Hierarchical Human Activity Recognition Using GMM

Prateek Srivastava
Department of Electrical and Computer Engineering,
National University of Singapore (NUS), Singapore
prateek.k.srivastava@gmail.com

Wai-Choong Wong
Department of Electrical and Computer Engineering,
National University of Singapore (NUS), Singapore
elewwcl@nus.edu.sg

***Abstract*— In this paper, we propose a hierarchical human activity recognition system using Gaussian mixture models (GMMs) on continuous daily activities. The system recognizes the human activities by making use of tri-axial accelerometer and bi-axial gyroscope. We use different features such as mean, variance, root mean square, pitch, and roll for activity classification. Comparative performance assessments are carried out using the publicly available Wearable Action Recognition Dataset (WARD). The hierarchical recognition happens in two steps. First, the test data is classified into two broad clusters – static activity and dynamic activity. Second, the recognition is carried out within the identified class. For continuous activity recognition, our proposed system is able to achieve a recognition accuracy of 86.92% which is 2.63% above the baseline system. The new algorithm also provides more flexibility for better feature selection for different sets of activities.**

Keywords-Human Activity recognition; wearable sensors; pattern recognition; Gaussian Mixture modeling (GMM).

I. INTRODUCTION

Due to the recent progress in ubiquitous and wearable computing, activity recognition has become a major contributor towards many monitoring and interaction applications. Human action/activity recognition is an emerging field of research. Physical activity can be defined as "any bodily movement produced by skeletal muscles that result in energy expenditure above the resting level" [1]. The objective of any human activity recognition system is to recognize any human activity using its observed sensor/visual data. Generally, human activities can be classified into two broad categories – static (which involves minimal movement of body parts such as standing or sitting) and dynamic (which involves some motion in the body parts such as boxing or walking).

A popular approach to activity recognition is based on the use of visual data [1, 2], which is high-dimensional and dense. The visual data can sometimes become intrusive and disruptive, and hence, it poses a challenge to personal privacy. Moreover, the vision-based activity recognition systems are very sensitive to ambient lighting conditions and occlusion. With the recent miniaturization of simple sensors such as accelerometers and gyroscopes, researchers have begun to adopt the use of these low-powered, unobtrusive sensors.

A. Motivation

Due to the increase in the aging population, the world will soon have an increased number of aging baby boomers. As the existing and the future heath care sectors cannot effectively serve all the baby boomers, there is an increased demand for health monitoring and support of elderly-care units using assisted living systems. Consequently, the need for remote health care systems for patient monitoring is gradually growing; see Ibrahim et al. [3].

In this paper, an inertial sensor framework is used for human activity recognition because it provides an unobtrusive, low-powered and cost effective solution for many applications such as daily assisted living of elder care, virtual-real world interaction, sports training, and long term monitoring purpose. Longer term monitoring would reveal the subject’s activity levels with respect to metabolic energy expenditure, associated with different activities such as walking, standing, etc. Medical professionals believe that one of the best ways to detect an emerging medical condition before it becomes critical is to look for changes in the activities of daily living (ADLs), instrumental ADLs (IADLs), and enhanced ADLs (EADLs); see Tapia et al. [4].

B. Objective

In this work, our objective is to build a fast and accurate system for recognizing human activities. This system uses the data from the accelerometers and gyroscopes so as to recognize continuous activities.

C. Organization

This is paper is organized as follows. Section II discusses some prior works in the field of human action/activity recognition. In Section III, we discuss the approach and the procedure of our algorithm. In Section IV, the hierarchical action recognition approach is introduced. In Section V, the results are presented and discussed. Finally, Section VI concludes the paper with some recommendations about future work.

II. LITERATURE REVIEW

Recent developments in sensor technology have led to miniaturized inertial sensors – accelerometer and gyroscope.

 ISBN: 978-1-61208-235-6

These inertial sensors are widely used by the wearable activity recognition researchers. Bao et al [5] and Eric et al [6] have studied activity recognition using multiple sensors at different locations on the body. Thomas et al [7] have used the multimodal approach of activity recognition by combining motion sensors with ultrasonic sensors for continuous activity recognition. The application domains of activity recognition systems are diversity with examples such as Ernst et al [9] to recognize moves in martial arts, while David et al [10] built a mixed reality car parking game based on human computer interaction.

A number of activity recognition algorithms have been explored and these include:

1. Decision Trees: It finds a set of thresholds for a pattern-dependent sequence of features. Bao et al [5] used C4.5 decision tree classifier.
2. Nearest Neighbor (NN): It assigns patterns to the majority of class among k nearest neighbor using a performance optimized value of k. Ravi et al [8], Bao et al [5], and Maurer et al [11] used NN classifiers for activity recognition.
3. Naïve Bayes: It is a Bayes theorem based probabilistic classifier. Bao et al [5], and Maurer et al [11]
4. Support Vector Machine (SVM): Ravi et al [8] have used SVM classifiers.
5. Hidden Markov Model: one of the most popular statistical model for capturing temporal patterns in the data. It is extensively used in activity recognition. Yamato et al [1] used HMM to recognize different tennis strokes.
6. Gaussian Mixture model (GMMs): GMMs are parametric representations of any probability density function. Allen et al [12] used GMMs for transactional activities with an accuracy of 76.6 %. Ibrahim et al [3] recognize simple activities with small set of activities with 88.76 %. In this paper, Gaussian mixture models are also used for recognition.

Pattern recognition highly depends on the kinds of features used to model the patterns. Thus, features are very crucial for any recognition system. Some time domain features that have been commonly used include:

1. Mean: The mean value feature has been used by Ravi et al [8], and Bao et al [5].
2. Variance: The variance feature has been used by Ravi et al [8]
3. Root Mean Square (RMS): The RMS feature has been used by Maurer et al [11].

In the literature, researchers have tried different window lengths for activity recognition such as 6.7s in Bao et al [5], and 5.12s in Ravi et al [8]. In this paper, we use a window length of 1s is used to classify the activities. We have chosen a lower window length to facilitate real-time recognition.

III. APPROACH AND PROCEDURE

A. Dataset

To present a general comparison of results, a public dataset 'WARD: A Wearable Action Recognition Database' collected by Yang et al [13, 14] is used. In the WARD database, sensor data of different continuous activities have been collected. It contains data corresponding to 20 different subjects and 13 different activities. It also contains non-transient human actions. In order to sufficiently sample the continuous movement of a non-transient action, each subject performs one trial of an action for more than 10 seconds. The sensors are placed at different locations of the subject's body. Each sensor contains a 3-axis accelerometer and a 2-axis gyroscope. The locations of the sensors on the body are shown in Figure 1.

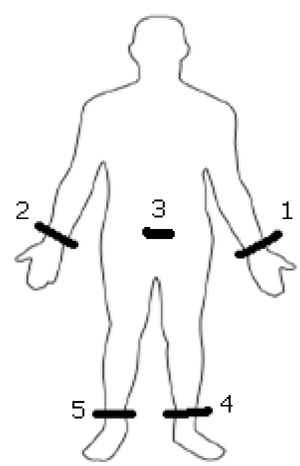

Figure 1. Sensor locations and orientation of a subject for WARD dataset, where the bold lines represent the sensor locations.

- Sensor 1: Outside center of the lower left forearm joint. The y-axis of the gyroscope points to the hand.
- Sensor 2: Outside center of the lower right forearm joint. The y-axis of the gyroscope points to the hand.
- Sensor 3: Front center of the waist. The x-axis of the gyroscope points down.
- Sensor 4: Outside center of the left ankle. The y-axis of the gyroscope points to the foot.
- Sensor 5: Outside center of the right ankle. The y-axis of the gyroscope points to the foot

This dataset contains the following 13 activities: 1. Rest at Standing (ReSt). 2. Rest at Sitting (ReSi). 3. Rest at Lying (ReLi). 4. Walk forward (WaFo). 5. Walk forward left-circle (WaLe). 6. Walk forward right-circle (WaRi). 7. Turn left (TuLe). 8. Turn right (TuRi). 9. Go upstairs (Up). 10. Go downstairs (Down). 11. Jog (Jog). 12. Jump (Jump) and 13. Push wheelchair (Push). For more details, please refer the WARD database manual [13].

B. System Overview

Our human activity recognition system recognizes the activities using the WARD data. In this system, a statistical recognizer for different activities is built. The system is divided into two phases, namely the training phase and the testing phase. In the training phase, the raw data for all activities is first collected using the 3-axis accelerometer and 2-axis gyroscope. The sampled data from the accelerometer and gyroscope are combined. Then, suitable time domain features are extracted and are used to model the activities using Gaussian mixture models (GMMs). The models of all the different activities are stored at the end of the training phase. In the testing phase, activity data of the test subject is first collected and the features are extracted. Then, the maximum probability of match of the test sample against all stored sample patterns is calculated. That pattern, which has the highest likelihood of match against the test pattern, is recognized as the correct activity. The overall recognition system is depicted in Figure 2.

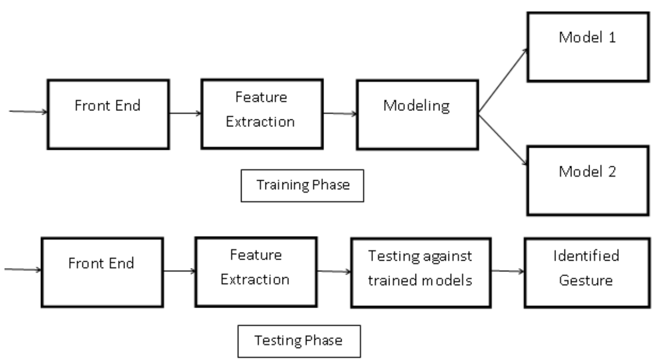

Figure 2. System overview of the activity recognition system.

C. Features for activity recognition

We have used time domain features for performing activity recognition. This is because our initial set of experiments suggests that time domain features perform better as compared to frequency domain features. The time domain features that we have used for the classification are mean, variance and root mean square (RMS).

In order to represent the angular information, the captured accelerometer and gyroscope data are used to calculate the pitch and roll. Since we are using 3-axis accelerometer and 2-axis gyroscope, the pitch and roll are calculated using the following equations.

roll = arctan(-acc_y/ -acc_z);
pitch = arctan(acc_x, sqrt(acc_y * acc_y + acc_z * acc_z));

where, acc_x, acc_y and acc_z are the x-axis, y-axis and z-axis accelerometer values respectively.

D. Activity Modeling using the GMM Algorithm

GMMs [15-16] are parametric representations of a probability density function. When trained to represent the distribution of a feature vector, GMMs can be used as classifiers. GMMs have proved to be a powerful tool for distinguishing time series data with different general properties. The use of GMMs for modeling activity is motivated by the interpretation that the (1) uni-variate Gaussian densities have a simple and concise representation, depending uniquely on two parameters, mean and variance, (2) they are capable of modeling arbitrary densities, (3) the Gaussian mixture distribution is universally studied and its behaviors are widely known, (4) a linear combination of Gaussian basis functions is capable of modeling a large class of sample distributions. In principle, the GMM can approximate any probability density function to an arbitrary accuracy.

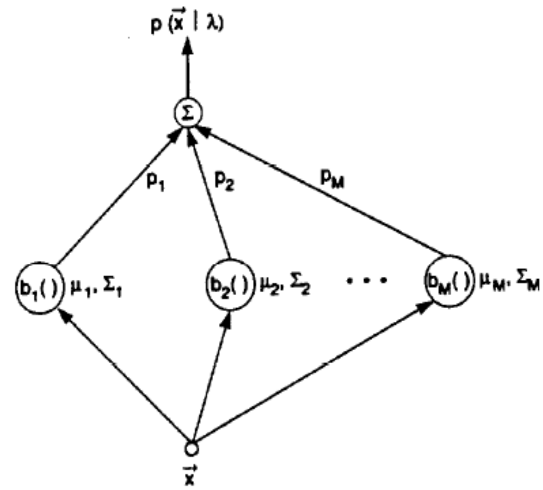

Figure 3. Depiction of an *M* component Gaussian mixture density [15].

A GMM is a weighted sum of M component densities as shown in Figure 3, given by the following equation:-

$$P\ (x_t\ |\lambda_s) = \sum_{i=1}^{M} p_i\, b_i\ (x(t))$$

Here, x_t is a sequence of feature vectors from the activity data, *x(t)* is a feature vector having *D*-dimensions. $b_i(s)$ is the Gaussian probability distribution function (PDF) associated with the i^{th} mixture component and is given by:

$$b_i(x_t) = \frac{1}{2\pi^{D/2}\,|\Sigma_s^i|^{1/2}}\, e^{\left(-\frac{1}{2}\right)(x-\mu i)^{\wedge T}\, {\Sigma_i^s}^{-1}(x-\mu i)}$$

Here, μ_i is the mean vector and $\sum_i^s$ is the covariance matrix of the i^{th} mixture component.

The mixture weights are such that :

$$\sum_{i=1}^{M} p_i \quad = 1$$

Each trained activity is thus, represented by a Gaussian mixture model, collectively represented by

$$\lambda_s=\{\mu_i , \textstyle\sum_i^s \ , p_i \}$$

where i = 1,2 ,…M, and μ_i , $\sum_i^s$, p_i represent the mean, covariance and weights of the i^{th} mixture respectively.

III. HIERARCHICAL RECOGNITION OF HUMAN ACTIVITY

A hierarchical recognition approach is proposed for human activity recognition on continuous activities. The dataset has data corresponding to two different types of activities – static and dynamic. In static activities such as sitting, standing, and lying, the motion sensor values are expected to have minimal variations over time. Apart from sitting, standing and lying, all other activities are categorized as dynamic since some motion is involved.

In our initial set of experiments, we observed that a number of static activities were mis-classified as dynamic activities and vice versa. In order to minimize such confusion, we propose to adopt a simple but powerful hierarchical approach to classify the full activity recognition on the continuous dataset. The first stage of this hierarchical approach tries to classify the test data into either static or dynamic activity categories. The second stage then tries to identify the correct activity from the chosen activity category. Together with improving the overall recognition accuracy, this approach also speeds up the process of computation because after the test activity is classified as static or dynamic in the first step, it will then, only be compared against its activity class for final recognition. The recognition flow of this algorithm is as shown in Figure 4.

This algorithm provides two major advantages over the baseline one-step classification system. First, since the two activity clusters: static and dynamic differ a lot in their corresponding statistical properties, the accuracy of classifying into static or dynamic classes is high. Since the test data is only compared with activities of its identified class in the second stage, the possibility of inter-class misclassification is eliminated. Second, this hierarchical system provides greater flexibility to extract different features for different clusters so as to improve the overall system performance.

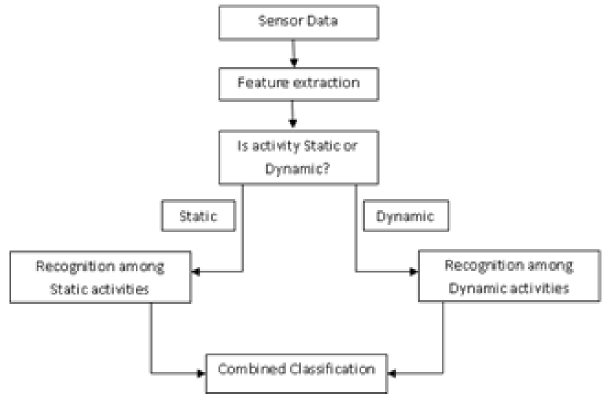

Figure 4. Flow chart of Hierarchical human activity recognition.

IV. RESULTS AND DISCUSSION

A. Continuous Activity Recognition (CAR) on WARD dataset

Different sets of experiments are performed to fine-tune the accuracy and speed of the recognition system. To achieve low latency, we use a window length of 1s in our recognition algorithm. The feature set includes pitch, roll, mean and variance from the accelerometer and gyroscope sensor data. To get the best number of mixture models, different numbers of mixture models are used to test the overall performance. The system with 32 mixture models gave the best overall performance, and thus we choose to model each activity using 32-GMMs. The results of the one-tier recognition system show an overall accuracy of 84.69 %. The confusion matrix of this system is shown in Table II.

From the confusion matrix of the one tier system it can be observed that a large number of static activities is misclassified as dynamic activities. For example, the misclassification accuracy of Rest at Sitting (Static activity) as Push Wheelchair (Dynamic Activity) is 12.2%.

A hierarchical algorithm using a two-step recognition process is proposed to classify the human continuous activities. The hierarchical recognition is proposed to minimize such inter-class misclassification as discussed above. To compare the results of the hierarchical algorithm over the one tier algorithm, we have used the same set of features {pitch & roll along mean and variance from accelerometer and gyroscope sensor data} and modeled them using 32-GMMs. The results of the cluster classification in the first step are shown in Table I. The overall accuracy of cluster classification is 96.58% where it is weighted against the number of test files for static and dynamic activities.

 ISBN: 978-1-61208-235-6

TABLE I. CLUSTER ACCURACY OF HIERARCHICAL RECOGNITION SYSTEM.

	Static	Dynamic
Static	95.5	4.5
Dynamic	3.0	97.0

In Table II, the accuracy of the one tier system with the same feature set and the same number of Gaussian mixtures as is used in the hierarchical system is 84.69 %. The accuracies of the hierarchical activity recognition system are shown in Table III. Here, we can observe that the overall accuracy of the hierarchical activity recognition system is 86.92%. This shows that the hierarchical algorithm improves the system performance by 2.63 % over the one tier system. It can also be observed from Table 3 that the misclassification of static activities as dynamic activities and vice versa is reduced. For example, the misclassification accuracy of Rest at Sitting (Static activity) as Push Wheelchair (Dynamic Activity) is 5.21% as compared to 12.2% in one tied system.

V. CONCLUSION AND RECOMMENDATIONS

In this paper, we have proposed a hierarchical human activity recognition system using Gaussian mixture models (GMMs). The results of the system are competitive as compared to prior activity recognition systems. The performance of the proposed system is tested using the publicly available WARD dataset to provide a better comparability. An overall system accuracy of 86.92 % is achieved using this hierarchical approach, which is an improvement of 2.63% over the baseline system. The proposed hierarchical algorithm also provides the flexibility to use different feature sets for the identification of different classes of activities. Since the static activities and dynamic activities differ a lot in their statistical properties, the best performing feature sets for these clusters can be used to obtain the best performance in each individual cluster.

Another important contribution of this paper is that the recognition is performed using less test data. In continuous activities, the activities are recognized using 1 sec of test data. Also, to capture the angular information, the pitch and roll using the accelerometer and gyroscope are used as the feature set. To further reduce the time of recognition, the concept of fast elimination of such patterns that certainly do not match a given behavioral pattern can be adopted. This concept has been proposed by Bajan [17].

In this paper, the sensors are placed at five different locations so as to capture full body motion statistics. The sensor placement is maintained as per the WARD Dataset so as to maintain the comparability of results. The orientation of sensors may not be optimum as this orientation of the sensors does not exploit the symmetry of body. In future, there is need to find the optimal number, orientation and placement of sensors required to perform human activity recognition.

TABLE II. CONFUSION MATRIX OF ONE TIER ACTIVITY RECOGNITION SYSTEM (IN PERCENTAGE).

	ReSt	**ReSi**	**ReLi**	**WaFo**	**WaLe**	**WaRi**	**TuLe**	**TuRi**	**Up**	**Down**	**Jog**	**Jump**	**Push**
ReSt	68.2	11.2	0	0	0.5	1.8	0	0	5.8	6.3	0	0	6.3
ReSi	0.5	74.7	0	0	0	0	0	0	4.5	4.1	0	4.1	12.2
ReLi	0	4.6	75.6	0	0	0	0	0	0	0	0	19.8	0
WaFo	0	0	0	84.8	3.2	3.8	0	0	1.3	1.9	0.6	4.4	0
WaLe	0	0	0	1.9	93.7	0	3.4	0	0	0	0	0.5	0.5
WaRi	0	0	0	2.5	0	86.4	3	2	1	2	2.5	0.5	0
TuLe	0	0	0	0	11.1	0	88.9	0	0	0	0	0	0
TuRi	0	0	0	1	0	5.6	0	88.3	0	3.1	1.5	0.5	0
Up	0	0	0	1.6	1.1	1.1	0	0	65.2	19.3	9.6	0	2.1
Down	0	0	0	0	1.2	0	0	0	1.8	84.4	10.8	1.2	0.6
Jog	0	0	0	0	0	0	0	0	2.9	0	97.1	0	0
Jump	0	0	0	0	0	0	0	0	0	0.6	0.6	98.8	0
Push	0.6	0	0	0	0	0	0	0	0.6	0.6	0	3.5	94.8

 ISBN: 978-1-61208-235-6

TABLE III. CONFUSION MATRIX OF OVERALL SYSTEM ACCURACY OF HIERARCHICAL RECOGNITION SYSTEM (IN PERCENTAGE)

	ReSt	ReSi	ReLi	WaFo	WaLe	WaRi	TuLe	TuRi	Up	Down	Jog	Jump	Push
ReSt	70.4	11.2	0	0	0	3.6	4	0.9	0	6.3	0	0	3.6
ReSi	6.6	83.9	0	0	0	0	0	0	0	4.3	0	0	5.2
ReLi	0	5.7	94.3	0	0	0	0	0	0	0	0	0	0
WaFo	0	0	0	84.2	3.2	3.8	0	0	1.3	1.9	0.6	4.4	0
WaLe	0	0	0	1.9	93.7	0	3.4	0	0	0	0	0.5	0.5
WaRi	0	0	0	2.5	0	87.3	3.1	1.5	1	2	2	0.5	0
TuLe	0	0	0	0	11.1	0	88.9	0	0	0	0	0	0
TuRi	0	0	0	1	0	5.6	0	88.8	0	3.1	1	0.5	0
Up	0	0	0	1.6	1.1	0.5	0	0	66.3	18.5	9.8	0	2.2
Down	0	0	0	0	1.2	0	0	0	1.8	84.4	10.8	1.2	0.6
Jog	0	0	0	0	0	0	0	0	2.9	0	97.1	0	0
Jump	0	0	0	0	0	0	0	0	0	0.6	0.6	98.8	0
Push	1.2	4.1	0	0	0	0	0	0	0	0	0	2.9	91.8

REFERENCES

[1] J. Yamato, J. Ohya, and K. Ishii, "Recognizing human action in time-sequential images using hidden markov model", International IEEE Conference on Computer Vision and Pattern Recognition (CVPR '92), 1992, pp. 379–385.

[2] T. Starner, "Visual recognition of american sign language using hidden markov models", Master's thesis, Massachusetts Institute of Technology, Cambridge, MA, USA (Feb 1995).

[3] R. Ibrahim, E. Ambikairajah, B. Celler, and N.H. Lovell, "Time-Frequency based features for classification of walking patterns",15th Int. Conf. on Digital Signal Processing (DSP '07), 1-4 Jul 2007, pp. 187–190.

[4] E. M. Tapia, S. S. Intille, and K. Larson, "Activity Recognition in the Home Using Simple and Ubiquitous Sensors", Pervasive Computing, LNCS, Springer, 2004, Vol. 3001/2004, pp. 158-175.

[5] L. Bao and S. Intille, "Activity recognition from user-annotated acceleration data", Pervasive Computing, LNCS, Springer, 2004, Vol. 3001/2004, pp. 1-17.

[6] E. Guenterberg, H. Ghasemzadeh, V. Loseu, and R. Jafari, "Distributed continuous action recognition using a hidden markov model in body sensor networks", Distributed Computing in Sensor Systems, LNCS, Springer, 2009, Vol. 5516/2009, pp. 145-158.

[7] T. Stiefmeier, G. Ogris, H. Junker, P. Lukowicz, and G. Tröster, "Combining motion sensors and ultrasonic hands tracking for continuous activity recognition in a maintenance scenario", 10th Int. Symp. on Wearable Computing (ISWC '06), 11-14 Oct 2006, Zurich, Switzerland, pp. 97-104.

[8] N. Ravi, N. Dandekar, P. Mysore, and M. Littman, "Activity recognition from accelerometer data", IAAI'05 Proceedings of the 17th conference on Innovative applications of artificial intelligence - Volume 3 pp 1541-1546.

[9] E. A. Heinz, K. S. Kunze, M. Gruber, D. Bannach, and P. Lukowicz, "Using wearable sensors for real-time recognition tasks in games of martial arts – An initial experiment", 2nd IEEE Symp. on Computational Intelligence and Games (CIG '06), 22-24 May 2006, Reno, NV, USA, pp. 98–102.

[10] D. Bannach, O. Amft, K.S. Kunze, E.A. Heinz, G. Tröster, and P. Lukowicz, "Waving real hand gestures recorded by wearable motion sensors to a virtual car and driver in a mixed-reality parking game", IEEE Symp. on Computational Intelligence and Games (CIG '07), 1-5 Apr 2007, Colchester, U.K., pp. 32–39.

[11] U. Maurer, A. Smailagic, D. Siewiorek, and M. Deisher, "Activity recognition and monitoring using multiple sensors on different body positions", BSN '06 Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks pp 113 - 116.

[12] F. Allen, E. Ambikairajah, N. Lovell, and B. Celler, "An adapted gaussian mixture model approach to Accelerometry-Based movement classification using Time-Domain features", 28th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBS '06), 30 Aug – 3 Sep 2006, New York, NY, USA, pp. 3600–3603.

[13] A. Yang, P. Kuryloski, and R. Bajcsy, "WARD: A Wearable Action Recognition Database", 27th Annual CHI Conference, 4-9 Apr 2009, Boston, MA, USA.

[14] A. Yang, R. Jafari, S. Sastry, and R. Bajcsy, "Distributed recognition of human actions using wearble motion sensor networks", Jour. of Amibient Intelligence and Sensor Environments, Vol. 1, no. 2, 2009, pp. 103-115.

[15] D. A. Reynolds, "A Gaussian mixture modeling approach to text independent speaker identification", Ph.D. thesis, Georgia Institute of Technology, Atlanta, GA, USA, September 1992.

[16] S. Roberts, D. Husmeier, I. Rezek, and W. Penny, "Bayesian approaches to gaussian mixture modeling", IEEE Trans. On Pattern Anal. Machine Intell., vol. 20, no. 11, Nov. 1998, pp. 1133–1142.

[17] J. Bazan, "Hierarchical classifiers for complex spatio-temporal concepts", Trans. On Rough Sets IX, LNCS, Springer, 2008, Vol. 5390/2008, pp. 474-750.