

Investigating Inter-Subject and Inter-Activity Variations in Activity Recognition Using Wearable Motion Sensors

BILLUR BARSHAN* AND ARAS YURTMAN

Department of Electrical and Electronics Engineering, Bilkent University, Bilkent, TR-06800 Ankara, Turkey

**Corresponding author: billur@ee.bilkent.edu.tr*

This work investigates inter-subject and inter-activity variability of a given activity dataset and provides some new definitions to quantify such variability. The definitions are sufficiently general and can be applied to a broad class of datasets that involve time sequences or features acquired using wearable sensors. The study is motivated by contradictory statements in the literature on the need for user-specific training in activity recognition. We employ our publicly available dataset that contains 19 daily and sports activities acquired from eight participants who wear five motion sensor units each. We pre-process recorded activity time sequences in three different ways and employ absolute, Euclidean and dynamic time warping distance measures to quantify the similarity of the recorded signal patterns. We define and calculate the average inter-subject and inter-activity distances with various methods based on the raw and pre-processed time-domain data as well as on the raw and pre-processed feature vectors. These definitions allow us to identify the subject who performs the activities in the most representative way and pinpoint the activities that show more variation among the subjects. We observe that the type of pre-processing used affects the results of the comparisons but that the different distance measures do not alter the comparison results as much. We check the consistency of our analysis and results by highlighting some of our activity recognition rates based on an exhaustive set of sensor unit, sensor type and subject combinations. We expect the results to be useful for dynamic sensor unit/type selection, for deciding whether to perform user-specific training and for designing more effective classifiers in activity recognition.

Keywords: wearable sensing; motion capture; motion sensors; inertial sensors; accelerometer; gyroscope; magnetometer; activity recognition and classification; inter-subject variation; inter-activity variation; feature extraction; feature reduction; dynamic time warping

Received 21 January 2015; revised 22 June 2015

Handling editor: Raif Onvural

1. INTRODUCTION AND RELATED WORK

Recent work on automatically recognizing daily activities focuses on machine learning algorithms that rely on simultaneous input from several different sensor modalities such as visual, inertial, acoustic, force, pressure, strain, physiological and kinetic sensors, among others [1, 2]. Collecting information about a user's activities for ambient-assisted living in smart homes and detecting abnormal behavior to assist the elderly or people with special needs are challenging research issues [3, 4]. These systems aim to maintain the user's

independence, enhancing their personal safety and comfort and delaying the process of moving to a care home. However, automatic monitoring of people performing daily activities should be done without restricting their independence, intruding on their privacy or degrading their quality of life.

A commonly used approach in designing smart environments involves the use of one or more types of external sensors in a complementary fashion (e.g. cameras and tactile sensors), usually with relatively high installation cost and heavy demands on computing power [5, 6]. If a single camera is used, the 3D

scene is projected onto a 2D one, with significant information loss. Other people or pets moving around may easily confuse such systems. Occlusion or shadowing of points of interest (by human body parts or objects in the surroundings) is resolved by using 2D projections from multiple cameras in the environment to reconstruct the 3D scene. Each camera needs to be individually calibrated and suffers from the correspondence problem. To resolve the latter, points of interest on the human body are pre-identified by placing special, visible markers at those points and the positions of the markers are recorded by cameras. Processing and storing camera recordings is costly and camera systems obviously interfere with privacy. Recorded data are highly sensitive to privacy breaches when transmitted or stored [7]. Continuous monitoring may cause stress and discomfort on the subject and may subsequently cause changes in his natural movements.

The main advantage of embedding external sensors in the environment is that the person does not have to wear or carry any sensors or devices. This approach may also eliminate problems related to misplacing sensors on the body, although some camera systems do require wearing/pasting on special tags or markers as mentioned above. Designing smart environments may be acceptable when the activities of the person are confined to certain parts of a building. However, when the activities are performed both indoors and outdoors and involve going from one place to another (e.g. riding a vehicle, going shopping, commuting, etc.), this approach becomes unsuitable. It imposes restrictions on the mobility of the person since the system operates only in the limited environment being monitored.

The use of wearable motion sensors in activity recognition has pervaded since this approach is superior to using external sensors in many respects. The required infrastructure and associated costs of wearable sensors are much lower than designing smart environments. Unlike visual motion-capture systems that require a free line-of-sight, wearable sensors can be flexibly used inside or behind objects without occlusion. They can acquire the required 3D motion data directly on the spot without the need for multiple camera projections. The 1D signals acquired from the multiple axes of wearable motion sensors are much simpler and faster to process. Because they are light, comfortable and easy to carry, wearable sensors do not restrict people to a studio-like environment and can operate both indoors and outdoors, allowing free pursuit of activities without intruding on privacy.

Wearable systems are criticized mainly because people may forget, neglect or not want to wear them. If they are battery operated, batteries need to be recharged or replaced from time to time. However, with the advances of the MEMS technology, these devices have been miniaturized. Their lightness, low power consumption and wireless use have eliminated the concerns related to carryability and discomfort. Furthermore, the algorithms developed can be easily embedded to a device

or accessory that the person normally carries, such as a mobile phone or a hearing aid. Wearable sensors are thus suitable for automatic monitoring and classification of daily activities, and we have chosen to follow this approach in our works [8–10].

Although numerous studies on activity recognition with wearable sensors exist (see [11, 12] for recent surveys), only a small number of them have looked into the minimal number and configuration of sensors to recognize activities. Minimizing the number of sensors would improve the user's comfort as well as reduce the complexity and energy consumption of the system since less data would have to be processed. Existing works report conflicting results on the placement of a single device such as inside the trousers' pocket, in a bag carried by the user, on the belt, chest or on the dominant wrist [11]. Some studies claim that the arms and the legs are not suitable for carrying a device since they are associated with higher accelerations in general.

In studies that use multiple sensors, the highest recognition accuracy is usually achieved when all of the sensors are employed and decreases when a subset of the sensors is omitted. For example, in [13], bi-axial accelerometers are placed on the individual's hip, wrist, arm, ankle, thigh and combinations of them. They conclude that with only two accelerometers (i.e. either thigh and wrist or hip and wrist) the recognition performance drops only slightly and such a configuration is sufficient to recognize ambulation and other daily activities. In [14], the effect of the number of sensors on recognizing 10 manipulative upper-body activities of assembly-line workers in a car production environment is investigated. Sensors are selected according to their contribution to the classification accuracy while training the system. The accuracy tends to increase with the number of sensors used. The highest accuracy of 98% is achieved when all of the 19 sensors are included and drops to 97% with only three sensors out of 19. Considering that most studies on activity recognition use their own customized datasets (usually not publicly available) with different characteristics and the classification techniques differ widely, it is not possible to reach a general conclusion about the number and placement of the sensors. It appears that the best sensor configuration depends on the application, the type of activities to be recognized and the desired level of recognition accuracy.

In many of the studies on human activity recognition, the acquired activity data vary considerably among subjects, which we refer to as *inter-subject variability*. More specifically, the amplitude and pace of performing activities vary from subject to subject according to their personal styles and anthropometry (i.e. physical attributes). This kind of variability may also be caused by movement disorders. The variation in time is often nonlinear and may be difficult for an artificial system to perceive. Another type of variation in the data is *intra-subject variability*, where the nature of a given activity performed by the *same* subject at different times shows variations. Possible reasons for this type of variability are natural variations

in the movements, the random nature of an activity (such as playing basketball), changes in the subject or the environment (such as motivation and physical energy level, clothing, shoes worn, surface on which the subject performs activities), deviation in sensor positions and orientations, and measurement errors.

This study is motivated by the contradictory statements in the literature related to user-specific training. In [13], an average accuracy of 80% is achieved with decision-tree-based classifiers in recognizing the daily activities considered, and the authors claim that user-specific training is not necessary to achieve good recognition accuracy. In [15], an activity dataset is acquired from 25 subjects whose physical properties are homogeneous. In activity recognition, leave-one-subject-out cross validation performed significantly better than user-specific training, even when the size of the training set was equalized in both cases by randomly sampling the training set in the former. This is explained by the homogeneity of the subjects and contradicts most of the existing studies where leave-one-subject-out cross validation performs worse. For example, in [16], where physical therapy exercises are classified using feature vectors obtained from accelerometer data, three cases are considered: using the training data of (i) the subject being tested (within-subject cross validation), (ii) all subjects (across-subjects cross validation) and (iii) subjects other than the one being tested (leave-one-subject-out cross validation). It is shown that the accuracy achieved in the last case is significantly lower than in the first two. As expected, classification accuracy degrades significantly if one attempts to recognize the activities of a given subject with a system trained on other subjects' data. For this reason, most systems designed, for example, for physical therapy and rehabilitation, are individually trained for each subject so that the reference data are directly acquired from the subject who will use the system [16–18].

Building a system that can perform well when presented with an unknown subject's data (or first-seen data) for testing is challenging. Trying to build a subject-independent classifier by including a variety of subjects in the training set may improve classifier accuracy by encompassing all common variations and capturing more of the personal styles. Another solution may be to develop an individual compensation method for each subject based on his physical parameters. However, measuring and recording the parameters of the subjects may not be trivial. This approach is followed in some studies on energy expenditure estimation that use wearable sensors, such as [19]. In a system that estimates the joint angles of the knees from inertial sensor data, physical properties of the legs of each subject are calculated from captured photos and provided to the system to get more accurate results [20]. Although these systems are individually calibrated for each person, personal variations in the properties and their effects on the system has not been investigated.

Although inter-subject variability of vision-based systems has been examined to some extent [21–23], there are very few

studies that explore inter- and intra-subject variability of wearable sensor data in activity recognition. In [24], the intra-subject variability of the accelerometer data of free-living activities performed by 17 male patients with chronic obstructive pulmonary disease is studied, and observed to be low. In [25], intra- and inter-subject variations in the walking behavior of 24 subjects with multiple sclerosis are investigated. Based on the integrated signals acquired from accelerometers on the waist and the ankle, a quantity called 'movement count' is estimated and observed to be highly correlated with the walking speed, which is considered to be a subject-specific property. In both of these studies, the recorded data are from rare disease populations that cannot be generalized to people free from movement disorders.

The purpose of this study is to investigate inter-subject and inter-activity variability of wearable sensor data and provide some definitions to quantify such variability [26, 27]. We demonstrate our analysis and results on our publicly available daily- and sports-activities dataset acquired using five wearable motion sensor units from eight subjects while performing 19 types of activities. Each sensor unit contains three tri-axial devices: an accelerometer, a gyroscope and a magnetometer. We calculate the average inter-subject and inter-activity distances of wearable sensor data in various forms by employing absolute, Euclidean and dynamic time warping (DTW) distance measures to quantify the similarity of the time sequences. The comparisons are based on raw and pre-processed time-domain data as well as on feature vectors. First, we calculate the average inter-subject distances per subject and per activity. Based on the minimum average inter-subject distance per subject, we identify the subject who performs the activities in the most representative way. We also specify the activities that show more variation among the subjects. Then, we present the average inter-activity distances and their standard deviations per subject, per sensor unit and per sensor type. We summarize the results in various forms and discuss the effects of using different types of pre-processing and distance measures on the results.

The rest of this article is organized as follows: In Section 2, we provide a brief description of our publicly available activity dataset. In Section 3, we present the three distance measures used to quantify the similarity between time sequences and feature vectors. Section 4 describes the pre-processing of the time sequences and feature vectors. In Section 5.1, we calculate the average inter-subject distance per subject, and based on this quantity, identify the subject who performs the activities in the most representative way. We calculate the average inter-subject distance per activity in Section 5.2, and in Section 5.3, we investigate the average inter-activity distances per subject, per sensor unit and per sensor type separately. To verify some of our results, in Section 6, we provide activity classification results based on an exhaustive set of sensor unit, sensor type and subject combinations. We draw conclusions and indicate directions for future research in Section 7.

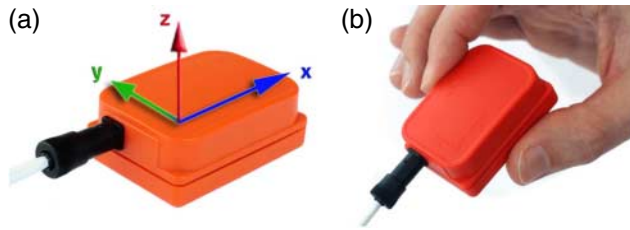


FIGURE 1. (a) MTx with sensor-fixed coordinate system overlaid. (b) MTx held between two fingers (both parts of the figure are reprinted from [29]).

2. DAILY- AND SPORTS-ACTIVITIES DATASET

The dataset used in this study was acquired by our research team and was made publicly available in July 2013 [28]. In the experiments, eight subjects wearing five motion sensor units, performed 19 daily and sports activities, each activity lasting for 5 min. The subjects were free from any movement disorders. The physical attributes of the subjects can be found in [8]. Figure 1 depicts the sensor units and Fig. 2 illustrates their configuration on the body. Each sensor unit contains three tri-axial devices: an accelerometer, a gyroscope and a magnetometer, resulting in nine sensor axes. The orientations of the sensor axes are shown in Figs 1a and 3b. The activities are the following:

Sitting (A_1), standing (A_2), lying on back and on right side (A_3 and A_4), ascending and descending stairs (A_5 and A_6), standing still in an elevator (A_7) and moving around in an elevator (A_8), walking in a parking lot (A_9), walking on a treadmill in flat and 15° inclined positions at a speed of 4 km/h (A_{10} and A_{11}), running on a treadmill at a speed of 8 km/h (A_{12}), exercising on a stepper (A_{13}), exercising on a cross trainer (A_{14}), cycling on an exercise bike in horizontal and vertical positions (A_{15} and A_{16}), rowing (A_{17}), jumping (A_{18}) and playing basketball (A_{19}).

The first four activities are considered to be static, whereas the remaining ones are dynamic. Dynamic activities can be further classified as quasi-periodic, such as walking, cycling or rowing, and those that contain random elements, such as playing basketball.

For each of the eight subjects and 19 activities, we acquired 45 ($= 5$ units $\times 9$ sensors) recordings, each one of five minutes' duration. Each recording corresponds to a discrete-time sequence $x_{...}[n]$ that we obtain by uniformly sampling a continuous-time signal $x_{...}(t)$ at a rate of 25 Hz, and consists of $N = 7500$ samples:

$$x_{p,a,u,s}[n] = x_{p,a,u,s}(t)|_{t=n/25} \quad (1)$$

where $0 \leq t \leq 300$ s, $1 \leq n \leq N$, $p \in [1, 8]$ is the subject index, $a \in [1, 19]$ is the activity index, $u \in [1, 5]$ is the unit index and $s \in [1, 9]$ is the sensor index. The number of subjects, activities,

units and sensors are $N_p = 8$, $N_a = 19$, $N_u = 5$ and $N_s = 9$, respectively.

As an alternative to using time sequences, to represent the data with its statistical properties and reduce the amount of data to be processed, we extract feature vectors based on the time sequences in exactly the same way as described in [9]. Since each activity of each subject is recorded for 5 min and the recording is divided into 5-s segments, we extract $N_k = 60$ [$= (5 \times 60)/5$] feature vectors for each activity of each subject, and a total of 9120 ($= 60$ feature vectors $\times 19$ activities $\times 8$ subjects) feature vectors of dimension 1170×1 . Each feature vector contains the features extracted from the same 5-s segment of all 45 discrete-time sequences of a particular activity of a particular subject. The features extracted from each segment of each axis are the following: the minimum, maximum, mean, variance, skewness and kurtosis values, the auto-correlation sequence, and the peaks of the discrete Fourier transform with the corresponding frequency values. Feature vectors are denoted by $\mathbf{v}_{p,a}\{k\}$, where $k \in [1, N_k]$ is the segment index. (Please refer to [9] for further details on the feature vectors).

3. DISTANCE MEASURES

The choice of the distance measure depends on the invariances required by the domain of application [30]. Motion capture data typically requires invariance to warping [30]. We employ three commonly used distance measures to calculate the distance between two discrete-time sequences $\mathbf{x} = \{x[1], x[2], \dots, x[N]\}$ and $\mathbf{y} = \{y[1], y[2], \dots, y[N]\}$:

(1) absolute (taxicab) distance:

$$d_{\text{abs}}(\mathbf{x}, \mathbf{y}) = \sum_{n=1}^N |x[n] - y[n]| \quad (2)$$

(2) Euclidean distance:

$$d_{\text{Euc}}(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{n=1}^N (x[n] - y[n])^2} \quad (3)$$

(2) DTW distance:¹

$$d_{\text{DTW}}(\mathbf{x}, \mathbf{y}) = \text{DTW}(\mathbf{x}, \mathbf{y}) \quad (4)$$

Here, $d_{\mathcal{D}}(\mathbf{x}, \mathbf{y})$ denotes the distance between the two time sequences calculated according to the distance measure \mathcal{D} . Regarding Equation (4), there is no closed-form expression for the DTW distance between the two sequences. In this type of

¹ The basic step pattern is used to calculate the DTW distance, where the warping path can proceed to one of the three adjacent cells in the horizontal, vertical or diagonal direction [31]. Costs for the three directions are assigned uniformly.



FIGURE 2. Positioning of the MTx units on the body.

distance measure, the sequences \mathbf{x} and \mathbf{y} are matched by elastically transforming or warping their time (or sample) axes to make them resemble each other as much as possible. In this way, similar patterns in the sequences and features, such as local minima and maxima, can be matched [31]. In the end, the Euclidean measure (as in this study) or another distance measure is used, whose square provides the DTW distance between the matched time sequences. Note that for the absolute and

Euclidean distance measures, the two time sequences must be of the same length, whereas there is no such constraint for the DTW distance measure. Therefore, the DTW distance measure can be used in the more general case where the sequences \mathbf{x} and \mathbf{y} have different lengths. An example from our dataset that illustrates the difference between the Euclidean and DTW distance measures is provided in Fig. 4. The former makes a one-to-one alignment, whereas the latter allows a one-to-many

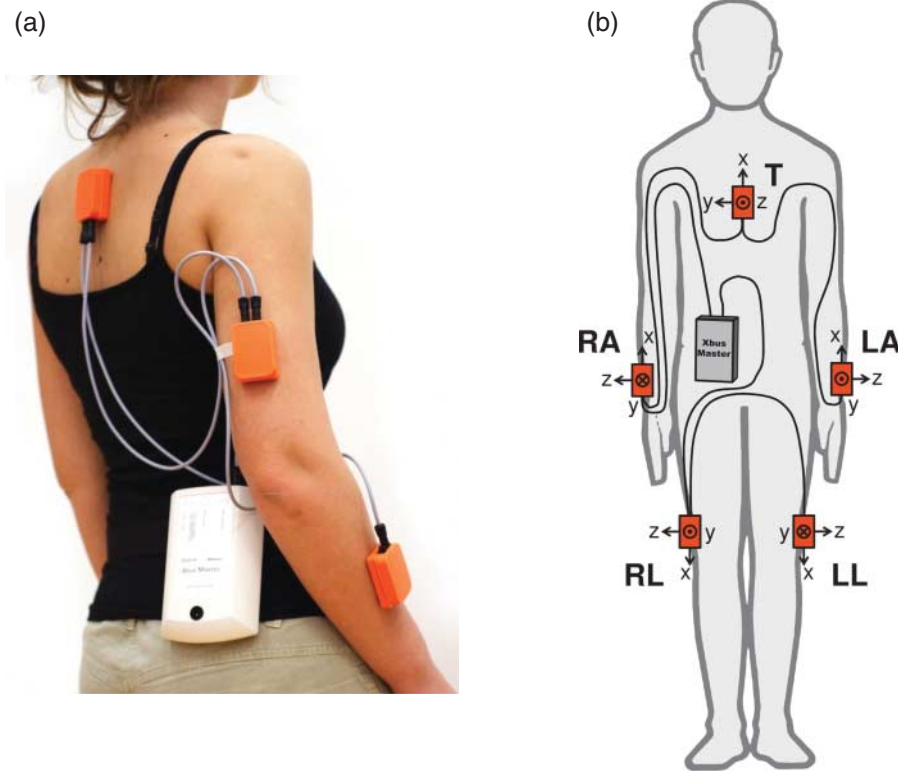


FIGURE 3. (a) MTx units and Xbus Master [29] and (b) connection diagram of MTx units (the body drawing in the figure is from <http://www.clker.com/clipart-male-figure-outline.html>; the cables, Xbus Master and sensor units were added by the authors).

alignment. We note here that the DTW distance is comparable to the square of the Euclidean distance. In fact, Euclidean distance is a special case of the DTW distance: If no warping is performed in DTW, the DTW distance is exactly equal to the square of the Euclidean distance. If warping is involved, it is less than the square of the Euclidean distance.

The computational complexity of the DTW distance measure is proportional to the product of the lengths of the two sequences to be compared [31]. If the two sequences have the same length, the computational complexity of DTW is $\mathcal{O}(N^2)$, whereas it is $\mathcal{O}(N)$ for the Euclidean distance measure. Thus, when the DTW distance measure is used to calculate the distances, calculations last about $N = 7500$ times longer for time-domain data and $N_k = 60$ times longer for feature vectors, compared with using the Euclidean distance measure. However, by performing a *subsequence search* or *subsequence monitoring task* to compare time-domain sequences [31–33], the speed of DTW may be improved considerably so that it becomes slower than the Euclidean distance by less than a factor of two and can be even further optimized [34].

DTW can be generalized to the multi-dimensional case (which is the case in this study) in two different ways: In the first (DTW_I: independent DTW), sequences acquired from multiple data channels are warped independently, whereas,

in the second (DTW_D: dependent DTW), they are all warped in the same way. Whether a given dataset should be handled using DTW_I or DTW_D is difficult to determine and, in general, it contains a mixture of exemplars that belong to either [35]. Reference [35] proposes an adaptive methodology (DTW_A: adaptive DTW) which is at least as accurate as the better of DTW_I or DTW_D. In this study, we have chosen to employ DTW_I because we believe that the sequences comprising our dataset do not have strong dependence.

4. PRE-PROCESSING OF THE ACQUIRED SEQUENCES

The wearable system is self-calibrating in that the sensor units are auto-calibrated by the driver of the Xbus kit before and during the data acquisition process. However, recorded time sequences still contain bias error that may be constant or time varying. The level of the bias error is, in general, different for each axis of a given sensor unit. This error will affect the time average of the sequence that may be an informative statistical feature about the activity being performed. The bias error is an important and complicating issue in interpreting sensor readings and its effects should be investigated.

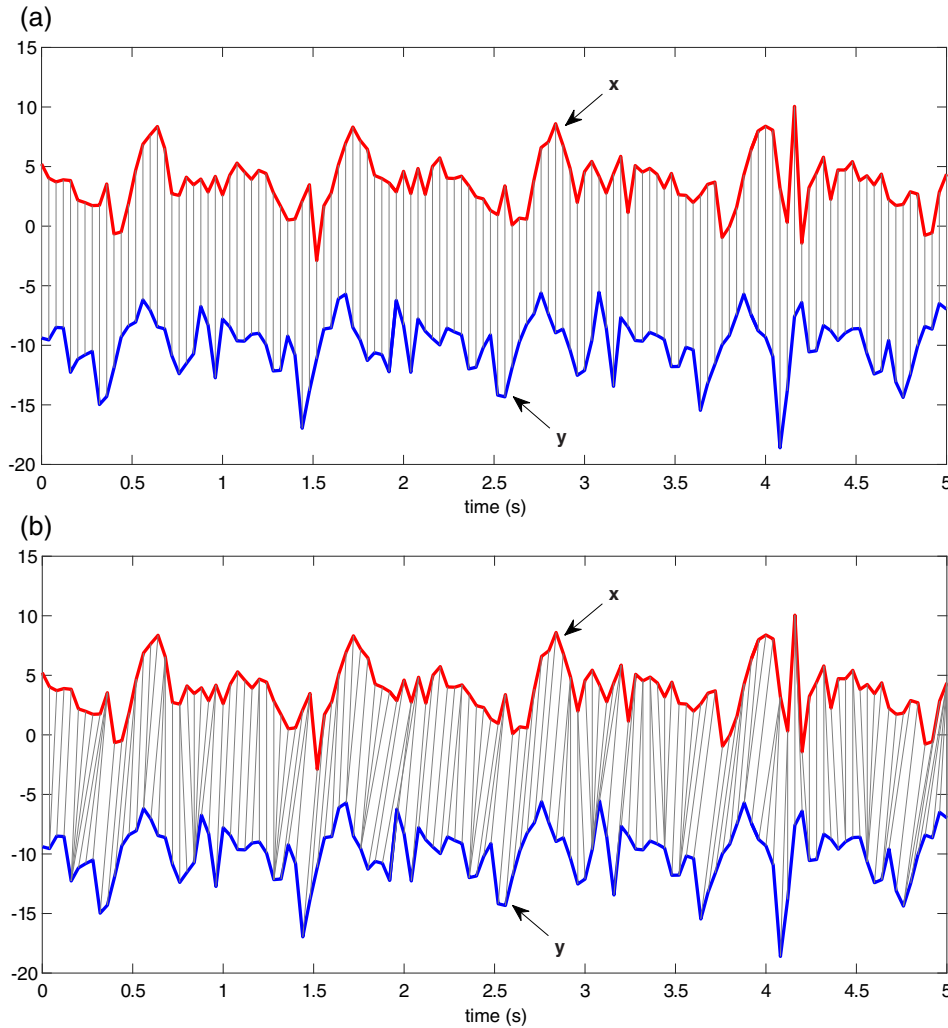


FIGURE 4. Comparison of the Euclidean and DTW distance measures. (a) The Euclidean measure compares the samples at the same time instants, whereas (b) the DTW measure compares samples with similar shapes to minimize the distance. The sequences \mathbf{x} and \mathbf{y} are recorded by the x -axis accelerometers on the right legs of subjects 1 and 2, respectively, while performing activity A_{10} .

If two sequences with different bias levels are compared, the distance between them depends on the type of distance measure used. Assuming constant bias levels, suppose that the elements of the sequences \mathbf{x} and \mathbf{y} are related by $y[n] = x[n] + E$, for $1 \leq n \leq N$, where $E > 0$ is a constant bias error. Then, according to the three distance measures, the distance between \mathbf{x} and \mathbf{y} is given by

$$d_{\text{abs}}(\mathbf{x}, \mathbf{y}) = \sum_{n=1}^N |E| = NE \quad (5)$$

$$d_{\text{Euc}}(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{n=1}^N E^2} = \sqrt{NE^2} = \sqrt{N}E \quad (6)$$

$$d_{\text{DTW}}(\mathbf{x}, \mathbf{y}) \leq NE^2 \quad (7)$$

We observe from Equations (5)–(7) that for all three distance measures, the distance between the two sequences is a function of the amount of bias error E and the sequence length N . As a numerical example, for $N = 100$ and $E = 0.01$, $d_{\text{abs}}(\mathbf{x}, \mathbf{y}) = 1$, $d_{\text{Euc}}(\mathbf{x}, \mathbf{y}) = 0.1$ and $d_{\text{DTW}}(\mathbf{x}, \mathbf{y}) \leq 0.01$. Thus, depending on the values of N and E , there can be several orders of magnitude difference between the distances calculated by using different measures.

As an attempt to eliminate the constant part of a bias error, in the first type of pre-processing, we remove the time average of the original time sequence $\mathbf{x}_{p,a,u,s}$ to obtain the sequence $\bar{\mathbf{x}}_{p,a,u,s}$:

$$\bar{\mathbf{x}}_{p,a,u,s} = \mathbf{x}_{p,a,u,s} - \langle \mathbf{x}_{p,a,u,s} \rangle_n \quad (8)$$

Besides removing the time average, the sequence values may be scaled with their standard deviation, calculated over time, to have unity standard deviation. Thus, the sequence obtained as a

result of this second type of pre-processing is given by

$$\tilde{\mathbf{x}}_{p,a,u,s} = \frac{\bar{\mathbf{x}}_{p,a,u,s}}{\text{std}_n \bar{\mathbf{x}}_{p,a,u,s}} \quad (9)$$

In the third type of pre-processing, we limit the sequence to the interval $[-1, 1]$ by shifting and scaling it, a common procedure especially before applying DTW. The resulting sequence is given by

$$\hat{\mathbf{x}}_{p,a,u,s} = 2 \frac{\mathbf{x}_{p,a,u,s} - \min_n \mathbf{x}_{p,a,u,s}}{\max_n \mathbf{x}_{p,a,u,s} - \min_n \mathbf{x}_{p,a,u,s}} - 1 \quad (10)$$

Similar pre-processing is also applied to the feature vectors extracted from the time sequences.

5. INTER-SUBJECT AND INTER-ACTIVITY VARIATIONS IN THE ACTIVITY DATASET

5.1. Average inter-subject distance per subject

In this section, we calculate the average distance between the time sequences/features acquired from one subject and those of all the other subjects to identify the person who performs the activities in the most representative way. We apply the three different distance measures to both raw and pre-processed time sequences as well as to raw and zero-mean feature vectors.

We propose a measure of similarity between two subjects p_1 and p_2 based on their activity data in the time domain as follows: For each activity and for each sensor of each unit, we calculate the distance between the time sequences of the two subjects and average them out over all activities, units and sensors. The resultant value is considered to be the average inter-subject distance between the activities of the two subjects:

$$d_{\text{inter-subject}, \mathcal{D}}^{\text{time-domain}}(p_1, p_2) = \frac{1}{N_a N_u N_s} \sum_a \sum_u \sum_s d_{\mathcal{D}}(\mathbf{x}_{p_1,a,u,s}, \mathbf{x}_{p_2,a,u,s}) \quad (11)$$

where the subscript \mathcal{D} denotes one of the three distance measures defined in Section 3.

To compare the same two subjects based on their feature vectors, we average out the distance between the feature vectors of the two subjects over all time segments and activities:

$$d_{\text{inter-subject}, \mathcal{D}}^{\text{features}}(p_1, p_2) = \frac{1}{N_a N_k} \sum_a \sum_k d_{\mathcal{D}}(\mathbf{v}_{p_1,a}\{k\}, \mathbf{v}_{p_2,a}\{k\}) \quad (12)$$

To identify those subjects who are most similar to the others in the average distance sense, we first calculate the inter-subject distances between all subject pairs using Equations (11) and (12). Then, for each subject, we average out the distances from that subject to all the others, resulting in the average inter-subject distance of that person to the others. If

time sequences are used, for the subject with index p_1 , this is given by

$$d_{\text{avg-subject}, \mathcal{D}}^{\text{time-domain}}(p_1) = \frac{1}{N_p - 1} \sum_{p \neq p_1} d_{\text{inter-subject}, \mathcal{D}}^{\text{time-domain}}(p_1, p) \quad (13)$$

If feature vectors are used, the corresponding expression is

$$d_{\text{avg-subject}, \mathcal{D}}^{\text{features}}(p_1) = \frac{1}{N_p - 1} \sum_{p \neq p_1} d_{\text{inter-subject}, \mathcal{D}}^{\text{features}}(p_1, p) \quad (14)$$

According to this scheme, the subject with the smallest average inter-subject distance to the others (in other words, the subject who performs the activities in the most similar way to the others) is considered to be the person who performs them most representatively, and is represented by index p^* in

$$(p^*)_{\mathcal{D}}^{\text{time-domain}} = \arg \min_p d_{\text{avg-subject}, \mathcal{D}}^{\text{time-domain}}(p) \quad (15)$$

or

$$(p^*)_{\mathcal{D}}^{\text{features}} = \arg \min_p d_{\text{avg-subject}, \mathcal{D}}^{\text{features}}(p) \quad (16)$$

depending on whether time sequences or feature vectors are used.

Although the proposed method to identify the subject who performs the activities most representatively may appear to evaluate the subjects' performances, this is not the case. The most representative subject p^* is not necessarily the person who performs the activities most properly or correctly; he is the one in the middle, in the sense that he does not perform the activities in any extreme way. In other words, the selection depends on how similarly a subject performs the activities compared with all the others. For instance, assuming that the dataset consists of only the walking activity and the subjects differ only in their walking speed, this approach would identify the subject who walks nearest to the average speed as the most representative.

The results for identifying the most representative subjects are summarized in Fig. 5, where the calculations of the expressions in Equations (13) and (14) are presented for each of the eight subjects. From the figure, we observe that pre-processing the time sequences and feature vectors decreases the average inter-subject distance by a factor of ~ 5000 . This is more visible when we use the DTW distance, because for the DTW algorithm to warp the two sequences properly, they must be of comparable amplitude scale. If the amplitude levels and amplitude ranges of the sequences differ too much, DTW cannot match their similar parts because DTW cannot scale or shift the sequences' amplitude values; it only warps their time (or sample) axes. For this reason, applying the DTW distance measure to the raw sequences is not expected to result in small distance values. When we compare the time sequences normalized between -1 and 1 by using the absolute, Euclidean and DTW distance measures, we observe that Subject 2 is identified to be

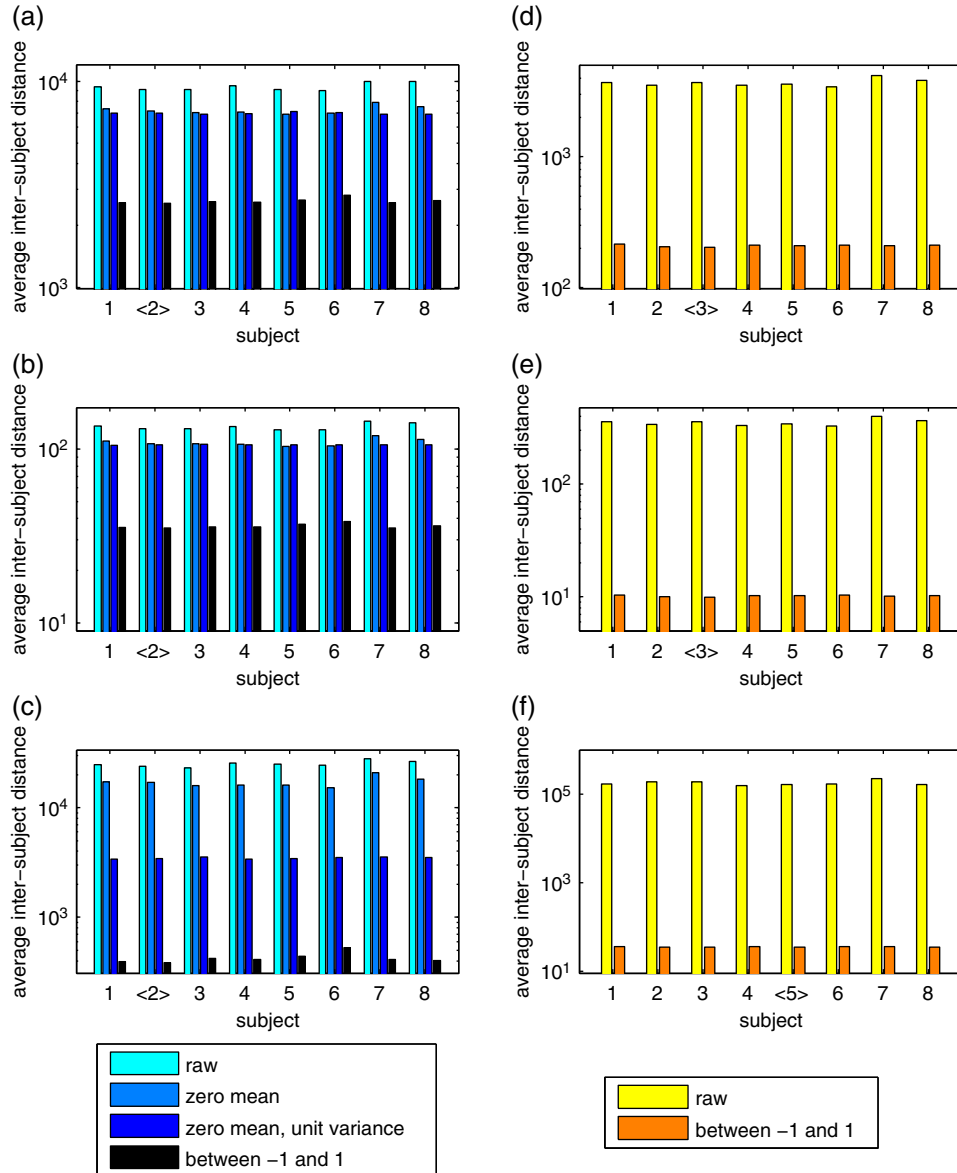


FIGURE 5. Average inter-subject distance in terms of the (a), (d): absolute, (b), (e): Euclidean, and (c), (f): DTW distances when (a)–(c): raw and pre-processed time sequences and (d)–(f): feature vectors are used. The subject index with the smallest distance (for the sequences normalized between -1 and 1) is enclosed in brackets.

the most representative (Fig. 5a–c). On the other hand, when we compare the feature vectors normalized between -1 and 1 by using the first two distance measures, Subject 3 is identified to be the most representative (Fig. 5d and e). When we use the DTW distance measure to compare the same, Subject 5 is identified as the most representative (Fig. 5f).

5.2. Average inter-subject distance per activity

The average inter-subject distance for a given activity is a measure of how differently a specific activity can be performed by different subjects. This quantity can be considered as a

measure of the inter-subject variability of the activity. For a given activity a , the average inter-subject distance for the distance measure \mathcal{D} is

$$d_{\text{avg-activity}, \mathcal{D}}(a) = \frac{1}{N_p(N_p - 1)} \sum_{p_1} \sum_{p_2 \neq p_1} \left[\frac{1}{N_u N_s} \sum_u \sum_s d_{\mathcal{D}}(\mathbf{x}_{p_1, a, u, s}, \mathbf{x}_{p_2, a, u, s}) \right] \quad (17)$$

In the equation above, the term in the square brackets is the inter-subject distance between two subjects p_1 and p_2 for

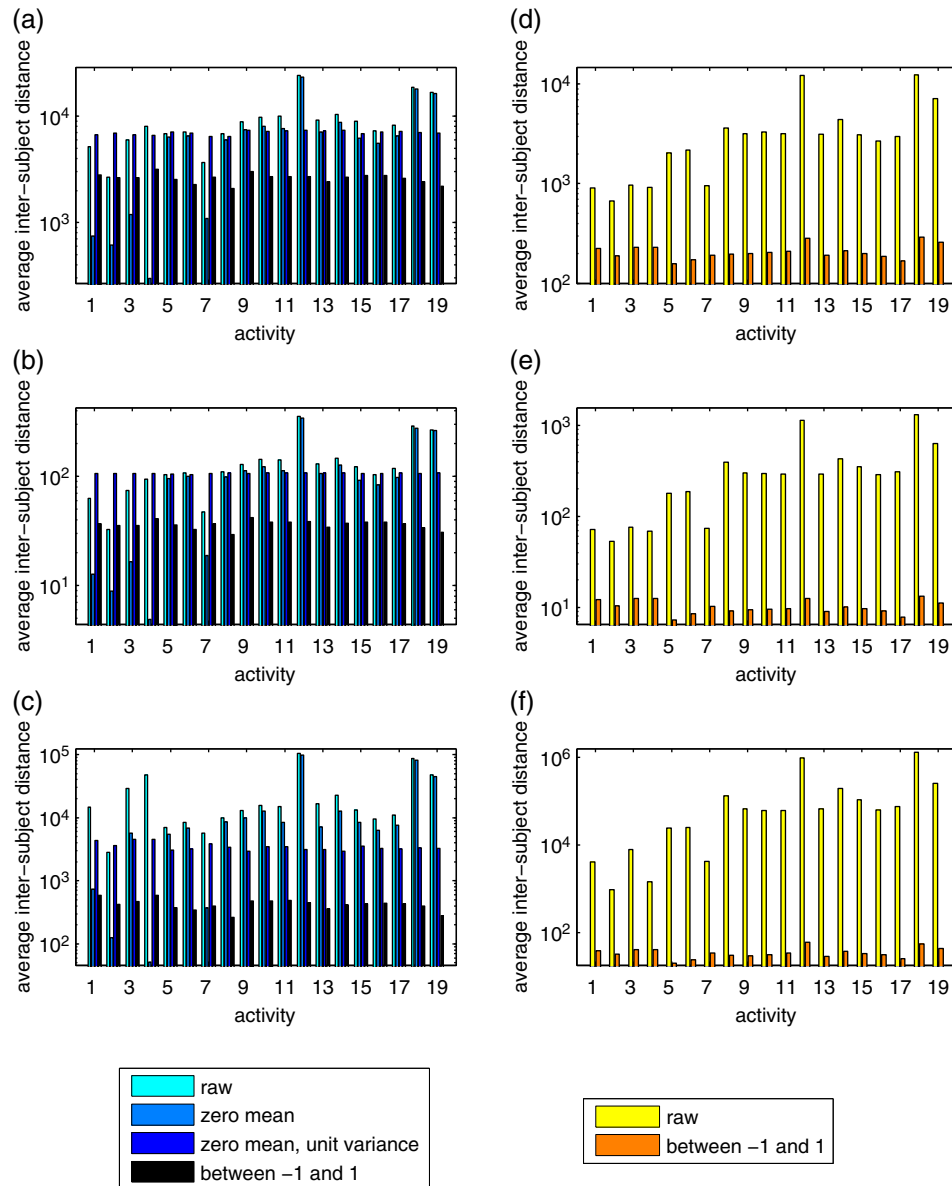


FIGURE 6. Average inter-subject distance for each activity in terms of the (a), (d): absolute, (b), (e): Euclidean, and (c), (f): DTW distances when (a)–(c): raw and pre-processed time sequences and (d)–(f): feature vectors are used.

activity a , averaged over all units and sensors. This expression is then averaged out over all distinct subject pairs, resulting in the average inter-subject distance for activity a .

The results are shown in Fig. 6 for the three distance measures applied to the raw and pre-processed data. From the figure, we again observe that in almost all cases, pre-processing reduces the calculated distance values considerably. This is especially significant when feature vectors are used (Fig. 6d–f). When the time averages of the sequences are removed, the average inter-subject distances of the sequences corresponding to the static activities (A_1 – A_4) are smaller than the other activities even though their raw (not pre-processed) forms have

larger distance values than some of the other activities. This is because these activities are static, unlike the others; thus, the inter-subject differences in the sequences are mostly caused by noise and the sensors' bias and drift errors. Hence, when the time-average values of the sequences are removed, they resemble each other more and the distances decrease considerably. The raw forms of the feature vectors obtained from these first four activities also result in relatively small distance values (Fig. 6d–f).

The inter-subject distance of the standing activity (A_2) varies the least among the participants in our study because standing is a stationary activity, where the body posture is about the same

in all subjects, and the subjects' physical attributes do not affect the sensor readings significantly. This is especially visible when feature vectors are used in their raw form (Fig. 6d–f). On the contrary, vigorous activities (A_{12}) or activities that contain random elements, such as jumping (A_{18}) and playing basketball (A_{19}), result in consistently larger inter-subject distances. For example, while playing basketball, subjects perform random motions such as dribbling or shooting at different times, resulting in large distance values. This is the case even when the DTW distance measure is used, which allows warping in time. Although the DTW algorithm can match random activities of different subjects to some extent to maximize similarity, it penalizes unmatched subsequences and large amounts of warping, still resulting in a relatively large DTW distance. On the other hand, DTW can relatively easily match dynamic activities of a quasi-periodic nature, leading to smaller distances, as expected.

The average inter-subject distance per activity obviously depends on the type of pre-processing used. For example, when the absolute distance measure is used and the average values of the sequences are removed, activity A_4 (lying on right side) has the smallest average inter-subject distance. However, when the same distance measure is used but the sequences are normalized between -1 and 1 , activity A_8 (moving around in an elevator) has the smallest average inter-subject distance.

5.3. Average inter-activity distance per subject, unit and sensor type

In this section, we calculate inter-activity distances (i.e. distances between two distinct activities) using pre-processed time sequences. This is a measure of how different the two activities are and can be useful in providing guidance for their classification.

For a given subject, unit, sensor and distance measure \mathfrak{D} , we first form a 19×19 activity distance matrix $\mathbf{D}_{p,u,s,\mathfrak{D}}$ for the pairwise distances between the 19 activities. The rows and columns of this matrix are indexed by the different activities, and the (a_1, a_2) th element of the matrix is calculated using the zero-mean time sequences

$$d_{p,u,s,\mathfrak{D}}(a_1, a_2) = d_{\mathfrak{D}}(\bar{\mathbf{x}}_{p,a_1,u,s}, \bar{\mathbf{x}}_{p,a_2,u,s}) \quad (18)$$

where (a_1, a_2) is an activity pair, and p, u and s are the subject, unit and sensor indices, respectively. Since distance measures have the symmetry property, this is a symmetric matrix. In addition, the diagonal elements are zero because the distance of a sequence to itself is always zero. Hence, the essential part is the upper-triangular submatrix with $N_a(N_a - 1)/2 = (19 \times 18)/2 = 171$ elements, corresponding to the pairwise distances between the 19 activities. For eight subjects, five units and nine sensors, there are 360 ($= 8 \times 5 \times 9$) such distance matrices indexed by p, u and s , and considering the distances in the upper diagonal part of each matrix, a total of 61 560 ($= 171 \times 360$) pairwise distance values.

Averaging the values of the elements in the upper-triangular part of one of these activity distance matrices, corresponding to a given subject, sensor and unit (indexed by p, u and s), provides us the average inter-activity distance over all distinct activity pairs

$$\bar{d}_{p,u,s,\mathfrak{D}} = \frac{2}{N_a(N_a - 1)} \sum_{a_1=1}^{N_a-1} \sum_{a_2=a_1+1}^{N_a} d_{p,u,s,\mathfrak{D}}(a_1, a_2) \quad (19)$$

and the corresponding standard deviation is

$$\tilde{d}_{p,u,s,\mathfrak{D}} = \sqrt{\frac{2}{N_a(N_a - 1)} \sum_{a_1=1}^{N_a-1} \sum_{a_2=a_1+1}^{N_a} [d_{p,u,s,\mathfrak{D}}(a_1, a_2) - \bar{d}_{p,u,s,\mathfrak{D}}]^2} \quad (20)$$

In the following subsections, further averages of these quantities are taken to calculate the average inter-activity distance and its standard deviation for each subject over all units and sensors, for each unit over all subjects and sensors, and for each sensor over all subjects and units.

5.3.1. Average inter-activity distance and standard deviation per subject

To find the average inter-activity distance and its standard deviation for a given subject over all units and sensors, we calculate the averages of Equations (19) and (20) over all units and sensors

$$\bar{d}_{p,\mathfrak{D}} = \text{average}_{u,s}(\bar{d}_{p,u,s,\mathfrak{D}}) = \frac{1}{N_u N_s} \sum_u \sum_s \bar{d}_{p,u,s,\mathfrak{D}} \quad (21)$$

and

$$\tilde{d}_{p,\mathfrak{D}} = \text{average}_{u,s}(\tilde{d}_{p,u,s,\mathfrak{D}}) = \frac{1}{N_u N_s} \sum_u \sum_s \tilde{d}_{p,u,s,\mathfrak{D}} \quad (22)$$

The results are shown in Fig. 7. We observe that among the eight subjects, the sixth one has the smallest average inter-activity distance for all three distance measures. The average inter-activity distance is consistently the largest for Subject 7, indicating that he or she performs the activities in an extreme way, for example, with a larger amplitude or more distinctively, unlike Subject 6.

From the figure, we observe that the average standard deviations obtained using the DTW distance measure are, in general, relatively larger fractions of the average inter-activity distances compared with the other two distance measures.

5.3.2. Average inter-activity distance and standard deviation per unit

To find the average inter-activity distance and its standard deviation for a given unit over all subjects and sensors, we calculate the averages of Equations (19) and (20) over all subjects and

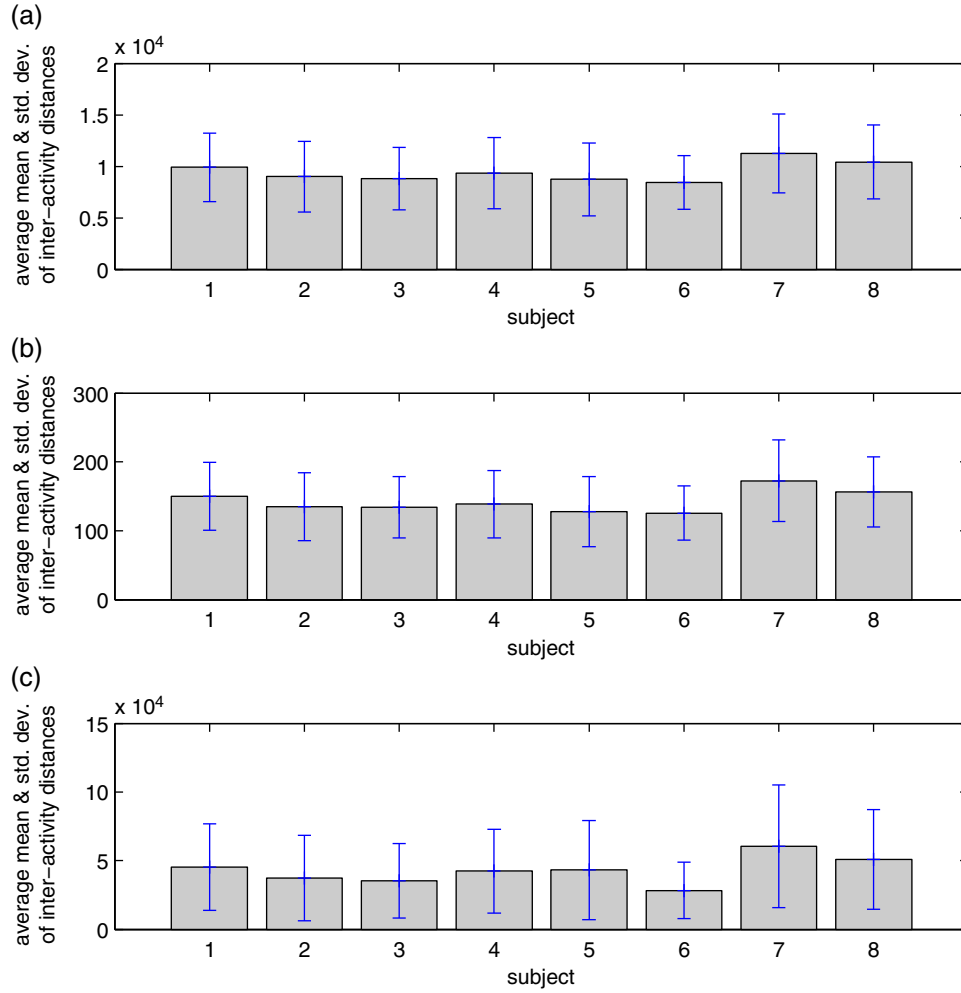


FIGURE 7. Average mean and standard deviation of inter-activity distances for each *subject* in terms of the (a) absolute, (b) Euclidean and (c) DTW distances calculated using zero-mean time-domain data. Gray bars represent the average distance values $\bar{d}_{p,\mathcal{D}}$, whereas the vertical lines indicate one standard deviation around the average, ranging between $\bar{d}_{p,\mathcal{D}} \pm \tilde{d}_{p,\mathcal{D}}$.

sensors

$$\bar{d}_{u,\mathcal{D}} = \text{average}_{p,s} (\bar{d}_{p,u,s,\mathcal{D}}) = \frac{1}{N_p N_s} \sum_p \sum_s \bar{d}_{p,u,s,\mathcal{D}} \quad (23)$$

and

$$\tilde{d}_{u,\mathcal{D}} = \text{average}_{p,s} (\tilde{d}_{p,u,s,\mathcal{D}}) = \frac{1}{N_p N_s} \sum_p \sum_s \tilde{d}_{p,u,s,\mathcal{D}} \quad (24)$$

The results are presented in Fig. 8. As expected, average inter-activity distances of the sensor units on the subjects' legs are the largest, followed by the arms, for all three distance measures. For most of the dynamic activities considered in this study, the limbs move a lot more than the torso; therefore, the accelerations recorded at the limbs are larger on the average. The torso unit has the smallest average inter-activity distance for all three distance measures. We do not observe any significant difference

between the average inter-activity distances of the right and left limbs, indicating that on the average they are used equally during activities.

Again, the average standard deviations obtained using the DTW distance measure are, in general, relatively larger fractions of the average inter-activity distances compared with the other two distance measures.

5.3.3. Average inter-activity distance and standard deviation per sensor type

To find the average inter-activity distance and its standard deviation for the axes of a given sensor type over all subjects and units, we calculate the averages of Equations (19) and (20) over all subjects and units

$$\bar{d}_{s,\mathcal{D}} = \text{average}_{p,u} (d_{p,u,s,\mathcal{D}}) = \frac{1}{N_p N_u} \sum_p \sum_u \bar{d}_{p,u,s,\mathcal{D}} \quad (25)$$

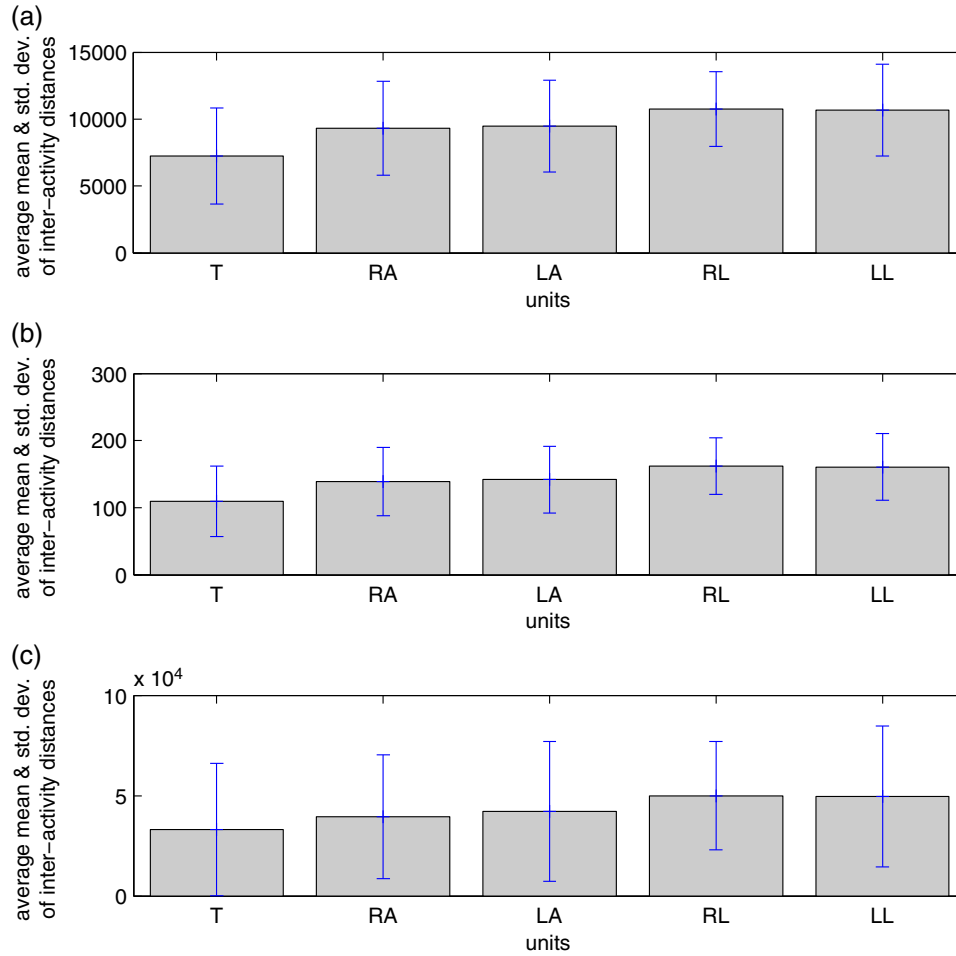


FIGURE 8. Average mean and standard deviation of inter-activity distances for each *unit* in terms of the (a) absolute, (b) Euclidean and (c) DTW distances calculated using zero-mean time-domain data. T, torso; RA, right arm; LA, left arm; RL, right leg; LL, left leg. Gray bars represent the average distance values $\bar{d}_{u,\mathcal{D}}$, whereas the vertical lines indicate one standard deviation around the average, ranging between $\bar{d}_{u,\mathcal{D}} \pm \sigma_{u,\mathcal{D}}$.

and

$$\tilde{d}_{s,\mathcal{D}} = \text{average}_{p,u}(\tilde{d}_{p,u,s,\mathcal{D}}) = \frac{1}{N_p N_u} \sum_p \sum_u \tilde{d}_{p,u,s,\mathcal{D}} \quad (26)$$

The results are illustrated in Fig. 9 for the three sensor types (accelerometer, gyroscope and magnetometer). The x -, y -, z -axes displayed in the figure for each sensor type point in the same directions as the axes of the torso unit and correspond to the three perpendicular axes of the human body. (The x -axis of the torso unit points up along the vertical axis, its y -axis points to the right along the transverse axis and its z -axis points to the front along the sagittal axis of the human body, as shown in Fig. 3b.) In other words, the measurements made by the x -, y -, z -axes of the sensor units have been projected onto the axes of the torso unit to be able to interpret the results in a unified way. Comparing the average inter-activity distances recorded by the sensor axes with respect to the three perpendicular axes

of the human body is more meaningful, since this removes the effect of the initially different orientations of the sensor units (Fig. 3b). Therefore, the average inter-activity distances displayed in the figure are caused by the differences in the nature of the activities only and help us identify those directions along which the variation between the activities is the largest.

The average inter-activity distances along the three perpendicular axes of the human body are quite different for each of the three sensor types. For the accelerometers, the activities show greater variation along the x -axis, followed by the z - and y -axes. This indicates that the inter-activity distances along the vertical axis are the largest, followed by the sagittal and transverse axes of the human body. For the gyroscopes, the corresponding order is y, x, z , indicating that the angular rate about the transverse axis shows the greatest variation among the different activities. Smallest variation being along the z -axis for gyroscopes indicates that during the activities, the body does not rotate much about the sagittal axis. For the magnetometers, average

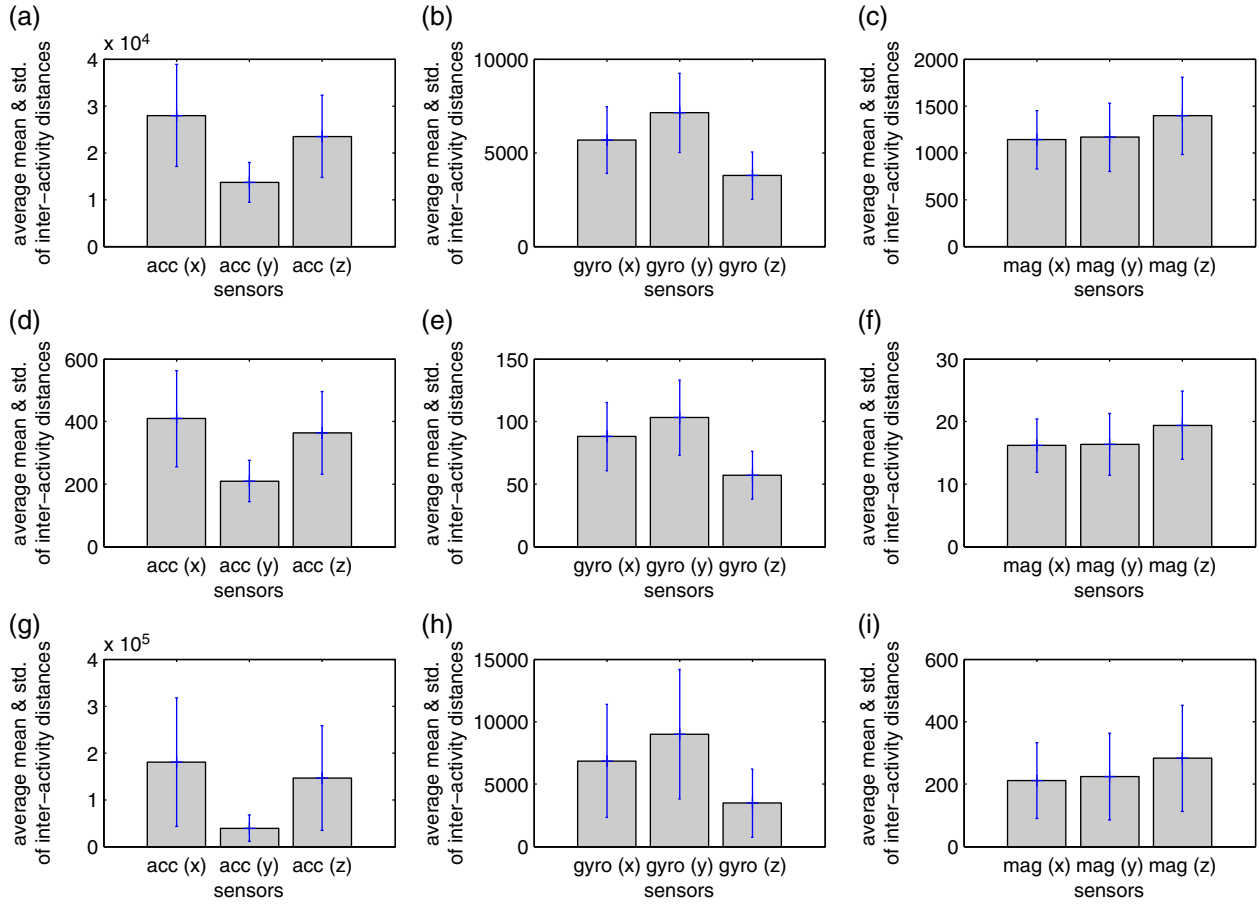


FIGURE 9. Average mean and standard deviation of inter-activity distances for each *sensor* in terms of the (a)–(c): absolute, (d)–(f): Euclidean and (g)–(i): DTW distances calculated using zero-mean time-domain data. Gray bars represent the average distance values $\bar{d}_{s,\mathcal{D}}$, whereas the vertical lines indicate one standard deviation around the average, ranging between $\bar{d}_{s,\mathcal{D}} \pm \tilde{d}_{s,\mathcal{D}}$.

inter-activity distances are sorted as z, y, x in decreasing order where x - and y -axis distances are about the same. The x -axis of the magnetometers have the smallest average inter-activity distance because the body posture is vertical (as in standing) in most of the activities, and the Earth’s magnetic field vector, pointing to the magnetic North, lies on the horizontal y – z plane which is perpendicular to the vertical (x) axis of the body. We observe that the average inter-activity distance values recorded by the three magnetometer axes are closer to each other than the accelerometers and gyroscopes. This may be because the Earth’s magnetic field vector has an almost constant norm, whereas the norms of the acceleration and angular rate change over time depending on the motion of the sensor units.

The average inter-activity distances of the different sensor types are not comparable because the sensors’ operating ranges, measured quantities and sensitivities are different. We observe that the average inter-activity distances of different sensor types can differ up to a factor of 1000.

Once again, we observe that the average standard deviations obtained using the DTW measure are, in general, relatively

larger fractions of the average inter-activity distances compared with the first two distance measures. Furthermore, one can observe that smaller average inter-activity distances are associated with smaller average standard deviations in general.

6. CLASSIFICATION RESULTS WITH THE DATASET

We have implemented a number of classifiers for activity recognition based on the dataset that we analyze here. Among these are the naïve Bayesian (NB) classifier, Bayesian decision-making (BDM), dissimilarity-based classifier (DBC), the least-squares method (LSM), the k -nearest neighbor algorithm (k -NN) (with $k = 7$), support vector machines (SVM), random forest decision tree (RF-T), artificial neural networks (ANNs), Gaussian mixture models (GMMs) with one-to-four components and a classifier based on the DTW distance measure. The last classifier can, in fact, be considered to be a one-nearest-neighbor (1-NN) classifier that uses the DTW distance instead

TABLE 1. Different sensor unit combinations and the corresponding correct classification rates for various classifiers using 10-fold cross validation [9, 10].

Units used	NB	BDM	DBC	LSM	<i>k</i> -NN	SVM	RF-T	ANN	GMM	NN-DTW
T	73.5	96.6	79.5	79.0	92.8	95.7	89.4	95.3	96.3	92.9
RA	72.8	94.5	77.3	72.5	88.4	95.1	89.8	95.5	94.8	87.5
LA	72.5	93.7	77.1	75.3	87.8	96.2	88.5	92.6	95.1	84.4
RL	87.0	97.3	87.1	82.3	91.0	96.8	93.2	97.6	97.4	87.3
LL	87.7	94.8	86.3	79.1	96.7	97.6	93.2	98.2	97.7	97.5
RA + LA	83.9	97.6	84.5	83.4	95.5	97.5	94.9	95.5	96.9	94.8
RL + LL	91.3	98.8	89.6	84.0	96.6	98.1	96.6	98.5	98.5	95.6
RA + LA + RL + LL	94.4	99.0	93.2	88.0	98.3	99.1	98.6	99.1	99.0	98.5
T + RA	86.2	98.1	86.8	84.0	95.8	97.8	95.2	97.5	96.7	95.7
T + LA	87.8	98.1	87.3	87.2	95.6	97.9	95.7	97.7	97.5	94.6
T + RL	88.1	98.4	98.5	85.4	97.2	98.3	96.4	98.4	98.1	97.6
T + LL	89.3	98.4	89.3	85.4	97.3	98.4	96.6	98.6	98.7	97.5
T + RA + LA	91.6	98.5	89.6	86.8	97.5	98.5	97.0	98.0	97.9	97.4
T + RL + LL	93.3	99.0	92.0	86.3	97.7	98.8	97.7	98.8	98.8	97.7
T + RA + LA + RL + LL	96.6	99.2	94.8	89.6	98.7	99.2	98.6	99.2	99.0	98.5

T, torso; RA, right arm; LA, left arm; RL, right leg; LL, left leg.

The lower part of the table displays the results when the torso unit is added to the sensor unit combinations in the upper part.

of the Euclidean distance. Since it classifies a test vector into the class of the training vector that is nearest in terms of the DTW distance, we abbreviate it as NN-DTW.

For effective sensor selection, we have considered an exhaustive set of sensor unit and sensor type combinations. In Table 1, we highlight some of our sensor unit combination results for the 10 classifiers mentioned above. Background information on the classifiers, details of the implementations and more extensive results can be found in our earlier works [8–10]. The classification results given in Table 1 correspond to the average values over all activity types and are consistent with the results in Fig. 8. Since the average inter-activity distances of the sensor units on the subjects' legs are the largest, followed by the arms, according to this figure, using only the units on the legs results in higher classification rates than those on the arms, in general (rows 2–5 of Table 1). When both arm or both leg units are used, the results improve compared with using a single arm or a single leg unit (rows 6 and 7). In this case, units on the legs are still more informative compared with those on the arms. Although the torso unit has the smallest average inter-activity distance for all three distance measures (Fig. 8), it provides quite good classification results when used on its own (first row), usually better than using a single arm unit but worse than using a single leg unit. This may be because recordings made at the chest (torso) filter out the large fluctuations in acceleration recorded at the limbs which may not always be informative or meaningful. When the torso unit is added to the sensor combinations given in the first column of the upper part of Table 1, improvement is achieved in all of the classification results. Best classification results are obtained when all five units are

used (last row). Among the different classifier types, BDM, SVM, ANN and GMM (with two mixture components) seem to outperform the others. As expected, the results of BDM are significantly better than NB results, especially when a small number of sensor units is used. As the recordings of more and more units are included, the difference between BDM and NB decreases from 23% down to ~3%.

In Fig. 10, we provide a comparison of combining different sensor types for some of the classifiers in terms of correct differentiation rates, using 10-fold cross validation [10]. When a single sensor type is used, magnetometers provide the highest classification rates, followed by accelerometers and gyroscopes. When two sensor types are used together, combinations that involve the magnetometer result better. When all three types of sensors are used, the results improve slightly compared with using only magnetometers, or magnetometers used in conjunction with another sensor type.

Using some of the better performing classifiers (BDM, *k*-NN and SVM), we calculated the correct classification rates using data from subsets of the subjects [8]. An exhaustive set of subject combinations are considered and those that result in the highest correct classification rates are reported in Table 2 when 10-fold cross validation is used. According to the table, performances of all three classifiers are comparable. Highest classification rates are observed for all three classifiers when data from two subjects are used. Using data from more than two subjects causes a slight decrease in classification performance, as expected.

In Section 5.1, Subjects 2, 3 and 5 were identified to be the most representative subjects depending on the distance

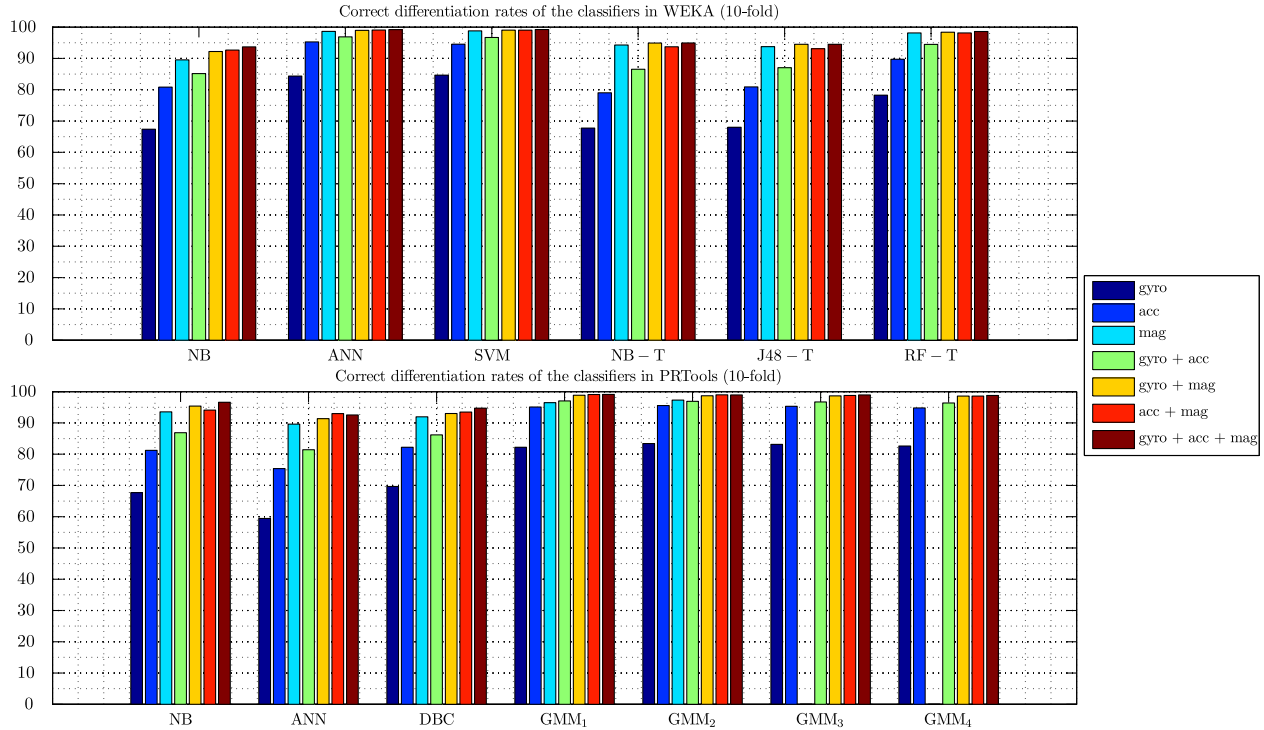


FIGURE 10. Comparison of combining different sensor types for some of the classifiers in terms of correct differentiation rates using 10-fold cross validation [10]. The sensor type combinations represented by the different colors in the bar chart are identified in the legend (gyro, gyroscope; acc, accelerometer; mag, magnetometer). WEKA and PRTools are two commonly used open source machine learning environments.

TABLE 2. Combinations of subjects resulting in the highest correct classification rates using 10-fold cross validation [8].

BDM		<i>k</i> -NN		SVM	
Subject no.	%	Subject no.	%	Subject no.	%
5	99.0	1	98.9	5	98.5
2,5	99.6	1,2	99.4	1,2	99.4
2,5,6	99.5	1,2,5	99.3	1,2,5	99.4
1,2,4,6	99.5	1,2,5,6	99.1	1,2,5,6	99.3
2,4,5,6,7	99.4	1,2,3,5,6	99.0	1,2,5,6,7	99.1
1,2,3,5,6,7	99.4	1,2,3,4,5,6	98.9	1,2,3,4,5,6	99.0
1,2,3,4,5,6,7	99.2	1,2,3,4,5,6,8	98.8	1,2,3,4,5,6,7	98.9

measure and whether raw time sequences or feature vectors are used. Consistent with this finding, we observe that Subjects 2 and 5 appear in almost all of the subject combinations that result in the highest classification rates. Subject 3 also appears in about one-third of the combinations, especially when data from five or more subjects are included. Subject 1 turns out to be a participant who appears in many of the successful combinations.

7. CONCLUSIONS AND FUTURE WORK

We investigated inter-subject and inter-activity variability based on a publicly available activity dataset that our research

team acquired [28] and provided some definitions that quantify such variability. Because of their many advantages, we have chosen to employ wearable sensors as opposed to external sensors that function as part of a smart environment. We considered pre-processing the acquired time sequences and feature vectors in different ways and used three different distance measures to compare them. We presented the inter-subject distances between the recorded activity sequences of distinct subjects in the dataset by averaging them out for each activity and subject. We identified the subject who performed an activity in the most representative way based on the minimum inter-subject distance. We also pinpointed the activities that show more variation among the subjects and presented the average inter-activity distances for each subject, unit and sensor.

We observe that the type of pre-processing used affects the results of the comparisons.

Although the average distance values that provide information about the similarity between the sequences are different for the three distance measures, the sorting of the distance values rarely changes, indicating that using different distance measures does not alter the comparison results as much as the type of pre-processing does.

We have verified some of our findings related to the dataset analyzed in this article by providing classification results based on the same dataset. In particular, we have considered an exhaustive set of sensor unit, sensor type and subject combinations and various classifiers.

Quantifying the inter-subject distances may be useful in a setting where a subject is training others to perform some activity (such as in dance or sports), teaching others to use a tool or an instrument or teaching rehabilitation exercises to a patient. Even though the trainer should perform the activity properly, trainees will often deviate from the proper motion during the learning phase. The average inter-subject distance can be used as a measure of the errors or deviations of the trainees during the learning process and can be provided as feedback to improve performance. Since subjects' physical parameters and personal styles vary, it might be better to acquire the reference data from the subjects themselves (under supervision) rather than from the trainer.

Average inter-activity distances can be used to identify activities that are easier or more difficult to distinguish. We expect our analysis and definitions on inter-activity distances to provide guidance in designing more effective classifiers for activity recognition. Identifying the sensor units and sensor types that are more informative for activity recognition is useful in dynamic sensor unit/type selection which will reduce the amount of data and processing time as well as optimize power consumption and minimize the number of units to be worn, while achieving a desired level of activity recognition accuracy [14]. Sensor unit/type selection may be performed online specific for the on-going activity. Reference [36] provides a framework for multi-dimensional time series classification by weighting each classifier's track record with a self-reported confidence score that is adjusted online. Our ongoing work in this area investigates sensor-activity relevance in human activity recognition with wearable motion sensors using the mutual information criterion [37].

The definitions of inter-subject and inter-activity variability provided in this work are sufficiently general and can be applied to a broad class of datasets that involve time sequences or features, acquired using wearable sensors. In future work, the methodology proposed here may be used to analyze other types of activity datasets such as motion data from patients with movement disorders or from dance/sports training. The effects of subjects' physical properties may be compensated for by developing techniques to reduce inter-subject variability of the sensor data as an attempt to build subject-independent classifiers.

ACKNOWLEDGEMENTS

The authors are grateful to Dr Kerem Altun, Mr Murat Cihan Yükses and Mr Orkun Tunçel for collecting the dataset and their contributions on the classifiers. The authors also thank the eight volunteers who participated in our study for their efforts, dedication and time.

FUNDING

This work was supported by The Scientific and Technological Research Council of Turkey (TÜBİTAK) under grant number

EEEAG-109E059, which participated in MOVE (COST Action IC0903).

REFERENCES

- [1] Logan, B., Healey, J., Philipose, M., Tapia, E.M. and Intille, S. (2007) A Long-Term Evaluation of Sensing Modalities for Activity Recognition. In Krumm, J. *et al.* (eds), *Proc. 9th Int. Conf. on Ubiquitous Computing*, Innsbruck, Austria, September 16–19, Lecture Notes in Computer Science 4717, pp. 483–500. Springer, Berlin, Heidelberg, Germany.
- [2] Turaga, P., Chellapa, R., Subrahmanian, V.S. and Udreă, O. (2008) Machine recognition of human activities: a survey. *IEEE Trans. Circuits Syst. Video Technol.*, **18**, 1473–1488.
- [3] Mileo, A., Merico, D., Pinarđi, S. and Bisiani, R. (2010) A logical approach to home healthcare with intelligent sensor-network support. *Comput. J.*, **53**, 1257–1276.
- [4] Ros, M., Cuéllar, M.P., Delgado, M. and Vila, A. (2013) Online recognition of human activities and adaptation to habit changes by means of learning automata and fuzzy temporal windows. *Inf. Sci.*, **220**, 86–101.
- [5] Poppe, R. (2010) A survey on vision-based human action recognition. *Image Vis. Comput.*, **28**, 976–990.
- [6] Moeslund, T.B., Hilton, A. and Krüger, V. (2006) A survey of advances in vision-based human motion capture and analysis. *Comput. Vis. Image Underst.*, **104**, 90–126.
- [7] Karaođlan, D. and Levi, A. (2014) A survey on the development of security mechanisms for body area networks. *Comput. J.*, **57**, 1484–1512.
- [8] Altun, K. and Barshan, B. (2010) Human Activity Recognition Using Inertial/Magnetic Sensor Units. In Salah, A.A., Gevers, T., Sebe, N., Vinciarelli N. (eds), *Proc. 1st Int. Workshop on Human Behavior Understanding (HBU 2010)*, Lecture Notes in Computer Science 6219, Istanbul, Turkey, pp. 38–51. Springer, Berlin, Heidelberg, Germany.
- [9] Altun, K., Barshan, B. and Tunçel, O. (2010) Comparative study on classifying human activities with miniature inertial and magnetic sensors. *Pattern Recognit.*, **43**, 3605–3620.
- [10] Barshan, B. and Yükses, M.C. (2014) Recognizing daily and sports activities in two open source machine learning environments using body-worn sensor units. *Comput. J.*, **57**, 1649–1667.
- [11] Lara, O.D. and Labrador, M.A. (2013) A survey on human activity recognition using wearable sensors. *IEEE Commun. Surv. Tutorials*, **15**, 1192–1209. Third Quarter.
- [12] Bulling, A., Blanke, U. and Schiele, B. (2014) A tutorial on human activity recognition using body-worn inertial sensors. *ACM Comput. Surv.*, **46**, article no: 33.
- [13] Bao, L. and Intille, S.S. (2004) Activity Recognition from User-Annotated Acceleration Data. In Ferscha, A. and Mattern, F. (eds), *Proc. 2nd Int. Conf. Pervasive Computing*, Lecture Notes in Computer Science 3001, Vienna, Austria, pp. 1–17. Springer, Berlin, Heidelberg.
- [14] Zappi, P., Lombriser, C., Stiefmeier, T., Farella, E., Roggen, D., Benini, L. and Tröster, G. (2008) Activity Recognition from On-Body Sensors: Accuracy-Power Trade-Off by Dynamic Sensor Selection. In Verdone, R. (ed.), *Proc. 5th European Conf.*

- Wireless Sensor Networks*, Lecture Notes in Computer Science 4913, Bologna, Italy, pp. 17–33. Springer, Berlin, Heidelberg, Germany.
- [15] Dalton, A. and ÓLaighin, G. (2013) Comparing supervised learning techniques on the task of physical activity recognition. *IEEE J. Biomed. Health Inform.*, **17**, 46–52.
- [16] Taylor, P.E., Almeida, G.J.M., Kanade, T. and Hodgins, J.K. (2010) Classifying Human Motion Quality for Knee Osteoarthritis Using Accelerometers. In *Proc. 32nd IEEE Annual Int. Conf. of the Engineering in Medicine and Biology Society*, Buenos Aires, Argentina, August 31–September 4, pp. 339–343. IEEE, Piscataway, NJ.
- [17] Tormene, P., Giorgino, T., Quaglini, S. and Stefanelli, M. (2009) Matching incomplete time series with dynamic time warping: an algorithm and an application to post-stroke rehabilitation. *Artif. Intell. Med.*, **45**, 11–34.
- [18] Yurtman, A. and Barshan, B. (2014) Automated evaluation of physical therapy exercises using multi-template dynamic time warping on wearable sensor signals. *Comput. Methods Programs Biomed.*, **117**, 189–207.
- [19] Altini, M., Penders, J. and Amft, O. (2012) Energy Expenditure Estimation Using Wearable Sensors: A New Methodology for Activity-Specific Models. *Proc. Conf. on Wireless Health*, San Diego, CA, October 23–25, pp. 1–8. article no: 1. ACM, New York, NY.
- [20] Dejnabadi, H., Jolles, B.M. and Aminian, K. (2005) A new approach to accurate measurement of uniaxial joint angles based on a combination of accelerometers and gyroscopes. *IEEE Trans. Biomed. Eng.*, **52**, 1478–1484.
- [21] Aggarwal, J.K. and Ryoo, M.S. (2011) Human activity analysis: a review. *ACM Comput. Surv.*, **43**, 16.1–16.43.
- [22] Sheikh, Y., Sheikh, M. and Shah, M. (2005) Exploring the Space of a Human Action. *Proc. 10th IEEE Int. Conf. on Computer Vision (ICCV 2005)*, Beijing, China, October 17–21, vol. 1, pp. 144–149. IEEE, Piscataway, NJ.
- [23] Veeraraghavan, A., Chellappa, R. and Roy-Chowdhury, A. (2006) The Function Space of an Activity. *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, New York, NY, June 17–22, vol. 1, pp. 959–968. IEEE, Piscataway, NJ.
- [24] Moy, M.L., Matthes, K., Stolzmann, K., Reilly, J. and Garshick, E. (2009) Free-living physical activity in COPD: assessment with accelerometer and activity checklist. *J. Rehabil. Res. Dev.*, **46**, 277–286.
- [25] Motl, R.W., Sosnoff, J.J., Dlugonski, D., Suh, Y. and Goldman, M. (2010) Does a waist-worn accelerometer capture intra- and inter-person variation in walking behavior among persons with multiple sclerosis. *Medical. Eng. Phys.*, **32**, 1224–1228.
- [26] Yurtman, A. and Barshan, B. (2012) Inter- and Intra-Subject Variations in Activity Recognition Using Inertial Sensors and Magnetometers. *The 5th Int. Conf. on Cognitive Systems*, Collection of Posters, Vienna, Austria, February 22–23, p. 8. Technical University of Vienna, Vienna, Austria.
- [27] Yurtman, A. and Barshan, B. (2012) Investigation of Personal Variations in Activity Recognition Using Miniature Inertial Sensors and Magnetometers (Minyatür Eylemsizlik Duyucuları ve Manyetometrelerle Aktivite Tanımada Kisiler Arası Farklılıkların İncelenmesi). *Proc. IEEE 20th Signal Processing and Communications Applications Conference*, (in Turkish), Fethiye, Muğla, Turkey, April 18–20. IEEE, Piscataway, NJ.
- [28] Barshan, B. and Altun, K. (2013) Daily and Sports Activities Dataset. University of California Irvine Machine Learning Repository, University of California, Irvine, School of Information and Computer Sciences. Available online: <http://archive.ics.uci.edu/ml/datasets/Daily+and+Sports+Activities> (accessed on 5 November 2015).
- [29] Xsens Technologies B.V. (2015) Enschede, The Netherlands, *MTi, MTx, and XM-B User Manual and Technical Documentation*, <http://www.xsens.com> (accessed on 5 November 2015).
- [30] Batista, G.E.A.P.A., Wang, X. and Keogh, E.J. (2011) A Complexity-Invariant Distance Measure for Time Series. *Proc. SIAM Int. Conf. on Data Mining*, Mesa, AZ, April 28–30, pp. 699–710. SIAM, Philadelphia, PA.
- [31] Müller, M. (2007) *Information Retrieval for Music and Motion*, Chapter 4, pp. 69–84. Springer, Berlin, Heidelberg, Germany.
- [32] Rakthanmanon, T., Campana, B., Mueen, A., Batista, G., Westover, B., Zhu, Q., Zakaria, J. and Keogh, E. (2012) Searching and Mining Trillions of Time Series Subsequences Under Dynamic Time Warping. *Proc. 18th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, Beijing, China, August 12–16, pp. 262–270. ACM, New York, NY.
- [33] The UCR Suite, funded by NSF IIS-1161997 II, University of California Riverside, <http://www.cs.ucr.edu/~eamonn/UCRsuite.html> (accessed on 5 November 2015).
- [34] Rakthanmanon, T., Campana, B., Mueen, A., Batista, G., Westover, B., Zhu, Q., Zakaria, J. and Keogh, E. (2013) Addressing big data time series: mining trillions of time series subsequences under dynamic time warping. *ACM Trans. Knowl. Discovery Data*, **7**, article no: 10.
- [35] Shokoohi-Yekta, M., Wang, J. and Keogh, E. (2015) On the Non-Trivial Generalization of Dynamic Time Warping to the Multi-Dimensional Case. *Proc. SIAM Int. Conf. on Data Mining*, Vancouver, British Columbia, Canada, April 30–May 2. SIAM, Philadelphia, PA.
- [36] Hu, B., Chen, Y., Zakaria, J., Ulanova, L. and Keogh, E. (2013) Classification of Multi-Dimensional Streaming Time Series by Weighting Each Classifier’s Track Record. *Proc. IEEE 13th Int. Conf. on Data Mining*, Dallas, TX, December 7–10, pp. 281–290. IEEE, Piscataway, NJ.
- [37] Dobrucalı, O. and Barshan, B. (2013) Sensor-Activity Relevance in Human Activity Recognition with Wearable Motion Sensors and Mutual Information Criterion. In Gelenbe, E. and Lent, R. (eds), *Information Sciences and Systems 2013, Proc. 28th Int. Symposium on Computer and Information Sciences*, Paris, France, October 28–29, pp. 285–294. Springer International Publishing, Cham, Switzerland.