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

Position-independent Physical Activity Monitoring: Development and Comparison with Market Devices

Joana Silva, Diana Gomes, Francisco Nunes, Dinis Moreira, José Alves, Ana Pereira, Inês Sousa
Fraunhofer Portugal AICOS
Porto, Portugal
joana.silva@fraunhofer.pt

***Abstract*—The last years have seen great growth in wearable devices for activity monitoring. Fueled by miniaturization and market demand, many devices can accurately estimate physical activity. However, and despite the variety of solutions, existing wearables are tied to classifying activity from a single body position. Once users move their wearable, accuracy is affected. This paper describes the development and validation of a position-independent activity monitoring solution, the GoLiveClip. The performance of GoLiveClip is comparable to position-dependent devices, such as Fitbit Flex 2, Mi Band 2, Garmin Vivoactive HR, and Jawbone Up 24, for activity recognition, step detection, and distance estimation. These results motivate the development of position-independent activity monitoring solutions.**

***Index Terms*—wearable, inertial sensors, activity monitoring, activity tracker, activity classification, machine learning**

I. Introduction

Lack of physical activity is the fourth risk factor leading to death worldwide[1]. While there are many ways to perform physical activity, it is hard to keep track of one's progress and meet recommended levels of exercise. Pervasive activity monitors can help users better understand their physical activity and are thus becoming increasingly common in the market. The demand for accurate wearable activity monitors motivates exploring solutions that are more flexible and less obtrusive.

Existing wearable devices need to be attached to a specific body part, such as the wrist or the waist. Wrist-based devices became very common probably because they can enable physiological signals monitoring (e.g. pulse), watch-like features, and an interface for displaying app notifications. Nevertheless, if physical activity monitoring is the main goal, wrist motion data is usually associated with more noise than motion from other body parts, due to the variety of movements of great amplitude enabled by the wrist, elbow, and shoulder joints. To improve accuracy, some wearables were adapted to fit several on-body positions. When starting a workout session, for example, users are asked to attach the wearable to a certain body part, using a clip, clasp or sports band, and to tell the system where they are wearing the wearable or the physical activity they will engage in (e.g. running, swimming, etc.). This means that these devices lack complete position-independence and/or automatic activity recognition. Moreover, wearables frequently aggregate activities (metrics are presented to the user as overall active time or activity intensity levels) and require a minimum time to pass (often several minutes) before they can automatically classify the on-going activity. These handicaps impact recognition accuracy [1] and may decrease user acceptance and adherence, motivating the development of a truly position-independent and automatic physical activity monitoring framework.

Pervasive position-independent activity monitoring has explored the use of smartphones as the sensing device [2]–[10], and to less extent wearables [11]–[13]. However, these position-independent wearable solutions have not been released to the market nor have they been compared against market devices.

This work presents a framework for automatic physical activity classification (and associated metrics extraction) which processes data from a wearable device that can be placed in the chest or the belt, without prior knowledge of its on-body position. Our solution was compared with similar purpose and commercially available wearable devices, in a benchmark validation, obtaining equivalent or even better results than top seller position-dependent devices.

This paper has three sections besides the introduction. The Methodology section describes the activity monitoring framework and the benchmark validation protocol and used devices. The Results section presents the training accuracy of the activity classification, steps detection, and distance estimation, as well as the comparative accuracy of the developed algorithms compared to market devices. The main results discussion and conclusions are detailed in the last section.

II. Methodology

A. Wearable Activity Monitoring Framework

This section describes the development of a position-independent activity monitoring framework to automatically classify the activities of the user, and to retrieve, for dynamic activities as walk and run, the number of steps and travelled distance. Duration and energy consumed in each classified activity is also calculated. The framework was designed to process inertial sensor data from a wearable device placed in the chest or in the belt; however, the user does not need to inform the framework where the wearable is fixed. The

Supported by project Symbiotic technology for societal efficiency gains: Deus ex Machina (NORTE-01-0145-FEDER-000026) supported by North Portugal Region Operational Programme (NORTE 2020), Portugal 2020 and the European Regional Development Fund (ERDF) from European Union.

[1]https://www.who.int/dietphysicalactivity/pa/en/

978-1-5386-8428-3/19/$31.00 ©2019 IEEE

wearable device is connected to a smartphone, where the processing of inertial data and display of the information is done. To train the several models that constitute the framework, we have collected a laboratory dataset of activities, with annotated number of steps and travelled distance.

1) Framework description: The activity monitoring framework is based on previous work by our group [14], [15] that used smartphones' internal sensors. The smartphone app was adapted to receive inertial data from a wearable sensor, and the activity monitoring framework was changed to process data from multiple on-body positions. We considered two main body positions: chest and belt. For training the framework, data collected with the wearable located on the waist and on the belt have been used without distinction, to ensure independence of the framework from the wearable position.

First, the framework discriminates between several physical activities and stances, using a sequential classification pipeline with an activity classifier followed by a stance classifier (Fig. 1). Having detected the activity, if the user is walking or running, the framework retrieves the number of steps and travelled distance, as previously described in [14]. The framework also retrieves the energy consumed during the activities and stances, by means of an energy estimation algorithm [14].

The activity classifier is based on a decision tree algorithm, that discriminates between *Walk*, *Run*, *Tilt* (random movements, e.g. removing the device from the body and placing it on the table, or shaking the device on the hand when it is being transported) and *Inactive* (considers a combination of stances, e.g. stand, sit, lay, and events of not using the device, like charging or leaving it on the table). When the activity classifier detects inactivity, it is followed by a stances classifier. This classifier is also based on a decision tree and discriminates between *Not Using* and *Still*. These models were based on time-domain features of low computational power, extracted from the accelerometer's data sampled at 33Hz, and segmented in 5-seconds non-overlapping windows. The framework has been trained with accelerometer signals acquired from the wearable device located on both waist and belt positions, for ensuring better generalization of framework's algorithms.

The features extracted were statistical features, namely, skewness, kurtosis, mean, maximum, minimum, standard deviation (stdev), root-mean-square (RMS), interquartile range and median absolute deviation, from the three axis x, y, z and magnitude of the accelerometer data. Through forward feature selection methods using a decision tree algorithm and 10 fold cross validation, the best features for both stance and activity classifiers were selected. For the activity classifier, the features selected are: standard deviation (stdev) of the x, y, z axis and magnitude; mean of the y axis and magnitude and root mean square of the x, y axis. For the stances classifier the features selected are the stdev of the y axis and magnitude and mean of the x axis. Then, a decision tree classifier algorithm was trained with data from the training dataset, Section II-A2, using only the selected features. The algorithm was then validated and compared with some other commercial solutions, Section II-B.

The step detection and distance estimation algorithms described in [14] were adapted for samples recorded from wearable devices. Mainly, the coefficients of the linear regression used for estimating the stride length were adjusted for the samples collected with the wearable devices and considering different carrying positions.

The developed activity monitoring framework was deployed on the commercially available GoLivePhone (version 4.2.3.0) and GoLiveClip (GLC), commercialised by Gociety Solutions[2]. The GLC includes a 3-axis accelerometer that streams data to a smartphone using Bluetooth. Data is processed by the activity monitoring framework running on the smartphone. Computed metrics – activity time, energy consumed, steps, and travelled distance – are displayed on the smartphone's interface and synchronized to a cloud server.

2) Training dataset: To train the models for activity classification, steps detection and distance estimation, a dataset was collected from 44 persons and divided into train set and test set. The train set comprises 23 users (7 females, avg. 27.9 ± 4.5 y.o.) and the test set comprises 21 other users (5 females, avg. 25.4 ± 2.65 y.o.). All users engaged in walking (at 4 km/h and 5 km/h), running (at 7 km/h and 8 km/h), sitting, standing, lying, and tilting (only for the train set) with one GLC attached to the chest and another to the belt. We have also collected samples when the GLC was not used, e.g. while charging. The number of steps was manually annotated for walk and run activities, and the travelled distance was annotated based on the treadmill's screen. Overall, the train set was finally comprised of approximately 36h of data and the test set approximately 21h.

B. Benchmark validation

1) Devices: To assess GLC performance for activity classification and assess accuracy of estimated metrics, we designed an experimental protocol that directly compared GLC against top selling devices. After pre-testing several devices, the Fitbit Flex 2[3], Jawbone Up 24[4], Garmin Vivoactive HR[5], and Mi Band 2[6] were chosen for comparison (latest versions at the time of the study). All these devices can monitor walking and running activities[7], and present activity-related metrics such as duration of activity, number of steps, travelled distance or expended energy. Apart from GLC, none of the analysed devices distinguishes sitting or standing postures. All market devices were designed and implemented to be worn on the wrist, with the exception of GLC, that supports belt and chest usages. Each device may work differently from each other, especially for segmenting and classifying different activities. Some of the market devices are not able to classify activities automatically, only presenting accumulated data for the whole

[2] https://www.goliveclip.eu/

[3] https://www.fitbit.com/eu/flex2

[4] https://www.cnet.com/reviews/jawbone-up24-review/

[5] https://buy.garmin.com/en-US/US/p/538374

[6] https://www.mi.com/global/miband2/

[7] Garmin HR and Jawbone Up 24 do not distinguish walking from running.

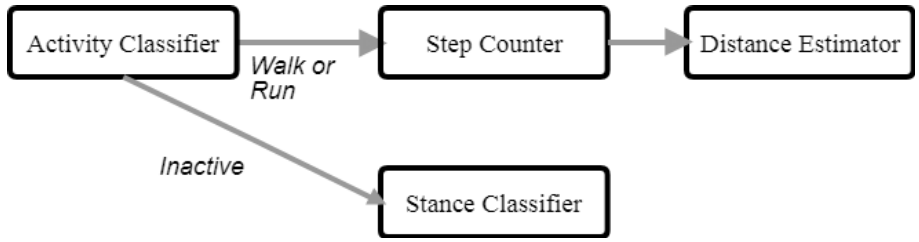

Fig. 1. Activity monitoring framework phases.

day, or require an activity to exceed at least a minimum amount of time in order to be recognized.

2) Experimental Protocol: The protocol was defined after consulting each device's specifications and after testing devices. The sequence of activities of the protocol included: 12 minutes standing, 12 minutes walking at 3.6 km/h, 12 minutes running between 5.5 and 6.5 km/h and 12 minutes sitting. During the experiment, each activity duration was measured using a chronometer, the number of steps was manually counted by visual inspection of the participant, and walking and running speeds were controlled by the treadmill machine. Overall, the dataset has almost 16 h (4 h of each activity). The total number of annotated steps was 22278 for walking and 35196 for running, and the total annotated distance was 13815 m for walking and 27125 m for running.

The protocol was conducted with 20 participants, all males, with an average age of 34.2 ± 12.19 years old, ranging from 24 to 73 years old. Personal information such as age, height, weight and dominant hand was collected and entered into each device's application (when applicable). Each participant wore one GLC on the chest and one GLC on the belt. Fitbit Flex 2, Jawbone Up 24, Garmin Vivoactive HR and Mi Band 2 were placed on the wrist of the users - two of them were placed in the right wrist and the remaining ones on the left wrist. The devices placed on each wrist were randomly attributed to each participant, in order to ensure variability across different users and devices.

III. Results

A. Wearable Activity Monitoring Training

The results obtained for the development of the activity classification hierarchical pipeline, using the test set, are presented in Table I.

The amount of samples collected for each activity was balanced to obtain equal number of class samples for each one of the classifiers, with the exception of *inactive* samples, which include samples of *still* and *not using*, and are represented in higher number. As observed at the top of Table I, the global accuracy for activity classification was 96.32%. Some *running* samples were misclassified as *tilting*, which could impair real-time performance during treadmill running.

Regarding the stance detection, at the bottom of Table I, all samples correctly classified as *inactive* by the first classifier (10190 samples) went on to the second classifier to be discriminated between *still* (which combines stand and sit

TABLE I
Confusion matrix for the activity (top) and stance (bottom) classifiers. Precision, Recall and Accuracy (Acc) in %.

	Real			
Predicted	**Walk**	**Run**	**Inactive**	**Precision**
Walk	**2293**	18	1	99.18
Run	41	**2166**	1	98.10
Inactive	70	71	**10190**	98.64
Tilt	92	229	34	-
Recall	91.87	87.20	99.62	Acc=96.32

	Real		
Predicted	**Still**	**Not using**	**Precision**
Still	**4814**	125	97.47
Not using	230	**5021**	95.62
Recall	95.44	97.57	Acc=96.52

postures) and *not using*, with very high accuracy (96.52%), an evidence of appropriate classification performance.

The results regarding steps detection and distance estimation were compared with the annotated values, and the accuracy of the step detection algorithm was 92.05%, considering 53783 annotated steps, while the accuracy of the distance estimation was 87.40%, in a total annotated distance of 39920 meters.

B. Benchmark Validation

Besides GLC, no other tested market device enabled the automatic classification of inactive periods; therefore its performance comparison against other devices cannot be performed. However, comparing GLC with the manual annotation with the chronometer revealed that the GLC failed to classify still for most of the users. We believe this was due to the system entering the battery optimization mode, if no motion is detected for a given period by the accelerometer[8]. During this state, the system will automatically go into sleep mode and no accelerometer data stream is provided to the smartphone, when there is no motion from the user.

As for the remaining activities, namely walking and running, all tested devices were able to identify and monitor metrics such as duration, number of given steps or travelled distance. The user interfaces are different for all devices, thus comparisons are restricted to the values presented on the table. Moreover, Mi Band 2 and Fit Bit Flex 2 can,

[8]https://www.invensense.com/products/motion-tracking/9-axis/mpu-9250/

TABLE II
ACCURACY (MEAN ± STANDARD DEV., %) FOR ACTIVITY CLASSIFICATION, STEPS AND DISTANCE ESTIMATION, FOR WALK (W) AND RUN (R).

	Activity classification			**Step detection**			**Distance estimation**		
	Walk	**Run**	**W+R**	**Walk**	**Run**	**W+R**	**Walk**	**Run**	**W+R**
GLC chest	96 ± 5	91 ± 18	95 ± 8	96 ± 5	90 ± 19	94 ± 10	89 ± 10	84 ± 17	88 ± 12
GLC belt	95 ± 5	94 ± 23	96 ± 9	94 ± 7	94 ± 23	94 ± 10	88 ± 12	84 ± 24	88 ± 15
GLC (avg. chest & belt)	95 ± 5	92 ± 21	95 ± 8.5	95 ± 6	92 ± 21	94 ± 10	88 ± 11	84 ± 21	88 ± 14
Fitbit Flex 2	93 ± 4	93 ± 4	98 ± 2	93 ± 7	89 ± 5	91 ± 5	-	-	-
Mi Band 2	89 ± 7	82 ± 13	88 ± 3	88 ± 16	90 ± 13	92 ± 6	82 ± 28	84 ± 13	89 ± 10
Garmin Vivoactive HR	-	-	72 ± 20	-	-	98 ± 2	-	-	50 ± 20
Jawbone Up 24	-	-	92 ± 12	-	-	95 ± 8	-	-	63 ± 21

automatically, distinguish walking from running while Jawbone Up 24 cannot (it accumulates data over the whole day without distinguishing different activities). The available information about the Garmin Vivoactive HR also claimed that it should automatically differentiate activities such as walking or running, even though this scenario was not observed in this trial (it usually aggregated walking and running as a single activity). When a device failed to detect an activity or to calculate the related features, the sample was also considered for analysis, and the error was included in its performance.

Having this in mind, and in order to provide the best possible comparison among the selected devices, an average accuracy for estimated total activity duration, number of given steps and travelled distance was calculated for the walking activity, running activity and the combination of both. Results are presented in Table II.

IV. DISCUSSION AND CONCLUSION

The results of the offline framework validation for automatic activity and stance classification achieved very high accuracy on an independent test set with data from users which did not integrate the training set. Reporting an accuracy of over 96%, our framework appears to outperform the position-independent solution of [11] in accuracy and the baseline (not position-aware) method of [12] in terms of precision and recall for the activities of interest. Step detection performance also proved to be very accurate (92.05%) with respect to the annotated ground truth. Distance estimation accuracy was, as one can expect, slightly lower (87.40%), due to the dependence on the previous stages of classification and step detection which resulted in accumulated errors.

The benchmark validation shows that GLC performs better than the remaining tested devices in walking activity classification, steps detection, and distance estimation. Fitbit Flex 2 was the only device which outperformed GLC in running (and, consequently, the joint combination of walking and running) classification. Further and more direct comparisons could not be performed, since Garmin HR and Jawbone Up 24 do not automatically distinguish these activities, nor do they present the same type of extracted metrics per activity, and the Fitbit Flex 2 does not calculate travelled distances. Table II also shows that running-related metrics present higher standard deviations than those related to walking, as there were more misclassified samples of the first. This conclusion was also supported by the offline validation results, on top of Table I.

Comparing the activity classification performance of the tested market devices, Fitbit Flex 2 revealed the highest accuracy, followed by GLC and, then, Jawbone Up 24. Garmin HR was the least accurate device, as it frequently failed to detect either walk or run. This device requires a minimum of 10 consecutive minutes of activity in order to automatically classify and consider such activity[9]; since the protocol only included 12 min of each activity, failing to detect the 10 minutes of activity is not improbable and it might have hindered its performance.

Regarding steps detection, Garmin HR was the most accurate device compared to manual annotation, immediately followed by Jawbone Up 24 which has been showing appropriate performance against other market devices in step detection ability [16]. In distance estimation, the accuracy of Garmin HR and Jawbone Up 24 was very low compared to GLC; Mi Band 2 revealed the highest accuracy, comparable to GLC. The distance estimation accuracy of Garmin HR was also likely influenced by its high error of activity classification.

GLC's performance results across the supported chest and belt positions reveal similar performance for both sensor placements, an evidence of coherence when we claim position-independence. Overall, GLC demonstrated better performance for both chest and belt positions over the remaining tested devices, worn on the wrist. As acknowledged by the prior state of art [1], sensor positioning on the body plays a crucial role in the accuracy of activity classifiers. This is why this work implemented a thorough validation protocol, that enables a fair comparison of these particular devices.

Besides the conceptual contribution of this work, in suggesting a position-independent framework for activity classification, this paper contributes to the understanding of classification performance and travelled distance estimation of wearable devices from the market. Prior work often focused on comparing how solutions failed in counting steps, but activity classification and distance estimation were not thoroughly investigated [17].

Even though this and previous works considered different device positions, the issue of position-independent physical activity monitoring is not fully addressed yet [1], and the

[9]https://support.garmin.com/en-US/?faq=zgFpy8MShkArqAxCug5wC6

position of the device has proved to be indeed relevant to the system's accuracy. Thus, as future work, we foresee the inclusion of the wrist position in the activity monitoring framework to assess the impact of considering this extra position in the discussed metrics, and increase the number of body positions supported by the system.

ACKNOWLEDGMENTS

The authors thank all participants from this study, as well as all research scientists, who made the study possible.

REFERENCES

[1] J. Qi, P. Yang, A. Waraich, Z. Deng, Y. Zhao, and Y. Yang, "Examining sensor-based physical activity recognition and monitoring for healthcare using internet of things: A systematic review," *J Biomed Inform*, 2018.

[2] B. Almaslukh, A. Artoli, and J. Al-Muhtadi, "A robust deep learning approach for position-independent smartphone-based human activity recognition," *Sensors*, vol. 18, no. 11, p. 3726, 2018.

[3] J. Saha, S. Chakraborty, C. Chowdhury, S. Biswas, and N. Aslam, "Designing device independent two-phase activity recognition framework for smartphones," in *WiMob*, pp. 257–264, IEEE, 2017.

[4] A. Anjum and M. U. Ilyas, "Activity recognition using smartphone sensors," in *2013 IEEE CCNC*, pp. 914–919, Jan 2013.

[5] S. Reddy, M. Mun, J. Burke, D. Estrin, M. Hansen, and M. Srivastava, "Using mobile phones to determine transportation modes," *ACM TOSN*, vol. 6, no. 2, p. 13, 2010.

[6] R. Mohamed, M. N. S. Zainudin, M. N. Sulaiman, T. Perumal, and N. Mustapha, "Multi-label classification for physical activity recognition from various accelerometer sensor positions," *J. Inf. Commun. Technol*, vol. 18, pp. 209–231, 2018.

[7] A. M. Khan, M. H. Siddiqi, and S.-W. Lee, "Exploratory data analysis of acceleration signals to select light-weight and accurate features for real-time activity recognition on smartphones," *Sensors*, vol. 13, no. 10, pp. 13099–13122, 2013.

[8] D. Shi, R. Wang, Y. Wu, X. Mo, and J. Wei, "A novel orientation- and location-independent activity recognition method," *PERS UBIQUIT COMPUT*, vol. 21, no. 3, pp. 427–441, 2017.

[9] S. A. Antos, M. V. Albert, and K. P. Kording, "Hand, belt, pocket or bag: Practical activity tracking with mobile phones," *Journal of neuroscience methods*, vol. 231, pp. 22–30, 2014.

[10] P. Siirtola and J. Röning, "Ready-to-use activity recognition for smartphones," in *IEEE Symposium on CIDM, 2013*, pp. 59–64, IEEE, 2013.

[11] A. M. Khan, Y.-K. Lee, S. Lee, and T.-S. Kim, "Accelerometers position independent physical activity recognition system for long-term activity monitoring in the elderly," *MBEC*, vol. 48, no. 12, pp. 1271–1279, 2010.

[12] T. Sztyler and H. Stuckenschmidt, "On-body localization of wearable devices: An investigation of position-aware activity recognition," in *IEEE International Conference on PerCom, 2016*, pp. 1–9, IEEE, 2016.

[13] A. Pereira and F. Nunes, "Physical activity intensity monitoring of hospital workers using a wearable sensor," in *PervasiveHealth'18*, 2018.

[14] B. Aguiar, J. Silva, T. Rocha, S. Carneiro, and I. Sousa, "Monitoring physical activity and energy expenditure with smartphones," in *IEEE-EMBS BHI*, pp. 664–667, June 2014.

[15] C. Figueira, R. Matias, and H. Gamboa, "Body location independent activity monitoring," in *BIOSIGNALS*, 2016.

[16] J. J. Chow, J. M. Thom, M. A. Wewege, R. E. Ward, and B. J. Parmenter, "Accuracy of step count measured by physical activity monitors: The effect of gait speed and anatomical placement site," *Gait & posture*, vol. 57, pp. 199–203, 2017.

[17] K. R. Evenson, M. M. Goto, and R. D. Furberg, "Systematic review of the validity and reliability of consumer-wearable activity trackers," *IJBNPA*, vol. 12, no. 1, p. 159, 2015.