This full text paper was peer-reviewed at the direction of IEEE Instrumentation and Measurement Society prior to the acceptance and publication.

Design of Activity Recognition Systems with Wearable Sensors

Igor Khokhlov, Leon Reznik
B. Thomas Golisano College of Computing and Information Sciences
Rochester Institute of Technology
Rochester, USA
ixk8996@rit.edu, lr@cs.rit.edu

Justin Cappos
Department of Computer Science and Engineering
New York University
New York, USA
jcappos@nyu.edu

Rohit Bhaskar
B. Thomas Golisano College of Computing and Information Sciences
Rochester Institute of Technology
Rochester, USA
bxb7772@g.rit.edu

***Abstract*—Wearable sensors are widely utilized in human activity monitoring and recognition systems. Not only do these sensors come in different form factors but the software that comes bundled with them also varies from device to device and is constantly evolving. Also, multiple types of sensors on these devices are used to recognize human activity. Owing to the flexible form factor, these devices can also be mounted on a plethora of different positions on the human body. With all the aforementioned variables, it becomes imperative that the quality of data provided by wearable sensors needs to be evaluated. This paper describes an empirical study resulting in evaluating the accuracy of human activity recognition by wearable sensors based on the type of sensor, the physical mounting position of the sensor on the human body, their type of activity being monitored and the type of device being used. The paper further delves into assessing the results of this study. It provides guidelines for designing better wearable sensor systems for human activity recognition.**

Keywords— wearable sensors; activity recognition; machine learning.

I. Introduction

Over the last few years, a human activity recognition with wearable sensors has moved from an academic arena to an industrial practice with multiple devices and computer apps emerging on the market [1]. According to Forbes [2], wearable market is going to exceed four billion US dollars in 2017 and more than 125 million devices are expected to be shipped in 2019. These devices recognize activities using either embedded or standalone sensors. They may have various design and implementation platforms, ranging from fitness bracelets with very limited resources to smartphones, and even bigger implementation platforms including desktops. They may have different purposes, from a simple activity recognition and detection of elderly people's falls [3], [4], [5] to guiding sports activities or estimating the worker's load at a factory [6]. An accelerometer, a gyroscope, and a pedometer are among the most popular sensors for an activity recognition [7], [8], [9]. Most commercial smartphones incorporate either all or, at least, two of them. Activity recognition applications, such as Google Fit, Samsung Health, Noom Walk for Android OS mobile devices and Human - Activity Tracker, Pacer for iOS devices, which collect data from embedded sensors and utilize them for a healthcare purposes, are available.

However, the recognition performance of those devices and applications is still low. Bender et al. [10] conducted an empirical study of various fitness devices such as Fitbit Flex, Fitbit Charge HR, Garmin vívoactive, and Apple Watch. Their study shows that results of different step counting devices used by the same person on the same distance may deviate up to 30% from each other. This example clearly demonstrates the need for improvement in the design of the data collection and utilization systems that incorporate wearable sensors. There exist numerous challenges [11] in a wearable sensor system design ranging from the physical implementation platform choice to the selection of the algorithms, which are employed to process the sensor data to recognize human movements and activities, and to count the number of steps. Many challenges are caused by form-factor, weight and energy constrains as well as by safety requirements.

This paper aims at developing recommendations that could help to address those challenges. In particular, we are addressing the following problems:

- How to mount a sensor on a human body? (here we are providing recommendations on the physical design of an implementation platform);
- How many sensors should a wearable system to employ (is many much better than one?) and which ones? (the problem of the structural and architectural system design);
- Does out-of-the-box ready-to-use system demonstrate much poorer performance than one that requires additional training to adapt to a particular user? (the problem of the machine learning strategy choice: generic design vs. individual adaptation);
- Which techniques to employ to train a system? (the problem of the pattern classifier design to recognize various human activities and body movements).

As additional criteria, our recommendations should reduce the cost of the platform by employing the minimum number of sensors and achieving the reasonable accuracy of the activity recognition. Also, they should not make the user's experience worse.

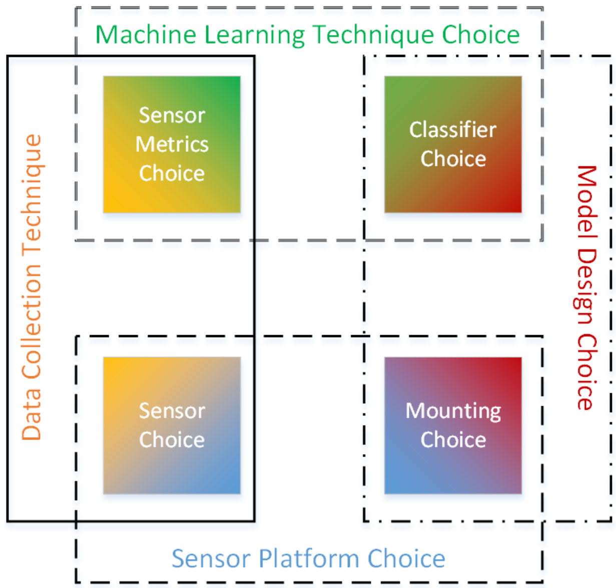


Fig. 1. Activity Recognition Design Steps

A system designer has to decide: which sensors to choose or what machine learning technique to implement as well as to make other important decisions. The design of activity recognition systems consists of several interconnected steps (see Fig. 1), which involve other choices:

- Data Collection Technique – which sensors to use and what data from these sensors to use;
- Sensor Platform Choice – which sensors to use and how these sensors should be mounted;
- Machine Learning Technique Choice – what features from sensor data to extract and what classifier fits the best;
- Model Design Choice – which classifier to use depends on how sensors are mounted.

The classifier design can be subdivided into two other steps: training the model and using it on a particular individual (see Fig. 2).

In order to develop design recommendations, we piloted an empirical study that utilized Android-based smartphones and sensors embedded therein and other wearable sensors in order to collect and analyze data received. This paper describes the results analysis of this study, conducted with various Android mobile devices (see Table I for the list). Accelerometer and gyroscope sensors were employed. Although in an activity recognition system design both unsupervised and supervised learning techniques had been researched [6] [12], in our study we compared four supervised machine learning algorithms: J48, Naive Bayes, Random Forest and K-Nearest Neighbor (KNN). In our study, we investigated the choice of:

- Number of sensors;
- Type of sensors;
- Classifier design;
- Type of the system mounting;
- Feature set used in a classifier design;
- Necessary number of measurements for accurate activity identifying;
- Which model to use: a model trained on a personal data or a generic model.

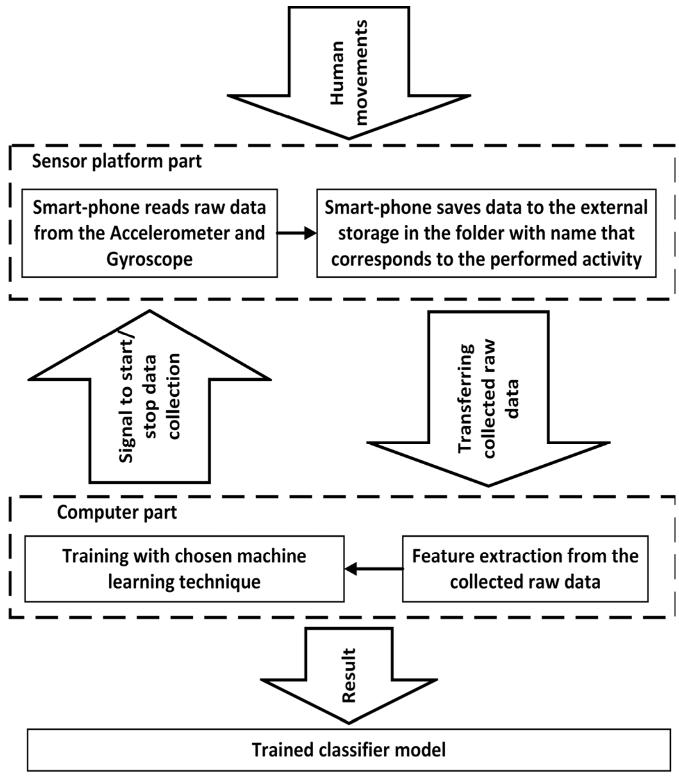


Fig. 2. Pipeline of the data collection and a classifier model training

The generic description and structure of an activity recognition design are presented in section II. An empirical study is described in section III. In section IV the following design problems and their solutions are discussed:

- How many sensors are needed? - see subsection A;
- Which machine learning techniques are a better fit for various data collection schemes? - see subsection B;
- The mounting choice is discussed in subsection C.

TABLE I. SMART-PHONE SPECIFICATION

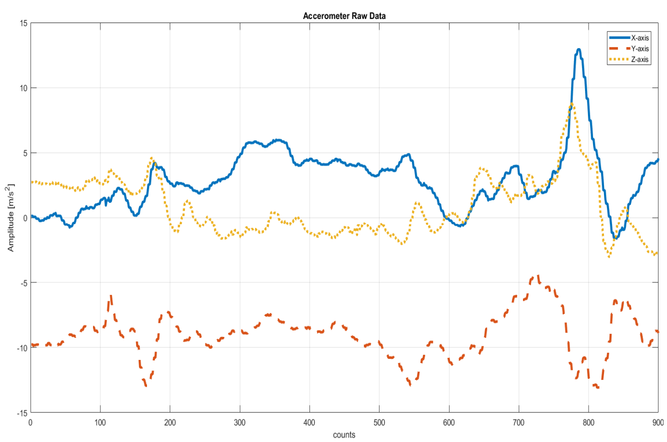
Smartphone model	Version of Android OS
HTC One	4.2 (Jelly Bean)
Google Nexus 4	5.1.1 (Lollipop)
Google Nexus 5	4.4 (Kit-Kat)
Samsung Galaxy S6	5.0 (Lollipop)
Samsung Note	4.4 (Kit-Kat)
Google Nexus 6P	6.0.1 (Marshmallow)

II. ACTIVITY RECOGNITION DESIGN PROCESS

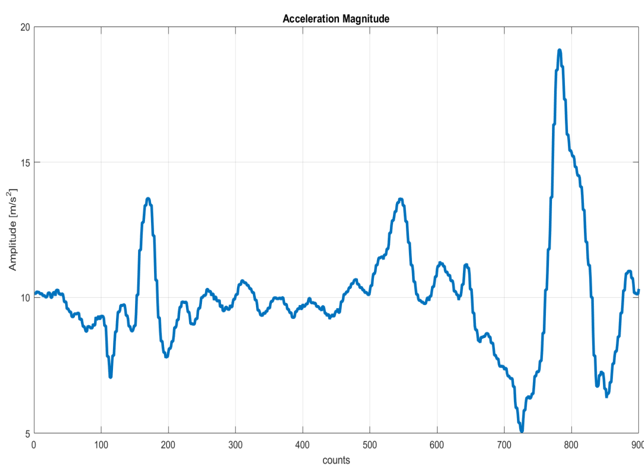
The activity recognition design typically includes the following steps (see Fig. 2):

- Data collection from sensors;
- Data pre-processing and feature extraction;
- Classifier choice and its design.

In order to identify an ongoing human activity, a series of measurement should be taken. This series of measurement composes a sample window. As no decision is made until the sample window is filled with measurements, its size is critical for conducting the activity recognition in a real time. Although previous studies [13] required a few hundred measurements, we developed the method that demonstrated a good recognition performance with the sample window of 150 measurements (counts) only. The measurement frequency depends on a sensor platform. In our study, we employed modern smartphones with the measurement frequency of 100Hz. Our approach allows reducing an activity recognition latency to 1.5 seconds in this case.



(a)



(b)

Fig. 3. Data recordings from the accelerometer for walking activity: (a) Raw data from x,y and z axes. (b) 3-dimensional acceleration values transformed into one value.

The initial step includes collecting data from one or several sensors of different types, such as an accelerometer, a gyroscope, a magnetometer [14] and others. In this paper, we examine an accelerometer and a gyroscope sensor data. An accelerometer is a sensor that measures acceleration along x, y and z-axes and a gyroscope is a sensor that measures rotation rate around x, y and z-axes. Fig. 5a presents an example of a raw data recorded from the accelerometer, where each color represents one of the three axes: blue color corresponds to x, red to y, and orange to z.

For further processing, three-dimensional acceleration values are transformed into one acceleration magnitude value that is shown in Fig. 5b. The magnitude value $\boldsymbol{\alpha}$ of the 3-dimensional acceleration vector has been calculated as a Euclidean magnitude:

$$\alpha = \sqrt{x^2 + y^2 + z^2} \tag{1}$$

Feature extraction is the next step in data processing that facilitates a classifier design. Academia and industry have investigated various feature sets as well as various classification techniques. Reddy et al. [15] employ mean, variance, energy, and the Discrete Fourier Transform (DFT) energy coefficients between 1-10Hz. These features were applied in the two classifiers: continuous Hidden Markov Model and two-staged classifier (Instance Classifier + Discrete Hidden Markov Model). Some researchers adopt less computationally expensive feature extraction methods than DFT. Kose et al. [16] employ average acceleration magnitudes over the decision window as features that are used in clustered KNN and Naive Bayes classifiers.

Our study examines three feature extraction methodologies. First includes the following features:

- *Fundamental Frequency,* which represents the average value of the first three dominant frequencies of the signal amplitude obtained by applying Fast Fourier Transform to the sample window.
- *Average Magnitude of the Acceleration,* which represents the arithmetic average value of the magnitude over the sample window.
- *Minimal Magnitude of the Acceleration,* which represents the minimum value of the magnitude over the sample window.
- *Maximal Magnitude of the Acceleration,* which represents the maximum value of the magnitude over the sample window.

In the second approach to the feature extraction, features listed above are complemented with additional three features (seven in total): the average value of the acceleration amplitude for each of the three axes calculated over the sample window. In the third approach, those seven features are complemented by the average rotation rate around the smartphone's each axis calculated over the sample window.

III. EMPIRICAL STUDY DESCRIPTION

In our experiments, we have employed six different models of smartphones running various Android OS versions (see TABLE I) to collect data representing walking, standing, and sitting activities. For the feature extraction, model training and classification, a laptop running Windows OS was used.

During the training stage, a remotely controlled smartphone collects raw data using built-in accelerometer and gyroscope sensors. For the data collection a specifically designed, dedicated JAVA-written application was developed and used. This application is remotely controlled from a laptop. Laptop sends corresponding commands to collect a specific activity type. This command information gives us a ground-truth knowledge about the performed activity. The application records raw data from the accelerometer and the gyroscope sensors and stores it in the smartphone's internal memory. After data is collected, it is transferred to a remote system where a JAVA application extracts features and trains the activity recognition model by using machine learning techniques. The pipeline of the training stage is shown in Fig. 2.

During our study, measurements from the accelerometer and gyroscope (three from each) were recorded every 10 ms over the period of approximately 13 minutes with each of the phones listed in TABLE I. Overall, 457500 records from the accelerometer and gyroscope were collected and processed.

Data collected was subdivided into a training part (70%) and a testing part (30%). As we develop a generic recognition model, training data represents all participants and testing data represents only one participant. This approach simulates a real-life scenario when an activity recognition device employs the software pre-trained on a generic data. In this case, the device can be used directly out of the box and will not require a consumer's involvement into its training and adaptation. However, in the experiment that is described in the section IV.B.1 the generic model was produced without using data from the person on which it is tested later on.

TABLE II. RECOGNITION PERFORMANCE IN % (CORRECT/TOTAL)

Algorithm name	4 Features only from the accelerometer	7 Features only from the accelerometer	10 Features from the accelerometer and gyroscope
J48	94.8%	96.5%	96.6%
Naive Bayes	76.2%	96.4%	96.5%
Random Forest	95.0%	97.5%	97.6%
KNN (k=1)	93.6%	97.6%	97.9%

IV. RESEARCH RESULTS AND DISCUSSION

TABLE II presents results of testing using various techniques employed in our study. We investigated three data acquisition and feature extraction methods (data recorded from the accelerometer only versus data collected from both the accelerometer and the gyroscope with the number of features ranging from four to ten), and four machine learning algorithms.

A. How many sensors are needed?

The previous study indicated [8] that an application of multiple sensors could drastically improve the recognition performance in case of recognizing complex activities such as climbing, meditation, etc. In a perfect case, as inferred from the TABLE II, increasing the number of features from four to seven results in a recognition performance improvement by more than 3%. When seven features are extracted from the accelerometer data, "K Nearest Neighbor" algorithm achieves the highest performance with 97.6% accuracy. Adding a gyroscope sensor might increase the recognition performance by 0.1-0.3% only. Also, a gyroscope sensor can help in eliminating errors in human activity recognition if the sensor is displaced during data collection as was shown by [17].

Our investigation demonstrates that complementing accelerometer data with data from a gyroscope for recognizing simple activities such as walking, sitting and standing, does not significantly improve the recognition performance (see TABLE II). In this scenario, KNN technique achieves 97.8% with a generic model and 99.8% using the individual trained model.

B. Which machine learning techniques are better fit to various data collection schemes?

1) Generic model and individual dataset for testing:

In this scenario, we derive a generic model that was trained with data collected from more than one person but does not contain data from the one who will be using this application. One should pay attention that this model differs from the one that was described above, where training set may include records from all persons. Such approach can be implemented, for example, in wearable activity trackers where the user cannot easily update the model. To test this scenario, a generic model was trained and later used for activity recognition of an earlier unseen person. Results are shown in TABLE IV (see generic model column).

2) Individual model and individual dataset for testing:

For smart devices, such as smartphones or fitness trackers with advanced features, another scenario can be applied. In this scenario, a user can train a model using their own data. Our investigation shows (see TABLE IV last two columns) that this approach may produce much more accurate results with the almost perfect recognition rate achieved with the KNN algorithm.

Both scenarios have advantages as well as disadvantages. A generic model simplifies the user interface and improves the user's experience with simple devices such as fitness trackers that can be used straight out of the box with no initial adjustment. On the other hand, using an individual model provides a significant increase in the recognition performance, but requires a more complicated setup process.

3) Sensor measurement uncertainty consideration:

During the training stage, data is collected in a controlled environment where sensors are considered properly calibrated, which might not hold valid in a real-life situation. In our study, we investigated the influence of the measurement uncertainty on the recognition performance. The possible uncertainty rates of 1%, 5%, and 10% were investigated with the recognition results shown in Fig. 6. The results indicate a possible drop in the recognition performance up to 50%. This drop could be decreased by employing both an accelerometer and a gyroscope sensor and extracting ten features instead of only four features [13]. Employing KNN technique allows achieving the highest recognition performance.

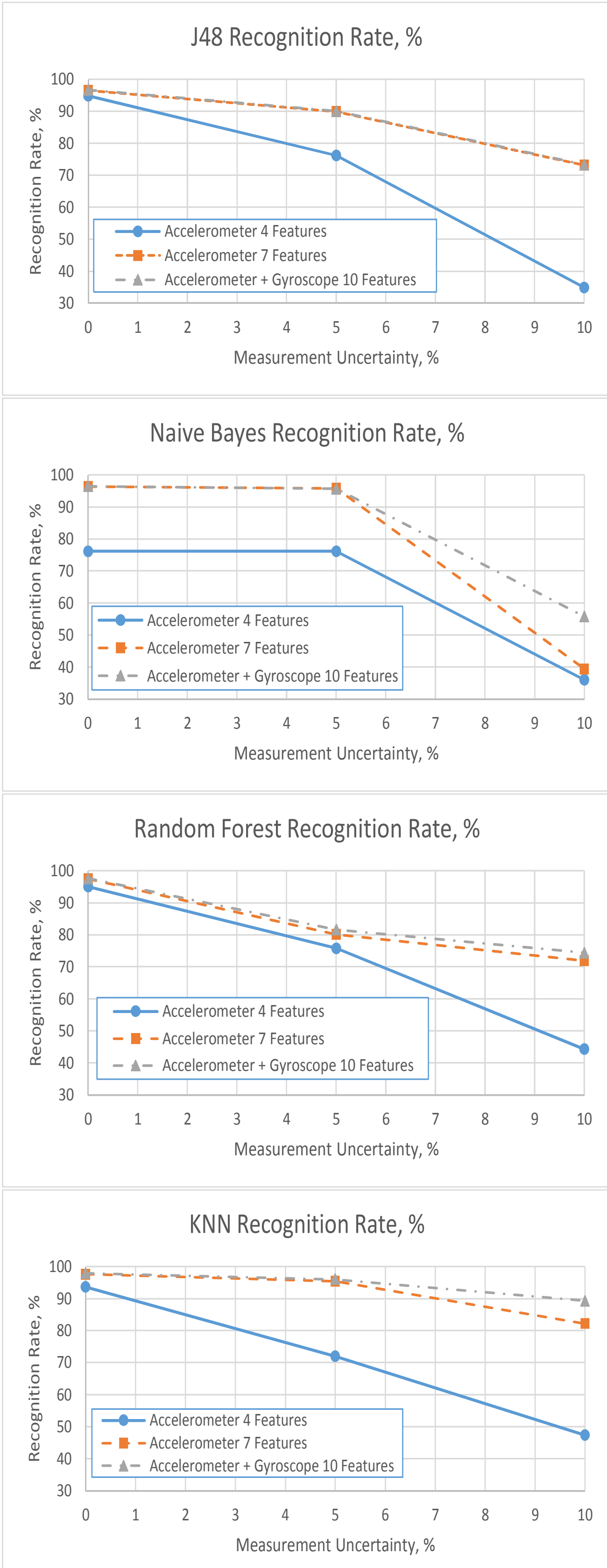


Fig. 4. Activity recognition performance for different classification techniques

TABLE III. FLEXIBLE MOUNTING VS. STRONG MOUNTING. ACCELEROMETR SENSOR ONLY, 4 FEATURES WERE USED

Algorithm name	**Flexible Mounting**	**Strong Mounting**
J48	66.6%	99.4%
Naive Bayes	66.7%	98.4%
Random Forest	66.7%	99.6%
KNN (k=1)	66.9%	99.8%

TABLE IV. RESULTS FOR INDIVIDUAL TESTS USING GENERIC MODEL AND INDIVIDUAL MODEL. ACCELEROMETR SENSOR ONLY, 4 FEATURES WERE USED

Algorithm name	**Generic model**	**Individual model, person 1**	**Individual model, person 2**
J48	98.4%	99.8%	99.4%
Naive Bayes	76.2%	72.2%	98.4%
Random Forest	95.0%	99.8%	99.6%
KNN (k=1)	93.6%	100%	99.8%

C. *Mounting choice*

Possible sensor movements due to its flexible mounting against the body during activity recognition can drastically decrease the recognition performance. In all experiments described above, a sensor platform was tightly fixed against the human body allowing movements not more than a few millimeters. In additional experiments described in TABLE IV (see “Flexible Mounting” column) a sensor could move for 5 mm to 5 cm against the body. Our study shows that the flexible mounting of the sensor can decrease recognition performance by about 33%. This result needs to be considered in wearable sensors platform physical design.

V. CONCLUSION

An accelerometer and a gyroscope sensors embedded in smartphones and other mobile devices collect data for multiple applications that could be used for recognizing human activities, counting steps and measuring traveled distances. However, the poor quality of recorded data may result in low recognition performance in standard mobile applications. Our research produced a number of recommendations on the design of activity recognition systems. The recognition technology developed in our investigation allows decreasing the number of measurements required for reliable recognition from a few hundred reported previously [13] to 150, that reduces the result’s latency to 1.5 seconds. This delay value makes our design acceptable for human activity recognition systems and allows for their quality improvement.

Our investigation shows that for simple human activities the high recognition performance (97.6%) can be achieved by using only a 3-axis accelerometer sensor with a generic trained model. It means that a simple high-performance activity recognition system can be implemented on a mobile device with only one accelerometer sensor. Avoiding a gyroscope employment may drastically prolong a device’s battery life. For example, according to the specification of the InvenSense MPU-6515 accelerometer/gyroscope chip [18], which is used in Google Nexus 5 phone [19], the accelerometer consumes almost seven times less current than a gyroscope (450μA against 3.2mA accordingly). However, in a real life, a typical recognition performance could drastically drop due to various factors, such as flexible mounting of the device against the body, the absence of a sensor calibration service and high measurement uncertainty characteristics. For example, our study demonstrated that possible sensor movement against human body due to a flexible mounting could decrease the recognition performance using a generic model by up to 30%.

We investigated and developed special techniques, which we could recommend to apply in order to prevent this degradation, such as:

- Employment of a gyroscope along with an accelerometer.
- Extending the features set used in the classifier design.
- Choosing the most suitable machine learning technique in a classifier design.
- Employing a model trained on the personal data instead of a generic model.

Our investigation has proved that all those methods increase the recognition performance, but require more resources for their implementation. The individual model trained and used on the same person may still achieve almost 100% recognition performance, but requires a more complicated setup process. On the other hand, a generic model may achieve a performance of about 95%, but a corresponding device could be used straight out-of-the box.

Our research demonstrates that smartphones with embedded accelerometer and possibly gyroscope sensors could be used as an activity recognition device without a significant sacrifice in accuracy against dedicated devices. In relation to the recognition performance, quality rigid mounting of the device on the human's body is more important than the choice of the data processing or feature selection algorithm.

ACKNOWLEDGMENT

This material is partially based upon work supported by the National Science Foundation under awards # ACI-1547301 and # ACI-1547290.

REFERENCES

[1] J. Duffy and A. Colon, "The Best Fitness Trackers for 2016," PC Magazine, 2016. [Online]. Available: http://www.pcmag.me/a/2404445. [Accessed 6 June 2016].

[2] B. Marr, "15 Noteworthy Facts About Wearables In 2016," Forbes, 18 March 2016. [Online]. Available: https://www.forbes.com/sites/bernardmarr/2016/03/18/15-mind-boggling-facts-about-wearables-in-2016/. [Accessed 27 July 2017].

[3] O. Aziz and S. N. Robinovitch, "An analysis of the accuracy of wearable sensors for classifying the causes of falls in humans}," vol. 19, no. 6, pp. 670-676, 2011.

[4] C. Ranhotigmage, Human activities and posture recognition: Innovative algorithm for highly accurate detection rate, Massey: Massey University, 2013.

[5] T. Shany, S. J. Redmond, M. R. Narayanan and N. H. Lovell, "Sensors-based wearable systems for monitoring of human movement and falls," pp. 1-13, 2011.

[6] T. M. Quan Kong, "Sharing training data among different activity classes," Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication, pp. 701-712, 2013.

[7] M. Zhang and A. A. Sawchuk, "Human daily activity recognition with sparse representation using wearable sensors," Biomedical and Health Informatics, IEEE Journal of, vol. 17, no. 3, pp. 553-560, 2013.

[8] S. Dernbach, B. Das, N. C. Krishnan, B. L. Thomas and D. J. Cook, "Simple and complex activity recognition through smart phones," in Intelligent Environments (IE), 2012 8th International Conference on, 2012.

[9] R. Liu, T. Chen and L. Huang, "Research on human activity recognition based on active learning," in Machine Learning and Cybernetics (ICMLC), 2010 International Conference on, 2010.

[10] C. Bender, J. C. Hoffstot, B. T. Combs, S. Hooshangi and J. Cappos, "Measuring the fitness of fitness trackers," in Sensors Applications Symposium (SAS), 2017 IEEE, 2017.

[11] S. C. Mukhopadhyay, "Wearable sensors for human activity monitoring: A review," IEEE sensors journal, vol. 15, no. 3, pp. 1321-1330, 2015.

[12] T. Maekawa, N. Daisuke, O. Kazuya and N. Yasuo, "Toward practical factory activity recognition: unsupervised understanding of repetitive assembly work in a factory.," Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 1088-1099, 2016.

[13] S. Das, L. Green, B. Perez and M. Murphy, "Detecting User Activities using the Accelerometer on Android Smartphones," 2010.

[14] K. Altun and B. Barshan, "Human activity recognition using inertial/magnetic sensor units," in Human behavior understanding, Springer, 2010, pp. 38-51.

[15] S. Reddy, M. Mun, J. Burke, D. Estrin, M. Hansen and M. Srivastava, "Using mobile phones to determine transportation modes," ACM Transactions on Sensor Networks (TOSN), vol. 6, no. 2, p. 13, 2010.

[16] M. Kose, O. D. Incel and C. Ersoy, "Online human activity recognition on smart phones}," in Workshop on Mobile Sensing: From Smartphones and Wearables to Big Data, 2012.

[17] K. Kunze and P. Lukowicz, "Dealing with sensor displacement in motion-based onbody activity recognition systems," in Proceedings of the 10th international conference on Ubiquitous computing, 2008.

[18] MPU-6500 Product Specification, Revision 1.0, InvenSense Inc., 2013.

[19] "Nexus 5 Teardown - iFixit," iFixit, [Online]. Available: https://www.ifixit.com/Teardown/Nexus+5+Teardown/19016. [Accessed 18 January 2018].

[20] Bosch Sensortec, BMI160 Datasheet (BST-BMI160-DS000-07), Bosch Sensortec, 2015, p. 2.