

# Validation of Automated Mobility Assessment using a Single 3D Sensor

Jiun-Yu Kao, Minh Nguyen, Luciano Nocera, Cyrus Shahabi, Antonio Ortega, Carolee Winstein, Ibrahim Sorkhoh, Yu-chen Chung, Yi-an Chen, and Helen Bacon

University of Southern California, Los Angeles, CA, USA

**Abstract.** Reliable mobility assessment is essential to diagnose or optimize treatment in persons affected by mobility disorders, e.g., for musculo-skeletal disorders. In this work, we present a system that is able to automatically assess mobility using a single 3D sensor. We validate the system ability to assess mobility and predict the medication state of Parkinson’s disease patients while using a relatively small number of motion tasks. One key component of our system is a graph-based feature extraction technique that can capture the dynamic coordination between parts of the body while providing results that are easier to interpret than those obtained with other data-driven approaches. We further discuss the system and the study design, highlighting aspects that provide insights for developing mobility assessment applications in other contexts.

**Keywords:** mobility assessment; 3D Sensor; Parkinson’s Disease; human performance; classification

## 1 Introduction

Observation of a person’s movements performing certain tasks is widely used in many contexts, such as early diagnosis and treatment of various diseases, sports and military applications. Example applications include estimating the risk of falls in elderly patients, adjusting medication levels for those being treated for musculo-skeletal disorders or evaluating movement for rehabilitation. It is common practice for physicians to assess mobility, e.g., gait and balance, by direct observation of patients performing standardized tasks. In rare cases highly specialized equipment providing kinematic measurements is used. Clearly, cost as well as personnel/equipment availability make it impossible to provide this kind of assessment more broadly. Thus, while in some situations it would be desirable to assess mobility frequently, and to do so at a patient’s home, this is not possible in practice.

Several factors motivate our work aimed at developing an automated mobility assessment systems based on depth sensors. The increased availability of low cost wearable sensors and other body sensing technologies provides the opportunity to unobtrusively and continuously sense and assess human mobility. Moreover, harnessing mobility analytics can lead to the development of a broad

range of applications. For example, wearable sensors that summarize activity in the field from measurements from various sensors (e.g., acceleration, gyroscope) have already been utilized in various applications [27], 3D sensors (e.g., Microsoft Kinect) have the potential to complement wearable sensor measurements at home or in the clinic providing detailed insights about motion characteristics, while having the advantage of being unobtrusive. In this work, we validate an automated mobility assessment system and provide study design insights in a specific context and highlight design aspects that can be generalized to other applications. For this, we focus on musculo-skeletal disorders as a case study in part because of the existence of abundant literature related to these disorders. Furthermore, there is potential for very significant impact for these automated systems, since common clinical practice already involves mobility assessment protocols carried out on simple activities.

A reliable monitoring system to evaluate musculo-skeletal disorders, e.g., Parkinson’s Disease (PD), can be beneficial to both patients and clinicians as the population of patients with these types of disorders is increasing. One example of such a system is introduced in [9] with the *POCM*<sup>2</sup> system, where a single 3D sensor, e.g., Microsoft Kinect, is used to enable frequent mobility assessments from the home. In this work we present a systems to automatically assess mobility. For this, we show that novel preprocessing, feature extraction and classifier selection is required. We present results that demonstrate the effectiveness of the proposed system to predict the medication state of persons with PD. Our proposed system is evaluated based on a real data set collected using MS Kinect camera from a group of subjects with PD.

Gait in persons with PD is characterized by Bradykinesia, instability, episodes of freezing and increased variability, and may be mitigated through medication. Optimized medication and rehabilitation plans require reliable and frequent mobility assessments. In current practice, trained physicians assess the mobility of persons with PD during visits at the clinic by visually observing patients as they perform standardized tasks. In addition to requiring patients to perform these tasks in the controlled environment of the clinic, current practice is limited due to the lack of quantifiable measurements. Current clinical scales, e.g., the Unified Parkinson’s Disease Rating Scale (UPDRS), also lack resolution and rely on visual observation.

Our hypothesis is that differences in performance under different medication conditions in subjects with PD can be distinguished from the skeleton data (i.e., joint positions over time) produced by the Kinect sensor. We consider the two medication states in persons with PD, *ON* condition corresponding to a medicated state and *OFF* condition corresponding to a non-medicated state. The *OFF* condition effectively simulates a state in which the medication is no longer effective and mobility may be affected. We further analyze which skeleton data features support classifying the medication state using various machine learning approaches. With our experiments, we focus on *subject-dependent classification* as we are interested in predicting the medication state of specific individuals, however, we provide a comparison with subject-independent classification to

show how these results generalize. We further report on performance for different classifiers to determine what approach is most suited for these data.

Results presented are based on a study carried out on 14 PD patients that perform a dual-task walking action, i.e., walking in a figure-of-eight pattern while counting backward, for which the Kinect sensor seems to be the most reliable. The cognitive task further challenges participants, therefore we expect the cognitive task to accentuate the difference between conditions. Previous work [9, 26] shows that some gait parameters, e.g., stepping time and step size, are significantly different between PD and non-PD subjects. This paper addresses the more challenging task of distinguishing between different medication levels. We show that more appropriate features (e.g., proposed graph-based features) and system design (e.g., proper normalization and combination of classifiers) are necessary in order to capture the fine movement differences between conditions.

To our knowledge this is the first study to use Kinect skeleton data to discriminate the medication state of persons with PD. Our results show the potential for Kinect type sensors to be used to quantify mobility performance in persons with PD and possibly other mobility related disabilities. The main contributions of this work are: (1) We develop and validate a method to automatically assess mobility using a single 3D sensor that can discriminate the medication state of PD subjects, (2) We propose a novel graph-based feature extraction technique to reveal the dynamic coordination between parts of the body that compared with purely data-driven techniques provide comparable classification performance but is significantly easier to interpret, and (3) We provide study design insights on the proposed system and discuss how they can inform achieve mobility assessment in other contexts.

The remainder of the paper is organized as follows. Section 2 presents a brief review of related work. Section 3 proposes the general methodology and the insight into key factors for deploying a successful automated mobility assessment system. Section 4 describes the features (including the proposed graph-based features) and classification algorithms implemented. Section 5 presents the experimental setup and methods, and experimental results. Finally Section 6 concludes the paper and provides future directions.

## 2 Related Work

New depth sensing technologies [1] capable of providing accurate and reliable measurements of human motion, have prompted researchers to explore how data from these devices can be used to recognize and quantify movement for a wide range of medical applications. One such device, primarily designed for gaming, is Microsoft Kinect, which infers depth information using a stereo sensor and projected structured infrared light. The resulting depth stream at a resolution of 320 x 240 can be utilized to fit a skeleton of 15 joints at each time frame and in real time [22]. Compared to other motion capture systems, e.g., optical-passive techniques using retro-reflective markers or optical-active LED markers that are tracked by infrared cameras such as the Vicon motion capture system, Kinect of-

fers a passive and non-invasive alternative but at the cost of a lower accuracy [19, 23]. Occlusions and self occlusions generate joint coordinate measurement errors that are not filtered out by Kinect as it appears that there is no mechanism to enforce strict rigidity constraints. Thus, developing skeleton features and defining tasks that are robust to noise becomes critical and challenging. In general, gesture recognition remains a difficult problem and for complex tasks the skeleton data is pre-processed so as to segment the movement into simpler repeated movements referred to as skeletal action units (SAU), e.g., in [26] a walking task is segmented into steps and it is shown that these steps can in turn be summarized, thereby reducing noise and providing a representative action unit.

Several previous studies have used Kinect data for action recognition. The approach proposed in [28] extracts the histogram of 3D joint positions in a spherical coordinate system originating at the hip joint. In [16], Sequence of the Most Informative Joints (SMIJ) is proposed by automatically selecting a few skeletal joints at each time stamp that are most informative with respect to the current action according to some interpretable parameters, such as the mean variance of joints angles or the angular velocity of joints. In [29], Yang and Tian develop a representation for actions based on the position differences between joints. They compute the position differences of all pair of joints in one frame, of each joint between two consecutive frames and of joints in any frame and corresponding joints in the initial frame, which captures both spatial and temporal information of the human actions. This study found that the Kinect was good at measuring timing of movements and spatial characteristics of gross movements.

Kinect has previously been used in applications related to PD, primarily as an intervention tool [5, 18] where game-play supports motivation and a game score is used as a measure of performance. In [5] a game was developed to train dynamic postural control and the accuracy of Kinect to measure on the spot walking, stepping and reaching. A comparison of the Kinect system against a Vicon motion capture system is presented in [4].

Kinect has also been used to supplement inertial sensor data as part of a home monitoring system for detecting freezing of gait [24]. The *POCM*<sup>2</sup> system [9] used the raw Kinect skeleton data on pilot data to detect pauses and discriminate between a non-PD person and a person with PD using dimensionality reduction techniques. A similar comparison was reported by [14] using a Vicon system and showing a difference between PD patients and healthy controls at similar age both in angle changes and in spatiotemporal parameters of gait.

As the field of signal processing on graphs is emerging, graph-based approaches to represent the motion capture data have been carried out. In [10], spectral graph wavelet transform is applied to a spatial-temporal graph constructed based on the motion data, which provides an over-complete set of features. These features are shown to be good in terms of action recognition accuracy for three state-of-the-art datasets. In [8], graph Fourier transform is applied to a skeletal graph constructed according to the natural connections

between human limbs and the generated basis shows the ability to exploit the correlations between body parts.

In this paper, we build on previous work (in particular [9, 26]) to tackle the difficult problem of assessing fine changes in mobility induced by medication in PD patients.

### 3 System Design

In this section, we propose a general methodology for deploying an automated mobility assessment system based on cost-effective 3D sensors. We provide insights into key factors that can lead to the success or failure of the deployment of this type of systems. Key aspects of this methodology will be validated in our evaluation section based on experimental results from our case study.

**Hardware and environment:** One of the most important factor is the usage and limitation of the 3D sensors in use. Taking Microsoft Kinect sensor as example, its horizontal Field of View has a practical range of 1.2 to 3.5 meters while its vertical Field of View has a practical range of 0.8 to 2.5 meters. Therefore, when accounting for the environment to deploy the system and the tasks to be performed by the subjects, not to have subjects exceed these ranges is important and critical to the system robustness. For example, an outdoor open environment may not be suitable for deploying this system. Also, requesting the subjects to perform certain tasks such as climbing up stairs or running far away may lead to system failure, as the device Field of View could be exceeded. Furthermore, as the estimation for the 3D positions of skeletal joints with Kinect SDK have much larger estimation errors under certain situations, e.g., walking away from the sensor at the periphery of the Field Of View or taking a turn, the environment and task should limit the occurrence of these situations.

**Task:** The tasks that the subjects are asked to perform should be properly chosen to satisfy the following criteria. First, activities that can fully exploit and examine the mobility of all parts of the body are preferable to those that place explicit constraints on certain parts of body. For example, a task requiring the subject to count silently while walking can be better than a walking task where the subject is required to hold a tray, as the latter activity has less potential to exploit the upper body mobility, while in the former the walking may be more “natural” as the subject has to focus on counting. Secondly, the level of difficulty in performing activities will affect the capability of discriminating subjects’ states. An over-simplified task will make it easier for the subjects to control the mobility performance under different states/conditions, which would make it more difficult for the system to distinguish between states/conditions based on the assessed mobility. Finally, it is better to have each task performed repeatedly by each subject in order to improve the robustness of the assessment.

## 4 Feature Design and Classification

A preliminary statistical analysis on a partial dataset indicated that there were no significant differences between medication states in gait kinematics. Additionally, it appears that when participants performed the task in the OFF state, they took shorter steps and had increased hip flexion, the latter, likely due to increased trunk forward lean. These results prompted us to focus on *subject-dependent classification* and extend the features considered to include angular gait measurements (e.g., angles extracted from the skeleton joints and angular speed) and graph-based features.

Mobility data are often normalized when studied. For example in the context of action recognition [30, 10] proposes a normalization scheme, which is applied first to the captured skeleton data by estimating the expected lengths of skeleton limbs (connections) across subjects from the training data, and then adjusting the locations of joints to achieve the same lengths of limbs, with the limb direction vectors being preserved. We compare the effects of normalization on the subject-independent classification performance in Section 5.5.

We provide hereafter details about the features used for classification.

**Gait Measurements** To measure the gait statistics, after segmenting the strides we extract a set of spatial-temporal parameters including *step lengths*, *stride time* and *stride width* as suggested in [13]. Since each stride consists two steps, i.e., a left step and a right step, For consistency and across-subjects analysis, we labeled steps based on the most/least affected side (as opposite to using left/right). Consequently, step feature are named as *most-affected-side step length* and *least-affected-side step length*.

**Angular Statistics** We extract a set of angular parameters associated with each joint as features, inspired by the SMIJ method [17]. Angles corresponding to each joint at each time stamp are computed by evaluating the dot product of the vectors defined by the limb segments connecting that joint. Assuming the angle corresponding to the dot product is  $\alpha$ , we have two possible choices for the joint angle:  $\alpha$  and  $2\pi - \alpha$ . The actual value used is defined by taking into account the type of joint (e.g., elbow or knee) and the direction of motion. We consider 19 angles in total: one angle for elbow, knee, and neck joints and two angles for the hip and shoulder joints. To capture temporal variations we further consider the following five statistics: *average*, *standard deviation*, *min*, *max*, and *angular speed*. The resulting feature vector of angular statistics has a dimension of  $19(\text{angles}) \cdot 5(\text{statistics}) = 95$ .

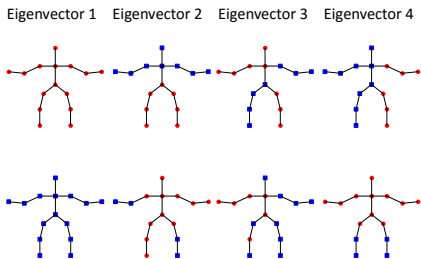
**Graph-based Features** In order to extract a set of features which can capture and evaluate more global properties in motions, e.g., the coordination between two body parts, we adopt and modify the graph-based method proposed by Kao et al. [8]. First, the human skeletal structure is modeled as a fixed undirected graph  $G = (V, E)$  with the vertex set as  $V = \{v_1, v_2, \dots, v_{15}\}$  corresponding to the 15 joints detected and estimated by Kinect at each frame. The edge set  $E$  consists of the unweighted edges corresponding to the directly

connected physical limbs of the human body. Specifically speaking, an edge  $(i, j)$  exists in  $G$  with its weight set to unity only if there exists a physical limb directly connecting the  $i$ -th and  $j$ -th joint. Given  $G$  decided, the adjacency matrix  $A$  and the degree matrix  $D$  of  $G$  can be computed as well as the normalized Laplacian matrix  $\mathcal{L}$  with  $\mathcal{L} = I - D^{-1/2}AD^{-1/2}$ .

Furthermore, as a graph signal is a function  $f : V \rightarrow \mathbb{R}$  that assigns a value to each vertex, which can be represented as a vector  $\mathbf{f} \in \mathbb{R}^{|V|}$  lying on the graph, the coordinates of all the joints at each time frame in one SAU can be regarded as a graph signal lying on the above defined skeletal graph  $G$ . Specifically speaking, we calculate the difference of 3D position at each joint between two consecutive frames, i.e.,  $\mathbf{v}_{t,i} = \mathbf{p}_{t,i+1} - \mathbf{p}_{t,i} = [v_{x_i}^{(t)} v_{y_i}^{(t)} v_{z_i}^{(t)}]$ , where  $i \in \{1, \dots, 15\}$  and  $t \in \{1, \dots, T-1\}$ . By processing each axis of 3D space separately, a graph signal  $\mathbf{f}_a^{(t)} \in \mathbb{R}^{15}$  lying on the previously defined  $G$  can be defined so that  $\mathbf{f}_a^{(t)}(i) = \mathbf{v}_{t,i}(a)$  where  $a = \{1, 2, 3\}$  indicating the coordinate axis of choice. According to [2, 8], frequency analysis of a graph signal  $\mathbf{f}_a^{(t)}(i)$  can be performed by applying the graph Fourier transform (GFT) as  $\mathbf{F}_a^{(t)}(i) = U^T \mathbf{f}_a^{(t)}(i)$  where  $U$  comes from the eigendecomposition  $\mathcal{L} = U\Lambda U^T$ . For each frame  $t$ , repeating this for three coordinate axis, i.e.,  $a = \{1, 2, 3\}$ , leads to  $C^{(t)} = [\mathbf{F}_1^{(t)}, \mathbf{F}_2^{(t)}, \mathbf{F}_3^{(t)}] \in \mathbb{R}^{15 \times 3}$ . We vectorize  $C^{(t)}$  to a row vector  $\mathbf{c}^{(t)}$  with dimension as 45 and concatenate  $\mathbf{c}^{(t)}$  from  $t = 1$  to  $T-1$ , which leads to a matrix  $\mathbf{C} \in \mathbb{R}^{(T-1) \times 45}$  with all the transform coefficients.

As illustrated in [8], the GFT basis can capture global motion properties. For example, as shown in Figure 1, the second basis vector will be able to exploit the coordination between upper and lower body while the third basis vector can help capture and measure the degree of bilateral symmetry which is an important characteristic in walking movement.

Principal component analysis (PCA), whose variants are popular in representing spatial-temporal data such as motion capture data, is one special case of our proposed graph-based feature. Considering a fully connected graph with edge weights set according to the covariance in data, the resulted graph-based basis vectors are exactly the principal components in PCA method. However, constructing the graph without taking data into consideration, as proposed here, can lead to an easier interpretation of the results, as compared to PCA. Figure 1 shows a comparison of the structure of basis vectors obtained using our proposed graph-based features and PCA. We can observe that the component of data that is captured by each PCA basis vector is more difficult to interpret than the GFT basis vectors. For example, the third eigenvector of PCA includes an isolated vertex in the left leg and the component on the fourth eigenvector of PCA is hard to be interpreted as the coordination between upper and lower body. Furthermore, in the evaluation section, our proposed feature shows to be able to achieve comparable performance to PCA does. Finally, our proposed basis is not data-dependent while PCA highly depends on the training dataset. This lack of data-dependence makes it easier to compare results across different subjects, tasks, coordinate systems or datasets.



**Fig. 1.** 15 Kinect skeleton joints. Sign values of graph features basis vectors: blue (+), red (-). *Top:* the proposed graph-based features. Notice that zero-crossings between neighboring vertices increase as the eigenvector corresponds to higher eigenvalue (frequency). *Bottom:* PCA basis constructed with captured data of PD subjects on  $x$ -axis.

To cope with different number of steps in each sequence and to represent the temporal dynamics of the frame-by-frame coefficients, we adopt the temporal pyramid pooling scheme similar to [6, 12, 25]. We define an average pooling function  $\mathcal{F} : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{1 \times n}$  such that  $\mathbf{z} = \mathcal{F}(\mathbf{B})$  provides column-wise average to a matrix  $\mathbf{B}$ . Let  $K$  denote the maximum number of pyramid levels to be used. Then at level  $k \leq K$ , we compute the pooled coefficient vector as  $\mathbf{z}_k = [\mathcal{F}(\mathbf{B}_1), \dots, \mathcal{F}(\mathbf{B}_{2^{k-1}})]$ , where  $\{\mathbf{B}_i\}$  is a set of non-overlapping block matrices uniformly dividing the matrix  $\mathbf{C}$  which contains all the transform coefficients as calculated previously. A final feature vector  $\mathbf{d}$  is obtained as a concatenation of pooled coefficient vector at each level, i.e.,  $\mathbf{d} = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_K]^T$ , with the dimension as  $45 \cdot (2^K - 1)$ .

**Classification Methods** Several classification algorithms were used to categorize the medication state. Extracted features were labeled (with ON and OFF) and used for training. We evaluate our system using Naive Bayes, SVM, k-NN, Decision Tree, and Random Forest classifiers with WEKA [7]. Combining classifiers improves performance as reported in [11, 20, 15]. We therefore report performance measurements for different combinations of classifiers. We use two ensemble methods: *Average of Probabilities* [3] and *Majority Voting* [21]. The Average of Probabilities fusion method returns the mean value of probabilities of multiple classifiers. The Majority Voting returns the class which gets the most votes among multiple classifiers. In the following Section 5, we provide details on the evaluation setup and results.

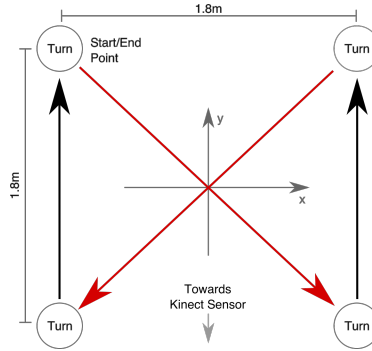
## 5 Experiments and Evaluation

### 5.1 Experimental Methodology

Fourteen adults with PD (9 men, disease duration  $8.66 \pm 7.48$  years, Hoehn and Yahr stage I-III) took part in the pilot study<sup>1</sup>. They each visited a Uni-

<sup>1</sup> The retrospective pilot study was approved by the University Institutional Review Board





**Fig. 2.** Walking tasks trajectory used in the experiments. Arrows indicate direction or movement. Only segments shown in red were used to classify the medication state.

versity laboratory 4 times. The first and last sessions consisted of qualitative interviews where 2-5 participants were interviewed in focus groups about their expectations and perceptions of the system. The remaining two sessions consisted of quantitative assessments (one each for ON and OFF medication state). Participants performed 7 standardized functional tasks: (1) walking, (2) walking whilst counting, (3) walking whilst carrying a tray, (4) walking around an obstacle, (5) sit-to-stand, (6) lifting a soda can and (7) lifting an object to a shelf. In this paper we focused our analysis on tasks (1) walking, (2) walking whilst counting tasks, and (3) walking whilst carrying a tray. We focused our attention on the walking-based tasks as we found skeleton data had a lower noise/signal ratio compared to other tasks. However, the walking whilst counting task was specifically singled-out because the added cognitive task is known to create an additional challenge in person with PD that we expect to accentuate differences between ON and OFF conditions.

Accelerometer, camcorder video and Kinect were recorded for all tasks, however in this work we consider only Kinect sensor data. Figure 2, shows the trajectories followed for walking-based tasks for which participants walked 5 times in a figure-of-eight pattern and repeated this at least twice.

## 5.2 Preprocessing and Methods

At each timestamp, we consider a subset of 15 joints (head, neck, torso, shoulders, elbows, hands, hips, knees and feet) in addition to the self-reported most affected side. As shown in the Figure 2, we exclude segments of the trajectory corresponding to turns and where the sensor did not have a good viewing angle as Kinect skeleton data corresponding to those portions of the trajectory are noticeably noisier. Note that the direction of the trajectory was chosen so that the segments used faced the sensor, to match Kinect’s skeleton reconstruction assumption.

Walking sequences were automatically segmented similar to [26] into Skeletal Action Units (SAU) each consisting of two steps and subsequently used to

derive linear and angular kinematic measurements. We also limited extracted angles maximum extents to what is bio-mechanically possible (Kinect does not constraint skeleton joins).

We extracted a total of 1521 SAUs. The total number of ON-labeled SAUs is 759 while the total number of OFF-labeled SAUs is 762. These numbers translate to an average of 109 SAUs per subject and 54 SAUs for each condition, ensuring a balanced dataset. Depending on the experiment, feature vectors for each SAU are generated based on the gait, angle and graph features extracted as described in Section 4. For the graph-based features (Section 4), the maximum number of pyramid levels is set to 3, i.e.  $K = 3$ . As customary, we further  $l_2$ -normalize the feature vector for robustness.

Evaluation results were obtained for Naive Bayes, SVM, k-NN, Decision Tree, and Random Forest primarily for subject-dependent classification. Results for across-subjects classification are also presented to allow comparing both approaches performance.

A 3-fold method is used for training and testing. In the subject-dependent approach, each subject’s set of SAUs is uniformly randomly separated into 3 non-intersecting folds; then, two folds are used for the training model and the remaining fold is used to test the trained classifier. In the subject-independent approach, all SAUs of all subjects are combined and then, separated into 3 folds. The 3-fold method steps are used in a similar way for the subject-dependent approach. We also provide the evaluation of combination of multiple classifiers. The performance metrics we report include accuracy, precision, recall and F-measure.

### 5.3 Feature Performance

The system performance is evaluated based on the walking whilst counting task data and using an SVM classifier trained with separate feature vectors for gait, angle, graph, and for a combination of all of the above. We also include PCA-based feature as a comparison to the proposed graph-based feature. The PCA-based feature is acquired from the training set and followed by exactly the same pyramid pooling scheme as graph-based feature is, which leads to the same dimension of feature vectors for both graph-based method and PCA.

Table 1 provides a summary of results. Our result shows that the combination of three proposed features sets achieve the highest performance metrics. Table 1 shows that the combination of sets in SVM performs the best with 84.79% of accurate rate, 85.43% of precision and 83.38% of recall. When only one type of feature is used the graph-based feature set outperforms the other two choices in terms of all four metrics. Besides, when doing evaluation, we notice that the worst-case accuracy with graph-based features (69.63%) also highly outperforms the worst-case accuracy with gait (39.53%) or angular features (53.58%), which shows that graph-based features are more robust in terms of exploiting the difference in motions between ON and OFF states. The possible reasons include: (1) its better ability to capture the characteristics in global coordination among

**Table 1.** SVM performance for various features. Accuracy is reported with the format as average accuracy (best accuracy/worst accuracy) across 14 subjects. A: Accuracy, P: Precision, R: Recall and F-M: F-measure. ALL: Gait, Angle, and Graph.

FEATURE	A (%)	P (%)	R (%)	F-M
GAIT	63.58 (88.71/39.53)	57.26	55.40	0.51
ANGLE	75.30 (92.22/53.58)	75.01	74.20	0.74
GRAPH	82.41 (95.68/69.63)	83.04	81.93	0.82
ALL	84.79 (93.95/71.23)	85.43	83.38	0.84
PCA	84.66 (95.32/71.99)	85.30	84.44	0.85

**Table 2.** Performance of single classifier and multiple classifiers combination. A: Accuracy, P: Precision, R: Recall, F-M: F-measure, AP: Average of Probabilities, MV: Majority Voting, S: SVM, k: k-NN, D: Decision Tree, R: Random Forest.

CLASSIFIER/COMBINATION	A (%)	P (%)	R (%)	F-M
SVM	84.79	85.43	83.38	0.84
RANDOM FOREST	83.09	83.68	83.09	0.83
k-NN	79.24	80.16	79.24	0.79
DECISION TREE	72.66	72.92	72.66	0.73
NAIVE BAYES	71.02	71.64	71.02	0.70
SkR (AP)	87.41	87.61	87.40	0.87
Sk (AP)	87.28	87.49	87.29	0.87
SkR (MV)	85.62	85.79	85.61	0.86
DSk (AP)	85.37	85.61	85.37	0.85
DSk (MV)	85.29	85.51	85.30	0.85

body parts during a motion, and (2) its ability to capture the temporal dynamics/evolution of the frame-based features.

Furthermore, results of PCA-based and graph-based features are comparable. As detailed in Section 4, the Graph-based features provide number of advantages including the interpretability of the body coordination, robustness to the noise and selection of the dataset, and comparability between different schemes.

## 5.4 Classifier Performance

To test the performance for different classifiers we consider the walking whilst counting dataset, a combination of Gait, Angle and Graph features (as we showed it provides the best performance in 5.3) and the subject-dependent approach. We report performance metrics for Naive Bayes, Decision Tree, k-NN, Random Forest and SVM. Table 2 provides the average performance metrics. According to Table 2, SVM performs the best with 84.79% of accuracy, 85.43% of precision, and 83.38% of recall while Naive Bayes gives the lowest performance rate with 71.02% - accuracy, 71.64% - precision, and 71.02% - recall. Both Random Forest and SVM achieve high accuracy (more than 83%). We also tested using combination of classifiers with two fusion methods: Average of Probabilities and

**Table 3.** Effect of normalization on subject-independent performance. A: Accuracy, P: Precision, R: Recall, F-M: F-measure.

NORMALIZATION	A (%)	P (%)	R (%)	F-M
WITH	72.91	72.32	72.32	0.72
WITHOUT	70.87	71.03	71.03	0.71

Majority Voting. Table 2 also presents the performance metrics of five best combinations of classifiers. Comparing to the best of the single classifier, i.e., SVM (2, 84.79% of accuracy), these results show that combining multiple classifiers can outperform single classifier. The best performance rate is 87.41% of accuracy (2.62% better than single SVM), 87.61% of precision, 87.40% of recall with combination of SVM, k-NN, and Random Forest using Average of Probabilities fusion method. Overall, the best five combinations have more than 85.00% of accuracy. It can also be seen that the best five correspond to the combination of the best performing single classifiers: SVM, k-NN, Decision Tree and Random Forest.

## 5.5 System Performance

We report the system performance using the subject-independent approach. Set-up is similar to the experiments of section 5.4. To assess the effects of normalization we compare results obtained when normalization is applied as suggested in [30] and without normalization using the SVM classifier with graph-based features. Results presented in Table 3 show that normalization does not provide a significant improvement. This counter-intuitive result may be explained by the fact that the effects of PD are highly person-dependent and not correlated to body size.

Table 4 shows the result of classification of five classifiers in this approach: Naive Bayes, Decision Tree, SVM, Random Forest, and k-NN using the combination of feature sets. The accuracy ranges from 60.09% to 76.86% while precision is from 60.50% to 77.10% and recall ranges from 60.10% to 76.90%. k-NN achieves the highest rate (76.86% of accuracy, 77.10% of precision, 76.90% of recall, and 0.77 of F-measure) and Naive Bayes gives the worst value in performance comparing to other classifiers (60.09% of accuracy, 60.50% of precision, 60.10% of recall, and 0.60 of F-measure). Both Random Forest and k-NN has more than 75% of accuracy.

Possible reasons for the lower performance of subject-independent results include the fact that our model only includes information of most/least affected side. The inclusion of other demographic factors such as age, gender, condition, how long medicated, affected limbs, etc. might improve subject-independent results.

We then test our system using fusion methods, similar to Section 5.4. The best three combinations of classifiers' performance metrics are reported in Table 5. Comparing to the best performed single classifier (k-NN with 76.86%

**Table 4.** subject-independent performance of single classifiers. A: Accuracy, P: Precision, R: Recall, F-M: F-measure.

CLASSIFIER	A (%)	P (%)	R (%)	F-M
NAIVE BAYES	60.09	60.50	60.10	0.60
DECISION TREE	62.98	63.00	63.00	0.63
SVM	67.13	67.10	67.10	0.67
RANDOM FOREST	75.67	75.90	75.70	0.76
k-NN	76.86	77.10	76.90	0.77

**Table 5.** Subject-independent combination performance. A: Accuracy, P: Precision, R: Recall, F-M: F-measure, AP: Average of Probabilities, MV: Majority Voting, S: SVM, k: k-NN, D: Decision Tree, R: Random Forest.

COMBINATION	A (%)	P (%)	R (%)	F-M
RK (AP)	77.32	77.50	77.30	0.77
SKR (MV)	76.92	77.00	76.90	0.77
RK (MV)	76.59	76.70	76.60	0.77

of accuracy), the three combinations reported has comparable result. The best accuracy rate is 77.32%, which is around 0.5% higher than that of single classifier, is from combination of Random Forest, and k-NN. It can be seen that the subject-independent approach does not work well in both single classifier and combination of multiple classifiers because each subject has different mobility traits (i.e., most affected side), and differences in mobility for both conditions.

## 5.6 Impact of Task Difficulty

To evaluate the impact of task difficulty on classification results we compare performance on three tasks: walking whilst counting, walking with holding a tray, and walking only using the same procedures, data size and features than what was used in Section 5.3. As summarized in table 6 we find that the average accuracy of the walking only task is 81.04%, slightly lower than for the other two dual tasks. Also, the worst accuracy across all subjects is much worse (48.99%) for the walking only task, compared to 71.23% with walking whilst counting task. This seems to corroborate the fact that dual tasks add cognitive load and increased coordination that accentuates the motion disabilities between conditions. Furthermore, we can observe that the average accuracy of the walking whilst holding a tray is comparable to that of walking whilst counting, however the worst accuracy across the subjects is much worse for walking whilst holding a tray. A possible explanation might be that the system is unable to capture changes in mobility between conditions in subjects for which the impairment of movement is mostly affecting the upper-body.

These results show that the task can significantly affect the system performance. Best performance seem to be achieved for tasks: (1) that do not constrain movement, and (2) that are sufficiently challenging.

**Table 6.** Performance results for three walking tasks. Accuracy is reported with the format as average accuracy (best accuracy/worst accuracy) across subjects. A, P, R and F-M denote respectively Accuracy, Precision, Recall and F-measure.

TASK	A (%)	P (%)	R (%)	F-M
COUNT	84.79 (93.95/71.23)	85.43	83.38	0.84
TRAY	82.04 (94.44/53.63)	82.19	90.00	0.85
WALK	81.04 (96.05/48.99)	81.63	87.75	0.83

## 6 Conclusion and Future Work

Mobility assessment is critical for several applications including rehabilitation, physical therapy, optimizing treatment, or performance in sport and military applications.

In this work, we propose a methodology to develop an automated mobility assessment systems based on motion data captured with a single cost-effective 3D sensor (i.e., Microsoft Kinect). We propose using three types of features that we show are capable to capture fine movement changes. In particular the proposed graph-based features can capture dynamic coordination between the different parts of the body.

We demonstrate the system with a pilot study involving 14 adults with PD (9 men, disease duration  $8.66 \pm 7.48$  years, Hoehn and Yahr stage I-III). Our results support the feasibility of using a Microsoft Kinect to recognize the medication state of persons with PD using a relatively small number of movements in the case of a dual-task, i.e., walking whilst counting. More specifically, we show that for a combination including gait, angle and graph-based features, it is possible to achieve subject-dependent classification performance rates of 87.41% of accuracy, 87.61% of precision and 87.40% of recall with a combination of SVM, k-NN, and Random Forest using and Average of Probabilities fusion method. It appears that among the features proposed, the graph-based features are more robust in terms of exploiting the difference in motions between medication states. Results obtained for subject-independent classification appear significantly worse. We also evaluate how different features, classifiers, approaches and tasks impact the system performance and discuss insights into the key performance factors and failure modes of the proposed system.

Future work will include extending the pilot study to a larger number of subjects to prove the statistical significance of specific features in discriminating between medical states and investigating methods that allow for a more fine grained mobility assessment. Furthermore, extending these results to the new Kinect One system can lead to significant improvements. Complementing the Kinect sensor with data from other wearable sensors could also lead to a boost in performance. Finally, extending the system to be capable of automatically measuring the degree of mobility impairment and deciding the most suitable tasks for subjects to perform can also be beneficial to clinical work.

## Acknowledgment

This project has been funded in part with Federal funds from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services, under Contract No. HHSN261201500003B and the USC Integrated Media Systems Center (IMSC).

## References

1. Chen, L., Wei, H., Ferryman, J.: A survey of human motion analysis using depth imagery. *Pattern Recognition Letters* 34(15), 1995–2006 (2013)
2. D. I. Shuman, Narang, S.K., Frossard, P., Ortega, A., Vandergheynst, P.: The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine* 30(3), 83–98 (May 2013)
3. Duin, R.P., Tax, D.M.: Experiments with classifier combining rules. In: *Multiple classifier systems*, pp. 16–29. Springer (2000)
4. Galna, B., Barry, G., Jackson, D., Mhiripiri, D., Olivier, P., Rochester, L.: Accuracy of the microsoft kinect sensor for measuring movement in people with parkinson's disease. *Gait & posture* 39(4), 1062–1068 (2014)
5. Galna, B., Jackson, D., Schofield, G., McNaney, R., Webster, M., Barry, G., Mhiripiri, D., Balaam, M., Olivier, P., Rochester, L.: Retraining function in people with parkinson disease using the microsoft kinect: game design and pilot testing. *Journal of neuroengineering and rehabilitation* 11(1), 60 (2014)
6. Gowayyed, M.A., Torki, M., Hussein, M.E., El-Saban, M.: Histogram of oriented displacements (hod): Describing trajectories of human joints for action recognition. In: *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*. pp. 1351–1357 (2013)
7. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The weka data mining software: an update. *ACM SIGKDD explorations newsletter* 11(1), 10–18 (2009)
8. Kao, J.Y., Ortega, A., Narayanan, S.: Graph-based approach for motion capture data representation and analysis. In: *Image Processing (ICIP), 2014 IEEE International Conference on*. pp. 2061–2065 (Oct 2014)
9. Kashani, F.B., Medioni, G., Nguyen, K., Nocera, L., Shahabi, C., Wang, R., Blanco, C.E., Chen, Y.A., Chung, Y.C., Fisher, B., et al.: Monitoring mobility disorders at home using 3d visual sensors and mobile sensors. In: *Proceedings of the 4th Conference on Wireless Health*. p. 9. ACM (2013)
10. Kerola, T., Inoue, N., Shinoda, K.: Spectral graph skeletons for 3d action recognition. In: *Computer Vision—ACCV 2014*, pp. 417–432. Springer (2014)
11. Kittler, J., Hatef, M., Duin, R.P., Matas, J.: On combining classifiers. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 20(3), 226–239 (1998)
12. Luo, J., Wang, W., Qi, H.: Group sparsity and geometry constrained dictionary learning for action recognition from depth maps. In: *Computer Vision (ICCV), 2013 IEEE International Conference on*. pp. 1809–1816 (2013)
13. McNeely, M.E., Duncan, R.P., Earhart, G.M.: Medication improves balance and complex gait performance in parkinson disease. *Gait & Posture* 36(1), 144–148 (May 2012)

14. Mirek, E., Rudzińska, M., Szczudlik, A.: The assessment of gait disorders in patients with parkinson's disease using the three-dimensional motion analysis system vicon. *Neurologia i neurochirurgia polska* 41(2), 128–133 (2006)
15. Nguyen, M., Fan, L., Shahabi, C.: Activity recognition using wrist-worn sensors for human performance evaluation. *The Sixth Workshop on Biological Data Mining and its Applications in Healthcare* (2015)
16. Ofli, F., Chaudhry, R., Kurillo, G., Vidal, R., Bajcsy, R.: Sequence of the most informative joints (smij): A new representation for human skeletal action recognition. *Journal of Visual Communication and Image Representation* 25(1), 24–38 (2014)
17. Ofli, F., Chaudhry, R., Kurillo, G., Vidal, R., Bajcsy, R.: Sequence of the most informative joints (smij): A new representation for human skeletal action recognition. *J. Vis. Comun. Image Represent.* 25(1), 24–38 (Jan 2014)
18. Palacios-Navarro, G., García-Magariño, I., Ramos-Lorente, P.: A kinect-based system for lower limb rehabilitation in parkinson's disease patients: a pilot study. *Journal of medical systems* 39(9), 1–10 (2015)
19. Pfister, A., West, A.M., Bronner, S., Noah, J.A.: Comparative abilities of microsoft kinect and vicon 3d motion capture for gait analysis. *Journal of medical engineering & technology* 38(5), 274–280 (2014)
20. Rokach, L.: Ensemble-based classifiers. *Artificial Intelligence Review* 33(1-2), 1–39 (2010)
21. Ruta, D., Gabrys, B.: Classifier selection for majority voting. *Information fusion* 6(1), 63–81 (2005)
22. Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., Cook, M., Moore, R.: Real-time human pose recognition in parts from single depth images. *Communications of the ACM* 56(1), 116–124 (2013)
23. Stone, E., Skubic, M.: Evaluation of an inexpensive depth camera for in-home gait assessment. *Journal of Ambient Intelligence and Smart Environments* 3(4), 349–361 (2011)
24. Takač, B., Català, A., Martín, D.R., Van Der Aa, N., Chen, W., Rauterberg, M.: Position and orientation tracking in a ubiquitous monitoring system for parkinson disease patients with freezing of gait symptom. *JMIR mHealth and uHealth* 1(2) (2013)
25. Wang, J., Liu, Z., Wu, Y., Yuan, J.: Mining actionlet ensemble for action recognition with depth cameras. In: *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. pp. 1290–1297 (2012)
26. Wang, R., Medioni, G., Winstein, C.J., Blanco, C.: Home monitoring musculo-skeletal disorders with a single 3d sensor. In: *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*. pp. 521–528. IEEE (2013)
27. Weiss, A., Sharifi, S., Plotnik, M., van Vugt, J.P., Giladi, N., Hausdorff, J.M.: Toward automated, at-home assessment of mobility among patients with parkinson disease, using a body-worn accelerometer. *Neurorehabilitation and neural repair* 25(9), 810–818 (2011)
28. Xia, L., Chen, C.C., Aggarwal, J.: View invariant human action recognition using histograms of 3d joints. In: *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*. pp. 20–27. IEEE (2012)
29. Yang, X., Tian, Y.: Eigenjoints-based action recognition using naive-bayes-nearest-neighbor. In: *Computer vision and pattern recognition workshops (CVPRW), 2012 IEEE computer society conference on*. pp. 14–19. IEEE (2012)



30. Zanfir, M., Leordeanu, M., Sminchisescu, C.: The moving pose: An efficient 3d kinematics descriptor for low-latency action recognition and detection. In: Computer Vision (ICCV), 2013 IEEE International Conference on. pp. 2752–2759 (Dec 2013)