Towards Detection of Bad Habits by Fusing Smartphone and Smartwatch Sensors

Muhammad Shoaib, Stephan Bosch,
Hans Scholten, and Paul J.M. Havinga
Department of Computer Science
University of Twente, The Netherlands
Email: {m.shoaib,s.bosch,hans.scholten,p.j.m.havinga}@utwente.nl

Ozlem Durmaz Incel
Department of Computer Engineering,
Galatasaray University
Istanbul, Turkey
Email: odincel@gsu.edu.tr

Abstract—Recently, there has been a growing interest in the research community about using wrist-worn devices, such as smartwatches for human activity recognition, since these devices are equipped with various sensors such as an accelerometer and a gyroscope. Similarly, smartphones are already being used for activity recognition. In this paper, we study the fusion of a wrist-worn device (smartwatch) and a smartphone for human activity recognition. We evaluate these two devices for their strengths and weaknesses in recognizing various daily physical activities. We use three classifiers to recognize 13 different activities, such as smoking, eating, typing, writing, drinking coffee, giving a talk, walking, jogging, biking, walking upstairs, walking downstairs, sitting, and standing. Some complex activities, such as smoking, eating, drinking coffee, giving a talk, writing, and typing cannot be recognized with a smartphone in the pocket position alone. We show that the combination of a smartwatch and a smartphone recognizes such activities with a reasonable accuracy. The recognition of such complex activities can enable well-being applications for detecting bad habits, such as smoking, missing a meal, and drinking too much coffee. We also show how to fuse information from these devices in an energy-efficient way by using low sampling rates. We make our dataset publicly available in order to make our work reproducible.

I. INTRODUCTION

Smartphones are being extensively used for activity recognition in recent studies, because they are carried by almost everyone and are equipped with various onboard sensors, such as an accelerometer and a gyroscope [1], [2]. Recently, wrist-worn devices, such as smartwatches, are coming into the market with such on-board sensors [3], [4], which can be used for human activity recognition as well [5]. The combination of these two devices provides us with richer information that can be used to detect various human activities.

There are some activities that cannot be detected reliably using a smartphone in the jeans' pocket position alone, because they mainly involve hand movements. Some of the examples of such activities are smoking, eating, writing, typing, drinking a cup of coffee, and giving a talk or presentation. These activities can be recognized by using a wrist-worn device, because of the various hand movements involved in these activities. Such hand movements can provide us with information that can be used either alone or in combination with the sensor information from the pocket position to detect different contexts, for example, having a cup of coffee while sitting. Various well-being applications can utilize this extra context information for detecting bad habits and giving a better context-aware

feedback to the users. For example, drinking too much coffee can be considered as a bad habit, which can be recognized with such smart devices. Smoking is a well-known bad habit and smoking detection information can be used by an individual or a health professional to regulate smoking habit by getting more insight in daily smoking usage. Another example of bad habit is missing meals or not taking them on time. With eating detection, a relevant feedback and meaningful insights can be given to the users about their eating habits. Moreover, activities like typing or writing can also be used to identify the right moments for feedback. For example, a user should not be interrupted while typing or writing but can be interrupted while having a cup of tea or smoking for a feedback message. The recognition of such complex activities can provide us with more insights into the daily bad habits of users, which in turn can be used in well-being applications. The possibility of recognizing these activities using various sensors is shown by a few studies [5], [6], [7], [8], [9], [10], which we discuss further in Section II. However, these studies do not consider the fusion of smartphone and smartwatch sensors. Moreover, some of these studies just consider the recognition of one of these activities. For example, in [8], the authors focus on eating activity only. Similarly, in [10], only smoking activity is considered.

In this paper, we explore the role of these two devices in recognizing 13 different human activities using three classifiers, such as SVM, KNN and decision tree (J48). We study the strengths and weaknesses of these two devices in terms of recognition performance. We are interested in their combination because they provide us with richer context information due to their different positions on the human body. We explore an intelligent fusion of these two devices in relationship with different sampling rates, that can lead to energy efficiency. We also study the impact of increasing window sizes while sampling the sensors for the recognition performance of these activities. Moreover, we study the effects that the possible synchronization delays can have on the recognition performance. We summarize the main contributions of this paper as follows:

- We evaluate the performance of the smartphone (pocket position) and smartwatch (wrist position) sensors for most common physical activities when used alone and in combination with each other.

- We evaluate the performance of these two devices in recognizing additional six complex activities: smok-

ing, eating, writing, typing on a laptop, drinking coffee, and giving a talk. This is done in five different sensors combinations from both devices.

- We explore the impact of smartphone and smartwatch sensor fusion on recognition performance with respect to three different sampling rates and three different window sizes.

- We make our dataset publicly available in order to make our work reproducible.

The rest of the paper is organized as follows. We describe the related work in Section II. In Section III, we discuss the data collection process and the simulation setup for the data analysis. In Section IV, we discuss various aspects of performance analysis of these two devices in different scenarios and the effects of synchronization delays on the recognition performance. Finally, in Section V, we present the conclusions and future work.

II. RELATED WORK

Activity recognition in general [11] and especially using smartphone sensors is well studied in recent years [1], [2] and it is still being studied extensively. Various such studies are summarized in many surveys [1], [12], [13]. There are also a few studies on activity recognition using wrist-worn devices. For example, in [14], the authors studied the role of smartwatch and smartphone sensors in activity recognition. They recognized nine physical activities using five classifiers. These activities were sitting, standing, walking, running, cycling, stair descent, stair ascent, elevator descent and elevator ascent. However, the authors studied these two devices separately and did not fuse sensor data from both these devices. They used accelerometer, magnetometer, gyroscope, and pressure sensors on the smartphone, and only an accelerometer on the smartwatch.

In [5], the authors used a wrist-worn sensor and a sensor on the hip to detect seven physical activities. They used logistic regression as a classifier. They showed the potential of using the wrist position for the activity recognition. However, they evaluated these two positions separately and did not fuse these two sensors. In [6], the authors use a single wrist-worn accelerometer to detect five physical activities. Similarly, a wrist-worn accelerometer was used in [7] to recognize 8 activities, including an activity of working on a computer.

In [8], the authors detect eating activity using a Hidden Markov Model (HMM) with a wrist-worn sensor. They use binary classification with eating as one class and all other activities as the other class. Similarly, in [9], the authors recognize the eating activity using a wrist-worn accelerometer and a gyroscope. A feasibility study on smoking detection using a wrist-worn accelerometer is done in [10], where the authors reported a user-specific accuracy of 70% for this activity.

Unlike these studies, we consider the fusion of both smartphone and smartwatch sensors at pocket and wrist positions. Moreover, we consider more complex activities in this work, such as eating, drinking a cup of coffee, giving a talk and smoking. Whereas, some of these studies only focus on one of these activities.

III. DATA COLLECTION AND SIMULATION SETUP

We used two datasets in this work: one for simple activities and one for complex activities. The first one is from our previous work [15] and its collection protocol can be found in [15]. In that dataset, ten users performed seven physical activities while carrying smartphones (Samsung Galaxy S2) in their right jeans pocket and on their right wrist position, thereby emulating a smartwatch. These activities were walking, jogging, biking, sitting, standing, walking upstairs, and walking downstairs. In sitting and standing activities, the user sat and stood still alone without talking and doing any other activity. The smartphones were used in same orientation on both positions. Because the new wrist-worn device are equipped with sensors like an accelerometer and a gyroscope, we simulate a smartwatch using Samsung Galaxy S2 on the wrist position. We collected data for multiple smartphone sensors, such as an accelerometer, linear acceleration, gyroscope, and magnetometer, but here we only consider an accelerometer and a gyroscope. The data was collected at 50 samples per second for these sensors. For this dataset, each activity was done for 3 minutes by all participants (30 minutes of data for each activity), thereby creating a balanced class distribution for training and testing.

In the second dataset, we asked five participants, who took part in our previous dataset, to perform 6 additional complex activities. The term *"complex"* is used for clarity in comparison purpose for these additional six activities. The duration for these activities varies because different participants took different times to complete them, such as smoking and drinking coffee. Moreover, we wanted to capture one complete cycle of these complex activities such as drinking one cup of coffee, and smoking one cigarette. These activities are mentioned below with the amount of time (for all participants) for which the data was recorded.

- Typing (21 minutes): all five participants typed on their laptops the introduction section of this paper.

- Writing (21 minutes): they wrote the same introduction section on a paper of the same size.

- Drinking coffee (24 minutes): they had a cup of coffee while sitting in office lounge.

- Giving a talk (16 minutes): they gave a talk in our meeting room about their research topic for 3-4 minutes.

- Smoking (25 minutes): Each participant smoked one cigarette while standing alone in the smoking area.

- Eating (23 minutes): For the eating activity, users were asked to eat soup or yogurt for 3-4 minutes in their natural style while sitting alone in lunch place. The soup cup was on a table while the participants used a spoon for eating.

For the other seven activities, the time duration was 15 minutes (for all participants), which we collected from the previous dataset from the same 5 participants. One of the participants was left-handed, so we put the smartwatch on his left hand, unlike others. We know that most people use watches on their left hands, however, in this study we explore the possibility of using a wrist-worn device in general so we used the right wrist

TABLE I.   Short Notations for Various Fusion Scenarios: (P stands for pocket position, W for wrist position, A for accelerometer, and G for gyroscope)

Fusion Scenarios	P_A	P_G	W_A	W_G
P_{AG}	✓	✓	✗	✗
W_{AG}	✗	✗	✓	✓
PW_A	✓	✗	✓	✗
PW_G	✗	✓	✗	✓
PW_{AG}	✓	✓	✓	✓

TABLE II.   Recognition performance for walking (A1), walking upstairs(A6) and downstairs (A7) using accelerometer with 50Hz sampling rate.

	Activity	P_A	W_A	PW_A	PW_A-max(P_A,W_A)	P_A-W_A
SVM	A1	0.97	0.92	0.98	**+0.01**	+0.05
	A6	0.80	0.80	0.95	**+0.15**	0.00
	A7	0.83	0.84	0.95	**+0.11**	0.00
IB1	A1	0.97	0.95	0.99	**+0.02**	+0.02
	A6	0.85	0.81	0.97	**+0.12**	+0.04
	A7	0.83	0.87	0.96	**+0.10**	-0.04
J48	A1	0.94	0.88	0.96	**+0.02**	+0.06
	A6	0.81	0.78	0.88	**+0.06**	+0.04
	A7	0.76	0.85	0.88	**+0.04**	-0.08

position. For the rest, it was always right jeans' pocket and right wrist position. The data was collected at 50 samples per second using the Android application, which we developed in our previous work [16], [15]. The previous dataset is available at [17] and this new dataset will be available at the same website.

In the preprocessing phase, we extracted two time domain features for both the accelerometer and gyroscope. They are mean and standard deviation. They are extracted over a window of 2 seconds. Moreover, we use a sliding window with the overlap of 50%. These decisions are based on our past experiences [15] and work done by other researchers [1]. To counter orientation effects, we calculate the magnitude for both these sensors and use it an extra dimension beside x, y, and z axis. This method has been used in other studies [18], [19], [20]. Then the two features are extracted for all four dimensions for each sensor, which makes it 8 features per each sensor.

For performance analysis, we used machine learning tool WEKA (version 3.7) [21]. We selected three classification methods which are commonly used for practical activity recognition and have been shown to be suitable for running on smartphones with reasonable recognition performance. These classifiers are decision tree (WEKA version: J48) [22], [23], [24], k-nearest neighbor (WEKA version: IB1) [18], [19], [25], and support vector machine (WEKA version: LIBSVM) [23], [13], [26]. We use these classification methods in their default modes in WEKA (version 3.7) in order to make this work easily reproducible. Moreover, we did not optimize the parameters for these classification methods, because we are more interested in the relative roles of the smartphone and smartwatch sensors. For the recognition performance, we use accuracy as a performance metric using 10-fold stratified cross validation.

IV.   Performance Analysis

We evaluated the recognition performance with smartphone and smartwatch sensors in various scenarios where these sensors are fused in various combinations. These scenarios are given in Table I with their short notations. These short notations are used in the rest of the paper.

As we mentioned, for performance analysis, we used two datasets. One of them was used for recognizing commonly performed simple activities and the other for more complex activities, such as drinking a cup of coffee and giving a talk. They are discussed in the next sections.

A.   Recognition of Simple Activities

In this section, we discuss the role of smartphone and smartwatch sensors in recognizing the commonly used phys-

ical activities, such as walking (A1), standing (A2), jogging (A3), standing (A4), biking (A5), walking upstairs (A6), and walking downstairs (A7). It has been shown in previous studies [27] that battery consumption can be improved by using low sampling rates. Therefore, we use three sampling rates namely: 50, 25 and 10 Hz, to evaluate wether reasonable recognition performance can be achieved by using lower sampling rates. However, we do not report the results for 25 Hz, because they were similar to those of 50 Hz. In the next subsections, we only discuss the results for 50 Hz and 10 Hz.

1) Using Only Accelerometer: We evaluate the performance with the accelerometer at pocket, wrist, and finally at the combination of the two positions. We show the results (rounded off to 2 decimals) for the walking, walking upstairs and downstairs activities in Table II. We observe a significant increase in the overall recognition performance for these three activities when the accelerometer data is combined from pocket and wrist positions, as shown in Table II. In this table, PW_A-max(P_A,W_A) shows the % improvement by PW_A compared to P_A and W_A. For the rest of the activities, there are negligible improvements when we combine sensors from both positions because their individual performances are very high, i.e. above 98 %. Based on our previous work [15], we have seen that such combinations only improve the overall performance if the individual performances of these sensors are not very high, so that there is room for performance improvement. Moreover, the performance difference between wrist and pocket positions are negligible for these four activities.

2) Using Only Gyroscope: We evaluated the gyroscope using three classifiers at pocket and wrist position, both alone and in combination with each other from these two positions. These results are shown in Table III. We observe the same general trends for the gyroscope as for the accelerometer in terms of performance improvements by fusing sensors from both devices. The only difference is that the improvements due to fusion are observed for almost all seven activities because the individual gyroscope performance is not very high for them. Moreover, unlike the accelerometer, the gyroscope at pocket position performs slightly better than that at the wrist position for all seven activities. This performance difference is shown in in Table III.

3) Using both Accelerometer and Gyroscope: In this scenario, we combined both accelerometer and gyroscope at pocket and wrist positions. Furthermore, they are first fused on these individual positions and then combined from both positions. The fusion is done in WEKA at raw level by combining the features of all these sensors. We observe the same trends in this situation as that of the accelerometer.

TABLE III.   Recognition performance for all seven simple activities using gyroscope with 50Hz sampling rate.

	Activity	P_G	W_G	PW_G	PW_G-max(P_G,W_G)	P_G-W_G
SVM	A1	0.93	0.76	0.94	**+0.01**	+0.17
	A2	0.08	0.01	0.09	**+0.01**	+0.07
	A3	0.94	0.94	0.98	**+0.04**	+0.00
	A4	0.98	0.94	0.96	**-0.02**	+0.05
	A5	0.95	0.94	0.96	**+0.01**	+0.01
	A6	0.74	0.75	0.89	**+0.14**	-0.01
	A7	0.73	0.82	0.93	**+0.11**	-0.08
IB1	A1	0.96	0.88	0.97	**+0.02**	+0.08
	A2	0.85	0.74	0.87	**+0.02**	+0.11
	A3	0.96	0.94	0.99	**+0.03**	+0.02
	A4	0.89	0.78	0.91	**+0.02**	+0.12
	A5	0.96	0.91	0.98	**+0.02**	+0.05
	A6	0.93	0.79	0.97	**+0.04**	+0.14
	A7	0.90	0.85	0.97	**+0.07**	+0.06
J48	A1	0.89	0.81	0.92	**+0.03**	+0.08
	A2	0.86	0.79	0.88	**+0.02**	+0.07
	A3	0.94	0.92	0.96	**+0.02**	+0.02
	A4	0.83	0.66	0.84	**+0.02**	+0.17
	A5	0.94	0.91	0.96	**+0.03**	+0.03
	A6	0.88	0.78	0.90	**+0.03**	+0.10
	A7	0.83	0.84	0.91	**+0.06**	-0.01

However, the absolute accuracy values are higher than that of the accelerometer in this case because of two extra sensors. This also makes the relative improvements by fusing the data from two positions lower compared to that of the accelerometer. For example, the average improvement due to fusion of wrist and pocket positions is 1%, which is negligible. All seven activities are recognized with very high accuracy, such as on average above 97% on wrist and pocket position, and above 99% when both data from both positions is fused. Therefore, we do not show the detailed results in this case. Based on these evaluations, we show that simple activities can easily be recognized with one pocket position but the fusion of smartwatch with the smartphone makes this process more reliable. Moreover, this fusion can enable the recognition of more complex activities in a reliable way, which will not be possible with the smartphone in the pocket position alone. We discuss this in Section IV-B.

4) Using Lower Sampling Rates: As we mentioned earlier, we repeated the above three scenarios for 10Hz to evaluate whether reasonable accuracy can be achieved with lower sampling rates. Our simulation results show that in most cases the recognition performance of 10Hz is almost the same as that of 50Hz. In terms of positions, 50Hz performs slightly better than that of 10Hz at the wrist and its combination with pocket position by an average of 1% for all three classifiers. However, this performance difference is 4% at pocket position, where we observe relatively higher performance drops for 10Hz, such as within a range of 0-19%. In terms of sensors combinations, this performance difference is 1% for accelerometer, 3% for gyroscope and 2% for their combination. For sitting, standing, jogging, and biking activities, this difference is almost zero on average. However, we observe an average performance drop of 3%, 5%, and 7% for walking, walking upstairs, and walking upstairs in all scenarios. One of the reasons for such performance drop in using 10Hz is that we have a lower number samples per window compared to that of 50Hz. This can be improved by increasing the window size. Moreover, an additional dimension of magnitude for these two sensors can be used to counter the orientation effects, which can improve the overall performance. For this purpose, we evaluated some of these scenarios with increasing window size of 5 seconds

and also by introducing an additional magnitude dimension. In both cases, we observe performance improvements for the situations, where we previously observed performance drops. For example, using an extra magnitude dimension improves the performance at pocket position by an average of 4%, with an improvement range of 2% to 17%. The possible reason for this improvement can be that this extra magnitude dimension is used to counter orientation effects and phone orientation in a pocket is not as fixed like that of the smart-watch's orientation. We do not report all these values here because of limited space.

5) The Effects of Synchronization: In our work, we assume that data from smartwatch is being sent in real-time. Hence, a delay can be observed in communication. For this purpose, we introduce intentional delay in our data by shifting the data from the smartwatch by a certain number of samples and see if it affects the recognition performance. We introduced a 100 millisecond delay in the data by removing some samples from pocket position at the beginning. This number was chosen as an extreme case to test the effects of a possible delay. Then we evaluated this new dataset for all three classifiers with accelerometer, gyroscope and their combination. For 50Hz, there was little or no performance drop: in the order of 0.5 to 2% on average. For 10 Hz, in some cases, we observed a very small performance drop of less than 1% but in some cases there was a small performance increase so we consider the overall effect negligible. We assume that the system experiences the same type of delay in the training phase as well, provided it is an online training on the phone. In that case, the effect is almost negligible. It is yet to be seen how these delays will affect the performance when the classifiers are trained offline without such delays and used online with data where such type of delays are observed. However, we still believe that its impact will be negligible, because these physical activities are repetitive and such minor delays might not affect the overall performance. Moreover, if there is any performance drop, then it can be recovered by doing online training on the phone. However, this needs to be further explored.

B. Considering Complex Activities

In this section, we consider some complex activities in addition to the seven activities in the previous section and explore the role of these two devices in their recognition. For this purpose, we focused on 13 activities, as discussed in Section III. For these activities, we considered 5 different sensor combination scenarios as shown in Table I: W_A, W_G, W_{AG}, PW_A, and PW_{AG}. Moreover, we use the extra magnitude dimension for these complex activities to counter the orientation effects, as we discussed in the previous section. We also use a higher sampling rate of 50Hz to recognize these complex activities and leave the analysis of 10Hz for this set of activities for future work.

Based on our evaluation results as shown in Table IV, we found that combining the accelerometer from both pocket and wrist positions perform the best in recognizing all these 13 activities. The combination of the accelerometer and gyroscope from both positions (PW_{AG}) performs the same as that of the accelerometer combination from both positions (PW_A) except for walking upstairs and walking downstairs activities. In PW_{AG} scenario, we observe a higher performance than that of PW_A by an average of 2% and 4% for the walking upstairs

TABLE IV. RECOGNITION PERFORMANCE FOR COMPLEX ACTIVITIES IN W_A, AND W_{AG} SCENARIOS. PW_A - MAX(W_{AG}, W_A) SHOWS THE OVERALL PERFORMANCE IMPROVEMENT BY PW_A.

	Activity	W_{AG}	W_A	$W_{AG} - W_A$	PW_A - max(W_{AG}, W_A)
	A1	0.96	0.96	+0.00	+0.02
	A6	0.93	0.87	+0.06	+0.05
	A7	0.98	0.87	+0.11	0.00
	type	1.00	1.00	0.00	0.00
SVM	write	0.95	0.96	-0.01	+0.04
	coffee	0.80	0.82	-0.03	+0.17
	talk	0.77	0.73	+0.04	+0.20
	smoke	0.91	0.92	-0.01	+0.05
	eat	0.94	0.94	+0.01	+0.06
	A1	0.98	0.96	+0.02	+0.01
	A6	0.96	0.85	+0.11	+0.03
	A7	0.99	0.91	+0.08	-0.02
	type	1.00	0.99	0.00	0.00
IB1	write	0.99	0.98	+0.01	+0.01
	coffee	0.93	0.91	+0.03	+0.06
	talk	0.86	0.79	+0.08	+0.11
	smoke	0.93	0.93	0.00	+0.05
	eat	0.97	0.92	+0.04	+0.03
	A1	0.93	0.90	+0.03	+0.03
	A6	0.91	0.80	+0.11	+0.01
	A7	0.94	0.86	+0.09	-0.05
	type	0.99	0.99	0.00	0.00
J48	write	0.95	0.94	0.00	+0.04
	coffee	0.87	0.85	+0.02	+0.12
	talk	0.84	0.78	+0.06	+0.11
	smoke	0.90	0.89	+0.01	+0.05
	eat	0.91	0.88	+0.03	+0.08

TABLE V. RECOGNITION PERFORMANCE FOR ALL ACTIVITIES AT WRIST USING GYROSCOPE WITH 2, 5, AND 10 SECOND WINDOW SIZE.

	Activity	W_{G2sec}	$W_{G(5sec-2sec)}$	$W_{G(10sec-2sec)}$
	A1	0.94	+0.00	+0.01
	A2	0.66	+0.11	+0.18
	A3	0.95	+0.04	+0.02
	A4	0.84	+0.04	+0.02
	A5	0.82	+0.07	+0.11
	A6	0.86	+0.06	+0.12
IB1	A7	0.87	+0.05	+0.09
	type	0.74	+0.16	+0.21
	write	0.79	+0.05	+0.11
	coffee	0.77	+0.01	+0.05
	talk	0.57	+0.10	+0.13
	smoke	0.55	+0.14	+0.24
	eat	0.72	+0.05	+0.10
	A1	0.86	-0.01	+0.00
	A2	0.62	+0.09	+0.14
	A3	0.90	+0.03	+0.04
	A4	0.78	+0.02	+0.03
	A5	0.74	+0.04	+0.07
	A6	0.83	+0.04	+0.13
J48	A7	0.79	+0.08	+0.14
	type	0.66	+0.14	+0.19
	write	0.68	+0.05	+0.08
	coffee	0.69	+0.03	+0.05
	talk	0.51	+0.07	+0.13
	smoke	0.50	+0.07	+0.21
	eat	0.64	+0.05	+0.10

and walking downstairs, respectively. We only show the results for PW_A scenario in Table IV and not for PW_{AG} because they have almost same recognition performance. The performance improvement of adding the two additional gyroscopes is not significant, especially taking into account its impact on the battery life, which is yet to be seen in practical implementations. Using only smartwatch sensors performs worse than its combination with the smartphone sensors. Therefore, it is better to combine these two devices for better recognition of these complex activities. Within smartwatch sensors, the combination of gyroscope and accelerometer performs better than their individual performances for some of the complex activities, such as eating, drinking coffee, giving a talk, walking upstairs and walking downstairs. The performance results of these different combinations using SVM, IB1 and decision tree classifiers are shown in Table IV. We do not show the results for sitting, standing, jogging, and biking, because their results are very high and we observe negligible performance improvements for them in different fusion scenarios. In general, the recognition performance of using the gyroscope alone was poor and that is why we do include it in these graphs. Moreover, the recognition performance of complex activities at pocket position alone was poor, so we did not report these results.

Though there was a pattern in gyroscope data for these additional six activities, it was not as frequent as in the case of walking, jogging, biking and using stairs. This less frequent pattern for complex activities resulted in poor performance for gyroscope. Therefore, we increased the window size from 2 seconds to 5 and 10 seconds respectively to capture these patterns. This did improve the recognition performance for gyroscope using IB and DT classifiers. Compared to the 2 second window, this performance improvement for the IB1 and J48 classifiers was on average 7% and 5% using a 5 second window and 11% and 10% using a 10 second window. This activity-wise improvements for these three time windows using IB1

and J48 classifiers are shown in Table V. The $W_{G(5sec-2sec)}$ and $W_{G(10sec-2sec)}$ shows % increase by window size of 5 and 10 seconds compared to 2 seconds. If the gyroscope is used alone for recognizing these activities, the window size must be large enough to capture a complete repetitive pattern. In this case, a window size of 10 seconds can produce a significant recognition performance. This performance could be improved further by using larger than 10 seconds windows. However, we did not evaluate bigger windows because of the small size of our dataset. This can be done as a future work.

We also tested the other four sensors combinations with increasing window sizes of 5 and 10 seconds. However, the recognition performance for those combinations was very high for 2 second window, and we did not observe any significant performance improvement by increasing the window size. That is why we do not report those results here.

Though the recognition performances for the complex activities, such as smoking, eating, having a cup of coffee and giving a talk, are encouraging, there can be many variants of these activities. For example, smoking while sitting can be different. Smoking in a group of friends or colleagues where all of them are talking, can be a completely different activity because in group the frequency of inhaling the smoke and the duration between two such inhaling might be different. The same problems arise with eating activity. For this study we considered eating soup or yogurt with a spoon while these food items were on the dining table. However, there can be different variants of the eating activity, such as eating a sandwich while sitting, standing or walking. This needs to be further explored. In this study, we explored the possibility of using the fusion of smartphone and smartwatch sensors for recognizing various interesting activities where hand movements are involved. However, these various hand movements also might cause problems in detecting some activities. It might be a good idea not to use wrist position for detecting simple activities because sitting activity can have so many variants and will be difficult to recognize with a smartwatch only.

V. Conclusion and Future Work

In this work, we evaluated the role of smartphone and smartwatch sensors in recognizing 13 different human activities. First, we explored the fusion of sensors from both these devices for recognizing seven commonly used physical activities, such as walking, jogging, sitting, standing, walking upstairs, walking downstairs, and biking. Then we considered six additional activities: smoking, eating, drinking coffee, giving a talk, typing and writing. We showed that these complex activities are recognized with a higher accuracy by combining sensors from the smartphone in the pocket position and a smartwatch. The recognition of these complex activities can enable new well-being applications for the detection of bad habits. We also showed that for simple activities the fusion of these two devices may not bring significant improvements in the recognition performance, but it does make the recognition process more reliable. We also show that we can achieve an acceptable accuracy using low sampling rates when using sensor fusion, which can save battery life.

This is the preliminary work on recognizing complex activities, such as smoking, drinking coffee, giving a talk, and eating. We intend to further explore these activities in more realistic scenarios. For example, smoking combined with other activities, and eating in a group. Moreover, we would like to consider more complex activities, where hand movements are involved. We will further explore the personalized classification for these complex activities for individual users. In this work, we have imbalanced classes for some activities, so we intend to collect more data on these activities to create a dataset with balanced class distribution.

Acknowledgment

References

[1] O. D. Incel, M. Kose, and C. Ersoy, "A review and taxonomy of activity recognition on mobile phones," *BioNanoScience*, vol. 3, no. 2, pp. 145–171, 2013.

[2] W. Z. Khan, Y. Xiang, M. Y. Aalsalem, and Q. Arshad, "Mobile phone sensing systems: A survey," *Communications Surveys & Tutorials, IEEE*, vol. 15, no. 1, pp. 402–427, 2013.

[3] G. Bieber, N. Fernholz, and M. Gaerber, "Smart watches for home interaction services," in *HCI International 2013-Posters Extended Abstracts*. Springer, 2013, pp. 293–297.

[4] B. J. Mortazavi, M. Pourhomayoun, G. Alsheikh, N. Alshurafa, S. I. Lee, and M. Sarrafzadeh, "Determining the single best axis for exercise repetition recognition and counting on smartwatches," pp. 33–38, 2014.

[5] S. G. Trost, Y. Zheng, and W.-K. Wong, "Machine learning for activity recognition: hip versus wrist data," *Physiological measurement*, vol. 35, no. 11, p. 2183, 2014.

[6] S. Chernbumroong, A. S. Atkins, and H. Yu, "Activity classification using a single wrist-worn accelerometer," in *Software, Knowledge Information, Industrial Management and Applications (SKIMA), 2011 5th International Conference on*. IEEE, 2011, pp. 1–6.

[7] F. G. da Silva and E. Galeazzo, "Accelerometer based intelligent system for human movement recognition," in *Advances in Sensors and Interfaces (IWASI), 2013 5th IEEE International Workshop on*. IEEE, 2013, pp. 20–24.

[8] R. I. Ramos-Garcia and A. W. Hoover, "A study of temporal action sequencing during consumption of a meal," in *Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics*. ACM, 2013, p. 68.

[9] Y. Dong, J. Scisco, M. Wilson, E. Muth, and A. Hoover, "Detecting periods of eating during free living by tracking wrist motion," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 4, pp. 1253–1260, 2013.

[10] P. M. Scholl and K. Van Laerhoven, "A feasibility study of wrist-worn accelerometer based detection of smoking habits," in *Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2012 Sixth International Conference on*. IEEE, 2012, pp. 886–891.

[11] M. Marin-Perianu, C. Lombriser, O. Amft, P. Havinga, and G. Tröster, "Distributed activity recognition with fuzzy-enabled wireless sensor networks," in *Distributed Computing in Sensor Systems*. Springer, 2008, pp. 296–313.

[12] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. T. Campbell, "A survey of mobile phone sensing," *Communications Magazine, IEEE*, vol. 48, no. 9, pp. 140–150, 2010.

[13] A. M. Khan, A. Tufail, A. M. Khattak, and T. H. Laine, "Activity recognition on smartphones via sensor-fusion and kda-based svms," *International Journal of Distributed Sensor Networks*, vol. 2014, 2014.

[14] J. J. Guiry, P. van de Ven, and J. Nelson, "Multi-sensor fusion for enhanced contextual awareness of everyday activities with ubiquitous devices," *Sensors*, vol. 14, no. 3, pp. 5687–5701, 2014.

[15] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. Havinga, "Fusion of smartphone motion sensors for physical activity recognition," *Sensors*, vol. 14, no. 6, pp. 10 146–10 176, 2014.

[16] M. Shoaib, H. Scholten, and P. J. Havinga, "Towards physical activity recognition using smartphone sensors," in *Ubiquitous Intelligence and Computing, 2013 IEEE 10th International Conference on (UIC)*. IEEE, 2013, pp. 80–87.

[17] "Dataset available at:," http://ps.ewi.utwente.nl/Datasets.php.

[18] S. Das, L. Green, B. Perez, M. Murphy, and A. Perring, "Detecting user activities using the accelerometer on android smartphones," Carnegie Mellon University (CMU), Tech. Rep., 2010.

[19] P. Siirtola and J. Röning, "Recognizing human activities user-independently on smartphones based on accelerometer data," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 1, no. 5, 2012.

[20] P. Siirtola and J. Roning, "Ready-to-use activity recognition for smartphones," in *Computational Intelligence and Data Mining (CIDM), 2013 IEEE Symposium on*. IEEE, 2013, pp. 59–64.

[21] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: an update," *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10–18, 2009.

[22] H. Lu, J. Yang, Z. Liu, N. D. Lane, T. Choudhury, and A. T. Campbell, "The jigsaw continuous sensing engine for mobile phone applications," in *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*. ACM, 2010, pp. 71–84.

[23] J. Frank, S. Mannor, and D. Precup, "Activity recognition with mobile phones," in *Machine Learning and Knowledge Discovery in Databases*. Springer, 2011, pp. 630–633.

[24] E. Miluzzo, N. D. Lane, K. Fodor, R. Peterson, H. Lu, M. Musolesi, S. B. Eisenman, X. Zheng, and A. T. Campbell, "Sensing meets mobile social networks: the design, implementation and evaluation of the cenceme application," in *Proceedings of the 6th ACM conference on Embedded network sensor systems*. ACM, 2008, pp. 337–350.

[25] S. Thiemjarus, A. Henpraserttae, and S. Marukatat, "A study on instance-based learning with reduced training prototypes for device-context-independent activity recognition on a mobile phone," in *Body Sensor Networks (BSN), 2013 IEEE International Conference on*. IEEE, 2013, pp. 1–6.

[26] V. Stewart, S. Ferguson, J. X. Peng, and K. Rafferty, "Practical automated activity recognition using standard smartphones." in *PerCom Workshops*, 2012, pp. 229–234.

[27] Y. Liang, X. Zhou, Z. Yu, B. Guo, and Y. Yang, "Energy efficient activity recognition based on low resolution accelerometer in smart phones," in *Advances in Grid and Pervasive Computing*. Springer, 2012, pp. 122–136.