# Gaussian Mixture Based HMM for Human Daily Activity Recognition Using 3D Skeleton Features

Lasitha Piyathilaka and Sarath Kodagoda
Centre for Autonomous Systems (CAS), University of Technology, Sydney, Australia
Email:Jayaweera.M.Piyathilaka@students.uts.edu.au

***Abstract*—Ability to recognize human activities will enhance the capabilities of a robot that interacts with humans. However automatic detection of human activities could be challenging due to the individual nature of the activities. In this paper, we present human activity detection model that uses only 3-D skeleton features generated from an RGB-D sensor (Microsoft Kinect $^{TM}$). To infer the human activities, we implemented Gaussian Mixture Modal (GMM) based Hidden Markov Model(HMM). GM outputs of the HMM were effectively able to capture multimodel nature of 3D positions of each skeleton joint. We test our model in a publicly available data-set that consists of twelve different daily activities performed by four different people.The proposed model recorded recognition recall accuracy of 84% with previously seen people and 78% with previously unseen people.**

## I. Introduction

Human Robot Interaction (HRI) is a sub field in robotics where interaction among human and robots is studied. It combines knowledge of many interdisciplinary fields such as robotics, Artificial Intelligence (AI), natural language processing and social sciences. The ability of robots to recognize the human activities and respond them accordingly is the key for better interaction between human and robots.

Our research focus is to develop robotic technologies to help and promote independent living for elderly people. It is motivated by the growing number of older people around the world and difficulty of finding enough care staff. Older people gradually lose the cognitive ability to keep track of their daily activities. In this context, an assistive robot that can recognize human daily activities will be immensely helpful. For example, an elderly person could be reminded of their medications in appropriate times and could follow it up until the activity has been completed. In addition, the robot may detect abnormal conditions like patient laying on the floor or sleeping longer than usual and notify the appropriate personnel.

Recent advancements in video game consoles have invented low cost RGB-D cameras like Microsoft Kinect. These cameras provide wealth of information like depth, which a normal video camera fails to provide. In addition, RGBD data from the Kinect sensor can be used to generate a Skeleton model of humans with semantic matching of 15 body parts. Since human activities are a collection of how different body parts move across each time period, these information can be used to detect human activities

In this research, we focus on the development of a probabilistic graphical model based human daily activity detection system using an RGB-D camera. We are interested in detecting daily human activities performed in home or office environments such as drinking water, brushing teeth, cooking etc. Often two individuals might perform the same activity in two slightly different ways. These variations make it difficult to generalize a machine learning technique that can train on one person and test on another.

In this research we utilized Gaussian Mixture Model (GMM) based HMM for human activity detection using only skeleton features provided by a RGB-D camera. Gaussian Mixtures are capable of clustering data into different groups as a collection of multinomial Gaussian distributions. Human actions are a collection of how different human body poses sequentially transfer at different times. Therefore, each body pose can be devised as a collection of multinomial distribution and HMM can model the intra slice dependencies between each time period. We implemented the proposed GMM based HMM using the Byasian Network ToolBox in Matlab and used Cornel activity Detection Dataset [8] to train and test the accuracy of our model.

## II. Related Work

Automatic detection of human activities is not a new research area with a large body of previous work. One popular approach uses data gathered from various sensors installed in smart environments to learn and predict human activities. Such sensor networks include motion sensors, door sensors to detect human movements and sensors installed in equipment and cupboards [4] [5]. Major challenge with such a system is the requirement of large number of sensors and their adaptability to new environments. In addition, due to the low resolution discrete nature of the data gathering the low level activities can not be observed.

Various researchers have also focused on the use of wearable sensors like gyros as a way of continuous activity detection [6]. However, the uses of wearable sensors are too intrusive discouraging the applicability. Even some forgetful old people or patients with Alzheimer might not wear the sensors nor recharge the battery, complicating the applicability.

Activity detection based on 2D video streams have been heavily exploited in the computer vision research. However, in general clutter contributes to low detection accuracies [7]. In addition, video data can be obscured due to lighting conditions, type of costumes and background colors. Further video recordings raise privacy issues in many scenarios.

978-1-4673-6322-8/13/$31.00 ©2013 IEEE

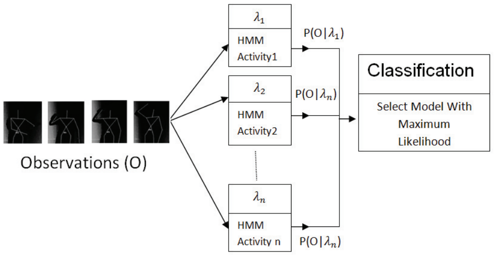


Fig. 1. Block diagram of the recognition process

In [8] two-layered maximum entropy Markov model with set of sub-activities is used to detect human activities. In that research both skeleton and 3D point cloud data are used extracting 715 features. However, their algorithm is heavily dependent on a particular sequence of sub activities to form human activities. This affects the generalization as individuals may have own ways of carrying out activities.

Use of probabilistic graphical model is one of the most popular techniques that has been used by automatic human activity detection. In [9], researchers used coupled HMM to detect human two hand activities whereas some other researchers utilized motion template together with HMM to recognize human activities [10]. However these graphical models fail to incorporate all skeleton joints in to their HMM models. Therefore such approaches can not be generalized for activities where most of the human body parts are required.

The success of any feasible human activity detection system will depend on ability to be trained on one person and test on another. These problems make automatic human activity detection very challenging. However, similar challenges exist in automatic speech recognition (ASR) research and Hidden Markov Model (HMM) has been successfully used to overcome these difficulties [11], [12]. Similar to the human activity detection ASR systems based on HMM should able to model different pronunciations that a single word can have and ASR system trained on some set of people should work reasonably well when tested on unseen set of people. Inspired by this success of the HMM in ASR, we decided to use HMM based system for automatic human activity detection.

## III. Activity Detection Model

Fig. 1 shows the overall process which is utilized in the proposed GMM based HMM for human activity detection. We trained separate HMM for each activity in the data-set. Once a sequence of skeleton features have been captured, these previously trained models produce likelihood estimation, from which the maximum is selected.

### A. Feature Selection

When HMM is applied to human activity detection, it is important to select appropriate set of observerable features. These observed features should be invariant to differences in individual sizes and body shapes. Microsoft Kinect RGB-D camera is capable of simultaneously track 15 body positions and their rotational matrices. These positions of the body parts can then be used to construct the skeleton structure of a human body. Since the human activities are a mixture of how different body parts are sequentially interacted, position information of body parts from the Kinect's human skeleton can be used to construct the HMM for human activity detection.

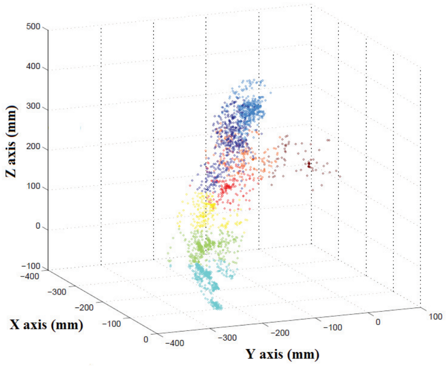


Fig. 2. (x,y,z) positions of the right hand when the action "drinking water" is performed

### B. HMM With Gaussian Mixture Outputs

Fig. 2 shows position information of the right hand when drinking water activity is performed by three people. It shows few distinguishable clusters. In addition, within each of these clusters, few sub clusters can also be observed. This is due to the subject related variations in performing even the same activity. Therefore it is challenging to model such activities. Although unimodel Gaussians are used in HMM to model continuous data, it is not capable of capturing multimodal nature of the joint movements and hence in this research we implemented HMM based on GMM.

A Gaussian mixture model is a weighted sum of M component Gaussian densities as given by the equation (1).

$$p(x|\lambda) = \sum_{i=1}^{M} w_i g(x|\mu_i, \Sigma_i) \tag{1}$$

where $x$ is a D-dimensional continuous-valued data vector , $w_i, i = 1, ..., M$, are the mixture weights, and $g(x|\mu_i, \Sigma_i), i = 1, ..., M$ are the component Gaussian densities. Each component density is a D-variate Gaussian function of the form,

$$g(x|\mu_i, \Sigma_i) = \frac{1}{(2\pi)^{D/2}\Sigma_i^{1/2}} \exp\{-\frac{1}{2}(x-\mu_i)'\Sigma_i^{-1}(x-\mu_i)\} \tag{2}$$

In our approach for human activity detection we modeled HMM as a Dynamic Baysian Network (DBN) as shown in Fig.3. A DBN is a directed acyclic graph, which represents the conditional independencies and the conditional probability distributions of each node [13]. Shaded nodes represent the observed continuous 3-dimensional joint positions ($J_n^t \; where \; 1 \leq n \leq 14, \;\; 1 \leq t \leq T$) and transparent squares represent the discrete hidden nodes. We assumed each human

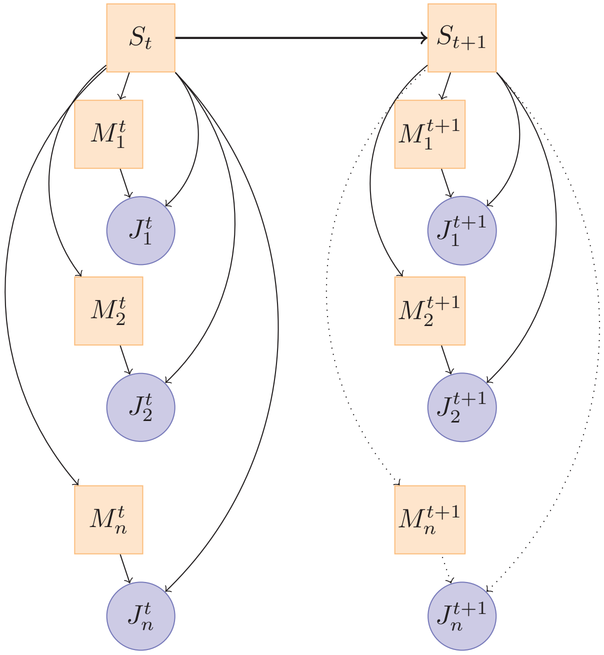


Fig. 3. 2-TBN representation of the GMM based HMM. Square nodes represent discrete hidden nodes and round nodes represent observed continuous 3-dimensional joint positions.

activity is a collection of different poses that evolves over time. Therefore, in the proposed model, top hidden node represents pose class and the middle hidden nodes represent mixture weight components. Pose classes are not directly observed as opposed to the joint positions, which can be directly measured from RGB-D camera's skeleton information.

HMM is generally parameterized by three probabilities A, B and $\pi$ as follows. First we define individual pose states as $S = \{S_1, S_2, ..., S_N\}$, the state at time t as $q_t$ and $K$ as the number of states. In the proposed model $a_{i,j}$ is the state transition probability from state *i* to state *j* and $b_t(i)$ represents the probability of the observation $O_t$ given the $i_{th}$ state of the pose nodes. Then initial state distribution, $\pi = \{\pi_i\}$ can be defined as

$$\pi_i = P(q_i = S_i),\ 1 \leq i \leq N \tag{3}$$

The observation probability distribution can be defined as ,$B = \{b_t(i)\}$ where

$$b_t(i) = P(O_t | q_t = S_i),\ 1 \leq i \leq K, 1 \leq t \leq T \tag{4}$$

$O_t$ is the joint observation at time t.

For the mixture of Gaussian approach the observation probability can be modeled as

$$b_t(i) = \prod_{n=1}^{N} \lceil \sum_{m=1}^{M_i^n} w^n_{i,m} N(O^n_t, \mu^n_{i,m}, \Sigma^n_{i,m}) \rceil \tag{5}$$

where N represents the total number of joints, $O^n_t$ the observation vector of the $n^{th}$ node at time *t*, $M^n_i$ is the number of mixture components in the joint $n$ and state $i$, and $\mu^n_{i,m}$, $\Sigma^n_{i,m}$, $w^n_{i,m}$ are the mean, covariance matrix, and mixture weight for the $n^{th}$ joint, $i^{th}$ state, and $m^{th}$ Gaussian mixture component, respectively.

Finally the state transition probability distribution can be defined as $A = \{a_{i,j}\}$

$$a_{i,j} = P(q_{(t+1)} = S_j | q_t = S_i), 1 \leq i, j \leq K \tag{6}$$

Since human activities are done in specific order additional constrained are placed on the state transition coefficient to make sure that large changes in the state indices do not occur. i.e no jump of more than one state. Also due to the repetitive nature of the human activities we only allowed sequential transition from states 1 → 2 → 3 → 1. The form of the state transition matrix is given below.

$$\begin{pmatrix} a_{11} & a_{12} & 0 \\ 0 & a_{22} & a_{23} \\ a_{31} & 0 & a_{33} \end{pmatrix} \tag{7}$$

## IV. Implementation

The proposed GMM based HMM is implemented using the Bayes Net Toolbox (BNT) for Matlab [14]. BNT is a public domain Matlab toolkit that is used to create and manipulate Bayesian networks. It supports both static and dynamic Bayesian networks. In BNT, HMM is implemented as a DBN. Therefore to specify a GMM based HMM , we defined the intra-slice topology (within a slice), the inter-slice topology (between two slices) and the parameters for the first two slices called a 2TBN. This 2TBN implementation allowed us to limit number of parameters that has to be defined. For our model we define three hidden states for pose node and three components for mixture node.

### A. Dataset

To test our proposed model we used Cornell Activity Dataset 60 (CAD 60) [8] which is published by the Cornell university. They have used the Microsoft Kinect RGBD sensor to record both depth and skeleton data of human daily activities. It is consisted of twelve unique activities done in five different environments: office, kitchen, bedroom, bathroom and living room. Data has been collected with four different people: two male and two females, recorded for about 45 seconds with each person, without compromising to any occlusion of arms and body. Therefore full skeleton was always observed throughout the activity. The Dataset also consists of a random activity of each individual, which is not similar to any other activity done before.

### B. Training GMM based HMM

When a DBN contains any hidden nodes expectation maximization (EM) algorithm can be used to train parameters [15]. However it is well known that EM algorithm only converges to a local optimum due to the none convex nature of the optimization equation. Therefore initial parameters need to be properly initialised to minimize the effect of local optimum issue. In our proposed DBN for activity recognition we used efficient method to initialize the parameters. For each activity in the dataset we considered each joint separately and trained single stream HMM with GMM output. K-mean algorithm [16] was used to initialised the mixture components and Gaussian parameters. Then we used Viterbi algorithm to determine the optimum sequence of the hidden joint states. This procedure is repeated for each joint of a given activity and optimum joint states sequences are concatenated to form a new feature vector. This feature vector is again clustered using K-mean algorithm according to the number of states of our proposed GMM based HMM. Then new sequence of observation vector was obtained by concatenating the observation vectors assigned to each state. For each state set of the pose nodes, mixture parameters and Gaussian parameters were initialized using K-mean algorithm. Finally with this initial parameters, we ran expectation maximization (EM) algorithm further fine tune our proposed model.

### C. Activity recognition

Once a HMM is trained for each action class, the idea is to select the most likely activity given some observation sequence. Therefore it can be formulated as follows. Given the observation sequence $O = O_1, O_2 ... O_t$, and model $\lambda = (A, B, \pi)$ how do we efficiently compute $P(O/\lambda)$, the probability of the observation sequence once the model is given? It is implemented with following steps.

- For each action A that need to be recognized an HMM is built, I.e., we must estimate the model parameters $(A, B, \pi)$ that optimize the like hood of the training set given observation sequences.
- Once the observation sequence is received ($O = O_1, O_2 .... O_t$), model likelihoods ($P(O/\lambda)$ need to be estimated for all possible models . Then select the maximum likelihood as the most probable activity.
  The log-likelihood calculation is done using the forward algorithm [17] for HMM. For the dataset with $V = 14$ activities, $N = 3$ states and $T = 1000$ (average) observation sequence a total of $V.N^2.T = 126,000$ computation is required for activity recognition. Clearly this amount of computation is modest as compared to the capabilities of most modern computers. Therefore the proposed classification can be done real time.

## V. Experiment

We have used two scenarios for examining the accuracy of the algorithm.

1) Unseen person: i.e. Leave one out cross validation is performed.( model was trained on three of the four subjects and tested the model on the data from the fourth person).
2) Previously seen person: One half of the dataset is used for training while other half is used for testing. The training data and testing data are not overlapped.

In the Cornel activity dataset, joint positions and orientations are recorded with respect to the sensor. Therefore we have transformed the data w.r.t the torso coordinates to alleviate the effects of the sensor location.

Further, data was mirrored to train for both “Right Handed” and “Left Handed” people. Some of the activities in the data-set even expand over more than 1000 frames and most of the activities are repetitive in the nature. Therefore, it is computationally inefficient to wait until the person completes the activity before detection.Therefore we used sliding window with a width of 100 frames with 50% overlap.

## VI. Results and Discussions

Table I shows the results of the proposed GMM based HMM human activity classifier in both “previously seen” and “previously unseen” settings. In “previously unseen” setting K-fold cross validation is carried out. We used precision and recall measures to evaluate the accuracy of the classifier. The model recorded recall/precision measure of 84/73% and 78/70% in “previously seen” and “previously unseen ” settings respectively. The model performed well even in the “previously unseen” setting indicating that HMM activity classifier has good generalizability. In addition, skeleton features are more invariant to different individuals.

Another Strength of our model is that it performed very well when tested on both left and right-handed people. This is because we trained our system both with regular and mirrored data.

Previous work, [8] that used the same Cornell activity dataset has recorded recall/precision measure of 78/86% and 57.3/64.2% in “previously seen” and “previously unseen” settings. Though our accuracies are higher than that, we can not directly compare the results due to differences in implementation and testing procedures.

Fig. 6 and 7 show the confusion matrices between “previously seen” and “previously unseen” settings. According to the confusion matrices, the proposed activity detection model often confuses among very similar activities. For example, cooking-chopping is often confused with cooking-stirring as well as drinking water is often confused with talking on the phone.

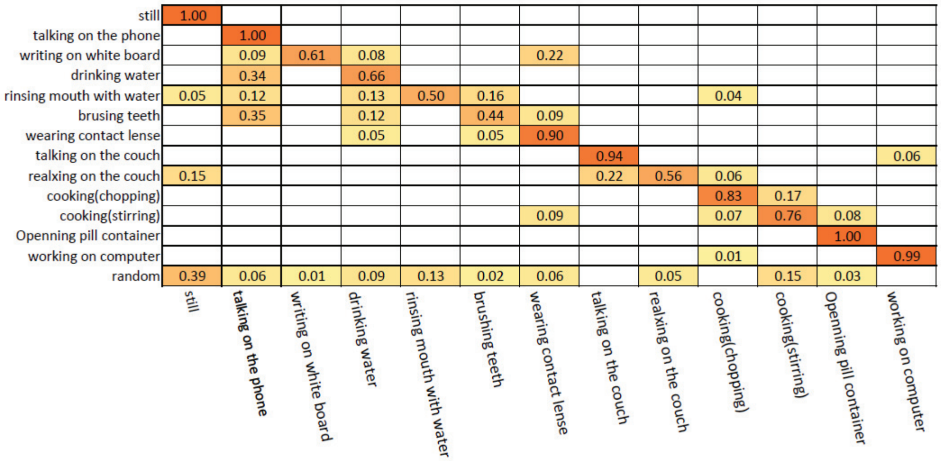
When the proposed model is tested with observation sequence that consists of random activities, as expected the algorithm classify those to the closely related model. It was observed that in most of the random activities, the subject was wandering in the room without much hand movements. Since joint positions were normalized with respect to the Torso the algorithm classify it to the most closely related model of “still”.

## VII. Conclusion

In this paper, we presented Gaussian mixture based HMM for detecting and recognizing human activities performed in an indoor environment by using only skeleton features generated from an inexpensive RGBD sensor (Microsoft Kinect). In the proposed model, Gaussian mixtures represent different position of skeleton joints and HMM models the sequential nature of the activities. We trained the proposed model for each

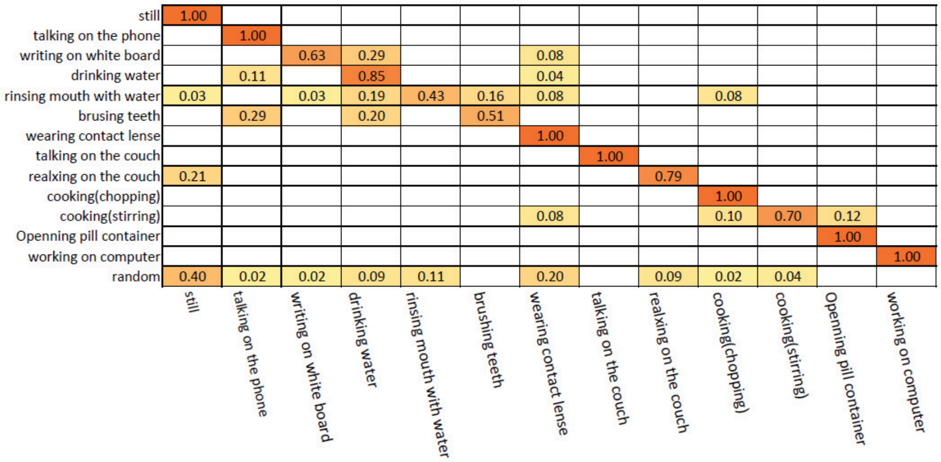
| Activity | Previously Seen | | Previously unseen | |
|---|---|---|---|---|
| | **Recall %** | **Precision %** | **Recall %** | **Precision %** |
| Still | 100 | 37 | 100 | 43 |
| Talking on the phone | 100 | 64 | 100 | 45 |
| Writing on white board | 63 | 92 | 61 | 88 |
| Drinking water | 85 | 44 | 66 | 55 |
| Rinsing mouth with water | 43 | 76 | 50 | 75 |
| Brusing teeth | 51 | 75 | 44 | 68 |
| Wearing contact lense | 100 | 31 | 90 | 49 |
| Talking on the couch | 100 | 100 | 94 | 83 |
| Realxing on the couch | 79 | 87 | 56 | 90 |
| Cooking(chopping) | 100 | 91 | 83 | 89 |
| Cooking(stirring) | 70 | 92 | 76 | 66 |
| Openning pill container | 100 | 60 | 100 | 77 |
| Working on computer | 100 | 100 | 99 | 93 |
| **Overall Average** | **84** | **73** | **78** | **70** |

TABLE I
PRECISION AND RECALL ACCURACIES FOR BOTH "PREVIOUSLY SEEN" AND "PREVIOUSLY UNSEEN" SETTINGS



| | still | talking on the phone | writing on white board | drinking water | rinsing mouth with water | brushing teeth | wearing contact lense | talking on the couch | realxing on the couch | cooking(chopping) | cooking(stirring) | Openning pill container | working on computer |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| still | 1.00 | | | | | | | | | | | | |
| talking on the phone | | 1.00 | | | | | | | | | | | |
| writing on white board | | 0.09 | 0.61 | 0.08 | | | 0.22 | | | | | | |
| drinking water | | 0.34 | | 0.66 | | | | | | | | | |
| rinsing mouth with water | 0.05 | 0.12 | | 0.13 | 0.50 | 0.16 | | | | 0.04 | | | |
| brusing teeth | | 0.35 | | 0.12 | | 0.44 | 0.09 | | | | | | |
| wearing contact lense | | | | 0.05 | | 0.05 | 0.90 | | | | | | |
| talking on the couch | | | | | | | | 0.94 | | | | | 0.06 |
| realxing on the couch | 0.15 | | | | | | | 0.22 | 0.56 | 0.06 | | | |
| cooking(chopping) | | | | | | | | | | 0.83 | 0.17 | | |
| cooking(stirring) | | | | | | | 0.09 | | | 0.07 | 0.76 | 0.08 | |
| Openning pill container | | | | | | | | | | | | 1.00 | |
| working on computer | | | | | | | | | | 0.01 | | | 0.99 |
| random | 0.39 | 0.06 | 0.01 | 0.09 | 0.13 | 0.02 | 0.06 | | 0.05 | | 0.15 | 0.03 | |

Fig. 4. Leave-one-out cross-validation confusion matrix for the "previously unseen" setting



| | still | talking on the phone | writing on white board | drinking water | rinsing mouth with water | brushing teeth | wearing contact lense | talking on the couch | realxing on the couch | cooking(chopping) | cooking(stirring) | Openning pill container | working on computer |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| still | 1.00 | | | | | | | | | | | | |
| talking on the phone | | 1.00 | | | | | | | | | | | |
| writing on white board | | | 0.63 | 0.29 | | | 0.08 | | | | | | |
| drinking water | | 0.11 | | 0.85 | | | 0.04 | | | | | | |
| rinsing mouth with water | 0.03 | | 0.03 | 0.19 | 0.43 | 0.16 | 0.08 | | | 0.08 | | | |
| brusing teeth | | 0.29 | | 0.20 | | 0.51 | | | | | | | |
| wearing contact lense | | | | | | | 1.00 | | | | | | |
| talking on the couch | | | | | | | | 1.00 | | | | | |
| realxing on the couch | 0.21 | | | | | | | | 0.79 | | | | |
| cooking(chopping) | | | | | | | | | | 1.00 | | | |
| cooking(stirring) | | | | | | | 0.08 | | | 0.10 | 0.70 | 0.12 | |
| Openning pill container | | | | | | | | | | | | 1.00 | |
| working on computer | | | | | | | | | | | | | 1.00 |
| random | 0.40 | 0.02 | 0.02 | 0.09 | 0.11 | | 0.20 | | 0.09 | 0.02 | 0.04 | | |

Fig. 5. Confusion matrix for the "previously seen" setting

activity in the data-set and maximum log-likelihood estimation is carried out in-order to select the most probable model for a given sequence of observations. The proposed algorithm was tested with both “unseen” and “previously seen” settings with very promising accuracies. It is also shown to be robust to inter individual and intra-individual variations even with left or right hand usage.

## VIII. Future Works

The proposed algorithm was tested on a publically available data set, which has apparently assumed full body observability. We are planning to collect a more challenging dataset with occlusions for further development of the algorithm.

## References

[1] T. Theodoridis, A. Agapitos, H. Hu, and S. M. Lucas, “Ubiquitous robotics in physical human action recognition: A comparison between dynamic anns and gp.” in *ICRA*. IEEE, 2008, pp. 3064–3069.

[2] Y. Demiris and A. Meltzoff, “The robot in the crib: A developmental analysis of imitation skills in infants and robots.” *Infant Child Dev*, vol. 17, no. 1, pp. 43–53, 2008.

[3] M. Lopes, F. S. Melo, and L. Montesano, “Affordance-based imitation learning in robots,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, USA, Nov 2007, pp. 1015–1021.

[4] S. Bang, M. Kim, S.-K. Song, and S.-J. Park, “Toward real time detection of the basic living activity in home using a wearable sensor and smart home sensors,” *Conference proceedings Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2008, pp. 5200–5203, 2008.

[5] E. M. Tapia, S. S. Intille, and K. Larson, “Activity Recognition in the Home Using Simple and Ubiquitous Sensors,” *Pervasive Computing*, vol. 3001, pp. 158–175, 2004.

[6] D. T. G. Huynh, “Human activity recognition with wearable sensors,” Ph.D. dissertation, TU Darmstadt, September 2008. [Online]. Available: http://tuprints.ulb.tu-darmstadt.de/1132/

[7] A. Gupta, P. Srinivasan, J. Shi, and L. S. Davis, “Understanding videos, constructing plots learning a visually grounded storyline model from annotated videos,” *2009 IEEE Conference on Computer Vision and Pattern Recognition*, vol. 21, no. 4, pp. 2012–2019, 2009.

[8] J. Sung, C. Ponce, B. Selman, and A. Saxena, “Human activity detection from rgbd images.” in *Plan, Activity, and Intent Recognition*, vol. WS-11-16. AAAI, 2011.

[9] M. Brand, N. Oliver, and A. Pentland, “Coupled hidden Markov models for complex action recognition,” *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 0, pp. 994–999, 1997.

[10] F. Martinez-Contreras, C. Orrite-Urunuela, E. Herrero-Jaraba, H. Ragheb, and S. A. Velastin, “Recognizing Human Actions Using Silhouette-based HMM,” *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*, pp. 43–48, 2009.

[11] A. V. Nefian, L. Liang, X. Pi, X. Liu, and K. Murphy, “Dynamic Bayesian Networks for Audio-Visual Speech Recognition,” *EURASIP Journal on Advances in Signal Processing*, vol. 2002, no. 11, pp. 1274–1288, 2002.

[12] A. V. Nefian, L. Liang, X. Pi, L. Xiaoxiang, C. Mao, and K. Murphy, “A coupled hmm for audio-visual speech recognition,” in *in International Conference on Acoustics, Speech and Signal Processing (CASSP?02*, 2002, pp. 2013–2016.

[13] K. Murphy, “Dynamic bayesian networks: Representation, inference and learning,” Ph.D. dissertation, UC Berkeley, Computer Science Division, Jul. 2002.

[14] K. P. Murphy, “The bayes net toolbox for matlab,” *Computing Science and Statistics*, vol. 33, p. 2001, 2001.

[15] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the em algorithm,” *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, vol. 39, no. 1, pp. 1–38, 1977.

[16] J. A. Hartigan and M. A. Wong, “A K-means clustering algorithm,” *Applied Statistics*, vol. 28, pp. 100–108, 1979.

[17] L. Rabiner, “A tutorial on hidden markov models and selected applications in speech recognition,” *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257 –286, Feb. 1989.