attributes.

### 2.2.1 Computing similarities

We first construct a feature vector that concatenates the features extracted from sensor data obtained from a training user’s accelerometers for each time slice. Then, we compute the activity similarity of each pair of training users by using their feature vector sequences that correspond to the activity class. We compute activity similarities between training users simply by using a Gaussian mixture model (GMM). Assume that we wish to compute an activity similarity between the ‘walk’ activities of training users A and B. We regard user A as a base user and user B as an object user, and compute the similarity between the object user’s ‘walk’ activity sensor data and the base user’s ‘walk’ activity model. We first model the ‘walk’ activity of base user A...
Table 1. Activities performed in our experiment.

<table>
<thead>
<tr>
<th>A</th>
<th>stand</th>
<th>D</th>
<th>sit</th>
<th>G</th>
<th>bicycle</th>
<th>J</th>
<th>use PC</th>
<th>M</th>
<th>play pingpong</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>walk</td>
<td>E</td>
<td>ascend stairs</td>
<td>H</td>
<td>brush teeth</td>
<td>K</td>
<td>draw on whiteboard</td>
<td>N</td>
<td>vacuum</td>
</tr>
<tr>
<td>C</td>
<td>run</td>
<td>F</td>
<td>descend stairs</td>
<td>I</td>
<td>wash dishes</td>
<td>L</td>
<td>write in notebook</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

with the GMM by using feature vectors extracted from her ‘walk’ sensor data obtained from her accelerometers. We employ the EM algorithm to estimate the GMM parameters [3]. There are 5 mixtures in our implementation. Then, we compute the GMM likelihood of each feature vector extracted from the ‘walk’ sensor data obtained from the accelerometers of object user B. We assume that the average likelihood over the feature vectors corresponds to the similarity between the ‘walk’ activity of base user A and that of object user B. We compute the similarity for each pair of training users. By doing so, we can obtain the activity similarity between the ‘walk’ activities of a training user (base user) A and those of each other training user. That is, we can rank training users from the computed similarities. This ranking reflects the similarities of ‘walk’ activities to those of user A. The leftmost column in Fig. 2 shows an example of the ranking of ‘walk’ activities when we focus on the right hand.

2.2.2 Computing attributes and learning model

We construct a model that estimates a certain kind of physical characteristic by using the above rankings and the physical characteristics of training users. Here, assume that we construct a model that estimates an end user’s gender. We focus on a training user A and compute a pair consisting of an attribute set and an answer. The answer corresponds to the gender of user A. We compute the attribute set by using the gender of similar users in the activity similarity rankings shown in Fig. 2. This is based on our idea that when the activity sensor data of two users are similar, the physical characteristics of the two users may also be similar. In each ranking, we compute attribute values by using the physical characteristics of the top-n similar training users. Here we employ a simple majority voting protocol. When n is 3 and the top-3 similar users consist of two males and one female, for example, we use ‘male’ as an attribute value. In our implementation, we use four n values (1, 3, 5, and 9), and compute an attribute for each n value.

Table 2. Physical characteristic information used in our experiment.

<table>
<thead>
<tr>
<th>name</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>dominant hand</td>
<td>(right, left)</td>
</tr>
<tr>
<td>gender</td>
<td>(male, female)</td>
</tr>
<tr>
<td>frequency of dish washing</td>
<td>(usually(3), sometimes(2), rarely(1), never(0))</td>
</tr>
<tr>
<td>touch typing capability</td>
<td>(yes(2), somewhat(1), no(0))</td>
</tr>
<tr>
<td>bicycle type</td>
<td>(upright, folding)</td>
</tr>
</tbody>
</table>

Table 3. Accuracies (percentages) of nominal physical characteristics.

<table>
<thead>
<tr>
<th></th>
<th>dominant hand</th>
<th>gender</th>
<th>bicycle type</th>
<th>frequency of dish washing</th>
<th>touch typing capability</th>
</tr>
</thead>
<tbody>
<tr>
<td>naive</td>
<td>88.5</td>
<td>52.5</td>
<td>50.8</td>
<td>60.7</td>
<td>46.0</td>
</tr>
<tr>
<td>proposed</td>
<td>98.4</td>
<td>91.8</td>
<td>93.4</td>
<td>62.3</td>
<td>57.4</td>
</tr>
</tbody>
</table>

2.3 Physical characteristic estimation

Unlabeled acceleration data obtained from an end user are given. Fig. 1 (b) shows the outline of this procedure. In this procedure, (1) we extract features from the sensor data, (2) an activity recognition system [1] labels the feature vectors, (3) we compute activity similarities between the end user and each training user according to each activity class and each sensor variation by using the labeled data, and construct rankings of the end user that include training users in ascending order of similarity to the end user, (4) we compute an attribute set from the rankings, and (5) we then estimate a certain kind of physical characteristic by using its corresponding model learned in the Preparation procedure and the attribute set. The first, third, and fourth sub-procedures are identical to those in the Preparation procedure. In the fifth sub-procedure, by using the attribute set, the model learned in the Preparation procedure estimates the value of the end user’s physical characteristics.

3 Experiment

3.1 Dataset

We used sensor data collected with wireless sensor nodes equipped with three-axis acceleration
sensors and sampling rates of 30 Hz. Each subject wore the sensor nodes on the wrists of both hands, the waist, and the right thigh. The subjects performed data collection sessions that included a random sequence of the activities listed in Table 1. Each subject also filled out a questionnaire that asked for information about the physical characteristics. We selected various kinds of physical characteristics related to the subjects ranging from basic information such as weight and gender to information about the activities listed in Table 1. The activity-related information included the dominant hand used in several activities, and the frequencies with which several activities were performed in their daily lives. The information about physical characteristics also included the types of objects used in the experiment, i.e., types of bicycle and types of vacuum cleaner. The characteristics used in our experiment are listed in Table 2.

### 3.2 Evaluation methodology

We evaluated our method using ‘leave-one-subject-out’ cross validation. That is, we regarded one subject as an end user and the remaining subjects as training users. We iterated the procedure so that each subject became an end user once, and we computed the estimation accuracies of the physical characteristics listed in Table 2.

We prepare naive methods for estimating physical characteristics and then compare them with our methods. A naive method for estimating nominal characteristics simply outputs the major nominal value among all training users. For example, when we want to estimate ‘gender,’ and the respective numbers of male and female training users are 10 and 20, the method outputs ‘female.’

### 3.3 Results

Table 3 tabulates the classification accuracies of nominal physical characteristics for each method. The accuracy means the percentage of correctly classified instances (end users).

#### [Dominant hand] With ‘dominant hand,’ the accuracy of our method was very high, and better than those of the naive methods as shown in Table 3. The features computed from the ‘draw on whiteboard’ and ‘write on notebook’ activity similarities contributed greatly to estimating a subject’s dominant hand when writing.

#### [Gender] We then focus on ‘gender’ results. We had considered it very difficult to estimate a subject’s gender using only accelerometers. Contrary to our expectation, the accuracy of our method was good and much higher than the accuracy of the naive method. The accuracy was approximately 92%. Female’s ‘run’ and ‘walk’ sensor data from the waist sensor were somewhat different from those of male’s data.

#### [Type of object] As regards ‘bicycle type,’ the accuracy of our method was good as shown in Table 3. The features computed from the ‘bicycle’ activity similarities provided an important contribution. Because the wheel size of an upright bicycle is larger than that of a folding bicycle, the dominant frequency of the acceleration data obtained from the thigh when a subject rides an upright bicycle may be different from that for a folding bicycle.

#### [Daily activity related property] As regards ‘frequency of dish washing,’ the accuracy of our method was poor and it is not very different from that of the naive method. This may be because the problem is the four-class classification problem. Also, it may be very difficult to find difference between sensor data from subjects with ‘usually’ property and those with ‘sometimes’ property. With ‘touch typing capability,’ while the accuracy of our method was poor, it was somewhat better than that of the naive method.

### Conclusion

This paper proposes a method for estimating physical characteristics by using body-worn accelerometers. As a part of future work, we plan to modify the method for calculating sensor data similarity. Also, we plan to estimate numerical characteristics such as height and weight.  

### Acknowledgement

We thank Dr. Yasushi Sakurai at NTT Communication Science Laboratories for supplying acceleration sensor data.

### References

