34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

An Investigation into Non-Invasive Physical Activity Recognition using Smartphones

Daniel Kelly[1] and Brian Caulfield[1]

***Abstract*— Technology utilized to automatically monitor Activities of Daily Living (ADL) could be a key component in identifying deviations from normal functional profiles and providing feedback on interventions aimed at improving health. However, if activity recognition systems are to be implemented in real world scenarios such as health and wellness monitoring, the activity sensing modality must unobtrusively fit the human environment rather than forcing humans to adhere to sensor specific conditions. Modern smart phones represent a ubiquitous computing device which has already undergone mainstream adoption. In this paper, we investigate the feasibility of using a modern smartphone, with limited placement constraints, as the sensing modality for an activity recognition system. A dataset of 4 subjects performing 7 activities, using varying sensor placement conditions, is utilized to investigate this. Initial experiments show that a decision tree classifier performs activity classification with precision and recall scores of 0.75 and 0.73 respectively. More importantly, as part of this initial experiment, 3 main problems, and subsequently 3 solutions, relating to unconstrained sensor placement were identified. Using our proposed solutions, classification precision and recall scores were improved by +13% and +14.6% respectively.**

I. INTRODUCTION

Monitoring the quantity, quality and variety of physical activity of the elderly and patients with particular chronic diseases can be a key component in identifying deviations from normal functional profiles and providing feedback on interventions aimed at improving health. However, the potential of physical activity measurements in the treatment and promotion of health and wellbeing cannot be fully realized until objective and accurate methods of measuring physical activity can be developed. Physical activity has traditionally been assessed by questionnaires but these have limitations due to their inherent subjective nature and caution must be taken when using questionnaires to identify activities [1]. A solution to this problem is to measure physical activity using body-worn sensors to automatically detect and measure particular activities. However, in order implement activity analysis in real world scenarios such as health and wellness monitoring, the activity sensing modality must unobtrusively fit the human environment rather than forcing humans to adhere to sensor specific conditions. One potential unobtrusive modality which is already being used naturally in the human environment is the mobile phone. A number of works have conducted research into the problem of physical activity classification. Liu et al. [2] and Gao et al. [3] propose multi-sensor systems in order to monitor physical activity. While results were very positive in both these works, the main downside was that subjects were required to wear invasive sensing devices. There have been other studies on developing activity recognition systems using a phone, or single motion sensor, but most of these works assume the motion signals are recorded from a known fixed device location and orientation [4]. This means that limitations and/or specific conditions are forced on the user of the system, requiring users to use their phone in a manner that is un-natural for them and could result in low adherence/participation in the activity monitoring application. Lester at al. [5] and Henpraserttae et al. [6] have addressed the problem of unconstrained sensor placement, but neither works address the problem of position variation and orientation variation at the same time. In this work, we investigate the effect of placing sensors at arbitrary locations and orientations has on activity classification.

II. METHODOLOGY

A. Sensor

In order to collect phone motion data, we use a Samsung Nexus S smartphone running the Android 2.3 operating system. Acceleration data, A_x, A_y and A_z along with gyroscope data, G_x, G_y and G_z, are recorded from the phone. A Kalman filter is used to calculate orientation angles θ and ϕ by measuring orientation from the gyroscope and utilizing the accelerometer to minimize any drift error that the gyroscope creates. A quaternion representation of the device orientation, defined as $q = \{q_w, q_x, q_y, q_z\}$, is calculated in order to overcome ambiguities which are inherent in the Euler angle measurements. We also define the overall magnitude of the acceleration as $A^m = \sqrt{A_x^2 + A_y^2 + A_z^2}$ and the overall magnitude of the angular velocity as $G^m = \sqrt{G_x^2 + G_y^2 + G_z^2}$.

B. Activity Data

In order to evaluate how an activity recognition system would perform with sensor data from varied phone positions and orientations, we collected data from 4 subjects performing 7 different activites using two phones per subject. The 7 chosen activities are as follows: (1) Standing (SN), (2) Sitting (ST), (3) Transition Down (TD), (4) Transition Up (TU), (5) Stairs Up (SU) (6) Stairs Down (SD) and (7) Walking (W). We chose these activities due to our overall application goal which is to accurately assess the patterns of daily activity of patients. One of the most important aspects of this is identifying sedentary behavior, therefore the need to distinguish between walking, standing and sitting/lying

*This work was supported by Science Foundation Ireland and the Irish Council for Science, Engineering and Technology

[1]With Clarity Center for Sensor Web Technologies, University College Dublin, Dublin, Ireland daniel.kelly@ucd.ie

is important. Detecting changes in stair usage could also be key in identifying loss or gain of physical function. We introduce the activities, 'Transition Up' and 'Transition Down' as transitional activities which could potentially be utilized to improve the accuracy of distinguishing between sitting or standing. The selection of these 7 activities are also consistent with other works in the area where an average of 6 activities are used [7], [8], [9], [6]. For each subject, one phone was placed somewhere on the torso and one phone was placed on the lower body. It should be noted that the goal is not to combine data collected from both phones to make a single classification. We utilize two phones in the data collection in order to collect a data set of activities that included data from varied body locations. For each subject the orientation and positioning of the phones were varied. Torso position varied the most due to more varied pocket positions of upper body clothing. Figure 1 shows an overview of the different phone placements. Each box represents the placement of a phone.

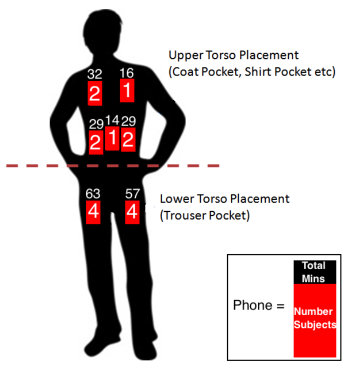

Fig. 1. Phone Locations

In order to collect a labeled set of activity data to train and test the different models, we developed a 'Remote Labeler' App for an Android Phone. A researcher controls the 'Remote Labeler' while monitoring the subject performing the different activities in a natural manner. The researcher presses pre-set activity buttons when the subject performs one of the corresponding pre-determined activities. The 'Remote Labeler' App then sends these labels over Bluetooth to the Sensor App, and the Sensor App logs these labels. We extracted features from the data set at intervals of 0.25 seconds (i.e. 4 classifications per second). The average duration of the activities which were recorded for each subject are as follows: Walking - 13 mins, Sitting - 3 mins, Standing - 6 mins, Stairs Up - 3 mins, Stairs Down - 3 mins, Transition Up - 1 min, Transition Down - 1 min. A total of 2 hours of data was collected across all subjects.

C. Feature Extraction

In order to measure the users activity at a given time t, a sliding window system is used. A time t, a number of different features are calculated based on the accelerometer signals A, the gyroscope signals G and the orientation q. Features are extracted from windows of different lengths in order to capture information on activities of different durations. We calculate a set of features based on the following measurements:

$\mu(x)$: Mean of signal x. $\sigma(x)$: Variance of signal x - helps discriminate between static and dynamic activities. $RMS(x)$: Refers to the Root mean squared of signal x and discriminates between different Static Activities. $IQR(x)$: Refers to the Interquartile range of signal x. IQR of Gyro is important for identifying transitional activities. $MFC(x)$: Refers to the Main frequency component of signal x as computed by a Fast Fourier Transform. $LPF_{<2.5Hz}(x)$: Refers to the signal x which has been passed through a Low Pass Filter. Discriminates between different dynamic activities. $BPF_{1.6-4.5Hz}(x)$: Refers to the signal x which has been passed through a Band Pass Filter. Discriminates between different dynamic activities. $\widehat{E}(x)$: Refers to the energy of signal x. $Corr(x, y)$: Refers to the correlation of signals x and y as calculated using the Pearson Correlation. Correlation between vertical and horizontal acceleration is important for identifying stair climbing.

A combination of these measurements, calculated from different motion signals, makes up a feature vector. Applying the 9 measurements to the 4 Acceleration signals, 4 Gyroscope signals and 2 orientation signal results in a set 90 features ($9 \times (4 + 4 + 2)$). We calculated these 90 features for windows lengths of 32, 64, 128 and 256, resulting in a total set of 360 features. In order to reduce this, we used an Information Gain Attribute Evaluation (IG) technique in order to evaluate the worth of each feature by measuring the information gain of each feature for each activity class [10]. Each feature was ranked by information gain and the top 50 features where chosen.

III. EXPERIMENT

In this section we investigate how an activity recognition system performs when subjected to data recorded from phones at different location and orientations. For each experiment described, a Leave One Subject Out (LOSO) evaluation protocol is used where each classifier is trained using data from 3 of the 4 subjects and tested using data from the remaining subject. This process is repeated for all combinations of subjects and the average performance is reported. Works on activity recognition report that identifying the best classification algorithm is dependent on the type of sensors, features and activities being used [11][12]. We therefore performed a preliminary experiment in order to identify a suitable model for the classification of data from a phone with unconstrained placement conditions. The preliminary experiment involved evaluating the following models on our data set: C.45 Decision Trees, MultiLayer Perceptrons, Logistic Regression, Bayes Network and Support Vector Machine (with a RBF kernel). Results showed that a C.45 Decision Tree was the best performing model and therefore is the model used in the following experiments.

A. Unmodified Features Evaluation

In the first experiment we evaluate how the classifier performs when trained and tested on all data collected from our data set. Feature vectors, comprising of the 50 features selected using IG feature selection, were computed for all data in the data set and then used to train and test the decision tree. Results detailed in Table I show that the classifier performs well when identifying walking while the remainder of the activities were classified with poorer performance. From these results we can see that the classifier has difficulty in learning how to discriminate between activities with more subtle differences. This can most likely be attributed to the fact that features, which could potentially discriminate between certain activities, have a large variation across the entire dataset due to the unconstrained position and orientation of the sensor.

1) Position Specific Classifier: In an attempt to reduce variation in the features caused by the unconstrained location conditions of the sensor, the data set is split up into 2 position specific data sets; 1) data from lower body and 2) data from the torso. Two position specific classifiers were then trained using the corresponding position specific data. Results, as detailed in Table I, show that the lower body classifier performs much better than the torso specific classifier. This can be attributed to the fact that less motion occurs in the upper body and therefore results in more discriminate patterns for particular activities. While distinct motion patterns for each activity do occur in the torso, the unconstrained sensor, and in particular the unconstrained orientation, results in the loss of discriminatory information such as movement in a specific direction. A second observation made from these results is that the performance of identifying sitting relating activities is, for both classifiers, very poor. Upon further investigation of the classification results it was discovered that a significant amount of sitting related activities were being mis-classified as standing. This is caused by the fact that no reference point exists in order to discriminate between a standing orientation and a sitting orientation due to the unconstrained orientation of the sensor.

TABLE I

SPECIFIC RECOGNITION PERFORMANCE

	All Positions:		Torso:		Lower Body:	
Activities	Precision	Recall	Precision	Recall	Precision	Recall
W	0.884	0.868	0.87	368	0.878	0.984
SN	0.753	0.7	0.847	0.442	0.775	0.983
TD	0.748	0.433	0.764	0.6	0.861	0.311
ST	0.669	0.716	0.552	0.916	0.628	0.766
TU	0.504	0.61	0.599	0.364	0.358	0.6
SU	0.343	0.579	0.194	0.954	0.746	0.617
SD	0.651	0.45	0.432	0.168	0.8	0.399
Total	0.75	0.736	0.725	0.533	0.851	0.835
			Combined Precision:		**0.788**	
			Combined Recall:		**0.684**	

B. Solutions

In this section we propose solutions to each of three problems identified in Section III-A followed by final experiment in order to evaluate the potential solutions.

(Problem 1): Large variations in potential discriminatory features occur due to difference in movement patterns between torso and lower body. To solve this problem we suggest that a two stage classification system be used where two position specific classifiers learn to classify activities from their corresponding positions. A location classifier, which we will evaluate in a later experiment, can be used to first classify the location of the sensor. The appropriate position specific classifier can be used to classify the activity.

(Problem 2): Unconstrained sensor orientation results in the loss of useful information such as movement in a particular direction. In order to solve this problem, a measure of motion, independent of sensor orientation, is required. This is performed by computing a global reference frame in order to measure acceleration with respect to gravity as opposed to a local reference frame which measures acceleration with respect to the sensor device. A rotation matrix $R_{\theta\phi}$ is computed from the orientation quaternion and the global acceleration frame is defined as $\bar{A} = A \times R_{\theta\phi}$, where $A = \{A_x, A_y, A_z\}$. The acceleration vector $\bar{A}$ now represents acceleration relative to gravity. The vertical component of the acceleration A^v can now be defined as $A^v = \bar{A}_y$. Yaw information is not utilized is this work due to the presence of noise in he yaw signal caused by ferromagnetic interference. We therefore have no representation of which direction the phone is pointing on the horizontal plane and therefore cannot disambiguate the remaining signals, $\bar{A}^x$ and $\bar{A}^y$, into dorsoventral and mediolateral (forward and sideward) directions. We must therefore combine these signals into a single horizontal acceleration component $A^h = \sqrt{(\bar{A}_x)^2 + (\bar{A}_z)^2}$.

(Problem 3): Discriminating between sitting and standing is difficult due to the fact that we do not have a constant reference point as to what orientation refers to standing and what refers to sitting. To overcome this problem we propose that an additional feature, $\overline{Diff(q, q^{stand})}$, be used. This feature refers to the mean difference between the current quaternion q and a dynamic reference quaternion q^{stand} over the period of a feature window. The dynamic reference point, q^{stand}, is a quaternion which is automatically updated to represent the current best estimate of a standing orientation. If a 'Walking' activity is detected then we know the user must be in a standing position. We can therefore use the orientation signals during the 'Walking' period in order to compute a representation of a standing orientation. The standing orientation, q^{stand}, is calculated by analyzing all features since the current walking activity was first detected in order to find a 0.25 second window with the lowest acceleration variation. We perform this analysis in order to find the most stable stance phase during the walking period between each heel strike and toe off. This process is particularly important when the user has placed the phone on the lower body as the stance phase of a walking gait is the period which most resembles a lower body standing posture. Using the window with the lowest acceleration variation, q^{stand} is then defined as the average quaternion from all orientations within that window

C. Solution Evaluations

In order for the solution to problem 1 to be applicable, the location of the sensor must first be identified such that the appropriate position specific classifier can be invoked. An experiment was conducted in order to evaluate how a classifier would perform when identify the location of the phone. Particular motion features from the torso and lower body differ during periods of walking activity and this can be taken advantage of in order to detect the phones location. The location of the phone is classified by training a single classifier to discriminate between the activities: 'Torso Walking' (WT), 'Lower Body Walking' (WL) and 'Not Walking'

(NW). The goal here is to detect not only if a walking activity is occurring, but also to identify if the walking signal was recorded from a phone placed on the torso or the lower body. If an 'Other' activity is detected then it is assumed that the position of the phone has not changed since the last phone location detected and a location specific classifier is used to detect what the 'Other' activity is. Location classification results, detailed in Table II, indicate that even though there are some errors in the classification process, the identification of the phone locations is robust with only 1% of the classified WT or WL positions receiving a false positive classification.

TABLE II

POSITION RECOGNITION PERFORMANCE

	Classified As:		
Position	NW	WL	WT
NW	**0.95**	0.02	0.03
WL	0.049	**0.941**	0.009
WT	0.087	0.0011	**0.932**

Finally, in order to evaluate the effect that the proposed solutions have on the overall classification process, the new measurements are first incorporated into the overall set of features. Vertical and horizontal acceleration and standing difference are calculated over windows length of 32, 64, 128 and 256 producing an additional set of 12 features. These 12 features where appended to the already existing set of 360 possible features. As with the original feature vector, the final set of features to be used in the classification process was identified by using a information gain feature selection process. Each feature was ranked by information gain and the top 50 features where chosen. The features are evaluated using the same protocol as discussed in Section III-A, where an overall classifier and two position specific classifiers are trained and tested. Table III details the results of the evaluations. The improved performance (+3.6% Precision, +6.4% Recall) of the position specific classifiers, compared to the overall classifier, show the positive effect our solution to problem 1 has on the classification performance. The solutions to problem 2 and 3 also have a positive effect on the classification performance. It can be seen that there is a significant performance increase (+9.2% Precision, +19.8% Recall) in the position specific classifiers when compared to the position specific classifiers evaluated on features with no global reference or dynamic standing features. In particular, there is a significant performance increase (+32.7% Precision, +11.3% Recall) in classifying sitting activities. Finally, the poor performance of the torso specific classifier has also been significantly improved upon (+10.5% Precision, +29.7% Recall).

TABLE III

RECOGNITION PERFORMANCE

	All Positions:		Torso:		Lower Body:	
Activities	Precision	Recall	Precision	Recall	Precision	Recall
W	0.912	0.83	0.803	0.916	0.955	0.989
SN	0.892	0.923	0.92	0.9	0.999	0.973
TD	0.75	0.535	0.748	0.433	0.767	0.444
ST	0.937	0.86	0.896	0.909	0.939	0.999
TU	0.656	0.677	0.504	0.559	0.654	0.459
SU	0.32	0.532	0.861	0.374	0.748	0.696
SD	0.485	0.589	0.368	0.326	0.724	0.707
Total	0.844	0.818	0.83	0.83	0.93	0.934
			Combined Precision:		**0.88**	
			Combined Recall:		**0.882**	

IV. CONCLUSION

This paper investigates the feasibility of utilizing a mobile phone, with limited placement conditions, as a sensor for non invasive activity recognition. Through experiments, three main problems that occur, when classifying activities using data from a sensor with unconstrained position and orientation, where identified. Moreover, three potential solutions for these problems were suggested and experiment results show that by utilizing these solutions, the overall classification performance can be significantly increased. These results show that activity classification can be carried out without using invasive sensing modalities. Future work will involve increasing the number of subjects for evaluation, the number of activities to be classified as well as investigating techniques to further increase performance. Since human activities occur in a temporal sequence, certain temporal restrictions are placed on the order a set of activities can occur in. Future work will also include adding a temporal classification stage to the system in order to take advantage of the temporal nature of human activities.

REFERENCES

[1] F. Pitta, T. Troosters, V. S. Probst, M. A. Spruit, M. Decramer, and R. Gosselink, "Quantifying physical activity in daily life with questionnaires and motion sensors in COPD." *European Respiratory Journal*, vol. 27, no. 5, pp. 1040–1055, May 2006.

[2] S. Liu, R. Gao, D. John, J. Staudenmayer, and P. Freedson, "SVM-based multi-sensor fusion for free-living physical activity assessment," in *Engineering in Medicine and Biology Society,EMBC, 2011 Annual International Conference of the IEEE*, 2011, pp. 3188–3191.

[3] L. Gao, A. Bourke, and J. Nelson, "A system for activity recognition using multi-sensor fusion," in *Engineering in Medicine and Biology Society,EMBC, 2011 Annual International Conference of the IEEE*, 2011, pp. 7869–7872.

[4] I. C. Gyllensten and A. G. Bonomi, "Identifying Types of Physical Activity With a Single Accelerometer: Evaluating Laboratory-trained Algorithms in Daily Life." *IEEE transactions on bio-medical engineering*, vol. 58, no. 9, pp. 2656–2663, Sep. 2011.

[5] J. Lester, T. Choudhury, and G. Borriello, *A Practical Appraoch to Recognizing Physical Activities*, ser. Pervasive Computing. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, vol. 3968.

[6] A. Henpraserttae, S. Thiemjarus, and S. Marukatat, "Accurate Activity Recognition Using a Mobile Phone Regardless of Device Orientation and Location," in *Body Sensor Networks (BSN), 2011 International Conference on*, 2011, pp. 41–46.

[7] N. Kawaguchi, Y. Kawahara, Y. Sumi, N. Nishio, Y. Yang, T. Yang, N. Ogawa, Y. Iwasaki, K. Kaji, T. Terada, K. Murao, and S. Inoue, "HASC2011corpus," in *Proceedings of the 13th international conference on Ubiquitous computing - UbiComp '11*, Jan. 2011, pp. 571–2.

[8] U. Maurer, A. Smailagic, D. P. Siewiorek, and M. Deisher, "Activity Recognition and Monitoring Using Multiple Sensors on Different Body Positions," *International Workshop on Wearable and Implantable Body Sensor Networks (BSN'06)*, pp. 113–116, Jan. 2006.

[9] Y. Hanai, J. Nishimura, and T. Kuroda, "Haar-Like Filtering for Human Activity Recognition Using 3D Accelerometer," *2009 IEEE 13th Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop*, pp. 675–678, Jan. 2009.

[10] Y. Yang, "A comparative study on feature selection in text categorization," *MACHINE LEARNING-INTERNATIONAL*, 1997.

[11] S. J. Preece, J. Y. Goulermas, L. P. J. Kenney, D. Howard, K. Meijer, and R. Crompton, "Activity identification using body-mounted sensors–a review of classification techniques." *Physiol Meas*, vol. 30, no. 4, pp. R1–33, Apr. 2009.

[12] J. Pärkkä, M. Ermes, P. Korpipää, J. Mäntyjärvi, J. Peltola, and I. Korhonen, "Activity classification using realistic data from wearable sensors." *IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 10, no. 1, pp. 119–128, Jan. 2006.