GENERALIZATION CAPABILITY OF A WEARABLE EARLY MORNING ACTIVITY DETECTION SYSTEM

Cheol-Hong Min, Nuri F. Ince, and Ahmed H. Tewfik

Department of Electrical and Computer Engineering, University of Minnesota
200 Union St. SE, 55455, Minneapolis, U.S.A.
phone: + (1) 612-625-5006, fax: + (1) 612-625-4583, email: cmin@umn.edu

ABSTRACT

In this paper, we study the generalization capability of a classifier system which can detect, classify and monitor the activities of daily living for assisting patients with cognitive impairments due to traumatic brain injuries. Generalization implies that the system does not need subject specific training or minimal training, if needed, when the system is deployed in a home setting. We briefly describe the infrastructure of a cost-effective system and show initial applications in detecting activities executed in the early morning. A set of in-home fixed wireless sensors and wearable wireless sensors were used to detect the activity of the user. Both time and frequency-domain features were extracted and used to classify activities using Gaussian Mixture Models post processed with a Majority Voter. We show promising experimental results from 7 subjects while completing washing face, shaving face and brushing teeth activities. We compare results from intra subject classification study with inter subject classification study and show the generalization capability of our wearable system to detect several early morning activities.

1. INTRODUCTION

According to the US Center for Disease Control (CDC), Traumatic Brain Injury (TBI) contributes substantially to the death and permanent disabilities annually. It is estimated that 2% of US population which is about 5.3 million people are suffering from long term disabilities and require assistance with their Activities of Daily Living (ADL). TBI can be classified as “mild” or “severe” depending on the injury and its impact on the patient. Severe injuries can cause long term amnesia or unconsciousness while the mild injuries cause short term amnesia or unconsciousness. Statistics show that about 75% all TBI cases are classified as mild forms of injuries [1].

The top 3 leading causes for TBI are falls, motor vehicle related accidents and struck by/against accidents. During these accidents, TBI is caused by a sudden blow, jolt or a penetrating injury to the head that damages cognitive functions of the brain. Thus, TBI patients have difficulties with remembering, concentrating or making decisions. They also get lost and are easily confused due to the disability in cognitive functions [1, 2]. Patients diagnosed with Alzheimer’s diseases or elderly population also have similar cognitive disabilities [1] where they have difficulties in remembering events or instructions.

Most patients are diagnosed and receive treatment at hospitals but when return to their homes, they cannot receive the amount of care and the necessary treatment due to the high cost. The cost is directly related to shortage of professionals, caregivers and also family burn outs. Therefore, the assistive system we propose to develop will decrease the above mentioned cost and allow TBI/cognitive impaired patients to lead an independent life.

To assist the people with cognitive disabilities, several scheduler or reminder systems that give instructions were developed. They were implemented on mobile devices that deliver reminders to the patients based on scheduled times. Once the reminder is issued, its roll as an assistant is completed. Therefore, it does not know if the patient has actually executed the task. To overcome this problem, researchers have designed a robot, called Autominder, that lives and moves with the patient in a home setting, which can track limited activities with its on-board sensors [4]. But, since this is an intrusive system some people may prefer not to have this kind of system. Thus, we have proposed to work with set of sensors which are non-intrusive to track whether the patient has executed the task or not [5, 6].

By detecting and monitoring the activities, the system can give instructions to the cognitively disabled only “when it is needed”. Static sensors detect the location and coarse level of activity, while the wearable sensors are used to monitor the physical parameters and execution of different activities at a fine level. Finally the data collected from both sensor networks are processed by intelligent algorithms which can drive a reminder system to give dynamic feedback to the subject only when needed. Such a feedback can range from reminding the patient to take their medicine, finish shaving or lunch preparation to giving an emergency call to their caregiver. In this study, we focus on detecting several early morning activities such as washing face, brushing teeth and shaving.

The main purpose of this paper is to extend the previous studies by exploring the generalization capability of the proposed system. Thus, we briefly describe our system architecture and current data collection methodology. We extend our previous work to extended number of subjects. As the number of subjects increase, more variation in data and overlap across activities are introduced. We discuss the problems

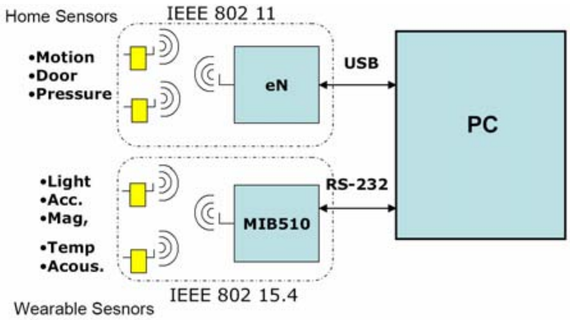
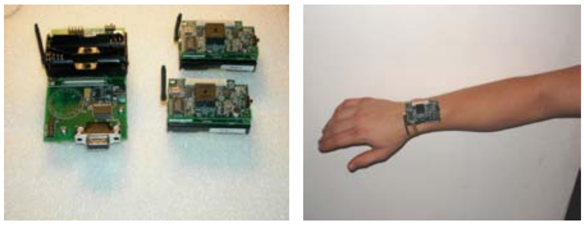


Figure 1 - The schematic diagram of the data acquisition and monitoring system. Shown on the left are the wireless networks to collect in home (eN) and wearable sensor data (xBow). The sensor readings are transmitted to a PC via the eN base station and MIB510 gateway.



(a) (b)

Figure 2 - The xBow wireless kits (a) on the left the serial gateway, on the right the MicaZ motes with MTS310 sensor boards installed. (b) mote kit attached to the wrist to collect data.

encountered during the study and methods to solve the problems and explore the generalization capability of our classification system. Lastly, we discuss about post processing the data by analyzing the behavioural patterns observed across the subjects while executing an activity. In Section 2, we give a brief overview of our system. In section 3, we discuss classification approach and show initial experimental results for generalization for data we collected and analyzed from healthy subjects to prove the concept. Finally, we close the paper with conclusion and remarks.

2. SENSOR SYSTEM INFRASTRUCTURE

The system developed by our group shown in Fig. 1. The Data Acquisition (DAQ) platform combines two sensor systems. The first sensor system is a collection of static wireless sensors while the second sensor system relies on wearable sensors that provide data which complements those collected by the static sensor system. The data recording method we used was "trial based" where subject's each activity was recorded independently based on different tasks and trials. Trial method was chosen since subjects did not feel comfortable with video and audio recordings of the system which was initially proposed in [5].

The use of in-home static sensors such as motion and contact sensors can give the location of the individual and coarse information of events involving the individual. In our system, we used the eNeighbor™ system (eN) that was recently developed by RedWing Technologies [6] and which is currently marketed under the name Healthsense™. The system is equipped with several sensors such as a motion, bed, chair, door, flow and contact sensors which enables to track most of the daily activities in broad range. Each sensor communicates with the base station only in the case of an event and this binary output is received by the base station and exported in real time through the USB port to an external device for storage. We developed a driver to capture the messages transmitted from the base station and save these messages with a time stamp to synchronize with the other sensors in the remaining system.

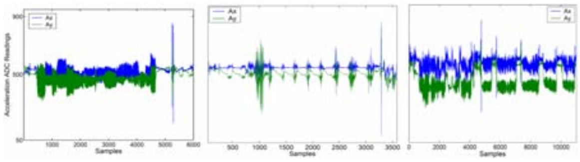
To obtain detailed information about the activity of the subject, we use wearable 2-axis accelerometer which is attached to the right wrist. We use the MTS-310 and MICAZ mote kits to collect and wirelessly transmit accelerometer data developed by Crossbow Technology Inc. [7]. The accelerometer on the MTS-310 multisensor board also have other sensors such as: light, temperature, 2-axis magnetometer and a microphone (see Fig. 2 (a)). Although, we had collected all sensor data, in this study, we only analyzed accelerometer data. Data was sampled at 50Hz, digitized with 10-bit on-board A/D, transmitted to the gateway and transferred to the PC from MIB-510 gateway (base station) in real-time. The mote kits are small and can be attached to body and wrists to record movements (see Fig. 2(b)).

3. DETECTION, CLASSIFICATION AND MONITORING OF ACTIVITIES

We developed a system that uses hierarchical approach for detecting, classifying and monitoring the activities of daily living (ADL). First, we localize the subject within a specific room in a home using the motion sensors. Then, we might be able to detect that the person is in front of a sink running the water which can be detected by the flow sensors. This allows us to constrain the list of most likely activities that the subject can be engaged in that particular location. Finally, the wearable sensors detect the body and/or limb motions while the system analyzes, classify and monitor the progress of ADL.

In particular we focused in the classification of 3 ADLs. Recorded tasks were washing face, brushing teeth and shaving face, which are closely related to personal hygiene. Authors in [8] classified ADLs into 3 different categories such that personal hygiene (bathing, toileting, etc) and personal nutrition (eating) were classified as Basic ADL which every people living independently must be capable of performing. We elected these 3 activities above others since they were shorter in duration and were more suitable for initial studies.

Data collection from TBI patients is difficult and it is known that most of TBI patients are not physically disabled. Therefore, in this initial phase of our work, we elected to use healthy subjects to study the feasibility of the concept. After successful initial study conducted on 2 healthy subjects [9], we extended the study for 7 healthy subjects with the system described above and explore the possibility for generalization capability of the system. We plan on continuing to design, refine and test the system with data collected from healthy



(a) (b) (c)

Figure 3 - Typical recordings obtained from 2 channel accelerometer sensor attached to the right wrist. (a) Brushing-Teeth (b) Washing-Face and (c) Shaving-Face. In each figure, horizontal axis correspond to time samples and vertical axis correspond to amplitude of the recorded waveforms.

subjects. Once we achieve an acceptable performance and level of friendliness, we will test on TBI patients and collect additional data to tailor the system to them.

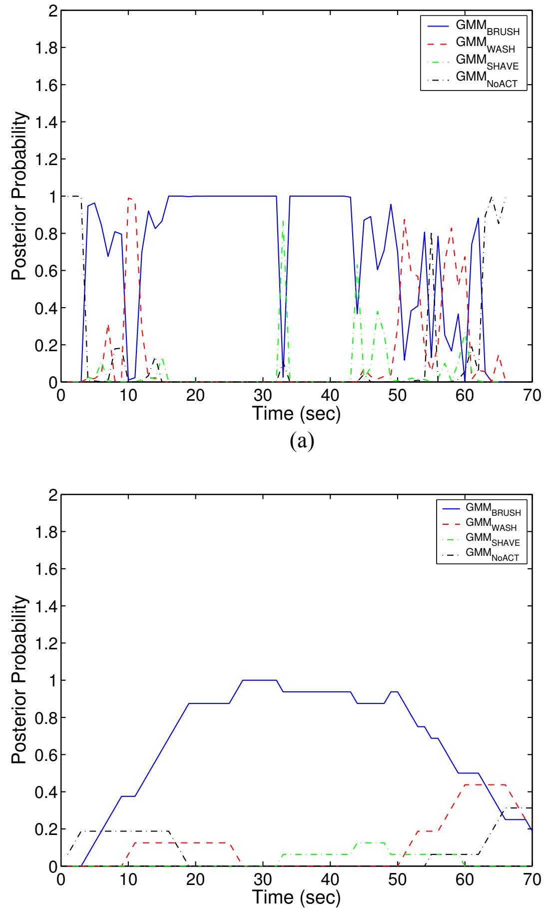
During the data collection, each subject freely executed the task and in total, 129 brushing, 142 washing and 91 shaving trials were recorded for this study. In addition to the 3 distinct tasks, the subjects also entered the bathroom and executed other types of activities that did not correspond to the 3 tasks listed above. These can be for example changing a towel, combing, drying hair, applying lotion to one's face or arranging items on the sink etc. 20 trials were recorded and are categorized under Other-Activity (OAct).

3.1 Classification Model

There are several methods for generating activity state models and ADL classification methods [10, 11]. But for this study, we used Gaussian mixture models (GMM), to model the above mentioned 4 classes [9]. We extracted 6 time domain (TD) and 10 frequency domain (FD) features for each task and combined them with GMM. For each accelerometer data, mean, root mean square and number of zero crossings were calculated in 64 sample time segments as TD features. Then, we extracted FD features by analyzing energies in 5 different frequency bands in which Fourier Transform was used for the same 64 sample time segments. FFT window was shifted with 50% overlap across the signal. For each segment, we calculated the energy in dyadic frequency bands as indicated in [9]. The FD features are then converted to log scale and combined with TD features. This resulting feature vector now has a dimension of 16 for each 64 sample time segments. Separate GMMs were generated for each activity. Finally, the outputs of all GMM were post processed by a Majority Voting (MV) procedure as indicated in [9]. The MV uses 16 points (≅10s) window to decide whether the observation sequence is related to any of the tasks.

3.2 Generalization Studies

Here, we study 2 different classification methods. The first method is using Leave One Trial Out (LOTO) cross validation method which we call intra subject training. Basically, for a single subject we use one trial for testing and all remaining trials for training the classification system. The second method is Leave One Subject Out (LOSO) cross validation method. In this method, data from a subject is left out and other subject's data is used to train the system.



(a)

(b)

Figure 4 - (a) The outputs of the GMMs related to brush activity of S1. Note that number of local FPs exists in the beginning and at the end of the task. (b) The output of MV post processing stage. Note that all local errors in (a) are now removed.

Study conducted on 2 subjects showed the False Positives (FP) mostly occurred at the beginning and end of each activity. Beginning and end stage of each activity is likely to overlap since the subject is getting ready to start or finish an activity where not much distinct motion of the arm is observed or where motions overlap across activities during these times. This was also verified by observing the overlap in the feature domain. Therefore, classification distances between different activities are greatest when the subjects are in the middle of executing the activity (see Fig. 4). We used this information to modify the MV to maximize the number of true positives (TP) while minimizing FP. We also noticed that the arm movements related to each task is composed of many stages where the activity is not locally related to the task under execution. For example, brushing is composed of applying toothpaste to the brush, actual brushing action and finally rinse-off stage. Similarly, shaving is composed of applying shaving cream to one's face, actual shaving and washing the face. In our studies, we vary the GMM model

Table 1 – MV output for intra subject classification in %

Mixture	Brush	Wash	Shave	Average
1	98.4	92.2	96.7	95.7
2	99.2	93.6	96.7	96.5
3	100	97.9	98.9	98.9
4	100	93.6	98.9	97.5
5	100	94.4	98.9	97.8

Table 2 – MV output for inter subject classification in %

Mixture	Brush	Wash	Shave	Average
3	96.1	64.1	78.0	79.4
6	95.3	64.8	85.7	81.7
9	94.6	68.3	83.5	82.1
12	94.6	66.9	90.0	83.9
15	95.3	66.9	85.7	82.6

order by varying the number of mixtures to obtain optimal classification accuracy across subjects and activities.

4. RESULTS

In this section, we summarize our findings in 3 different subsections through observations and experiments and discuss the generalization capability of a wearable early morning activity detection system.

4.1 Optimal Number of Mixtures

It is known that a low number of mixtures poorly represent the geometry of the activity in a D-dimensional space, whereas a high number of mixtures generally over fit the data. Therefore, when we increase the mixture number, then the classification accuracy starts to increase. After a certain number of mixtures the classification accuracy starts to drop. Based on our experiments for intra subject classification, we were able to confirm these statements and found that varying the number of mixtures from 1-5 was good enough to obtain the best classification accuracy. But for inter subject classification, best results were found at higher mixtures. This is because more variation had to be considered to represent the geometry in the feature domain which directly required increase in the number of mixtures. Therefore, in the inter subject classification, best classification accuracy was found at around 12 mixtures. In Table 1, we show the classification results for different mixtures. Here we observe that the data for each activity is highly correlated within a subject. Even at low number of mixtures, classification rate is already close to 100%. Therefore, we keep the number of mixtures low in order to reduce the complexity of the system. Currently, we are using only the accelerometer data but depending on the activity, such data might not be sufficient to detect the activity in execution. Thus, data from several sensors and platforms will need to be analyzed which will be costly in terms of complexity.

4.2 Observed Issues in System Generalization: Unique and Overlapping Patterns

Our classification results are quite promising as indicated Table 1 which provides classification results for intra subject training for 3 early-morning ADLs obtained from accelerometer data. It is in accordance to our previous study which

Table 3 – Classification result from S3. MV output for inter subject classification in number of trials.

Subject 3	Brush	Wash	Shave
Available Trials	12	9	11
Classification_1	12	1	9
Classification_2	12	6	10

Table 4 – Classification result from S4 for different MV window size.

Subject 4	Brush	Wash	Shave
Available Trials	17	17	-
MVwin=32	17	2	-
MVwin=16	17	15	-

Table 5 – Classification result from S5 for different MV window size.

Subject 5	Brush	Wash	Shave
Available Trials	13	16	10
MVwin=32	9	15	10
MVwin=16	4	16	8

involved 2 subjects [9]. In this paper, we extend our work and study the potential for generalization of the classification system. Therefore, analysis was also performed across subjects in Table 2, such that each subject data was left out during training stages and was only used for testing the system. This method is called inter subject training in our work. Our initial observation for inter subject training showed that the classification rates were much lower compared to that of the Table 1. We believe this is due to two major issues. First, it was due to the overlap in the feature domain especially during the beginning and end stages of various activities. But for several subjects, it was also observed during the execution of main-stage in the washing activity for subject 3. His circular arm motion while applying soap to his face was very similar to brushing motion of other subjects. This effect is shown as "Classification_1" in Table 3 in which only 1 of his 9 trials were classified correctly. Second issue was that the patterns of the main stage in a task were different from subject to subject, i.e. the uniqueness issue. Therefore when subject in testing data was left out, the model did not represent that subject's activity very well which drops the overall classification result.

In addition to uniqueness problem, we noticed that there are 2 different groups of people that have different patterns for washing the face. We verified the first hypothesis by dividing the washing activity into 2 groups. Although patterns are different from subject to subject, most washing tasks fall into 2 groups in which the subject fully draws water onto his hands before washing his face and others would simply splash the water after a quick drawing of the water. This was verified by collecting and observing wash data from additional 17 subjects. We trained the system using one group of data and applied to the other group for testing and vice versa. The overall classification rate was about 30% which proves that one group of washing data cannot represent the other group data while overall classification rate for within group training was 91%. This proves that if a similar pattern to a subject's data exists in the training set, it will increase the classification rate. But, if the pattern is unique for a given activity, similar patterns need to be in the training set for the system to classify different uniqueness in the activity. The solution is to obtain finite number of trials from the subject in testing. We tested the hypotheses on subject 3 by adding his data to the training set. Ratio of his data was about 5 % of total data in training set. Including his own data, we were able to increase the number of correct classification from 1 to 6 listed under "Classification_2". This shows possibility of collecting a finite number of subject's data to the system. The system then would be able to detect the unique or overlapping patterns of that subject and will be able to classify his activities which in tern can bring more confidence to classification result. Obtaining the amount of finite number is still an open question and authors are currently exploring the optimal number or ratio to be included for testing.

We are also studying and analyzing the behavioural patterns of subjects while they go through each activity. We observed that people go through different sequences of substages to start and complete an activity. For example, one subject would shake the shaving cream can from time to time, while others don't and other subject does not apply any shaving cream while shaving. As mentioned previously, these unique and overlapping patterns create problems for the classifier. Therefore, another possible solution is to break each activity into several sub-stages so that differences across the subject can be modelled to accommodate different sequences. This would allow the system to track the activity and observe the current stage of an activity. Through these processes, we expect the classification accuracy to improve over the current 83.9% that we achieved.

4.3 MV Window Length

The size of the MV window also plays an important role. We show their effect in Table 4. As we can see from the washing data for subject 4, most of them were unclassified due to the length of the window size. This corresponds to 20 seconds of time duration but since this subject quickly splashes the water onto the face, observing over a longer period time averages out the activity detected by the classification system. On the contrary, Table 5 shows a different result where longer window captures the global information and shorter window fails. As we have shown, short time windows are prone to local error and could miss global information while using longer time window local information can be lost. Finding optimal fixed parameters is a difficult iterative process. Therefore authors are currently studying dynamic window sizes to make final classification decisions.

Lastly, current activity detection is based on traditional brushing and shaving activities. If a subject uses electrical brush or shaver, arm movement detection using accelerometer will be a challenging problem. Thus we also explore add ing a different sensor platform for sound detection and feature extraction.

Finally, we show the confusion matrix in Table 6. Due to the reasons discussed in previous sub sections, only 95 trials out of 142 washing activities were correctly classified,

Table 6 – Confusion matrix for different activities for mixture 12.

Tasks	Brush	Wash	Shave	OAct
Brush	122	0	0	7
Wash	32	95	6	9
Shave	9	0	82	0
OAct	0	0	1	19

while 32 were misclassified as brush and 6 were misclassified as shave. The goal of the future studies would be to reduce the number of FPs while increasing TPs in classification.

5. CONCLUSION

In this paper, we described a system which is intended to assist people with cognitive impairments due to TBI. In particular, we focused on the problems of detecting, classifying and monitoring the progress of activities of daily living at home. We explored the potential for system generalization in which several issues and possible approaches were proposed to solve these issues. We showed experimental results from 7 subjects while completing washing face, shaving and brushing activities. Our preliminary results using accelerometers are quite promising and with full integration with additional post processing stage and using other sensors such as light, temperature along with additional sensor kits such as sound on other parts of the body will deliver more information of the activity in execution. Integration of proposed classification system with a reminder and planner module may open a new technology to assist TBI and other cognitive impaired patients by allowing them to continue their independent life.

REFERENCES

[1] National Center for Injury Control and Prevention, http://www.cdc.gov/ncipc/tbi/TBI.htm. 2006

[2] "Traumatic Brain Injury", Department of Veterans Affairs Employee Education System, 2004

[3] R. Levinson, "The Planning and Execution Assistant and Trainer". *Journal of Head Trauma Rehabilitation*, April, Aspen Press, 1997.

[4] M. E. Pollack, "Planning technology for intelligent cognitive orthotics". *In Proceedings of the 6th International Conference on Automated Planning and Scheduling*, 2002.

[5] one or more of the authors

[6] Redwing Technology, HealthSense http://www.healthsense.com

[7] Crossbow Technology Inc, http://www.xbow.com

[8] E. D. Mynatt, I. Essa and W. Rogers, "Increasing the Opportunities for Aging in Place", *Proceedings of ACM Conference on Universal Usability,* 2000

[9] one or more of the authors

[10] J. Parkka, M. Ermes, P. Korpipaa, J. Mantyjarvi, J. Peltola and I. Korhonen, "Classification Using Realistic Data From Wearable Sensors", *IEEE Trans. on Info. Tech. in Biomed.,* vol. 10, no. 1, Jan. 2006.

[11] J. Lester, T. Choudhury, N. Kern, G. Borriello and B. Hannaford, "A Hybrid Discrimitive/Generative Approach for Modeling Human Activities", *IJCAI 2005: 766-772*