2019 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom)

# Multi-Sensor Information Fusion and Machine Learning for High Accuracy Rate of Mechanical Pedometer in Human Activity Recognition

Michael Adjeisah
School of Information Science and Technology
Donghua University
Shanghai, China
madjeisah@hotmail.com

Guohua Liu
School of Information Science and Technology
Donghua University
Shanghai, China
ghliu@dhu.edu.cn

Douglas Omwenga Nyabuga
School of Information Science and Technology
Donghua University
Shanghai, China
dnyabuga@gmail.com

Richard Nuetey Nortey
School of Information Science and Technology
Donghua University
Shanghai, China
rn.nortey@yahoo.com

*Abstract*— **Many Human Activity Recognition (HAR) research works have been engineered, however, only a few studies combine Kinect and the smartphone. Also, most findings are geared towards wearable sensors, which however may develop wearer discomfort, predominantly for nonstop action applications where one may perhaps have to wear for a lengthy period. Additionally, while the novelties of these researches are encouraging, accuracy is subject to fluctuation over time. To address these concerns, this paper reports the fusion of sensor data to determine a robust accuracy. The system uses Microsoft Kinect Sensor (MKS) to detect angle change in the hip-knee-ankle and a gravitational accelerometer of a smartphone for step tracking while engaging users. In our study, three methods are compared: 1) Manually counted steps, 2) raw data from our algorithm and 3) implementation of parameterized classification algorithm classifiers, k-Nearest Neighbors (*k*NN) to improve the step counts. Best-first search hierarchy containing the maximum distance of the *k*NN algorithm is engaged. Varying *k*, an accuracy rate of 96.44% is achieved with $\sigma = 0.040$.**

***Keywords—Accelerometer; activity recognition; Kinect; parameterized learning; mechanical pedometer***

## I. INTRODUCTION

HAR is a dynamic study field in which approaches for understanding human activities are established by simply inferring features extracted from motion, location, physical signals and environmental information[1]. Most mobile phones[2]have inbuilt sensors like GPS, accelerometer, orientation, gravity, and rotation sensing capabilities, etc. This enables them to sense individual activities and assists in performing functions like location and forecast of environmental conditions. The footstep pawns technically referred to as mechanical pedometers consist of sensors that can enumerate the human motion[3]. Kinect has a device capability to precisely measure all your body proportions which engages randomize decision forest to perform an action in real time[4]. Its tracking technology has the capability of sensing up to six participants and 25 skeletal joints per person, comprising innovative joints of hand tips[5], [6]. If combined with other sensors, one can create a wide-ranging purpose and better scenarios. These sensors (Kinect and smartphone) are characterized by their capability of generating raw data[7] for detecting and gathering HAR information. While the ubiquitous nature and cutting-edge sensor technologies of the smartphones and the Kinect sensor devices makes them extremely suitable for in-the-wild research, the accuracy of human labelled submission remains underexplored.

This work was supported in part by Science and Technology Commission of Shanghai Municipal under Grant 17YF1427400 and in part by the Fundamental Research Funds for the Central Universities under Grant 17D111206.

Our main aims is to present a set of high quality research results reporting the current state-of-the-art of human accuracy using Kinect and smartphone. Hence the system applies a contextual information obtained through these commodity sensors for a robust data accuracy with a twist of custom algorithms and raw-data processing. In addition, we engaged *k*NN for additional steps that seem intuitive but not automatically detected by the accelerometer of the phone while tracking the steps. The *k*NN is not just the nearest neighbor but alternatively pruning of the best ***k***[8] that engages a maxima bound conforming to the extreme likelihood distance at which the nearest neighbor is certain to be initiated[9]. In this limited-scope application, we demonstrate that it is possible to combine these two sensors for effective mechanical pedometer for human action recognition model. Nevertheless, the system delivers a computationally well-organized approach which permits real-time data acquisition. The remaining part of this paper is organized as follows: Section 2 reviews various related works on Kinect-based human action recognition, joint tracking, and smartphone-based recognition approaches. Section 3 presents our proposed systems and a thorough description. Experimental results on the aforementioned framework are covered in Section 4 and we concluded the paper with the recommendation and future work in Section 5.

## II. RELATED WORK

Many research works sought to use machine learning and smartphone or Kinect sensor for their findings; nonetheless, only few combines the two. Meanwhile, when combined, can create wide-ranging capabilities on HAR with a higher accuracy and precision of real data acquisition to quantify human motion. One of the most recent attempts is the introduction of a face recognition coordination which was directed at assisting the visually impaired and blind in real-time[4]. The approach harbors the removal of Kinect's casket and used as a wearable device to implement face detection. To identify the person, a virtualized 3 dimensional (3D) estimate is developed; the system uses time-based consistency alongside a simple biometric technique to produce a sound linked to the location. User experience evaluation of the

978-1-7281-4328-6/19/$31.00 ©2019 IEEE
DOI 10.1109/ISPA-BDCloud-SustainCom-SocialCom48970.2019.00152

system was performed in the dark using depth stream information. The best accuracy rate was 56.23% (σ = 0.0148). Although the novelty and results of the findings are encouraging, however using the Kinect sensor as a wearable device may develop a weighty circumstance. Gaglio et al[10], in their paper, offered an approach for HAR by means of stream information recognition by an RGB-D camera. The method was based on the assessment of some applicable joints of the human anatomy using Kinect. Three different machine learning models were engaged; K-means clustering, support vector machines (SVM), and hidden Markov models (HMK), to practically sense the postures involved whereas engaging user to perform an activity. They further classify and modelled each activity as a spatiotemporal progress of recognized postures. Qiang et al[11], presented a queuing detecting system to support in managing to queue. QueueSense, as it is called, took advantage of the phone sensor to identify multi-queue lines and offers a standby time estimate for their clients. The system performed automatically with energy-efficiency, and accurate queuing detection connected to a cloud-based server for data collection of customer's information.

Mazomenos et al[12], presented an algorithm for the recognition of three fundamental upper limb activities using double MARG sensors. The MARG was attached to the wrist and elbow, from which they computed the joint angles position of the forearm and the upper arm by means of data fusion with respect to a quaternion-based gradient-descent approach and a double-link model of the upper limb. An inbuilt technology and sensors of a robot that allows precise quantification of motion gesticulations and agitations were reported by Hu et al[13]. The study was geared towards a duo crew of users: stroke recovery subjects and healthy active participant. The purpose was to derive portions in relation to upper limb performance and high spot changes in locomotion activities of an individual in medical rehabilitation. Even though the domino effect were encouraging, research on lower limb remain uncovered.

Evidently, only few studies contributed to the domain of fusion Kinect with a smartphone sensor[14]–[17]. Kinect and Bluetooth were combined[18] to build a smart conference system. The skeleton tracking of Kinect was activated to detect each person's skeleton with multi-function, and a tailored Bluetooth supported kit was actively used to recognize each member. A survey of Health Sensing by Wearable Sensors and Mobile Phones by Song et al[19], also outlines the important foundation of technologies in health rehabilitation sensing using a smartphone and wearable sensor. Although these findings have yielded tremendous results, we believe much are yet to be discovered.

## III. System Design Description

A general representation of the technical module of our proposed mechanical pedometer for activity recognition system and its principal components is depicted in figure 1. The system uses multi-sensor in terms of five technical modules, namely user engagement, feature extraction, machine learning model, action recognition and feedback/evaluation modules. The system takes participant action sequence captured by Kinect and the smartphone as input. The engagement of the user is made of two parts; Kinect and smartphone processing. While Kinect directly adopts the joint position prediction to obtain a structure of joints estimated skeletons, a custom algorithm is processed to communicate with accelerometer data of the smartphone. Feature extraction is processed in the next part based on the pose estimation and raw data of defining features. Our finding is specialized for the support of weight and the adaptation to the gravity of a user, therefore, the whole skeleton joint of the Kinect sensor was not activated. Attention is drawn to the lower limb of the human anatomy. Parameterized classification kNN, SVM, and K-means algorithm is carefully studied for more HAR processing and action classification is followed up through intersection hit testing for joints.

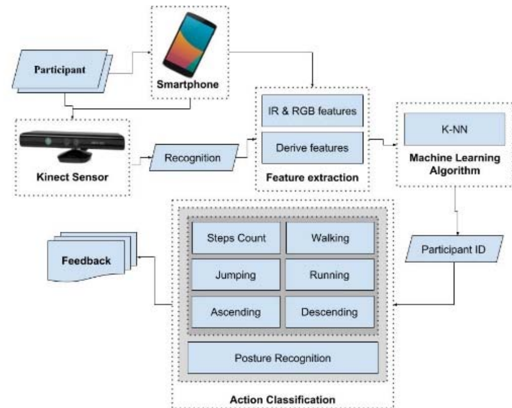



Fig. 1. Technical modules of the proposed human action detection system

## IV. Action Detection Methodology

### A. Stage One – Data Acquisition with Smartphone

To obtain raw data from the sensor, we have developed an algorithm based on MATLAB Support Package for Android OS Sensors. We initiate the process with the acquisition of the inbuilt sensor signals mainly accelerometer for steps counts. The accelerometer is made up of the x, y, and z coordinates where z hovers around 10.00 miles per seconds square. Based on the knowledge of the support library and how to deal with data acquisition via these sensors, we set up our MATLAB to connect with the smartphone.

The major process of data acquisition via the gravitational acceleration data (GAD) models is as follows. Given an input, we log and process data and stored in a two dimensional (2D) array [$u$, $v$] for the log and v is process data. The algorithm was made to compare the adjacent value with the current value when the acceleration becomes active. If the current value is greater than 1, a peak is considered. This was set to a local maximum height above 1 standard deviation. As a result, we set a threshold to a value that determined whether the value is high or low to be considered a peak. This was tuned experimentally to match a participant level of movement while performing an action. Iteration is set to loop through the entire samples to test whether the local maximum height above 1 standard deviation is greater than the minimum based on the following conditions; 1) to make sure it is greater than the threshold against the previous value and, 2) to make sure it is greater than the sample that comes after it. While computing the peak to the actual value, a time value at which a peak index occurs is also recorded.

### B. Stage Two – Human Activity Detection via Kinect

The skeleton tracking system of the Kinect Xbox one is composed of 25 joints and the orientation of respective joint is estimated with matching to confidence score. A single joint has 3 values: X, Y, and Z in a Cartesian coordinate system which is given in the real world coordinates with X directing to the horizontal axis (right), Y to the vertical axis (up), and Z aiming in the path with increasing depth with respect to MKS[20], [21].

In our case, all the 25 joints coordinate system of the Kinect is not used; attention is drawn to the lower limb of the human anatomy. The human leg is particularly for the sustenance of mass, adjusting to gravity, standing, and all forms of Kinesis including recreational such as dancing, running and establish a substantial ration of a person's mass[12], [22]. The approach is to calculate the angle between different tracking joint. The lower limb of Kinect skeleton engine is made up of nine skeleton points and three of these can be used to form a triangle. From the lower limb skeleton data, the knee becomes the center of the coordinate system, the ankle and waist/ hip joints establish the defining point of the angle as illustrated in Fig. 6(**left**). In Euclidean space, a Euclidean vector is a geometric entity which has equally a magnitude and a direction[23]. A vector can be visualized as an arrow in 6 (**right**). The magnitude is its length, and the bearing is the direction to which the arrow is pointed. A condition where the magnitude of a vector is p is represented by $\| p \|$. The scalar product of two Euclidean vectors ***p*** and ***q*** is defined by

$$p \bullet q = \| p \| \| q \| \cos(\theta) \tag{1}$$

where θ is the angle between ***p*** and ***q***.

Precisely, the angle between ***p*** and ***q*** is $90^{\circ}$ if they are orthogonal and $p \bullet q = 0$. On the other hand, if they are co-directional, the angle between them is $0^{\circ}$ and $p \bullet q = \| p \| \| q \|$. This infers that the dot product of a vector p with itself is $p \bullet q = \| p \|^2$ which gives $\| p \| = \sqrt{p \bullet q}$, based on the formula for the Euclidean length of the vector. The Euclidean distance between points, p and q is the length of the line segment connecting them $(\overline{pq})$. Assuming p and q are angular distances which in general, only gives the magnitude of the angle among two lines irrespective of the viewpoint, can be calculated from the dot product of the two line vectors that form the angle by;

$$\begin{aligned} p \bullet q &= pq\cos\theta \\ \cos\theta &= \frac{p \bullet q}{pq} = \frac{p_x q_x + p_y q_y + p_z q_z}{\sqrt{p_x^2 + p_y^2 + p_z^2}\sqrt{q_x^2 + q_y^2 + q_z^2}} \\ \theta &= \cos^{-1}\left( \frac{p_x q_x + p_y q_y + p_z q_z}{\sqrt{p_x^2 + p_y^2 + p_z^2}\sqrt{q_x^2 + q_y^2 + q_z^2}} \right) \end{aligned} \tag{2}$$

The return value of the inverse $\theta$ function ranges from $0 \le \theta \le \pi$, with a unit in radian (rad). Knowing the coordinate of each point implies that we know the length of each side but not the angle values using two vector quantities in equation 3. This is due to the fact that the sensor can generate points (joints), not angles.

$$V_{final} = V_{initial} + \Delta V \tag{3}$$

Equation (3) can be rewritten as;

$$\Delta V^2 = V_{initial}^2 V_{final}^2 - 2V_{initial} V_{final} \cos\theta, \tag{4}$$

Based on the original law of cosines. Shown in Fig. 2(**left**), is the Da Vinci Vitruvian Man and the possible position where angle change detection can be computed, (**middle**) the skeleton coordinate system and joint definition in the lower limb and the geometry of the hip, knee, and ankle used in our work (**right**).

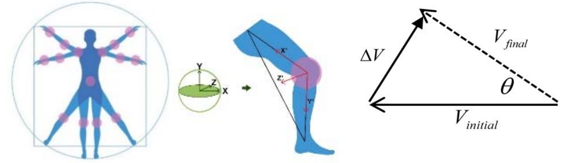



Fig. 2. (**left**) Possible positions where angle change detection can be computed in the lower limb. (**middle**) Is the original Cartesian coordinate system with axes [X, Y, and Z] rotated to the object-view coordinate system with axes and (**right**) is the representation of the elevation in geometry.

Figure 6b shows an example of the lower limb joint as the reference point, the axes of the original Cartesian coordinate system [X, Y, and Z] and the new axes of the object view coordinate system are notated as $[X'Y'Z']$. This can be represented by the elevation θ and the geometry of the hip-knee-ankle calculated with respect to the $X'-Y'$ plane and $Z'$ direction, respectively. Based on this, a matrix M is derived from the relationship between the original coordinate system of the joint orientation quaternion of the lower limb as;

$$[M][XYZ] = [X'Y'Z'] \tag{5}$$

which means M can be computed as

$$[M] = \begin{bmatrix} t_{11}t_{12}t_{13} \\ t_{21}t_{22}t_{23} \\ t_{31}t_{32}t_{33} \end{bmatrix} = [X'Y'Z'][XYZ]^{-1} \tag{6}$$

With respect to the axes, Euler angles (α, β, γ), for sequential rotation of the original coordinate system of M is

$$M(\alpha,\beta,\gamma) = M_{X(\alpha)} M_{Y(\beta)} M_{Z(\gamma)}. \tag{7}$$

where, α, β, γ can then be computed as[25]

$$
\begin{aligned}
\alpha &= \tan^{-1}\left(t_{21}/t_{11}\right) \\
\beta &= \tan^{-1}\left(-t_{31}\Big/\sqrt{t_{32}^2+t_{33}^2}\right) \\
\gamma &= \tan^{-1}\left(t_{32}/t_{33}\right).
\end{aligned}
\tag{8}
$$

During the setup, we placed the Kinect sensor near the edge of a flat, stable surface, mounting it 6 feet (1.8m) off the floor. The ToF can physically sense at a distance of 26 feet (8m), nevertheless, 14 feet (4.5m) is where you can reliably track body joints. Anything beyond 4.5m yields inconsistent results in body tracking. It also has a standard minimum operating distance of 2 feet (0.6m). Accordingly, a minimum of 0.6m and maximum of 4.5 distances were considered to avoid any drawback of our presented system regarding some of the activities. Also, the FoV is 70.6 degrees horizontally and 60 degrees vertically, therefore practicing different activities like running and walking were expressed in controlled scenario by gym facilities and activities like sitting, standing, upstairs and downstairs in real scenario.

## V. Experimental Results

This part presents the experimental results and findings on both the Kinect and the GAD and how the two were successfully fused in our research work as well as user experience of our system in total darkness.

### A. User-Experience Experiments

The primary experiment performed involved testing of the system within measured experimental environments. This was done to assist us not only to discover critical difficulties concerning issues relating to usability or bugs but also to find out information on how to improve the experiment further. In order to realize our goals of formalizing some of the partaker's responses emanating from the pilot testing, design, and usability heuristics were engaged accordingly. The test was set up as follows;

- A brief introduction to how the system operates before engaging 10 participants. They include 2 kids, 5 teenagers, and 2 adults and an old man.
- Among the 10, 2 of them were made to act weirdly, running so fast, combining multiple actions (running and walking/standing) at the same time and also moving in the Z direction of the Kinect that causes partial loss of the knee angle.
- The participants were engaged for about 1 to 2 min in a series of rounds depending on their strength capacity, especially for the old man.

This led to another experiment after the changes with fewer participants and only two rounds at this point. Below is a sample of steps and peaks plotted graph of a participant based on Kinect recognition. In the pilot testing, Kinect recognizes the action as walking, which is an activity performed by the user but few peaks were recorded in figure 3(**left**) whilst in (**right**) the running recognition is also fired with few steps and counts.

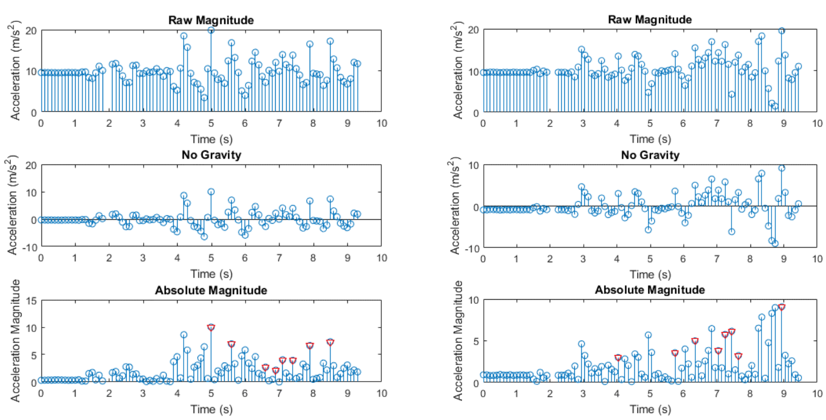



Fig. 3. Steps count data from GAD upon Kinect action recognition; few peaks recorded in our pilot test. The raw magnitude is to visualize the general changes in acceleration. Subtracting the mean from the raw data removes the gravity and finally, the local max maximum height above 1 standard deviation is treated as a peak

### B. Accuracy Performance Experiments

In the experiments, we notice a series of events like the number of steps tracked not accurate when compared to manual counting of steps which in many cases, is a characteristic of mechanical pedometers. This, therefore, is the reason for the application of machine learning algorithm to improve the number of peaks as well as the system's functionality. We engaged K-means clustering, ***k***NN and SVM but ***k***NN presented much more accurate and is known to be a good fit when processing distance information as shown in the figure 4. KNN is shown to be superior when compared to SVM and K-means. Therefore we engaged ***k***NN for additional steps that seem intuitive but not automatically recognized by the accelerometer of the smartphone while step recording. The features used to apply ***k***NN model were data collected from the smartphone. Values from the smartphone are labeled and given a target class of 0s and 1s (1 for 'peak' and 0 for 'no peak'). For example, in a 60 seconds engagement of a participant, data generated from the smartphone were used.

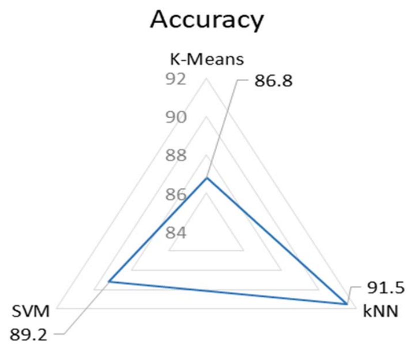



Fig. 4. Machine learning performance on the dataset to choose a good fit when processing distance information.

In performing the ***k***NN classifier, we iterate many models using different k, value sizes from 1 to 50 and plot out the error rate to see which of the values have the lowest error rate. A grid search for alternative pruning of the best hyper-parameter ***k*** for the *k*NN, with a variation of ***k*** from 1 to 50 was conducted. Accuracy performance of 96.44% with standard deviation of 0.040 is attained for ***k*** = 40 by means of random state variable (rSV) 101. We compared with different ***k*** values and rSV by parametric model, and the next closest ***k*** = 6, 7, 8, 11, 12, 13, and 18 with a random state of 45. Choosing ***k*** = 18, an accuracy rate of 93.5% with σ = 0.056 is attained. See figure 5. A grid search for alternative pruning of the best hyperparameter ***k*** resulted a 5% increment in accuracy.

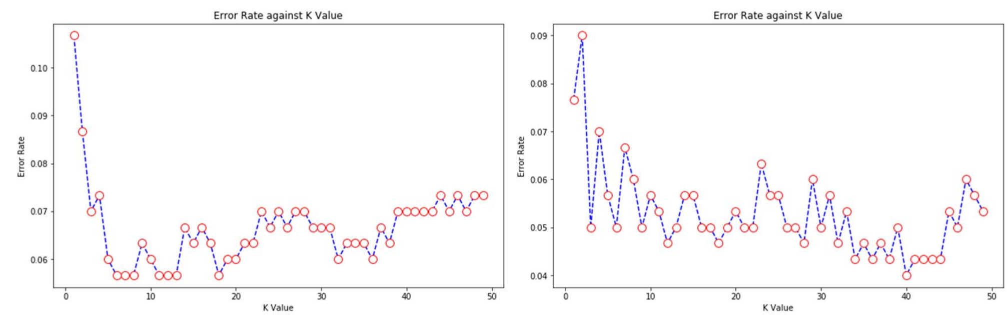
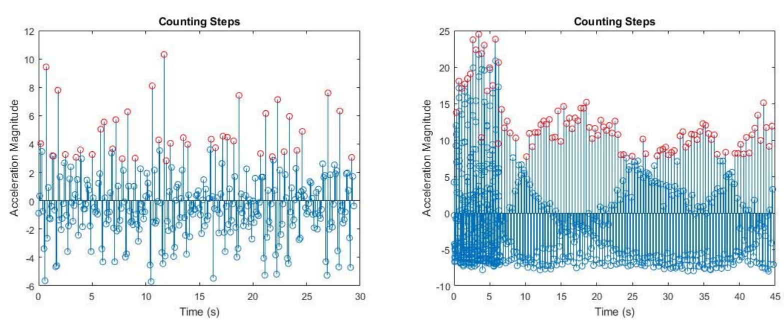



Fig. 5. Choosing the best ***k***-value with randomizes state of parameter adjusting to the hyperparameter variation. (**left**) Denotes an error rate against ***k*** value with a default rSV of 45. (**right**) A grid search for alternative pruning of the best hyperparameter ***k*** (where k = 40) for kNN, with performance of 96.44% and σ = 0.040 . Work on the background color is done

We replot the figure (6 below) to compare the activities performed and the number of peaks recorded in 30, 45, and 60 seconds to that of figure 3 in our pilot test. Clearly, there is a vast improvement in the number of step counts after the application of kNN. The figures are peaks recorded in 30, 45, and 60 sec for a running activity. As mentioned earlier on, in order not to get a small variation, a minimum magnitude is set and only tall peaks are considered. Peaks are only considered if the value is greater than 1. Comparing the first three figures to the last which is for a sitting activity, one can see red marks which are supposed to be recorded as peaks but got ignored by the model since it does not satisfied the criteria of the model.



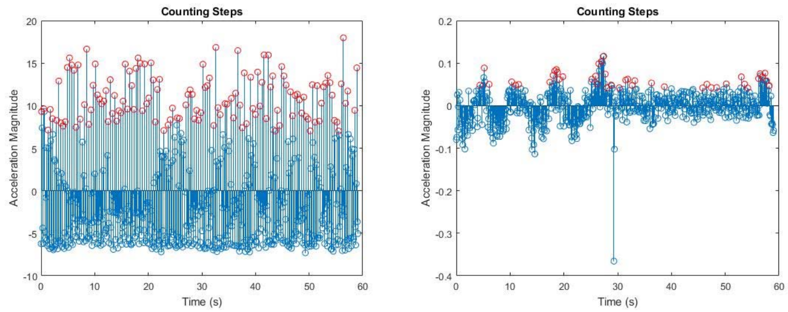



Fig. 6. Peak detection upon implementation of grid search for the best k; the first figures are peaks recorded in 30, 45, and 60 sec for a running activity. In order not to get a small variation, a minimum magnitude is set and only tall peaks are considered. The last is for a sitting activity.

To fuse information obtained from the sensors together, we engaged a time value to be able to identify values which are considered as peak and labeled them as the target class for the kNN while engaging the sensors simultaneously. The time value as a results helped in obtaining features from each sensor separately. The data from the sensor were then fused together for both tasks which finally resulted in accuracy improvement after application of kNN.

The process records the number of steps at a time value during a peak occurrence when Kinect fires an action recognition protocol. Therefore a T-pose is used as a default pose to track participant into the Kinect receptive field and initiate the accelerometer of the smartphone simultaneously. After a delivery of brief directives on how the system works, we mark a contextual center of which the participant could get tracked to the field of view, on T-pose. We demonstrate a correct movement to the user in terms of sideways (x and y-direction) and participants were made to perform each activity in an estimated time of 30, 45 and 60 seconds. Below are the results of the number of steps counted by the smartphone upon Kinect recognizing the action performed in Table I. This is basically peaks recorded in raw data and after applying ***k***NN. As mentioned, even though data were generated for sitting and standing activities, no peak were recorded since criteria was not satisfied.

TABLE I. COMPARISON OF STEP COUNT ON RAW DATA (PEAK) AND MACHINE LEARNING IMPLEMENTATION (KNN)

| Time/s / Actions | 30 sec | | 45 sec | | 60 sec | |
|---|---|---|---|---|---|---|
| | peak | kNN | peak | kNN | peak | kNN |
| Standing | - | - | - | - | - | - |
| Walking | 39 | 52 | 57 | 80 | 93 | 110 |
| Running | 83 | 85 | 97 | 120 | 131 | 154 |
| Upstairs | 35 | 41 | 49 | 78 | 83 | 102 |
| Downstairs | 34 | 45 | 53 | 72 | 79 | 107 |
| Sitting | - | - | - | - | - | - |

We expressed each activity results in terms of precision and recall measured according to the “participant” scenario, that is, by performing each activity in an estimated time from which data were collected, and analyzed.

TABLE II. EXPRESSION OF STEP COUNTS ACTIVITIES IN TERMS OF PRECISION AND RECALL MEASURED ACCORDING TO THE “PARTICIPANT”

| Activity | Time (sec) | “participant” | |
|---|---|---|---|
| | | Precision | Recall |
| Walking | 30 | 98.1 | 95.9 |
| | 45 | 99.2 | 96.8 |
| | 60 | 99.6 | 98.7 |
| | **Average** | **98.9** | **97.1** |
| Running | 30 | 92.4 | 89.3 |
| | 45 | 93.7 | 91.4 |
| | 60 | 96.0 | 94.2 |
| | **Average** | **94.0** | **91.6** |
| Standing | 30 | 98.6 | 97.4 |
| | 45 | 98.9 | 98.6 |
| | 60 | 99.7 | 99.4 |
| | **Average** | **99.1** | **98.5** |
| Sitting | 30 | 98.9 | 98.4 |
| | 45 | 99.4 | 99.1 |
| | 60 | 99.8 | 99.6 |
| | **Average** | **99.4** | **99.0** |
| Upstairs | 30 | 92.3 | 92.8 |
| | 45 | 94.3 | 93.7 |
| | 60 | 96.1 | 95.8 |
| | **Average** | **93.2** | **94.1** |
| Downstairs | 30 | 90.0 | 90.5 |
| | 45 | 92.6 | 91.9 |
| | 60 | 93.4 | 93.8 |
| | **Average** | **91.7** | **92.1** |
| **Overall Average** | | **96.4** | **95.4** |

Comparing the activity precision and recall from Table II, one can see clearly that an estimated time of 60 seconds records the highest accuracy. Sitting recorded the highest accuracy followed by standing since the angle change detection in these postures are stable, hence a stable calibration of the Kinect. The overall precision and recall of our approach h are 96.44% and 95.40%, respectively.

To perform user experience in the dark, we alternatively used IR image stream acquired with Kinect instead of RGB for action recognition built on the original method. Thus, we switched to what we call activity recognition in dim environs with IR image stream. The overall accuracy rate for Kinect in the dark is 71.4%. Comparing such value with Neto et al[4], with 56.23% accuracy, we can declare a superior user experience with an accuracy of 15.17% higher. Figure 7 shows RGB plots of some activities performed with Kinect.

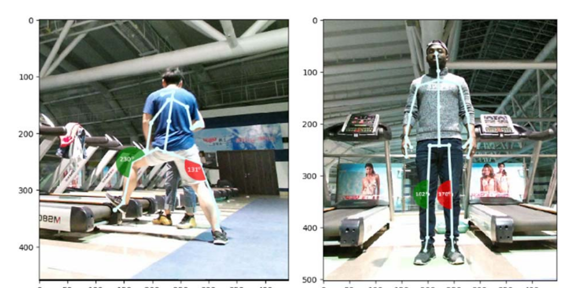


Fig. 7. Kinect receptive field images showing angle change detection in the hip-knee-ankle.

Also, figure 8 shows HAR in the dark via IR image stream in the dark environ. The left is a standing activity and a T-pose fired by Kinect and the right side is a sitting activity.

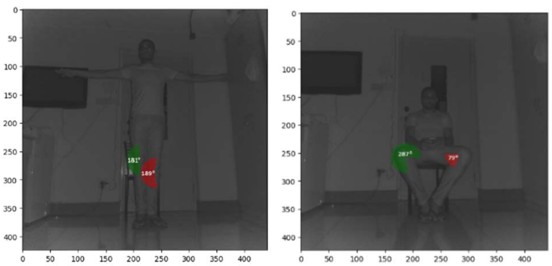



Fig. 8. Activity Recognition using IR-only information

## CONCLUSION

In this paper, we have shown that the ubiquitous nature and cutting-edge sensor technologies of the Kinect and the accelerometer of the smartphone makes them extremely suitable for obtaining contextual information to determine mechanical pedometer data accuracy. Our proposed system uses Microsoft Kinect sensor to detect angle change in the hip-knee-ankle and a gravitational accelerometer of a smartphone for step tracking while engaging users with a consistent performance. With parameterized classifier, an accuracy rate of 96.44% is achieved with $\sigma = 0.040$. The study successfully improved the steps tracked in all the activities performed with classification algorithm that learn patterns from their distance. Also, an overall accuracy rate of 71.4% is achieved from user experience tests conducted with IR image in the dark.

Information extracted from lower limb and the accelerometer of the smartphone would in the near future be used for Human Activity Recognition with a comparative simulation results involving the University of California (UCI) HAR dataset.

## REFERENCES

[1] J. L. Reyes-Ortiz, A. Ghio, X. Parra, D. Anguita, J. Cabestany, and A. Català, “Human Activity and Motion Disorder Recognition: towards smarter Interactive Cognitive Environments,” ESANN, 2013.

[2] N. D. Lane, E. Miluzzo, H. Lu, and D. Peebles, “A survey of mobile phone sensing,” Commun. Mag. IEEE, vol. 48, no. 9, pp. 140–150, 2010.

[3] Y. Jiang et al., “A personalized mobile sensing system for indoor air quality monitoring,” in Proceedings of the 13th international conference on Ubiquitous computing - UbiComp ’11, 2011, p. 271.

[4] L. B. Neto et al., “A Kinect-Based Wearable Face Recognition System to Aid Visually Impaired Users,” IEEE Trans. Human-Machine Syst., vol. PP, no. 99, pp. 1–13, 2017.

[5] A. Nava, L. Garrido, and R. F. Brena, “Recognizing Activities Using a Kinect Skeleton Tracking and Hidden Markov Models,” in 2014 13th Mexican International Conference on Artificial Intelligence, 2014, pp. 82–88.

[6] M. Adjeisah, Yi Yang, and Lian Li, “Joint Filtering: Enhancing gesture and mouse movement in Microsoft Kinect application,” in 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), 2015, pp. 2528–2532.

[7] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, “Activity recognition using cell phone accelerometers,” Acm Sigkdd Explor. Newsl., vol. 12, no. 2, pp. 74–82, 2011.

[8] H. B. T.-I. C. on I. A. and P. Samet 2003.proceedings, “Depth-first k-nearest neighbor finding using the MaxNearestDist estimator,” 2003, p. 486.

[9] H. Samet, “K-nearest neighbor finding using MaxNearestDist.,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 2, pp. 243–252, 2007.

[10] S. Gaglio, G. Lo Re, and M. Morana, “Human Activity Recognition Process Using 3-D Posture Data,” IEEE Trans. Human-Machine Syst., vol. 45, no. 5, pp. 586–597, 2015.

[11] Q. Li, Q. Han, and L. Sun, “Collaborative Recognition of Queuing Behavior on Mobile Phones,” IEEE Trans. Mob. Comput., vol. 15, no. 1, pp. 60–73, Jan. 2016.

[12] E. Mazomenos et al., “Detecting Elementary Arm Movements by Tracking Upper Limb Joint Angles with MARG Sensors.,” IEEE J. Biomed. Heal. Informatics, vol. 20, no. 4, pp. 1088–1099, 2017.

[13] M. C. Hu, C. W. Chen, W. H. Cheng, C. H. Chang, J. H. Lai, and J. L. Wu, “Real-Time Human Movement Retrieval and Assessment With Kinect Sensor,” IEEE Trans Cybern, vol. 45, no. 4, pp. 742–753, 2015.

[14] L. Kernec et al., “Multisensory Data Fusion for Human Activities Classification and Fall Detection,” in 2017 IEEE SENSORS, 2017, pp. 1–3.

[15] C. Jiang, M. Fahad, Y. Guo, J. Yang, and Y. Chen, “Robot-Assisted Human Indoor Localization Using the Kinect Sensor and Smartphones,” in 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2014, pp. 4083–4089.

[16] S. Moe and I. Schjolberg, “Real-time hand guiding of industrial manipulator in 5 DOF using Microsoft Kinect and accelerometer,” in 2013 IEEE RO-MAN, 2013, pp. 644–649.

[17] A. Akbari, X. Thomas, and R. Jafari, “Automatic Noise Estimation and Context-Enhanced Data Fusion of IMU and Kinect for Human Motion Measurement,” in 2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks (BSN), 2017, pp. 178–182.

[18] H. H. Hsu, C. Yi, Y. R. Chen, and T. K. B. T.-I. C. on N.-B. I. S. Shih, “Using Kinect to Develop a Smart Meeting Room,” 2013, pp. 410–415.

[19] L. Song, Y. Wang, J. J. Yang, and J. Li, “Health sensing by wearable sensors and mobile phones: A survey,” 2014, pp. 453–459.

[20] M. Camplani, T. Mantecon, and L. Salgado, “Depth-Color Fusion Strategy for 3-D Scene Modeling With Kinect,” IEEE Trans. Cybern., vol. 43, no. 6, pp. 1560–1571, 2013.

[21] A. Michael, Zhao Chen, Guohua Liu, and Yang Yi, “A new approach for tracking human body movements by kinect sensor,” in 4th International Conference on Smart and Sustainable City (ICSSC 2017), 2017, p. 15 (7 .)-15 (7 .).

[22] R. Colombo, I. Cusmano, I. Sterpi, A. Mazzone, C. Delconte, and F. Pisano, “Test-retest reliability of robotic assessment measures for the evaluation of upper limb recovery.,” IEEE Trans. Neural Syst. Rehabil. Eng., vol. 22, no. 5, pp. 1020–1029, 2014.

[23] R. O’Rahilly and F. Müller, Basic human anatomy : a regional study of human structure. Saunders, 1983.