Human Activity Recognition Using Smartphone Sensors With Two-Stage Continuous Hidden Markov Models

Charissa Ann Ronao
Computer Science Department
Yonsei University
Seoul, South Korea

Sung-Bae Cho
Computer Science Department
Yonsei University
Seoul, South Korea

Abstract—**Recognizing human activities from temporal streams of sensory data observations is a very important task on a wide variety of applications in context recognition. Especially for time-series sensory data, a method that takes into account the inherent sequential characteristics of the data is needed. Moreover, activities are hierarchical in nature, in as much that complex activities can be decomposed to a number of simpler ones. In this paper, we propose a two-stage continuous hidden Markov model (CHMM) approach for the task of activity recognition using accelerometer and gyroscope sensory data gathered from a smartphone. The proposed method consists of first-level CHMMs for coarse classification, which separates stationary and moving activities, and second-level CHMMs for fine classification, which classifies the data into their corresponding activity classes. Random Forests (RF) variable importance measures are exploited to determine the optimal feature subsets for both coarse and fine classification. Experiments show that with the use of a significantly reduced number of features, the proposed method shows competitive performance in comparison to other classification algorithms, achieving an over-all accuracy of 91.76%.**

Keywords-continuous hidden Markov model; random forests; activity recognition; smartphone

I. INTRODUCTION

Mobile devices are becoming more and more sophisticated with every new model release. Nowadays, smartphones normally incorporate many diverse and powerful sensors such as GPS, high-resolution cameras, microphones, light sensors, temperature sensors, magnetic compasses, gyroscopes, and accelerometers [1]. These mass-marketed devices provide a flexible, affordable, and readily-available tool to be able to automatically monitor activities of daily living (ADL) without imposing inconveniences to the user [2]. With a variety of sensors to our disposal, and with the prevalent ownership of smartphones in recent years, it is only natural to take advantage of these sensors and use them for context awareness, and in turn, perform context recognition.

Context recognition is the process of finding answers by deducting interpretation and abstraction of context, using data acquired from sensors, intermediate information, or even the context itself previously inferred. Context can be derived from single or multiple sensory data, with a price of higher power consumption for the latter. In this paper, we focus on human activity recognition (HAR)—basic activities people perform on a daily basis, namely: Walking, Walking Upstairs, Walking Downstairs, Sitting, Standing, and Laying [2].

There are several models that can be used to perform feature selection and classify ADL. Some of the more common ones are decision trees, naïve Bayes, support vector machines, and artificial neural networks. In this paper, we utilize two-stage continuous HMMs to perform the classification task, taking into account the sequential nature of acceleration and gyroscopic sensory data and the inherent hierarchical characteristics of activities [19]. Feature selection is performed with the aid of RF variable importance measures, since they work well with continuous, possibly highly-correlated variables [20].

The paper is organized as follows: we begin with the discussion of HAR related works, followed by the presentation of the proposed method using two-stage continuous hidden Markov models (CHMMs); after that, we explain the experiment process, which also includes the analysis of the dataset and a brief discussion of RF variable importance measures; and lastly, we present the test results, showing the feasibility and reliability of the proposed method.

II. RELATED WORKS

Most of the previous works on activity recognition utilized multiple accelerometer sensors placed on several parts of the body. In the paper by Bao and Intille [3], they collected data from 20 users using five biaxial accelerometers worn on the user's right hip, wrist, upper arm, ankle, and thigh. Using decision tables and naïve Bayes classifiers, they were able to recognize 20 different activities. Results from their experiments indicate that the accelerometer placed on the thigh has the most power in recognizing ADL.

Kwapisz, et al. used a single tri-axial accelerometer in an Android smartphone worn on the user's pants pocket, to recognize six basic activities [1]. Their results show that J48 decision trees can infer stationary activities with ease, while multilayer perceptrons are able to recognize moving activities with high accuracies. It is important to note that not one classifier produced good results for all activities, and that upstairs and downstairs movements are highly confused.

Lee and Cho also utilized a smartphone 3D accelerometer grasped by hand, performed by 4 volunteer subjects, to infer actions and activities with hierarchical hidden Markov models (HHMMs) [4]. Their results also showed difficulty in differentiating upstairs and downstairs movements, but emphasized that HHMMs were more efficient in classifying low- and high-level activities than simple HMMs and artificial neural networks.

A number of studies also made use of tri-axial accelerometers other than that of the smartphone and performed activity recognition using other popular methods besides HMMs. Sharma, et. al. used a specific type of tri-axial accelerometer and achieved 84% accuracy through neural networks [9]. Khan used the Wii Remote and its embedded 3D accelerometer (placed at the back of the waist) to distinguish between four basic activities and yielded very high accuracy results using J48 classifiers [10]. These studies, however, did not include walking upstairs and downstairs in their experiments, which were noticeably difficult to recognize compared to other activities according to previously published papers.

Various other sensors were also utilized in conjunction with the accelerometer to infer human activities [5][6]. Wu, et. al. experimented on accelerometer and gyroscope data for activity recognition and concluded that the addition of gyroscope proved to be more beneficial than solely relying on accelerometer readings [7]. Shoaib, et. al. also stated that not only does the accelerometer and gyroscope produces satisfactory results when used alone, they also significantly improve classification performance when used on top of each other [8].

Finally, Anguita, et al. [2], the contributors of the UCI HAR dataset, performed classification of six activities, which includes upstairs and downstairs movements, using multi-class SVMs with accelerometer and gyroscope data. However, SVMs lack result transparency; its parameters are difficult to interpret. Because of this, there is no way to discriminate important variables, which often leads to the use of a very large number of predictors and high memory overhead.

III. The Proposed Method

We propose a method using RF variable importance measures for feature selection and two-stage CHMMs to recognize activities. RF variable importance measures are suitable for continuous and possibly highly-correlated variables [20], while two-stage CHMMs are appropriate for the temporal characteristics of sensory data and the natural hierarchical structure of activities [19].

Features extracted from acceleration and gyroscopic sensor data collected from a smartphone are assessed through their importance scores outputted by RF, as shown in Fig. 1. The selected feature subset is then used for coarse classification to categorize stationary and moving activities into their subclasses. In their respective subclasses, a different feature subset, based also on RF, is fed to second-level CHMMs for fine classification, which performs the final stage of activity recognition.

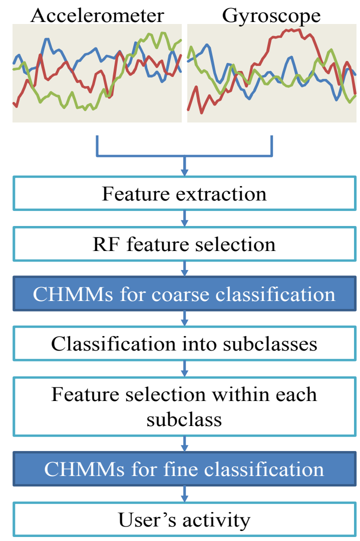

Figure 1. Activity classification hierarchy.

A. *Random Forests Importance Measures*

Dimension reduction should be performed not only for the purpose of reducing data volume, but also to find out which features matter the most for each activity. We address this by exploiting the importance measures produced by RF.

Random forests are an ensemble approach for classification introduced by L. Breiman and A. Cutler, which came from the technique of combining bagging and random feature selection to produce a collection of tree predictors with a controlled variance. RF is a classifier consisting of a collection of tree-structured classifiers, $h(x, \Theta_k)$, $k=1, \ldots$, where $\{\Theta_k\}$ are independent, identically-distributed random vectors, and each tree casts a unit vote for the most popular class at input x. The random subspace selection method has been proven to perform better than bagging alone when there are many redundant features present [12][13]. The method also gives useful internal estimates of error, strength, correlation, and variable importance [14].

The first step in getting the importance score of a variable Y_n is to fit a random forest to the data, and compute the out-of-bag error for each data point in the forest. Estimating the out-of-bag (OOB) error is similar to leave-one-out cross-validation, but is based on random resampling of data. Variable values of Y are then randomly permuted in an OOB_Y sample to get the perturbed sample denoted by $OOB^j{}_Y$. The error of predictor Y on the perturbed sample is then computed, giving the variable importance (VI),

$$VI(Y^j) = \frac{1}{ntree} \sum_t (errOOB^j{}_Y - errOOB_t), \qquad (1)$$

where *ntree* is the number of trees in the ensemble. In simple words, it is the average of the difference in out-of-bag error between the non-permuted and permuted values. The score is normalized by the standard deviation of the differences.

B. First-Level CHMMs for Coarse Classification

HMMs are probabilistic models based on simple Markov chains, but with hidden states. It is most suitable in handling time series data such as speech recognition and signal processing [11]. HMMs are based on an unobserved Markov chain $\{X_n\}$ which denotes the evolution of the state of a system. Given a realization of the state process $\{x_n\}$, the observed variables $\{Y_n\}$ are conditionally independent, which means that the distribution of each Y_n depends on the corresponding state x_n only. The random variables Y_n can take values in either a discrete or continuous space. Because sensory data are continuous in nature, we will focus on continuous HMMs on this paper.

The pair of processes $\{X_n\}$ and $\{Y_n\}$ is considered a continuous HMM if the observed process $\{Y_n, n \in N\}$ is real valued, or generally speaking, vector valued in a Euclidean space. CHMM is characterized by the following equation:

$$\lambda = (A, B, \pi), \tag{2}$$

where A is the transition probability matrix a_{ij}, B is the observation probability matrix $(\theta_1, \theta_2, \ldots, \theta_M)$, and π is the initial state distribution $(\pi_1, \pi_2, \ldots, \pi_M)$. A CHMM's initial state distribution is denoted by

$$\pi_i = P[X_0 = i],\ 1 \le i \le M, \tag{3}$$

where M is the number of probability distributions. The transition probability matrix of X_n, on the other hand, is signified by

$$a_{ij} = P[X_n = j \mid X_{n-1} = i],\ 1 \le i, j \le M. \tag{4}$$

Finally, the observation probability is represented by

$$b_i(y) = p_Y(y; \theta_i) = f(y; \theta_i), \tag{5}$$

where $p_Y(y; \theta_i)$ is the emission density of state *i*,

$$p_Y(y; \theta_i) = \frac{1}{\sqrt{2\pi\sigma_i}} \exp - \frac{(y - \mu_i)^2}{2\sigma_i^2}. \tag{6}$$

Equation (6) is the same as the probability density function of a Gaussian distribution with mean μ and variance σ^2.

Fig. 2 shows how data are passed through stages to classify activities into moving and stationary subcategories. Acceleration and gyroscopic data are preprocessed and fed to first-level CHMMs for training. Train data for moving activities are fed to the moving CHMM, and train data for stationary activities are directed to the stationary CHMM. All of these are done using a preselected feature subset and with two-state CHMMs. The number of states was determined according to the number of classes or activities to be classified

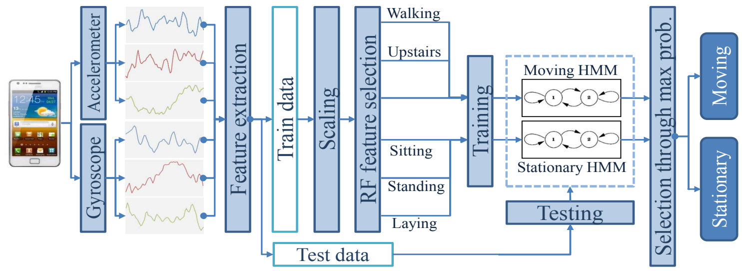

Figure 2. First-level CHMMs for coarse classification.

at each instance, as is often the case in activity recognition [15][17].

On this level, test data are fed to both moving and stationary CHMMs, and they are classified as either moving or stationary depending on which CHMM model produces the higher probability. Thus, for each first-level subclass $e \in E$, we build a CHMM λ^e and estimate the model parameters (A, B, π) that optimize the likelihood of the corresponding training observations.

With the feature test subset O, we estimate the likelihood of the observation belonging to first-level subclass $e \in E$. Using the relationship $P(O|\lambda^e)$, estimated across all CHMMs for each subclass, we select the activity with the highest probability, given by

$$e^* = \arg\max_{e \in E} P(O|\lambda^e). \tag{7}$$

$P(O|\lambda^e)$ is estimated using the forward-backward algorithm [12][16].

C. Second-Level HMMs for Fine Recognition

Once activities are categorized according to their respective subclasses on the first level, we then proceed to classify test data into their corresponding activity. Based on how the first level CHMMs classified the test data (as seen in Fig. 3), the process will continue to run either towards the moving subclass or the stationary subclass on the second level. We used three

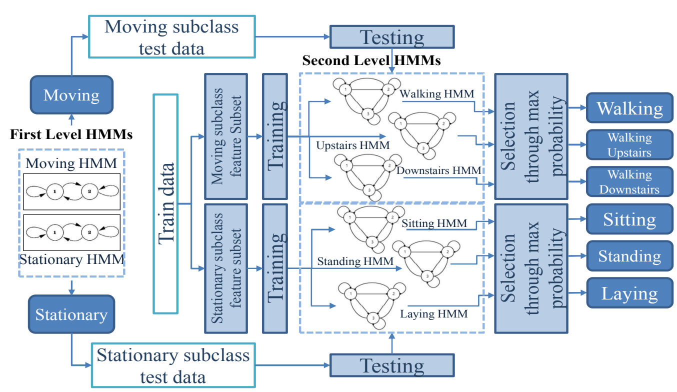

Figure 3. Second-level CHMMs for fine classification.

states for all CHMMs at this level since the two subclasses will be classified into 3 activities, and as previously mentioned, the number of states is determined by the number of classes [15][18].

The CHMMs for this level are also trained based on the subclasses, wherein training for moving activities use a completely different feature subset from stationary activities. We perform the same process on this level as the one applied on the first level. For each subclass, we build a CHMM λ^{em} for each moving activity and λ^{es} for each stationary activity, and estimate the model parameters $(A, B, \pi)_m$ and $(A, B, \pi)_s$, respectively, that optimize the likelihood of the corresponding training observations. We then proceed to follow the same procedure as we did on the first level, to arrive at the activity with the highest probability e_m* and e_s* for the moving and stationary subclasses, respectively.

IV. Experiments

The public domain UCI HAR data set, which was used throughout all our experiments in this paper, is a data set consisting of accelerometer and gyroscope xyz data values gathered from a Samsung Galaxy SII smartphone worn by 30 volunteer subjects. Each subject was instructed to perform a protocol of activities twice, first with the handset mounted on the left-side of the belt, and second with the handset placed according to the user's preference. This resulted into a data set with 561 features, all derived from sliding windows of 128 samples, and normalized to values between 1 and -1. The dataset was randomly partitioned into train and test data, with 70% allocated to the former (data from 21 subjects) and 30% for the latter (data from 9 subjects) [2].

Although the values on the data set are already normalized and scaled to the range [-1, 1], they are not z-scaled—the variables do not have a mean of approximately zero and a standard deviation of 1. HMM is a model that can handle only very short sequences; quantities would get extremely small as the sequence gets longer. To prevent this underflow phenomenon from happening when using HMMs, it is necessary to extract z-scores. For each variable Y_n, we can extract z-scores (or standardized scores) through

$$z = \frac{Y_n - \mu}{\sigma} \tag{8}$$

where z is the resulting z-scaled value, μ is the mean of all data points in a variable, and σ is the standard deviation of all variable points, as shown in Fig. 4.

All 561 features are fit to an RF to get the importance values of predictors. As observed on the importance measures output, variables with high importance scores for moving activities have substantially low importance scores for stationary activities, and vice versa. Evidently, a separation of moving and stationary activities would result to a boost in classification power and will also enable us to take advantage of using the least number of features on each subclass.

We adopt the method of RF feature selection similar to the one proposed by the Genuer, et. al. [21]. We ran repeated RF tests (using default settings) and computed the average of

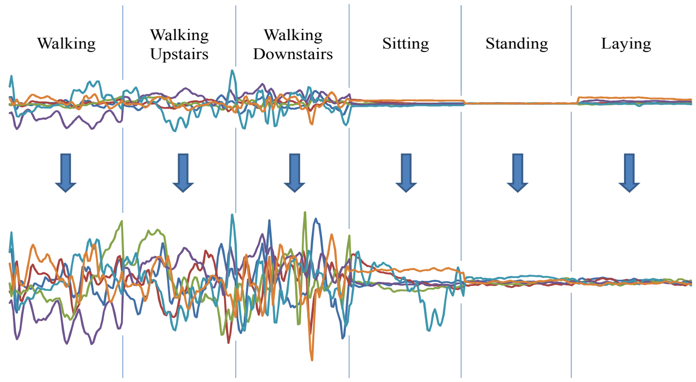

Figure 4. Scaling to obtain z-scores.

variable importance scores from all the runs. Variables with importance values higher than the mean importance value of all the variables are kept and ranked in descending order. We then performed a combined stepwise and 10-fold cross validation procedure on the training dataset and computed the error rates of the CHMM model starting from the two most important variables, and ending with the one involving all variables kept previously. The variables of the model with the smallest error rate are selected. The end result of this process is a feature subset of 119 variables, which is used as the feature subset for the first level.

The moving and stationary subclass feature subsets for the second level are derived from the first level feature subset. We performed the same cross validation procedure to determine the optimal feature subset for each subclass.

To make way for further evaluation of the variables, we plot the correlations of accelerometer and gyroscope xyz axes' values as shown in Fig. 5. As seen on the correlation heat maps, notable differences in Walking and Walking Upstairs lie on the gyro values. Incidentally, there is a dominance of acceleration-based variables in the feature subset for moving activities. Therefore, for a last attempt on variable evaluation, we focused on assessing gyro-based features. In addition to this, confounders (both acc-based and gyro-based variables) were also eliminated in the process. As a result, the new feature subset for the moving subclass consisting of 95 variables has a more balanced acc- and gyro-based feature ratio, from 41:8 to

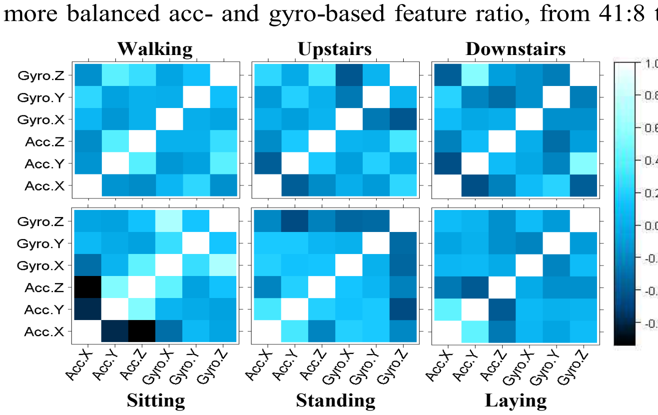

Figure 5. Correlation heat maps of sensor inertial values for each activity.

11:8. This last assessment also brought about a significant improvement in error rate on test data.

Referring again to the heat map, it can be observed that Walking Upstairs and Walking Downstairs activities exhibit almost the same acceleration correlation values. This strongly validates our stand on the need for gyroscope data in order to effectively classify upstairs and downstairs movements.

We have compared the proposed RF feature selection method with other dimension reduction techniques such as PCA (186 features), correlation (232 features, 93% threshold), and stepwise LDA (60 features) using CHMM. Error rate on test data is significantly lower for RF importance measures (95/5 features) as seen in Table I. It is apparent that RF variable importance measures feature selection method is a more suitable approach to be used in conjunction with CHMM.

Table II shows the confusion matrix of the two-stage CHMM classifier, as well as the results for precision and recall. For the stationary subgroup, the Laying activity achieved proper classification completely, while Standing surprisingly came out to be the hardest to classify. Moreover, Sitting achieved the lowest precision score—the latter activities being the ones greatly confused with each other. The confusion between the said activities might be attributed to the physical location of the device when the activities are performed (inside the pant pocket) [2]. In addition to that, although the accelerometer is enough to differentiate stationary movements, the extraction of more efficient features is necessary to precisely capture sensor behavior differences.

We have compared the proposed method with other statistical approaches such as naïve Bayes, J48 decision trees, artificial neural network, and conventional HMM, all using the derived RF feature subset (119 features), as shown in Fig. 6. Our two-stage CHMM classifier shows competitive performance when compared to other classifiers, with noticeably smaller standard errors in contrast to other techniques.

TABLE I. RESULTING ERROR RATES OF DIFFERENT FEATURE SELECTION TECHINIQUES USING CHMMS

	PCA	Correlation	Stepwise LDA	RF Imp score
No. of features	186	232	60	95/5
Walking	0.22	0.37	0.48	0.05
W. Upstairs	0.60	0.24	0.27	0.06
W. Downstairs	1.00	0.38	0.28	0.13
Sitting	0.84	0.69	0.96	0.11
Standing	0.40	0.41	0.22	0.14
Laying	0.20	0.61	0.06	0.00
Ave. Error Rate	0.54	0.45	0.38	0.08

V. CONCLUSION

We have shown that a two-stage continuous HMM classifier is a feasible method towards universal activity recognition on smartphones. Continuous HMMs are able to specifically handle time series data such as accelerometer and gyroscope sensor values, and the two-stage architecture enabled us to use a significantly smaller number of features at the same time utilizing the most effective ones. We have also demonstrated that random forest variable importance measures, in combination with proper domain knowledge, is an effective approach in uncovering the most useful features from a large feature set.

However, there is still a lot of room for development: future works will include experimenting on more efficient methods for feature extraction (e.g. genetic programming), as well as thorough analysis of the effectiveness of the features obtained. Furthermore, experimenting with different variants of HMM is another subject to explore. Also, the direct implementation of the CHMM classifier on the smartphone platform as well as venturing out on more complex activities to recognize are a few of the topics for future research.

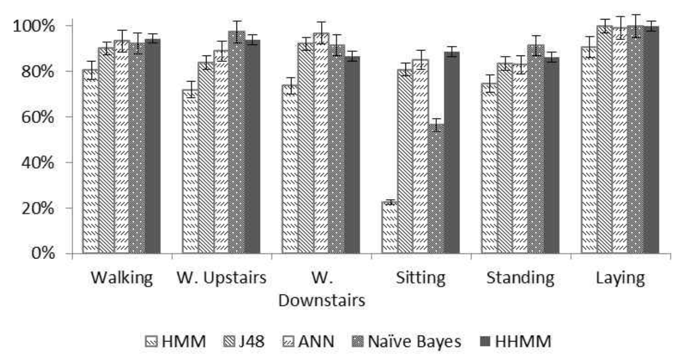

Figure 6. Comparison of classifiers.

TABLE II. CONFUSION MATRIX OF TWO-STAGE CONFUSION MATRIX OF TWO-STAGE HMMS

		Predicted Class						
		Walking	*Upstairs*	*Downstairs*	*Sitting*	*Standing*	*Laying*	*Recall*
Actual Class	*Walking*	469	7	20	0	0	0	94.56%
	W. Upstairs	16	443	12	0	0	0	94.06%
	W. Downstairs	28	27	365	0	0	0	86.90%
	Sitting	0	0	0	435	37	19	88.59%
	Standing	0	0	0	72	460	0	86.47%
	Laying	0	0	0	0	0	537	100.00%
	Precision	91.42%	92.87%	91.94%	85.80%	92.56%	96.58%	91.76%

REFERENCES

[1] J. Kwapisz, G. Weiss, and S. Moore, “Activity recognition using cell phone accelerometers,” SIGKDD Explorations, vol. 12, no. 2, pp. 74-82, 2010.

[2] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, “A public domain dataset for human activity recognition using smartphones,” European Symposium on Artificial Neural Networks (ESANN), pp. 437-442, 2013.

[3] L. Bao and S. Intille, “Activity recognition from user-annotated acceleration data,” Pervasive Computing, Lecture Notes in Computer Science, vol. 3001, pp. 1-17, 2004.

[4] S.-B. Cho and Y.-S. Lee, “Activity recognition using hierarchical hidden Markov models on a smartphone with 3D accelerometer,” Hybrid Artificial Intelligent Systems, IOS Press, vol. 6678, pp. 460-467, 2011.

[5] E. Garcia and R. Brena, “Real time activity recognition using a cell phone’s accelerometer and Wi-Fi,” Workshop Proceedings of the 8th International Conference on Intelligent Environments, pp. 94-103, 2012.

[6] P. Lukowicz, J.A.Ward, H. Junker, M. Stager, G. Tr¨oster, A. Atrash, and T. Starner. “Recognizing workshop activity using body worn microphones and accelerometers,” International Conference Pervasive Computing, pp 18-22, 2004.

[7] W. Wu, S. Dasgupta, E.E. Ramirez, C. Peterson, and G.J. Norman, “Classification accuracies of physical activities using smartphone motion sensors,” Journal of Medical Internet Research, vol. 14, no. 5, 2012.

[8] M. Shoaib, H. Scholten, and P. J. M. Havinga, “Towards physical activity recognition using smartphone sensors,” International Conference on Ubiquitous Intelligence & Computing, pp 80-87, 2013.

[9] A. Sharma, Y.-D. Lee, and W.-Y. Chung, “High accuracy human activity monitoring using neural network,” International Conference on Convergence and Hybrid Information Technology, pp 430-435, 2008.

[10] A. M. Khan, “Recognizing physical activities using Wii remote,” International Journal of Information and Education Technology, vol. 3, no. 1, pp 60-62, 2013.

[11] L. R. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” Proceedings of the IEEE, vol. 77, no. 2, pp. 257-286, 1989.

[12] T. K. Ho, “The random subspace method for constructing decision forests,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 8, pp 832-844, 1998.

[13] R. Bryll, R. Gutierrez-Osuna, F. Quek, “Attribute bagging: Improving accuracy of classifier ensembles by using random feature subsets,” Pattern Recognition, vol. 36, no. 6, pp 1291-1302, 2003.

[14] L. Breiman, “Random forests,” Machine Learning, vol. 45, no. 1, pp 5-32, 2001.

[15] R. Cilla, M. A. Patricio, J. Garcia, A. Berlanga, and J. M. Molina, “Recognizing human activities from sensors using hidden Markov models constructed by feature selection techniques,” Algorithms, vol 2, no.1, pp 282-300, 2009.

[16] K. Tran, I. A. Kakadiaris, and S. K. Shah, “Fusion of human posture features for continuous action recognition,” Trends and Topics in Computer Vision, Lecture Notes in Computer Science, vol. 6553, pp 244-257, 2012.

[17] H. Fang, R. Srinivasan, and D. J. Cook, “Feature selections for human activity recognition in smart home environments,” International Journal of Innovative Computing, Information and Control, vol. 8, no. 5(B), pp 3525-3535, 2012.

[18] D. Trabelsi, S. Mohammed, F. Chamroukhi, L. Oukhellou, and Y. Amirat, “An unsupervised approach for automatic activity recognition based on hidden Markov model regression,” IEEE Transactions on Automation Science and Engineering, vol. 10, no. 3, pp. 829-835, 2013.

[19] P. Natarajan and R. Nevatia, “Hierarchical multi-channel hidden semi Markov models,” International Joint Conference on Artifical Intelligence, pp. 2562-2567, 2007.

[20] K. Archer and R. Kimes, “Empirical characterization of random forest variable importance measures,” Computational Statistics & Data Analysis, vol. 52, no. 4, pp. 2249-2260, 2008.

[21] R. Genuer, J.M. Poggi, C. Tuleau-Malot, “Variable selection using random forests,” Pattern Recognition Letters, vol. 31, no. 14, pp. 2225–2236, 2010.