Non-Intrusive Robust Human Activity Recognition for Diverse Age Groups

Di Wang[†] and Ah-Hwee Tan[‡]
[†]Joint NTU-UBC Research Centre of Excellence in Active Living for the Elderly and [‡]School of Computer Engineering
Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798
Email: {wangdi, asahtan}@ntu.edu.sg

Daqing Zhang
Network & Services Department
Institut Mines-Telecom/Telecom SudParis
91011 Evry Cedex, France
Email: daqing.zhang@it-sudparis.eu

***Abstract*—Many elderly prefer to live independently at their own homes. However, how to use modern technologies to ensure their safety presents vast challenges and opportunities. Being able to non-intrusively sense the activities performed by the elderly definitely has great advantages in various circumstances. Non-intrusive activity recognition can be performed using the embedded sensors in modern smartphones. However, not many activity recognition models are robust enough that allow the subjects to carry the smartphones in different pockets with unrestricted orientations and varying deviations. Moreover, to the best of our knowledge, no existing literature studied the difference between the youth and the elderly groups in terms of human activity recognition using smartphones. In this paper, we present our approach to perform robust activity recognition using only the accelerometer readings collected from the smartphone. First, we tested our model on two published data sets and found its performance is encouraging when compared against other models. Furthermore, we applied our model on two newly collected data sets: one consists of only young subjects (mean age = 22.5) and the other consists of only elderly subjects (mean age = 70.5). The experimental results show convincing prediction accuracy for both within and across diverse age groups. This paper fills the blank of elderly activity recognition using smartphones and shows promising results, which will serve as the groundwork of our future extensions to the current model.**

I. Introduction

Aging-in-place (**AIP**) or "the ability to live in one's own home and community safely, independently, and comfortably, regardless of age, income, or ability level" [1] has received a significant amount of attention nowadays due to the aging of the world's population. A survey [2] shows that nearly 90% of persons aged 65 or above indicate that they want to stay in their homes as long as possible, among which four of five believe their current homes are where they will always live. Therefore, how to use modern technologies to ensure their safety presents vast challenges and opportunities.

Being able to know what the elderly are doing when they live alone in their own homes has significant advantages in various AIP scenarios. For example, being able to reliably distinguish between falling down and lying down enables a system to generate alarms for help whenever necessary without triggering false ones. More generally, being able to recognize the activities of daily living (ADLs) of the elderly would tell us more information, such as physical well-being, cognitive health, emotional stressfulness, social engagement, etc. [3]. However, no one likes being watched, even for one's own safety concerns. Instead of installing surveillance cameras or motion tracking sensors in the home environment, using sensors embedded in smartphones for autonomous activity recognition has already shown promising results (see Section II). We believe that using the widely owned commercial devices for human activity recognition is the future trend. Therefore, specially crafted wearable devices would be unnecessary and the recognition is non-intrusive in the sense that the subjects would not feel being watched.

Among all the existing work using sensors embedded in smartphones for activity recognition, not many of them tolerate the flexibility of the placement of the smartphones on humans with varying orientation (gravity). Some use many (sometimes redundant) sensory inputs for reliable recognition. Most importantly, to the best of our knowledge, in this field, no existing literature studied the behavioural difference between the youth and the elderly groups.

In this paper, we introduce our approach to perform robust human activity recognition using only the accelerometer (which is embedded in almost all modern smartphones) readings, but represented in different domains through feature extraction. We first tested it on two published data sets for benchmarking purposes and found the experimental results are encouraging. We then applied it on two newly collected data sets distinguished by the age groups (in average, 22.5 VS. 70.5). We conducted extensive experiments both within and across diverse age groups, analysed the results, and found the prediction accuracy of our model is convincing (F-score $= 0.9967$ for cross validation and 0.9425 for leave-one-user-out). Having this complete groundwork ready, we could make many extensions to the current approach, including more types of activities to be recognised and high-level activity transition recognition (e.g., stood up and sat down).

The rest of this paper is organized as follows. Section II reviews the related work. Section III introduces our model for robust human activity recognition. Section IV presents the performance of our model benchmarked against the others when applied on two published data sets. Section V shows how we collected data from subjects from 18 to 80 years of age. Section VI presents the experimental results on our collected data sets with ample discussions. Section VII concludes this paper and proposes future extensions.

II. Related Work

Human activity recognition could bring tremendous benefits to many people, especially to children and elderly who generally require more assistance in their daily living. Nam and Park [4] developed a wearable device that consists of several sensors (including accelerometer and barometer) to recognize eleven types of activities of ten children from 16 to 29 months of age. Chernbumroong et al. [5] developed a set of wearable devices (one chest strap for heart rate and two wrist watches with integrated accelerometer, altimeter and other sensors) to recognize twelve types of activities of the elderly aged 73 in average. Both papers show promising results. However, the usage of these specially crafted devices may be hindered due to the popularity reason. We prefer to use widely owned commercial devices for non-intrusive human activity recognition.

The penetration rate of smartphones is high in many developing and developed countries. For example, it is estimated that 83% of Singaporeans aged 55 and above own at least one smartphone [6]. In terms of human activity recognition using smartphones, accelerometer seems to be always in use. Sun et al. [7] extracted 22 features from the accelerometer readings to recognize seven types of activities. Anguita et al. [8] extracted 561 features from both accelerometer and gyroscope readings to recognize six types of activities. Their data sets are used in this paper for benchmarking purposes in Section IV, where the performance of their models is also compared against ours.

To incorporate the gravity or orientation information of the smartphone, Yang [9] used the raw accelerometer readings to estimate the vertical and horizontal components separately. Baek et al. [10] applied a second-order Butterworth high-pass filter to eliminate the gravity component. However, to possibly shorten the processing time, we want the computation to be simple but reliable, not complicated. Therefore, our model does not filter or transform the signals. Alternatively, we use the magnitudes of the raw triaxial readings to ease the necessity of the orientation information, as also used in [4], [5], [7].

Actually, more types of services can be delivered by using a richer set of the embedded sensors in smartphones. Other than activity recognition, Martin et al. [11] included light and proximity sensors in the framework to identify whether the phone is in the pocket or bag. Pei et al. [12] used both Wi-Fi and GPS signals for indoor localization. However, in this paper, we only focus on using the accelerometer readings to recognize the fundamental human activities.

Among all the afore-reviewed literature, only [7], [10], [11] do not restrict the placement, the orientation, or the varying deviations of the smartphone during data collection. In this paper, we also show the robustness of our model, which achieves promising accuracy when applied to the data sets collected in the most natural environment with minimum restrictions.

Support vector machine (SVM) is a well-known model for less over-fitting on the training data set. Among the afore-reviewed literature, in [4], [5], [9], [12], SVM has been identified as the best performing model compared against other statistical, decision tree, and neural network models. We also select SVM as the classifier for our activity recognition model.

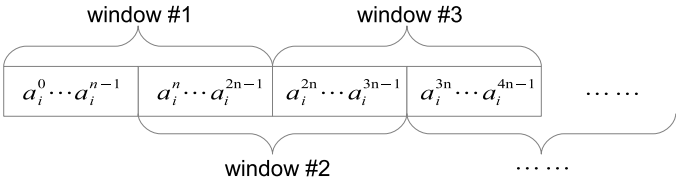

Fig. 1. Formation of half overlapping windows ($i \in \{x, y, z, ||x, y, z||\}$, see (2)). In this illustration, the frequency is n Hz and window size is 2 seconds.

III. Robust Activity Recognition Model

Modern smartphones are powerful (e.g., counting steps and estimating how many calories burned). Unfortunately, they do not provide activity recognition services yet. However, by using just the embedded accelerometer, we can extract features from the raw readings and apply machine learning algorithms to distinguish various human activities. In this paper, we refer our robust activity recognition model to the **RAR** model.

A. Feature Extraction

There are three readings ($A = \{a_x, a_y, a_z\}$) given by any triaxial accelerometer at any time (t) corresponding to the three orthogonal axes (x, y, z), respectively:

$$A^t = \{a_x^t, a_y^t, a_z^t\}. \tag{1}$$

Please note that the accelerometer readings used in this paper refer to the raw values queried directly from the sensor (i.e., when the phone is motionless, any reading may not equal to 0), rather than the linear acceleration values (i.e., when motionless, all readings equal to 0).

Because we only use accelerometer, which does not directly tell any orientation information. For robust activity recognition without knowing the orientation of the phone (e.g., from gyroscope or magnetometer), the prior work [7] empirically proved that the introduction of an extended dimension (acceleration magnitude) will increase performance. Therefore, at any time (t), we use four readings that are defined as

$$A^t = \{a_x^t, a_y^t, a_z^t, a_{mag}^t\}, \tag{2}$$

where mag denotes the acceleration magnitude ($||x, y, z||$) and $||x, y, z|| = \sqrt{(x^2 + y^2 + z^2)}$.

There are two parameters, window size and frequency, to be defined for data formation. After which, continuous data will be chopped into discrete windows (see Fig. 1). In this paper, all windows are half overlapping.

After data formation, assume each window consists of N readings. Then, in each dimension ($i \in \{x, y, z, ||x, y, z||\}$), we extract four features, namely mean (M), variance (V), energy (EE), and entropy (ET). These features are fundamental and have been commonly adopted by many existing models [4], [7], [8]. The first two features are defined as

$$M_i = \frac{1}{N} \sum_{n=0}^{N-1} a_i^n \tag{3}$$

and

$$V_i = \frac{1}{N-1}\sum_{n=0}^{N-1}(a_i^n - M_i)^2, \tag{4}$$

respectively. The latter two are computed in the frequency domain, where discrete Fourier transform (DFT) is applied:

$$F_i(k) = \sum_{n=0}^{N-1} a_i^n e^{-j2\pi nk/N}, k = 0, 1, 2, \ldots, N-1. \tag{5}$$

The energy ($L2$ norm) and entropy are then defined as

$$EE_i = \sqrt{\frac{\sum_{k=1}^{N-1}(F_i(k))^2}{N-1}} \tag{6}$$

and

$$ET_i = \sum_{l=1}^{N-1} -O_i(l)\ln(O_i(l)), \tag{7}$$

where

$$O_i(l) = \frac{|F_i(l)|}{\sum_{k=1}^{N-1}|F_i(k)|}, \tag{8}$$

respectively. Please note that since the DC component of (5) (mean) has already been used as an individual feature (see (3)), when computing the energy and entropy, the indices start from 1 instead of 0.

Up to now, all the extracted features are from the individual axes. To make the activity recognition more robust, covariances (COV) between the axes are also used:

$$COV_{p,q} = \frac{1}{N-1}\sum_{t=0}^{N-1}(a_p^t - M_p)(a_q^t - M_q), \tag{9}$$

where $p, q \in \{x, y, z, ||x, y, z||\}$ and $p \neq q$.

Therefore, after introducing six covariance measures, the total number of features extracted from the raw triaxial accelerometer is $4 * 4 + 6 = 22$.

Normalization is performed on all the extracted feature values before being processed by any machine learning algorithm. Let f denote the index of the feature ($f = 1, 2, \ldots, 22$), then the maximum and minimum values of all the observed data in each feature can be denoted as $\max_f$ and $\min_f$, respectively. Therefore, all values (both observed and unobserved data) will be normalized using the following equation:

$$v'_f = \frac{v_f - \min_f}{\max_f - \min_f}, \tag{10}$$

where v'_f denotes the normalized value in the fth feature and v_f denotes the original value in the fth feature.

B. Machine Learning Algorithm

Support vector machine (SVM) is a well-known model for less over-fitting on the training data set, because it minimizes the structural risk of the learnt model, especially when there are a limited number of training samples available. As introduced in Section II, SVM has been widely applied for human activity recognition [4], [5], [9], [12]. In this paper, we use the LibSVM package [13] for multi-class classifications.

For training samples (x_i, y_i), where $i = 1, 2, \ldots, l$, $x_i \in R^n$ (feature space of n dimensions), and $y \in \{-1, 1\}^l$ (binary classification, an h-class classification problem is solved by $h(h-1)/2$ binary classifications), SVM requires the solution of the following optimization problem:

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2}w^T w + C\sum_{i=1}^{l}\xi_i, \\ \text{subject to} \quad & y_i(w^T\phi(x_i) + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0, \end{aligned} \tag{11}$$

where C denotes the cost parameter ($C > 0$), which defines the amount of penalty on error. Function $\phi(x_i)$ maps the training vectors x_i into a higher dimensional space. In this paper, radial basis function (RBF) kernels are used:

$$\begin{aligned} K(x_i, x_j) &\equiv \phi(x_i)^T\phi(x_j) \\ &= \exp(-\gamma||x_i - x_j||^2), \end{aligned} \tag{12}$$

where γ denotes the gamma parameter ($\gamma > 0$), which controls the flexibility of the decision boundary.

Both C and γ control the level of generalization of the SVM model. Their optimal values should prevent both over- and less-fitting problems. The two parameters can be determined by a grid search using cross-validation. Alternatively, the default values are suggested as 1 and $1/l$, respectively [13]. The choices of C and γ in different experiments are introduced in the respective sections.

C. Performance Evaluation Metrics

After training, the learnt SVM model is applied to the testing data set for performance evaluation. To measure both Type-I and Type-II errors, precision and recall are defined as

$$\text{precision} = \frac{\sum \text{true positive}}{\sum \text{predicted positive}} = \frac{TP}{TP + FP} \tag{13}$$

and

$$\text{recall} = \frac{\sum \text{true positive}}{\sum \text{actual positive}} = \frac{TP}{TP + FN}, \tag{14}$$

respectively. The terms TP, FP, and FN denote true positive, false positive, and false negative, respectively.

The overall performance (F-score) is then computed as

$$F\text{-score} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}. \tag{15}$$

Please note that the measures among classes are averaged proportionally to the number of data samples in each class:

$$\text{Measure}^p = \frac{\sum_{h=1}^{H}\text{Measure}_h^p * N_h}{\sum_{h=1}^{H} N_h}, \tag{16}$$

TABLE I. PROFILE OF THE VPO DATA SET AND THE ACTIVITY RECOGNITION MODEL USED IN [7]

Category	Description
No. of subjects (gender info.)	7 (6 male and 1 female)
Age distribution (statistics)	23−46 (no statistics given)
Data collection device	Nokia N97
Smartphone placement	one of the six pockets near the pelvic
Sensor(s) used	accelerometer
Data sampling frequency	10 Hz
Window size (width)	varying (1 to 6 seconds)
List of activities	7 in total: stationary, walking, running, bicycling, ascending stairs, descending stairs, driving
List of features	22 in total, 5 different types
Classifier	SVM with RBF kernel
Experimental strategy	10-fold cross validation (grid search)

The 5 types of features: mean, variance, energy, entropy, correlation.

where p denotes the type of performance evaluation metrics (Measure^p $\in$ {precision, recall, F-score}), h denotes the hth class of activity, N_h denotes the total number of data samples in the hth class, and H denotes the total number of activities.

IV. PERFORMANCE BENCHMARKS

We applied our robust activity recognition (RAR) model to two published data sets. Here, we present the performance of RAR and compare it against the results reported in [7], [8] in the respective subsections.

A. Data Set Collected with Varying Positions and Orientations

In [7], the authors published their results of human activity recognition on their own collected data set. For data collection, the subjects may put the smartphone in any of the six pockets around the pelvic region (two front pockets on the coat plus two front and two rear pockets on the trousers). There is also no restriction on the facing direction and orientation of the phone placement. Moreover, because the size, shape, and orientation of the pockets vary, the orientations of the smartphone also vary during data collection (may be drastic when the activity involves more movements, such as walking and running) [7]. The varying positions and orientations of the smartphone are well catered by the recognition model [7]. We refer this data set to the varying positions and orientations (**VPO**) data set. The profile of the VPO data set and the activity recognition model used is introduced in Table I and more details can be found in [7].

We obtained the VPO data set from the authors of [7] through personal communication and applied our RAR model for performance evaluation. Following their strategy, we also form the data samples into different window sizes (varying from one to six seconds per window of half overlapping) and perform 10-fold cross validation. The performance comparisons are listed and visualized in Fig. 2.

Fig. 2 shows the F-score comparisons between our model and the one presented in [7] according to the different window sizes used to form the data samples. It is clearly shown that our model performs better than the counterpart (outperforms in the first five cases and is comparable in the sixth case). Furthermore, based on the performance trend, both models suggest a relatively larger window size works best on this data

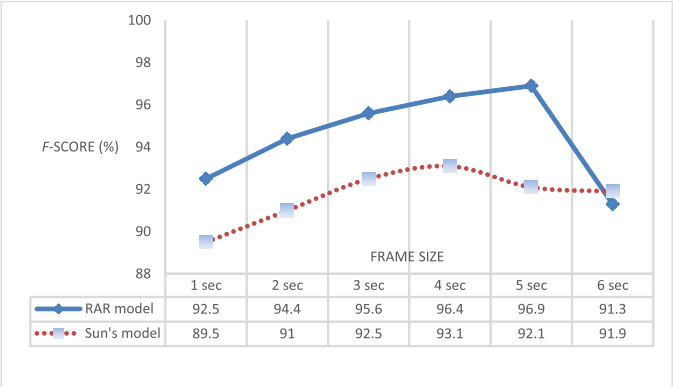

Fig. 2. Performance comparisons on the VPO data set [7]. Note that Sun's model refers to the model presented in [7] by Sun et al.

TABLE II. PERFORMANCE COMPARISONS ON THE VPO DATA SET

Model	RAR	Sun's [7]	Yang's [9]	Baek's [10]
F-score (%)	**96.9**	93.1	87.2	88.1

set (both achieved the highest performance when window size is four or five seconds). However, when the window size is larger than five seconds, both models' performance declines.

We apply the two models shown in Fig. 2 on the same data set (frequency and window size). Our model uses the same number of features and the same machine learning paradigm as used in [7]. However, we use covariance (see (9)) rather than the correlation (as used in [7]) between any two raw dimensions (see (2)). This would suggest that covariance between the orthogonal axes of the accelerometer (including the computed magnitude) provides more information towards more accurate human activity recognition than correlation does. In addition, another possible reason that causes the difference in performance might be due to the different grid search ranges used for the cost (C) and gamma (γ) parameters (see (11) and (12), respectively). In [7], the authors stated that they use grid search for optimal C and γ. However, they did not indicate the range of the search. We use the range $[2^{-5}, 2^{-4}, \ldots, 2^{5}]$ for both C and γ, which should be reasonably adequate and effective.

Table II lists the best F-score obtained (among the six different window sizes) by each model on the VPO data set. As reviewed in Section II, the last two models listed in Table II either transforms the accelerometer readings according to the estimated gravity [9] or eliminates the gravity component by applying a signal filter [10] to cater to the varying orientation of the smartphone. However, the fact that both RAR and Sun's model [7] outperform their models empirically proves that by including the computed magnitude (see (2)) and extracting informative features would achieve convincing performance without intentionally considering the gravity of the phone.

B. Data Set Collected with Fixed Placement

In [8], the authors published their results of human activity recognition on their own collected data set. Because the smartphone was always placed in the waist pouch during data collection, we refer this data set to the fixed placement (**FP**) data set (not introduced in [8], but we assume the position

TABLE III. PROFILE OF THE VPO DATA SET AND THE ACTIVITY RECOGNITION MODEL USED IN [8]

Category	Description
No. of subjects (gender info.)	30 (not given)
Age distribution (statistics)	19−48 (no statistics given)
Data collection device	Samsung Galaxy S2
Smartphone placement	always in the waist pouch
Sensor(s) used	accelerometer and gyroscope
Data sampling frequency	50 Hz
Window size (width)	fixed (2.56 seconds)
List of activities	6 in total: walking, walking upstairs, walking downstairs, standing, sitting, laying
List of features	561 in total, 17 different types
Classifier	SVM with Laplacian kernel
Experimental strategy	70% for training and 30% (unobserved) for testing, 10-fold cross validation is applied on the training data set to find optimal parameters (grid search)

The 17 types of features: mean, standard deviation, median absolute deviation, largest & smallest values, signal magnitude area, energy, interquartile range, entropy, regression correlation coefficients, index of the frequency component with the largest magnitude, weighted average of frequency components, skewness & kurtosis of the frequency domain signals, energy of a frequency interval, angle between vectors.

TABLE IV. CONFUSION MATRIX OF RAR MODEL ON THE FP DATA SET

Actual \ Predicted	wlk	ups	dns	std	sit	lay	rec
walking	**786**	59	0	0	0	0	93.0
walking upstairs	281	**520**	1	0	0	0	64.8
walking downstairs	4	70	**641**	0	0	0	89.7
standing	0	0	0	**495**	135	206	59.2
sitting	0	0	0	73	**825**	8	91.1
laying	0	0	0	29	5	**881**	96.3
precision (%)	73.4	80.1	99.8	82.9	85.5	80.5	**83.0**

Note: rec denotes recall (%) and the most bottom right number is the overall F-score.

and the orientation of the phone in the pouch still vary during data collection). The profile of the FP data set and the activity recognition model used is introduced in Table III and more details can be found in [8]. You may notice that there are a significantly large number of features (561) extracted from the raw readings of only two triaxial sensors, namely accelerometer and gyroscope.

We downloaded the data set used in [8] from the UCI machine learning repository [14]. To make the experimental results comparable, we follow their strategy (but with our feature extraction method) to form the data samples and apply our model afterwards (use the training and testing data sets indicated in [14] and run the same grid search on C and γ as introduced in Section IV-A on the training data set with 10-fold cross validation). Because they use both accelerometer and gyroscope data for activity recognition, from the downloaded raw readings, we only keep accelerometer ones (linear acceleration readings are also excluded). Furthermore, we use the fixed window size of 2.56 seconds (given the frequency is 50 Hz, so there are 128 readings in one window). After data formation, we follow (2) to (10) to extract 22 features. After obtaining the learnt model on the training data set, we apply it to predict the testing data set. The confusion matrix of our results is tabulated in Table IV (because the window size is fixed, only one confusion matrix is obtained and shown).

From Table IV, it is easy to notice that all confusions made are either among the moving activities (walking, up, and down) or the stationary ones (standing, sitting, and laying). This finding is as expected, but encouraging because without using features such as the signal magnitude area [15] (normally used to distinguish between stationary and moving activities, used in [8], see Table III), our model can differentiate those two groups of activities perfectly on the testing data set. To further justify the performance of our model, its F-score is compared against the other models' in Table V.

TABLE V. PERFORMANCE COMPARISON ON THE FP DATA SET

Model	RAR	MC-SVM [8]	MC-HF-SVM [8]
No. of features	**22**	561	561
F-score (%)	83.02	**89.3**	89.0

Note: MC denotes multi-class and HF denotes hardware friendly.

The MC-SVM and MC-HF-SVM listed in Table V refer to the standard multi-class SVM and the hardware friendly SVM (exploits fixed-point arithmetic to reduce the usage of the limited computational resource) [8], respectively. Here, we should state that the number of testing data samples listed in Table IV does not match the number listed in Table 1 of [8]. However, we used the exact data sets downloaded from [14] and followed the exact data formation process described in [8]. We presume there were some discrepancies between the data sets used in [8] and what the authors uploaded to [14]. Anyway, we still compare the results and present them in Table V.

It is shown in Table V that RAR performs worse than the others (around 6%). However, RAR requires much less amount of information or effort: 1) only raw acceleration instead of raw and linear acceleration with gyroscope readings used in [8], 2) no signal filtering required instead of a Butterworth low-pass filter being applied in [8], and 3) only 22 extracted features are used instead of 561 ($22/561 = 3.92\%$).

The fact that RAR uses significantly lesser amount of information to achieve tolerably worse prediction accuracy suggests the following findings: 1) including gyroscope in the selected features does not significantly improve the performance if some of the orientation information has already been incorporated (in our case, the magnitude of accelerometer readings and the covariance between features), 2) using more features does not necessarily boost up the performance, rather, those most distinctive ones should be selected, and 3) given limited resources (such as the phone's battery), a trade-off should be made between the number of sensory inputs to be collected and the satisfactory level of the performance.

We prefer to use less number of inputs and features to increase the usability of our model while maintaining a high level of accuracy. After benchmarking our model against the others, in the following sections, we introduce the data sets that we collected and how well our model performs on them.

V. DATA SETS OF DIVERSE AGE GROUPS

Human activity recognition has been studied using various data collection devices, various device placement strategies, various analysing techniques, and so on. However, to the best of our knowledge, no existing literature studied the difference between the youth and the elderly groups in terms of human

TABLE VI. VOLUNTEER PROFILES OF THE COLLECT DATA SETS

	Category	#1	#2	#3	#4	mean	std
Youth data set	Gender	F	F	F	M	-	-
	Age	18	20	22	30	**22.5**	5.26
Elderly data set	Gender	F	M	M	F	-	-
	Age	65	67	70	80	**70.5**	6.66

activity recognition using smartphones. In this section, we present how our diverse age groups data sets were collected.

We recruited eight volunteers (undergraduate students of School of Computer Engineering, Nanyang Technological University, Singapore, and their relatives), who did not receive any financial incentives, for the data collection. Their profiles are listed in Table VI. The age ranges from 18 to 80 and the mean age differs as 22.5 VS. 70.5. We refer the two data sets to the diverse age groups (**DAG**) data sets (youth and elderly).

Data collection was carried out in both indoor (flat cement floor) and outdoor (uneven soft ground) environments. Before data collection, all volunteers were orientated with the necessary knowledge: 1) how to use the data collection application installed on the smartphone, 2) the list of activities they would perform, and 3) where can they put the phone during data collection. Moreover, they could wear their own choices of trousers with at least one front pocket (during data collection, volunteers chose either the left or right pocket on their own, back pockets were not considered as they are impractical for sitting and lying). Therefore, the material, tightness, size, shape, and orientation of the pockets would vary. More importantly, during data collection, the volunteers were not restricted but allowed to fidget or move as they normally do (especially when sitting and standing). In this way, we might collect relatively noisier data. However, the data sets truly reflect the natural human behaviours.

After the volunteers got familiar with the data collection application and procedure, they decided the order of the activities that they would perform and where to perform each of them. The data collection procedure is depicted as follows:

1) The volunteer labels in the data collection application on the smartphone: (a) activity to perform, (b) pocket to place the phone, and (c) orientation of the phone (up or down and inward or outward). Please note that for (b) and (c), although we do not use the information during analysis, we still keep the record.
2) The volunteer then clicks the "start" button of the application, places the phone as indicated, and starts performing the activity as indicated.
3) The actual data collection will start after ten seconds (can be modified in app) with a beep sound. This preparation time is given for the volunteer to get ready and avoid collecting unwanted data.
4) The collection will end five minutes (can be modified in app) after the start beep and gives another beep sound to indicate the completion. We chose the five-minute interval because the data collection would be efficient and not drag the volunteers for too long.
5) The collected data are saved on the phone alongside with all the indication labels. They would be exported later for further pre-processing.

TABLE VII. ACTIVITIES COLLECTED IN THE DAG DATA SETS

Activity	Collected data (in mins)		
	Youth	Elderly	Total
Lying	79	20	99
Sitting	88	25	113
Standing	62	20	82
Walking	62	20	82
Running	41	15	56
Total	332	100	**432**

Note: Some young volunteers did not follow the default five-minute interval during data collection. One elderly volunteer was unfit to perform the running (even walk briskly) activity (therefore, not collected).

We use Samsung Galaxy S5 for data collection and the sampling frequency (0.1 second or 10 Hz) is defined using the Android Sensor Manage. However, the predefined frequency cannot be guaranteed because it might be interrupted by the operating system or other applications [16]. Therefore, based on the recorded timestamp, we sample the raw data at exactly 10 Hz before we further process them.

The list of activities and how many data samples we collected for each data set are shown in Table VII. We collected five types of activities in the DAG data sets, because we consider these are the most common and fundamental activities studied in the literature. Comparing our list to Table I, we did not collect ascending and descending the stairs, bicycling, and driving but we divided the stationary activity into lying, sitting, and standing. Similarly, comparing to Table III, we did not collect walking upstairs and downstairs but we collected running. In terms of walking upstairs and downstairs, or even using the lift to go up and down, we anticipate the barometer (not as common as accelerometer, only embedded in some models) would provide us more information (may also be useful to distinguish between a normal lied down and a fell down anomaly as part of our future extensions). The lying activity collected in the DAG data sets are either lying on bed or the Yoga mattress on the ground. However, both facing upwards (ceiling or sky). The volunteers chose their own preferred places for standing and chairs for sitting. We do not impose speed restrictions between walking and running, the volunteers just performed the activities as they normally do. However, we did not ask the elderly to run but to walk briskly (faster than normal walking). Therefore, naturally, due to the decline in the physical capability, the speed for both walking and running of the elderly is slower than that of the youth.

VI. EXPERIMENTAL RESULTS AND ANALYSIS

After collecting the DAG data sets[1], we applied our RAR model to first perform a cross validation on all the data samples in both data sets and then perform a series of leave-one-subject/group-out predictions. The results are presented and discussed in the following subsections.

A. Cross Validation on All Samples in the DAG Data Sets

In this section, we treat all the collected data (see Table VII) as a single data set to perform 10-fold cross validation. Similar to [7], we conduct experiments on varying window sizes (one

[1]We plan to publish the data sets either on our research centre's website or upload them to the UCI machine learning repository.

TABLE VIII. 10-FOLD CROSS VALIDATION ON THE DAG DATA SETS

Window size (sec)	1	2	3	4	5	6
F-score (%)	99.18	99.59	99.07	**99.67**	99.66	99.60

TABLE IX. CONFUSION MATRIX OF 10-FOLD CROSS VALIDATION (WINDOW SIZE = 4 SECONDS)

Actual \ Predicted	ly	sit	std	wlk	run	rec
lying	**2939**	5	0	0	3	99.73
sitting	4	**3533**	1	0	1	99.83
standing	0	1	**2429**	6	0	99.71
walking	0	9	3	**2566**	5	99.34
running	0	0	0	4	**1521**	99.74
precision (%)	99.86	99.58	99.84	99.61	99.41	**99.67**

Note: rec denotes recall (%) and the most bottom right number is the overall F-score.

to six seconds). The grid search range for both C and γ (see (11) and (12), respectively) is $[2^{-5}, 2^{-4}, \ldots, 2^{5}]$, as used before. The F-scores obtained are tabulated in Table VIII and the confusion matrix of the best performing configuration (window size = 4 seconds) is tabulated in Table IX.

Comparing Table VIII to Fig. 2, we can observe significant improvements of the obtained F-scores for all window sizes. This finding strongly supports our intention that the selected activities (see Table VII) are the most common and fundamental ones that we should be highly confident when recognizing them. This convincing result may also suggest a possibly high successful rate of our model's future extensions. For example, we are able to recognize motion transitions (e.g., stood up or sat down) only based on the prediction of the five fundamental activities from the continuous data stream. In that case, we only need to apply our model for the activity recognition and a reasoning algorithm for the transition detection, without the need to collect and train on the real transitional data.

Table IX shows highly convincing results in terms of recognition on all the activities (all accuracy measures are greater than 0.99). Unlike shown in Table IV that even for the prediction on unobserved data, there are no confusions between the moving and stationary activities. In Table IX, although the cross validation accuracy is high, there are still confusions between the moving and stationary activities. For example, the most number of errors made is incorrectly classified nine samples of walking into sitting. However, this proves that we did not collect overly clean data that during data collection, the volunteers were allowed to move (which makes real differences in raw readings, especially for sitting and standing).

From Table VIII, it is clearly shown that all F-scores are greater than 99%. However, to test the prediction capability of our model, we conduct extensive experiments and present the results with detailed discussions in the following section.

B. *Leave-One-Subject/Group-Out Predictions*

In this section, we define three types of leave-one-out (Loo) testing strategies as listed in Table X. In all these tests, we assign $C = 1$ and $\gamma = 1/22$ as the default values suggested in [13] (grid search for the highly diverse age groups suffers from over-fitting on the training data set, elaborated later in this section). The results are listed and visualized in Fig. 3.

TABLE X. DEFINITIONS OF LEAVE-ONE-OUT TESTING STRATEGIES

Name	**Training data set**	**Testing data set**
Loo-X	Entire DAG data sets except subject X	Subject X
Loo-Youth	Elderly	Youth
Loo-Elderly	Youth	Elderly

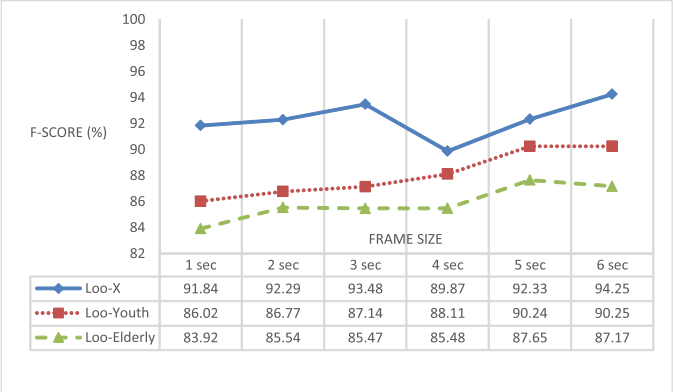

Fig. 3. Performance comparisons on the DAG data set.

TABLE XI. CONFUSION MATRIX OF LOO-ELDERLY (WINDOW SIZE = 5 SECONDS)

Actual \ Predicted	ly	sit	std	wlk	run	rec
lying	**372**	106	0	0	0	77.82
sitting	120	**477**	0	1	0	79.77
standing	0	0	**474**	4	0	99.16
walking	0	0	3	**474**	1	99.16
running	0	0	0	76	**282**	78.77
precision (%)	75.61	81.82	99.37	85.41	99.65	**87.65**

Note: rec denotes recall (%) and the most bottom right number is the overall F-score.

Results visualized in Fig. 3 are consistent that in terms of F-score for each window size, the following inequality always holds: Loo-X $>$ Loo-Youth $>$ Loo-Elderly. It is not surprise to find out Loo-X achieves the best prediction accuracy. However, although having more training data samples, the accuracy of Loo-Elderly is always lower than that of Loo-Youth. This suggests that behavioural differences do exist between the youth and the elderly groups (elaborated later in this section).

Although using different data sets and adopting different testing strategies (whether grid search is applied on the training data set and the percentage of unobserved data used for testing), we can still generally compare the F-score of Loo-X against those listed in Table V. The worst accuracy of Loo-X (window size = 4 seconds) is still better than all the results listed in Table V. This finding suggests that our human activity recognition model is robust that it achieves promising prediction accuracy on unobserved subjects with a huge age difference (18-80 compared to 19-48 introduced in [8]).

To further analyse the different behaviours between the youth and elderly groups, we tabulated the confusion matrix of Loo-Elderly (window size = 5 seconds) in Table XI. It is unexpected that the most number of confusions are between lying and sitting. It is probably because the elderly were too stationary during data collection for lying and sitting (not standing) and the placement and the orientation of the phone

TABLE XII. EVALUATIONS ON THE OVER-FITTING PROBLEM

Parameter values in use	Optimal F-score (%)	
	Loo-Youth	Loo-Elderly
Default: $C = 1$ and $\gamma = 1/22$	90.25	87.65
Grid search: both in $[2^{-5}, 2^{-4}, \ldots, 2^{5}]$	87.81	79.98
Performance difference (%)	-2.44	-7.67

Note: The optimal F-score is the highest among all six window sizes.

(in one of the front pockets of the trousers) for these two activities are highly similar. However, looking at Table IX, we are still highly confident in predictions as long as we already observed some behavioural patterns of the subject. The other highly confused recognition is that 76 running samples were classified as walking. This is within expectation that as mentioned in Section V, the elderly do not run (walk briskly) as fast or in the same way as the youth do.

As aforementioned, grid search on the control parameter values in the highly diverse age groups data sets may suffer heavily from over-fitting. To support the statement, we conducted the corresponding experiments and summarized the results in Table XII. The fact that the prediction accuracy on unobserved data gets worse if grid search is applied well demonstrates the over-fitting problem. In other words, using the suggested default parameter values suffers less from the over-fitting problem and makes more robust predictions.

VII. CONCLUSION

Non-intrusive human activity recognition could bring tremendous benefits in various aspects. For example, to ensure the safety of the elderly and to know how well they perform on their own. In this paper, we presented all the necessary details of our robust activity recognition model, which only uses accelerometer readings from smartphones to recognize the fundamental human activities. Moreover, we collected and introduced two data sets of diverse age groups. For benchmarking purposes, we first applied our model to two published data sets and compared its performance against the others. We found that our model, in the first case, performs better than the others (all use the same number of features), and in the second case, performs reasonably well when compared against the others (our model uses significantly lesser number of features). We then applied our model to the newly collected data sets and found the behavioural differences between the diverse age groups do exist. We conducted extensive experiments and presented the results with detailed discussions. The performance of our robust activity recognition model is convincing and the current model could lead us to many future extensions.

The possible promising future extensions to our model include but not limited to the following three directions: 1) incorporate more sensory inputs from the phone (e.g., barometer would be indicative to detect whether the subject is going up or down and suggestive to distinguish between a normal lied down and a fell down anomaly), 2) involve more wearable devices (e.g., with both smart watch and smartphone, a richer set of activities could be recognized, such as sitting while playing video games and standing while preparing food), and 3) autonomously recognize the fundamental activities and their transitions from a continuous data stream.

ACKNOWLEDGMENT

We thank Jia-Yi Wee and June Quak for their development of the data collection application, help during the data collection, and other work done in the project. This research is supported in part by the National Research Foundation, Prime Minister's Office, Singapore under its IDM Futures Funding Initiative and administered by the Interactive and Digital Media Programme Office.

REFERENCES

[1] Centers for Disease Control and Prevention. Healthy places terminology. [Online]. Available: http://www.cdc.gov/healthyplaces/terminology.htm

[2] American Association of Retired Persons. (2011, December) Aging in place: A state survey of livability policies and practices. [Online]. Available: http://www.aarp.org/home-garden/livable-communities/info-11-2011/Aging-In-Place.html

[3] D. Wang, B. Subagdja, Y. Kang, A.-H. Tan, and D. Zhang, "Towards intelligent caring agents for aging-in-place: Issues and challenges," in *Proceedings of IEEE Symposiumon on Computational Intelligence for Human-Like Intelligence*, 2014, pp. 1–8.

[4] Y. Nam and J. W. Park, "Child activity recognition based on cooperative fusion model of a triaxial accelerometer and a barometric pressure sensor," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 2, pp. 420–426, 2013.

[5] S. Chernbumroong, S. Cang, and H. Yu, "A practical multi-sensor activity recognition system for home-based care," *Decision Support Systems*, vol. 66, pp. 61–70, 2014.

[6] Blackbox Research Pte Ltd. (2012, May) Smartphones in Singapore: A whitepaper release. [Online]. Available: http://www.blackbox.com.sg/yka-smartphones-in-singapore/

[7] L. Sun, D. Zhang, B. Li, B. Guo, and S. Li, "Activity recognition on an accelerometer embedded mobile phone with varying positions and orientations," in *Ubiquitous Intelligence and Computing*, ser. Lecture Notes in Computer Science, Z. Yu, R. Liscano, G. Chen, D. Zhang, and X. Zhou, Eds. Springer, 2010, vol. 6406, pp. 548–562.

[8] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine," in *Ambient Assisted Living and Home Care*, ser. Lecture Notes in Computer Science, J. Bravo, R. Hervas, and M. Rodriguez, Eds. Springer, 2012, vol. 7657, pp. 216–223.

[9] J. Yang, "Toward physical activity diary: Motion recognition using simple acceleration features with mobile phones," in *Proceedings of International Workshop on Interactive Multimedia for Consumer Electronics*, 2009, pp. 1–10.

[10] J. Baek, S. T. Kim, H. D. Kim, J.-S. Cho, and B.-J. Yun, "Recognition of user activity for user interface on a mobile device," in *Proceedings of South East Asia Regional Computer Conference*, 2007, pp. 10.1–10.6.

[11] H. Martin, A. M. Bernardos, J. Iglesias, and J. R. Casar, "Activity logging using lightweight classification techniques in mobile devices," *Personal and Ubiquitous Computing*, vol. 17, no. 4, pp. 675–695, 2013.

[12] L. Pei, R. Guinness, R. Chen, J. Liu, H. Kuusniemi, Y. Chen, L. Chen, and J. Kaistinen, "Human behavior cognition using smartphone sensors," *Sensors*, vol. 13, no. 2, pp. 1402–1424, 2013.

[13] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 27, pp. 1–27, 2011.

[14] M. Lichman, "UCI machine learning repository," 2013. [Online]. Available: http://archive.ics.uci.edu/ml

[15] D. M. Karantonis, M. R. Narayanan, M. Mathie, N. H. Lovell, and B. G. Celler, "Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring," *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 1, pp. 156–167, 2006.

[16] "Android developers: Sensors overview." [Online]. Available: http://developer.android.com/guide/topics/sensors/sensors_overview.html