Human Daily Activity Recognition for Healthcare Using Wearable and Visual Sensing Data

Xi Liu[1], Lei Liu[2], Steven J. Simske[3], Jerry Liu[2]
[1] Michigan State University
[2,3] Hewlett Packard Laboratories
[1] East Lansing, MI, USA; [2]Palo Alto, CA, USA; [3]Fort Collins, CO, USA
liuxi4@cse.msu.edu, {lei.liu2, steven.simske, jerry.liu}@hp.com

***Abstract*—Wearable digital self-tracking technologies for monitoring individuals' health condition have become more accessible to the public in recent years with the development of connected portable devices, such as smart phones, smart watches, smart bands, and other personal biometric monitoring devices. Mining behavioural patterns from such wearable data along with other available sensory data, has the potential to offer an objective, insightful service in clinical professionals and healthcare. For example, accurate identification of human activities could help us provide a better patient recovery training guidance, or an early alarm of emergency that may happen to elder people, such as stroke, falls, etc. In this paper, we introduce an activity recognition system, which learns a nonlinear SVM algorithm to identify 20 different human activities from accelerometer and RGB-D camera data. Our early experimental results show that the proposed approach is promising and effective.**

I. Introduction

[1] Sensing technologies, such as human sweat sensing for non-invasive monitoring individuals' health state physiological information; wearable digital self-tracking technology for monitoring individuals' health condition; have become more accessible to the public in recent years with the development of connected portable devices, human biosensors, platform or systems that are particularly designed for monitoring, storing and analyzing human self-tracking sensory data. The proliferation of such technology has made it much easier than any time before to collect biological and physiological sensor signals, such as electrocardiogram (ECG), oxygen saturation (SpQ2), heart rate (HR), electroencephalogram (EEG), Galvanic Skin Response (GSR), blood pressure/oxygen level, body temperature, accelerometers, etc. By monitoring and analyzing such sensory data, it helps us better understand each individual's health condition. As a result, mining wearable sensory data for healthcare purpose has been gaining significant attention from both industry and academia in recent years.

Among all health issues, falls, stroke, obesity, cardiovascular disorders and musculoskeletal diseases are the biggest issues and the fastest rising categories of healthcare costs. Real time monitoring, analyzing and detecting human activities provides an opportunity of offering the patients better recovery guidance and early alarm of healthcare emergency.

In this paper, we aim to recognize 20 different human activities from wrist-worn accelerometer and RGB-D camera data, both of which are generated from mature sensing technologies, that are broadly built in off-the-shelf smart devices. The human activities that we considered can be generally cast into three groups: motions, stationary postures, and transition actions.

- **Motions**: activities requiring continuing movement, including *ascend stairs, descend stairs, jump, walk with load* and *walk*
- **Stationary postures**: times when the participants are stationary, including *bending, kneeling, lying, sitting, squatting* and *standing*
- **Transitions**: describe posture-to-posture transition activities, including *stand-to-bend, kneel-to-stand, lie-to-sit, sit-to-lie, sit-to-stand, stand-to-kneel, stand-to-sit, bend-to-stand* and *turn*.

Given wrist-worn accelerometer and RGB-D camera data that is associated with each single second, our goal is to predict the probability of each human activity within each target second. To summarize, we have the following five major challenges in this task.

1) Due to the nature of sensing platform, missing values appear in both accelerometer and RGB-D camera data (as shown in Figure 2 and Figure 3, respectively). The strategy of directly discarding seconds that contain missing values will leave us too few examples to train an effective predictive model. How to effectively handle such data missing issue is the first challenge we need to solve.
2) Human activities happen in continuous time, and a single second may be associated with more than one activities. Therefore, instead of assigning only a single activity to each second, we need to learn a model, which has the capability of assigning probability scores to all possible activities within each target second.
3) Temporal relations between different activities is an crucial information that we should not discard. For example, *lying* rarely happens next to *jumping*; *sitting* rarely happens next to *ascending stairs*; and *lie-to-sit* usually happens within *lie* and *sit*.How to effectively incorporate such temporal relationships in our algorithm is also a challenge.
4) Furthermore, location information is important to our task, and how to effectively import such prior knowl-

[1]This work was completed when Xi was an intern at HP Labs

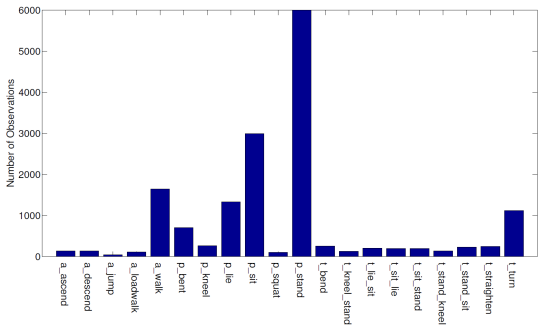


Fig. 1: The activity classes distribution.

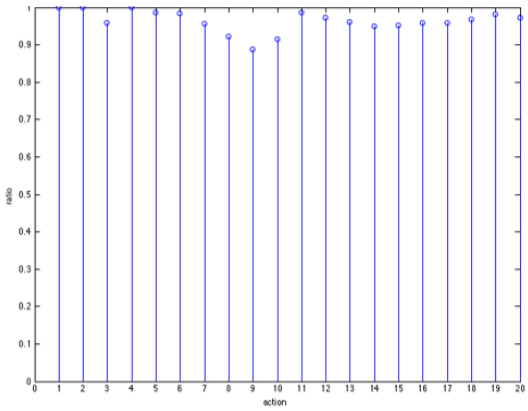


Fig. 2: Ratio of time where acceleration Data is available.

edge of activities at different locations is crucial. For example, it is more likely that the subject is ascending or descending the stairs rather than other activities on the stairs.

5) Finally, there exits serious class unbalance problem in such sensory data, because in a majority of time, the observations of activities belong to the a few largest classes such as *sitting, standing* than small classes, such as *squat, bend*. Take an example in Figure 1, which describes the class distribution over 20 classes. Due to this class unbalance problem, standard classifier tends to be more biased towards assigning observations into larger classes rather the smaller classes, which however are usually more of our interest. How to deal with such class imbalanced data is another challenge that must be addressed.

The remainder of this paper is organized as follows.Section II introduces the technology that we employed for activity recognition. Section III-A presents the preliminary of data exploration. Experimental results and discussions are provided in Section III Section IV presents the related work of this paper. We conclude our works and discuss a few future works in section V.

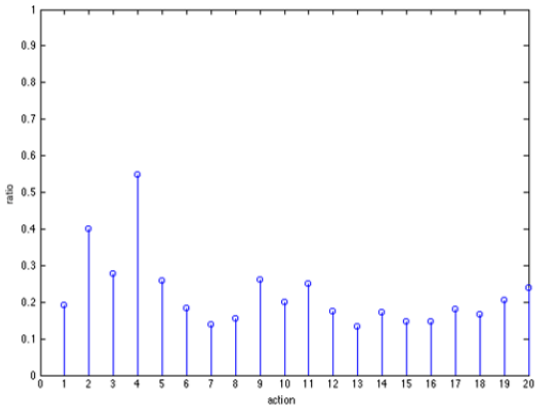


Fig. 3: Ratio of time where RGB-D Camera Data is available.

II. Methodology

A. Sensory Data Collection

In this paper, we investigated the feasibility of using data generated from two sensors namely accelerometers and RGB-D cameras for human activity recognition. Both sensors are cheap and widely embedded sensors in current smart phone, fitness band and motion sensing input devices, such as Xbox 360 and Asus Xtion Pro.

- **Accelerometers**: We use a triaxial accelerometer to record the real-time acceleration in all three spatial dimensions (X-Y-Z) in real time, which provides useful information for activity recognition.
- **RGB-D sensor**: RGB-D sensors combine RGB color information with per-pixel depth information. In this paper, we record 3D image sensory data, which includes both frames of RGB images and the depth of a subject. We can easily capture and extract the moving subject with a bounding box from the raw image frames with various libraries and SDKs, such as OpenCV [1], OpenNI[2], point cloud library (PCL)[3] and Microsoft Kinect SDK[4]. The change of the coordinates and shapes of the bounding box provides helpful information to differentiate the motion or posture of a subject.

B. Feature Extraction

In this paper, our target is predicting human activities within each single second. However, classification algorithms cannot be directly applied to raw time-series accelerometer and camera data. We firstly need to transform the data into data samples. To accomplish this we divide the data into 1-second-length segments of time series, and then generate features that is capable in discriminating different human activities. The features generated from each sensor are described below.

1) Accelerometer: We generated the following total of 12 features from raw triaxial accelerometer data.

- Kurtosis [5]: Describes the tailedness of a time-series data.

- Approximate Entropy [6]: Describes the unpredictability of fluctuations over time-series data.
- Top-10 Frequency by FFT [7]: Distribution of the resultant time-series energy in frequency domain.
- FFT distribution kurtosis: Describes tailedness of the resultant energy distribution in frequency domain.
- Average Jerk [8]: Average rate of change of acceleration on each axis.
- Average Absolute Value: Average absolute acceleration on each axis
- Average Value [7]: Average acceleration on each axis.
- Median: Meidian acceleration on each axis.
- Standard Deviation [7]: Standard deviation of acceleration on each axis.
- Maximum Value: Maximum acceleration on each axis.
- Minimum Value: Minimum acceleration on each axis.
- Maximum Absolute Value: Maximum absolute acceleration on each axis.

The above 12 features are selected because of their capability of distinguishing our 20 target human activities. For example, from Figure 4, we can see that class "a_jump" gets much higher maximum acceleration than any other activities. For another instance, while most of the activities don't repeat periodically, there are still some activities which show obvious repetitive patterns. With Fast Fourier Transform(FFT), it can be observed that their distributions of energy in frequency domain (Figure 8a and 8b) are apparently different from those who do not have periodic patterns (Figure 8c and 5d). Most of these repetitive activities get relatively higher energy in low-frequency bands, while others get comparable energy in all frequency bands.

2) RGB-D Camera: We use OpenNI [2] to extract both 2D and 3D bounding boxes of a human subject from raw RGB-D image frames. As shown in Figure 6, the coordinates of the centers reflect the detailed location of a human subject, and can be further used to compute displacements and speeds. The coordinates of the corners of bounding boxes can be used to compute the shape of a subject. It is quite straightforward to associate these features with human activities. For example, when the coordinate of the bounding box center moves fast, subjects are more likely to perform motions rather than staying in stationary postures; the shape of a subject is apparently different when the subject is standing compared against when he/she is sitting, etc.

Based on the above observations, we have extracted both 2D and 3D movement and shape features. The detailed camera features are provided in Table I.

The accelerometer and RGB-D camera have their distinguished advantages and disadvantages, which are irreplaceable by each other. The advantage of the accelerometer is that it receives signals all the time due to its portability; while the camera data is available only when the subject presents in the camera frame, introducing lots of missing data. In addition, an accelerometer measures much more precise acceleration than a camera. However, One of the disadvantage of accelerometer is that its three axes are not fixed, and in this way, we lost information about the direction of motion; while in this perspective, the cameras can capture the direction a subject moves. Another advantage of camera is that it captures location information of a subject, such as whether he/she is on the left or right, up or down.

TABLE I: 2D/3D Cemera Features

feature category	2D/3D	feature varaible	features
movement	2D	center coordinate	mean, std, gradient
	3D	center coordinate	mean, std, gradient
shape	2D	length	mean, std
		width	mean, std
		area	mean, std
	3D	length	mean, std
		width	mean, std
		height	mean, std
		volume	mean, std

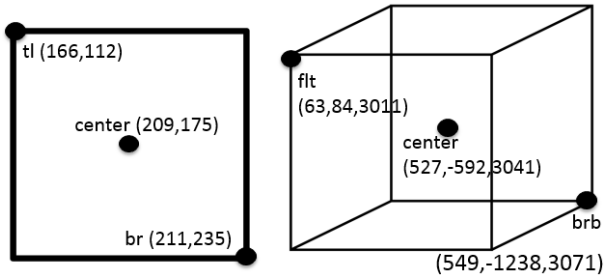


Fig. 6: The given coordinates of an example bounding box. 2D bounding box on the left, 3D bounding box on the right. (tl: top left; br: bottom right; flt: frong left top; brb: back right bottom.)

C. Annotation with Confidence

All activities are conducted with a series of continuous actions and labeled by multiple annotators, which suffers the issue of inter-annotator disagreement. The ratio of agreements are treated as confidence level, or probability score.

For instance, in Figure 7, *"a_walk"* starts from 2s to 8s. Among all these 7 seconds, only second 5 has *"a_walk"* with confidence 1, all the rest five seconds contain both *"a_walk"* and *"p_stand"* activities with different probability scores (as shown in Table II). Similar observation for activity of *"t_turn"* exists from 12s to 15s.

Since the task requires us to predict the activities for each single second, we treat the annotation probability score $\mathbf{P}_{ij}^{target}$ as the confidence score of an observation i assigned to class j. In this case, a cost sensitive objective function is built. We

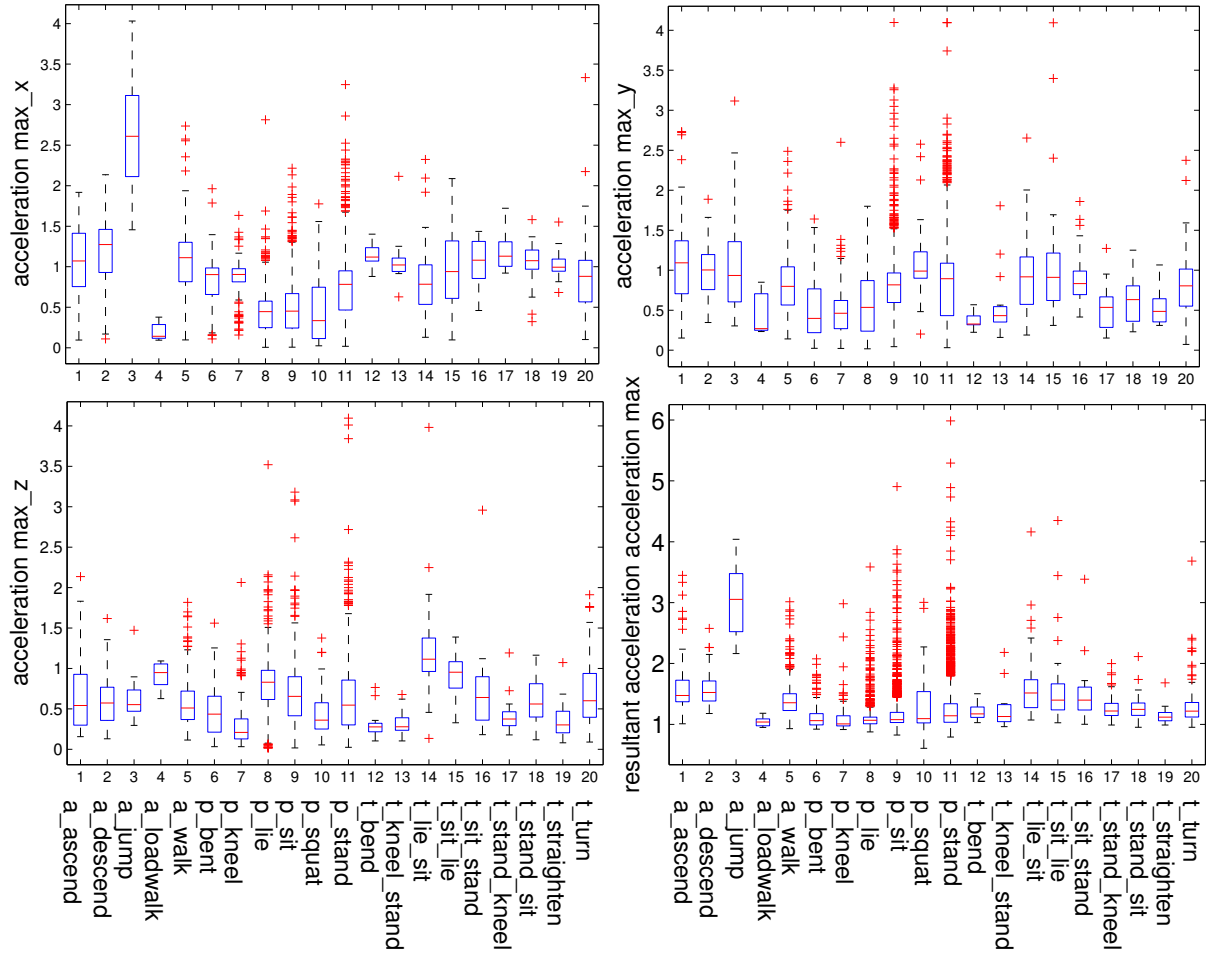


Fig. 4: Distribution of the maximum acceleration for each activity class.(The line in the middle of each box is the sample median; The tops and bottoms of each "box" are the 25th and 75th percentiles of the samples, respectively; The whiskers are lines extending above and below each box; Observations beyond the whisker length are marked as outliers)

TABLE II: Illustration of Activities Distribution

second	2	3	4	5	6	7	8	9
a_walk	0	0.5	0.7	1	0.8	0.7	0.3	0
p_stand	1	0.5	0.3	0	0.2	0.3	0.7	1

take the confidence score $\mathbf{P}_{ij}^{target}$ as the cost if observation i is not assigned to class j. An example of the revised objective function of Logistic Regression is shown as the following:

$$\min_{w_j} ||w_j||_1 + \sum_i \sum_j P_{ij}^{target} log(1 + exp(-y_{ij} w_j^T x_i)),$$

where $y_{ij} = 1$ if the observation i is labeled as class j with probability $\mathbf{P}_{ij}^{output}$.

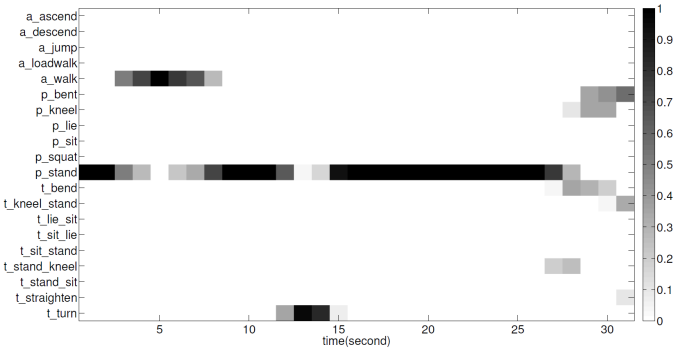


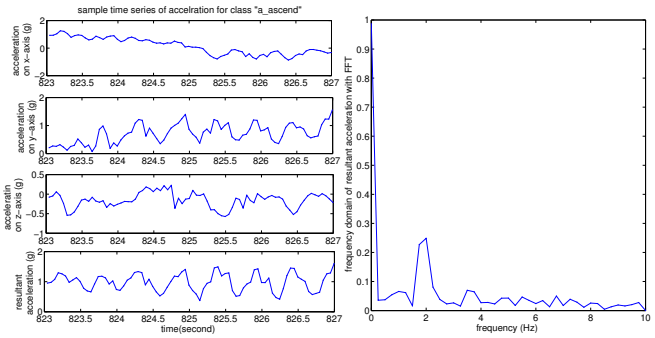
Fig. 7: A segment of ground truth activities.

III. Experimental Results

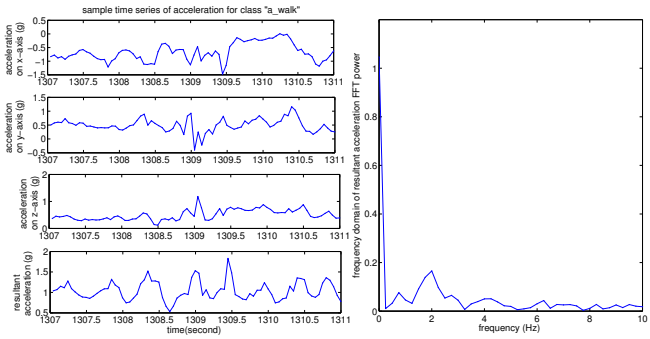
A. Overview of Data

We performed our experiments on real-world data set from "SPHERE Challenge: Activity Recognition with Multimodal Sensor Data" [9]. This data set is collected in a single house, which tracks accelerometer, RGB-D camera as well as passive environmental sensors (PIR) data. In this paper, we only use the accelerometer and RGB-D data for evaluation purpose:

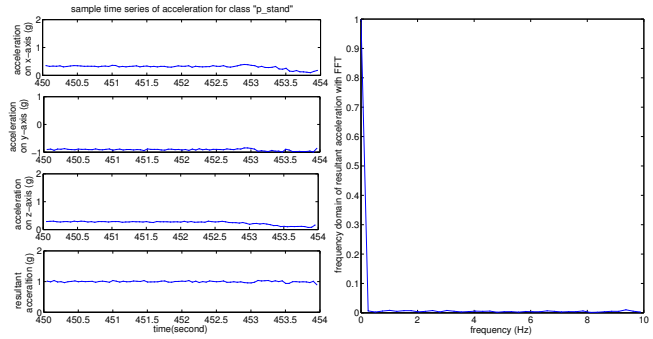
- An accelerometer on the wrist of a subject, which collects gravity distributed on three axes with a sampling rate at



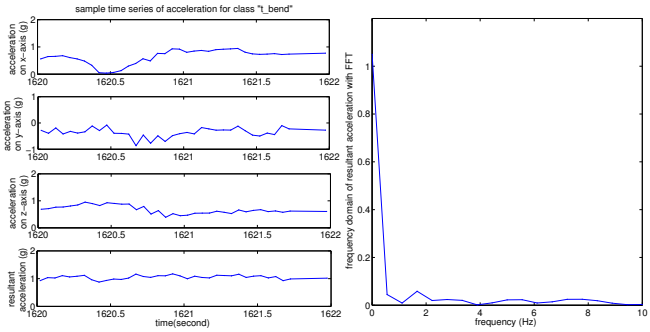
(a) Sample Illustration of Activity Class "a_ascend"



(b) Sample Illustration of Activity Class "a_walk"



(c) Sample Illustration of Activity Class "p_stand"



(d) Sample Illustration of Activity Class "p_bend"

Fig. 5: Time Domain and Frequency Domain of Activities.
(For each subplot: time domain on the left; frequency domain on the right)

TABLE III: Raw Signals

Sensor	Raw Data	
Acclerometer	gravity distribution on x, y, z axis	
RGB-D camera	2D bounding box	center (x,y) left top corner (x,y) right bottom corner (x,y)
	3D bounding box	center (x,y,z) front left top corner (x,y,z) back right bottom corner (x,y,z)

20 Hz.

- An RGB-D camera placed each in hallway, kitchen, and living room, at a sampling rate 20 Hz. The recorded videos were then processed for the coordinates of 2D and 3D bounding boxes.

The detailed raw signals given by these two sensors is summarized in Table III. This data set contains labeled 16,124 data instance, derived from10 sequence trials. The ground truth labels are given together with their probabilities score (or confidence level) within each second. We take the time series of each second as a data instance, and each data instance can be assigned to multiple classes with corresponding probability scores.

B. Evaluation Metric

In the rest of this article, the output of the algorithm is a predicted probability matrix $\mathbf{P}^{output} \in \Re^{N\times C}$, where N is the number of predicted observations, while C is the number of activity classes ($C = 20$ in this paper). Each element $\mathbf{P}_{ij}^{output}$ indicates the probability observation i is assigned to class j by the algorithm. Similarly, the ground truth probability matrix is denoted as $\mathbf{P}^{target}$ which represents the confidence of annotation.

Unfortunately, accuracy is not an appropriate metric for data sets with unbalanced classes. For example, if we have 99 instances belonging to the negative class and 1 instance from the positive class, then a classifier that predicts all 100 instances as negative class will have an accuracy of 99% even though it fails to detect any instance from the positive class. Instead of accuracy, we report the experimental results using both micro and macro-averaged F1 scores [10]. In addition, because the targets and predictions are probabilistic, in this paper, we also report the weighted brier score[11] to evaluate the classification performance.The metrics are formally defined as follows:

1) Weighted Brier Score:

$$\text{BS} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} r_j(\mathbf{P}_{i,j}^{output} - \mathbf{P}_{i,j}^{target})^2$$

where N represents the number of test samples, C is the number of activity classes, r_j is the weight for each class j. The lower brier score indicates the better performance of the

model achieves, with optimal performance achieved when a brier score is equal to 0.

2) Micro-average F1: Let TP_k, FP_k, and FN_k be the true positive, false positive, and false negative of class $k \in C$. The micro-average F1 score is defined as:

$$\begin{aligned} \text{Precision} &= \frac{\sum_{k \in C} TP_k}{\sum_{k \in C}(TP_k + FP_k)} \\ \text{Recall} &= \frac{\sum_{k \in C} TP_k}{\sum_{k \in C}(TP_k + FN_k)} \\ \text{Micro-average F1} &= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision+Recall}} \end{aligned}$$

Note that micro-average F1 is computed based on the average precision and recall values of all the test instances.

3) Macro-average F1: The macro-average F1 score is defined as:

$$\begin{aligned} \text{Precision}_k &= \frac{TP_k}{TP_k + FP_k} \\ \text{Recall}_k &= \frac{TP_k}{TP_k + FN_k} \\ \text{Macro-average F1} &= \frac{1}{|C|} \sum_{k \in C} \frac{2 \times \text{Precision}_k \times \text{Recall}_k}{\text{Precision}_k + \text{Recall}_k} \end{aligned}$$

The metric is computed by averaging the F1 scores for all the classes. When the class distribution is skewed, the macro-average F1 score will be highly influenced by the F1-score of the rare classes. In contrast, the micro-average F1 score will be dominated by the performance of the classifier on instances from the larger categories.

C. Baselines

The simplest baseline is to use the class distribution as prior probability for prediction, namely, $\mathbf{P}_{ij}^{output}$ is equal to the prior probability of class j. Furthermore, to validate that our proposed feature set is helpful, we compare the performance on the proposed feature set and that on the baseline feature set. The baseline feature set includes only the *"mean"*, *"median"*, *"std"*, *"max"* and *"min"* of the raw data which is shown in Table III. Finally, we also compare the performance of linear and nonlinear models on the proposed and baseline feature set. In this paper, L1-regularized Logistic Regression is employed as linear model, while SVM with RBF kernel as the nonlinear model.

D. Experiment Results

The experimental results reported in this section is obtained using 5-fold cross validation. For each run, we use 4 of the folds for training and the remaining fold as test data.

From the perspective of soft classification, we aim to recover $\mathbf{P}^{target}$ with $\mathbf{P}^{output}$ as well as possible. Soft classification is meaningful here, since a subject does postures or activities continuously, and it is common that two or more activities happen in a single integer second. The higher $\mathbf{P}_{ij}^{output}$ is, the more possible that the subject was doing activity j in second i.

TABLE IV: Weights Distribution over 20 Activity Classes

Activity	Class Weights	Activity	Class Weights
a_ascend	1.352985	p_stand	0.110181
a_descend	1.386846	t_bend	1.078033
a_jump	1.595874	t_kneel_stand	1.365604
a_loadwalk	1.353187	t_lie_sit	1.170241
a_walk	0.347784	t_sit_lie	1.193364
p_bent	0.661082	t_sit_stand	1.18037
p_kneel	1.047236	t_stand_kneel	1.344149
p_lie	0.398865	t_stand_sit	1.116838
p_sit	0.207586	t_straighten	1.080839
p_squat	1.505783	t_turn	0.503152

In this case, we will calculate the Weighted Brier Score [12]. Traditional brier score computes the average residual between a target probability matrix and an output probability matrix. In this way, we can evaluate the ability $\mathbf{P}^{target}$ is recovered by $\mathbf{P}^{output}$. However, if the algorithm outputs the probability of the biggest class to be 1 for all data instances, the brier score can still be small, but of course this is not an output we want. In this case, the weighted brier score brings in the unbalanced class distribution into the computing, with the residual of small classes suffering more penalty, while the residual of big classes suffering less penalty. The weight for each class is negatively correlated with its class distribution ratio, which is give in Table IV from [9].

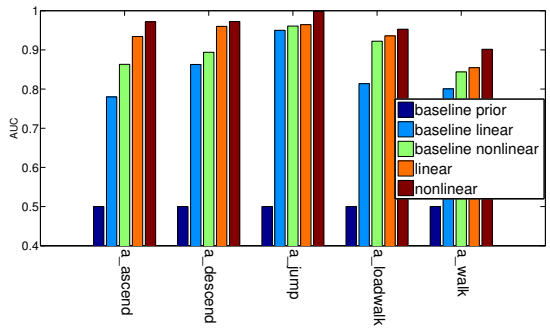
From another aspect, if we only allow the top-1 scored predicted class of each row in $\mathbf{P}^{output}$ as valid, we gets a hard classification solution. To evaluate the hard classification results considering the class unbalance problem, we will use both Macro F1 score and Micro F1 score as measurements.

1) Soft Classification: The Weighted Brier Score obtained by each algorithm is shown in Table V. On one side, the weighed brier score indicates that nonlinear model with the proposed data set gets the best overall performance with a weighted brier score obtained as 0.1700. From this table, we can also see that all the methods which take use of the sensor data are much better than the baseline which only use prior distribution. In addition, by using our proposed feature set, the weighted brier score of both linear and nonlinear algorithm decreases, which proves that our proposed feature set is effective to discriminate patterns from different activities. Furthermore, the nonlinear model always wins the linear model, which reveals that there are nonlinear relationship among features in this data set. . On the other side, in order to show the individual performance on each single class, we compare the Area Under Curve(AUC) of each class by different models in Figure 8. The same conclusion can be drawn on each class, namely, that the nonlinear model with our proposed feature set performs the best on each single class.

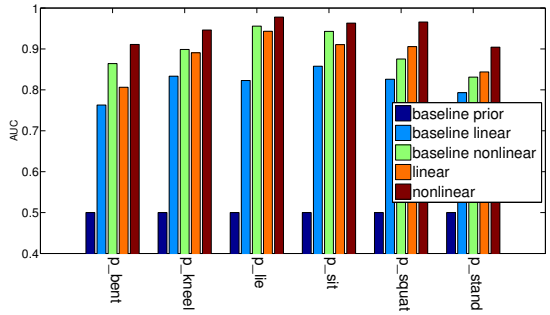
2) Hard Classification: By selecting only the top-1 scored activity as the prediction for each observation, we obtain a single predicted label for each observation. The hard classification performances are shown as Micro F1 score and Macro F1 score in Table V. To get a detailed knowledge of the performances on each class, we list the confusion matrix in Table VI. From this table, we see that the performance was greatly

TABLE V: Results Comparison

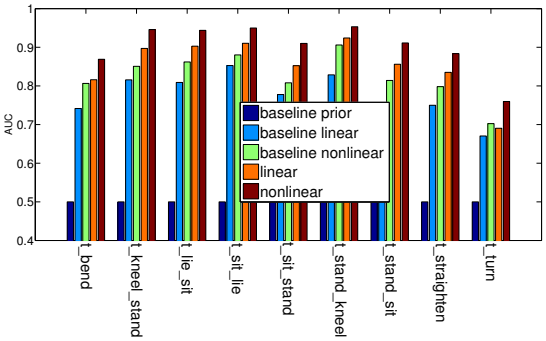
Method	Weighted Brier Score	Micro-average F1 score	Macro-average F1 score
Nonlinear model with proposed feature set	0.170001	0.650756	0.335296
Linear model with proposed feature set	0.195589	0.579377	0.256826
Nonlinear model with baseline feature set	0.194217	0.599603	0.265604
Linear model with baseline feature set	0.217490	0.528157	0.180234
Prior	0.256094	0.375149	0.027281



(a) AUC of each active class



(b) AUC of each motion class



(c) AUC of each transition action class

Fig. 8: Comparison of AUC from Different Models

affected by the data unbalance problem. Several biggest classes obtain quite high recall and precision, while observations from smaller classes are often prone to be assigned to bigger classes. For example, many misclassified observations are assigned to *"a_walk"*, *"p_sit"*, *"p_lie"* and *"p_stand"*. *"a_ascend"* and *"a_descend"* get relatively low recall and precision, and *"a_jump"* gets quite high scores, while they are all active postures. This is because *"a_jump"* usually gets outstanding maximum acceleration than other active postures (Figure 4), but all the rest active postures share similar acceleration values, and even similar acceleration frequency, since they are all rooted in walking.

IV. Related Work

Various sensors have been utilized to help elder people or patients who need special daily nursing. Some of the sensors are directly related to healthcare, such as EEG, ECG, EMG and so on. In [13], a electrocardiograph (ECG) sensor was used to assess the risk of falling, and prevent falling due to hypotension. In [14], NIR, EEG and ECG signals were acquired to build ICA, HRV and DOT engines as a portable brain-heart monitoring system.

Another way to serve healthcare, is to track the daily activities of the patients. Among all varieties of sensors, inertial sensors are most popularly used, especially accelerometer and gyroscope. Many inertial sensors are also embedded into smart phones for more popular usage. In [15], a waist-mounted smart phone was worn by subjects to collect data on mainly 6 activities. In [16], miniature inertial and magnetic sensors were employed to classify on totally 19 human activities. All of above experiments fail to involve recognition on transition actions. Besides inertial sensors, more traditional activity detection with RGB cameras are also popular [17] [18]. In [19], a novel robot centric activity recognition system is proposed, which aims to recognize interaction activities. Additionally, passive environmental sensors like PIR are often helpful when building a smart environment [20] [21].

V. Conclusion and Future Work

In this paper, we explore into a multimodal sensor data set for human activity recognition. Useful features are extracted from the data collected by accelerometer and RGB-D camera to describe the patterns of subjects' motions, stationary postures and transition actions. SVM with RBF kernel is employed to classify on 20 human activities. Our experimental results suggested that our proposed feature set is much more effective in recognizing activities than baseline approaches. Moreover, the nonlinear model gets better performance than linear model with a lowest weighted brier score 0.1700, a corresponding highest Macro F1 score of 33.53% and a corresponding highest Micro F1 score 65.08%.

However, as we can see from the confusion matrix in Table VI, the performances for rare classes are still not good enough, especially for transition activity classes. Next step, we plan to further enhance the performance by considering such class imbalance issue into our classification framework.

TABLE VI: Confusion Matrix by Nonlinear Model with Proposed Feature Set

	a_ascend	a_descend	a_jump	a_loadwalk	a_walk	p_bent	p_kneel	p_lie	p_sit	p_squat	p_stand	t_bend	t_kneel_stand	t_lie_sit	t_sit_lie	t_sit_stand	t_stand_kneel	t_stand_sit	t_straighten	t_turn	precision (%)
a_ascend	**16**	5	0	1	2	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	61.54
a_descend	2	**26**	0	2	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	72.22
a_jump	0	0	**24**	0	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	88.89
a_loadwalk	1	5	0	**7**	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	46.67
a_walk	76	68	5	79	**667**	11	3	2	11	0	104	10	0	0	3	6	0	4	17	92	0.576
p_bent	0	0	1	2	9	**237**	20	1	4	4	22	45	1	0	0	1	3	3	26	2	62.20
p_kneel	0	0	0	0	4	15	**104**	0	6	7	7	1	10	0	0	0	20	1	2	1	58.43
p_lie	1	0	0	0	5	19	11	**1034**	101	3	47	1	1	60	60	1	1	1	2	86	72.11
p_sit	0	0	1	1	20	34	13	153	**2359**	3	92	2	2	59	61	31	0	53	5	31	80.79
p_squat	0	0	0	0	0	1	0	0	0	**14**	1	0	0	0	0	0	1	0	0	0	82.35
p_stand	53	34	14	52	911	480	146	113	393	29	**5660**	157	80	41	25	119	83	124	150	547	61.45
t_bend	1	0	0	0	3	3	0	0	0	0	0	**11**	0	0	0	0	0	0	1	1	55.00
t_kneel_stand	0	0	0	0	0	0	2	0	0	0	1	0	**16**	0	0	0	0	1	0	0	80.00
t_lie_sit	0	0	0	0	0	0	0	6	7	0	3	0	0	**22**	3	1	0	1	0	0	51.16
t_sit_lie	0	0	0	0	0	1	0	1	6	0	1	0	0	4	**18**	0	0	2	0	2	51.43
t_sit_stand	0	0	0	0	3	0	0	0	6	1	3	0	0	0	0	**18**	0	0	0	0	58.06
t_stand_kneel	0	0	0	0	0	1	0	0	0	2	2	0	0	0	0	0	**16**	0	0	1	72.73
t_stand_sit	0	0	0	0	1	0	0	0	6	0	2	2	0	0	2	1	0	**26**	0	2	61.90
t_straighten	0	0	0	1	0	4	0	0	1	0	4	4	1	0	0	0	0	0	**17**	1	51.52
t_turn	0	1	0	11	57	10	3	43	22	2	28	4	0	10	2	6	2	5	5	**83**	28.23
recall (%)	10.67	18.71	53.33	4.49	39.47	29.04	34.44	76.42	80.73	21.54	94.63	4.64	14.41	11.22	10.34	9.78	1.27	11.76	7.56	9.76	65.08

REFERENCES

[1] "Open Source Computer Vision Library," http://opencv.org/.

[2] "OpenNI," http://www.openni.org/.

[3] "Point Cloud Library," http://www.pointclouds.org.

[4] "Microsoft Kinect SDK," https://developer.microsoft.com/en-us/windows/kinect.

[5] J. Baek, G. Lee, W. Park, and B.-J. Yun, "Accelerometer signal processing for user activity detection," in *Proceedings of International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, 2004, pp. 610–617.

[6] S. M. Pincus, I. M. Gladstone, and R. A. Ehrenkranz, "A regularity statistic for medical data analysis," *Journal of clinical monitoring*, vol. 7, no. 4, pp. 335–345, 1991.

[7] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in *Proceedings of International Conference on Pervasive Computing*. Springer, 2004, pp. 1–17.

[8] A. Weiss, T. Herman, M. Plotnik, M. Brozgol, N. Giladi, and J. Hausdorff, "An instrumented timed up and go: the added value of an accelerometer for identifying fall risk in idiopathic fallers," *Physiological measurement*, vol. 32, no. 12, p. 2003, 2011.

[9] "The sphere challenge: Activity recognition with multimodal sensor data," http://blog.drivendata.org/2016/06/06/sphere-benchmark/.

[10] Y. Yang, "An evaluation of statistical approaches to text categorization," *Information Retrieval*, vol. 1, no. 1, pp. 69–90, 1999.

[11] G. W. Brier, "Verification of forecasts expressed in terms of probability," in *Monthly weather review*, 1950, pp. 1–3.

[12] N. Twomey, T. Diethe, M. Kull, H. Song, M. Camplani, S. Hannuna, X. Fafoutis, N. Zhu, P. Woznowski, P. Flach, and I. Craddock, "The SPHERE challenge: Activity recognition with multimodal sensor data," *arXiv preprint arXiv:1603.00797*, 2016.

[13] P. Melillo, R. Castaldo, G. Sannino, A. Orrico, G. de Pietro, and L. Pecchia, "Wearable technology and ecg processing for fall risk assessment, prevention and detection," in *Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 7740–7743.

[14] C.-C. Fu, C.-K. Chen, S.-Y. Tseng, S. Kang, E. Chua, and W.-C. Fang, "Portable brain-heart monitoring system," in *Database Theory and Application, Bio-Science and Bio-Technology*. Springer, 2010, pp. 241–250.

[15] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones." in *Proceedings of ESANN*, 2013.

[16] K. Altun, B. Barshan, and O. Tunçel, "Comparative study on classifying human activities with miniature inertial and magnetic sensors," *Pattern Recognition*, vol. 43, no. 10, pp. 3605–3620, 2010.

[17] H. Zhang and L. E. Parker, "Code4d: color-depth local spatio-temporal features for human activity recognition from rgb-d videos," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 3, pp. 541–555, 2016.

[18] A. Farooq, A. Jalal, and S. Kamal, "Dense rgb-d map-based human tracking and activity recognition using skin joints features and self-organizing map," *KSII Transactions on internet and information systems*, vol. 9, no. 5, pp. 1856–1869, 2015.

[19] L. Xia, I. Gori, J. K. Aggarwal, and M. S. Ryoo, "Robot-centric activity recognition from first-person rgb-d videos," in *Proceedings of IEEE Winter Conference on Applications of Computer Vision*, 2015, pp. 357–364.

[20] N. Twomey and P. Flach, "Context modulation of sensor data applied to activity recognition in smart homes," in *Proceedings of Workshop on Learning over Multiple Contexts, European Conference on Machine Learning (ECML'14)*, 2014.

[21] M. Sathishkumar and S. Rajini, "Smart surveillance system using pir sensor network and gsm," *International Journal of Advnced Research in Computer Engineering & Technology*, vol. 4, no. 1, 2015.