Home Monitoring Musculo-Skeletal Disorders with a Single 3D Sensor

Ruizhe Wang[1], Gérard Medioni[1], Carolee J. Winstein[2], and Cesar Blanco[3]

[1]Computer Vision Lab, IRIS
[2]Division of Biokinesiology and Physical Therapy
[3]Alfred E. Mann Institute for Biomedical Engineering
University of Southern California
{*ruizhewa,medioni,winstein,cblanco*}*@usc.edu*

Abstract

*We address the problem of automated quantitative evaluation of musculo-skeletal disorders using a 3D sensor. This enables a non-invasive home monitoring system which extracts and analyzes the subject's motion symptoms and provides clinical feedback. The subject is asked to perform several clinically validated standardized tests (*e.g. *sit-to-stand, repeated several times) in front of a 3D sensor to generate a sequence of skeletons (*i.e. *locations of 3D joints). While the complete sequence consists of multiple repeated Skeletal Action Units (SAU) (*e.g. *sit-to-stand, one repetition), we generate a single robust Representative Skeletal Action Unit (RSAU) which encodes the subject's most consistent spatio-temporal motion pattern. Based on the Representative Skeletal Action Unit (RSAU) we extract a series of clinical measurements (*e.g. *step size, swing level of hand) which are crucial for prescription and rehabilitation plan design. In this paper, we propose a Temporal Alignment Spatial Summarization (TASS) method to decouple the complex spatio-temporal information of multiple Skeletal Action Units (SAU). Experimental results from people with Parkinson's Disease (PD) and people without Parkinson's Disease (non-PD) demonstrate the effectiveness of our methodology which opens the way for many related applications.*

1. Introduction

The population of patients with musculo-skeletal disorders (*e.g.* Parkinson's Disease, Stroke) has been continuously increasing worldwide over these years, with approximately 50,000 new cases of Parkinson's Disease each year. The musculo-skeletal disorder evaluation plays a key role in determining the patient's medication prescription as well as the rehabilitation plan and it is usually carried out by asking the subject to perform several standardized tests (*e.g.* walk back and forth, sit and stand that are components of the United Parkinson's Disease Rating Scale (UPDRS) [8]) while the clinicians observe the activity for stability, smoothness and coordination. Although currently the patient-clinician interactive evaluation mode shown in Fig 1(a) dominates, there is clearly room for better, more effective and efficient approaches due to the following six reasons. First, the clinician's evaluation is mostly subjective instead of quantitative. Though there are clinical scales such as the UPDRS, these tools suffer from low resolution and the need for significant training before one can obtain valid and reliable metrics. Second, some of the evaluation process is simple and is often repeated providing strong support for automatization. Third, the entire process can be

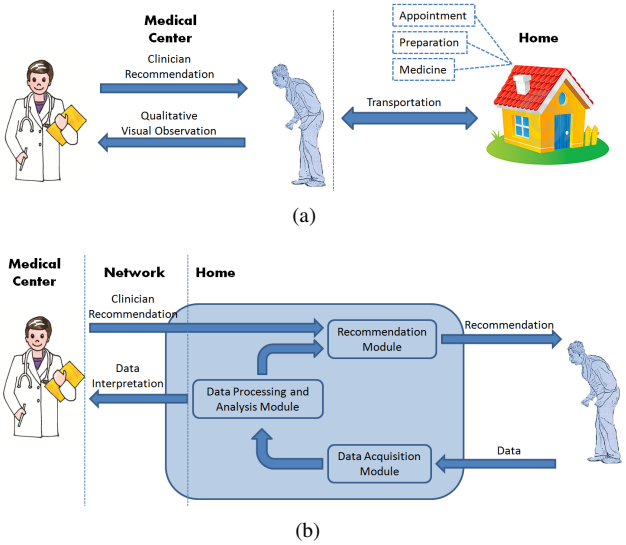


Figure 1. (a) Traditional patient-clinician evaluation mode (b) New home-based monitoring and evaluation mode

978-0-7695-4990-3/13 $26.00 © 2013 IEEE
DOI 10.1109/CVPRW.2013.83

time consuming considering the patient needs to travel to the appointment, prepare and even take medication in advance. Fourth, the patients usually prefer to stay at home instead of travelling to the clinician's office where the risk of injury could be higher. Five, the patients may behave differently when examined in the outpatient clinics [5]. Six, there are not enough medical resources to satisfy all patients' day-to-day requirements.

A reliable and accurate home monitoring and evaluation system is an attractive alternative (Fig 1(b)). The system acquires the subject's data and processes it to enable meaningful interpretations (*e.g*. objective measurements). The processed data is sent to the clinician via network for further clinical recommendations. The home monitoring system then provides feedback to the user combining the clinician's and the system's recommendations. Since the activity assessment is carried out in the patient's familiar environment, the data are more ecologically valid and the inconvenience and potential risks associated with a clinic visit are reduced. More importantly, the recommendations are now based on both the clinician's experience and the quantitative analysis. This enables several types of applications.

Rapid adjustment for more precise medication level. Patients with musculo-skeletal disorders may need to take medication everyday to maintain the mobility level. For patients with PD, excessive anti-Parkinsonian medication can cause overactivity while insufficient dosage could lead to Freezing of Gait (FOG). The system recommends a more precise medication level based on the extracted clinical measurements compared with the baseline.

Long term evaluation and prediction. The subject's functionalities fluctuate over time and inevitably decreases with disease progression. The system evaluates and predicts the trend long term by acquiring and analyzing the patients' data at certain intervals (*e.g*. once a week). In other words, the system builds a long term motion profile of the subject.

Rehabiliation. The patient's long term motion profile provides the clinician with rich information to adjust and manage the rehabilitation plan in a customized manner. The system also helps evaluate the patient's progress, such as a response to a specific rehabilitation program.

Several attempts have been made at automating and quantifying home-based musculo-skeletal disorders evaluation. The Objective Parkinson's Disease Measurement (OPDM) System [4] can extract the motor score of a patient by putting a combination of accelerometers, gyroscopes and magnetometers on the subject's sternum, wrists, ankles, and sacrum. After the subject performs several standardized tests, all motion data are processed to calculate a single motion score indicating the subject's mobility level. Several similar systems like Kinesia HomeView [1], Motus Movement Monitor [3] also exist.

While these existing systems have been clinically validated and partially solve the home monitoring challenge, they are quite intrusive by asking the subject to put on several sensors each time before use. Also, these systems provide single measurements instead of detailed analysis, which is a much richer representation. These incomplete solutions present an opportunity for developing better tools using computer vision. While traditional Motion Capture (MOCAP) systems can accurately track a subject, their cost and cumbersome set-up prevent wide applications outside the laboratory environment. The recent release of low-cost 3D sensors which generate depth streams provides a possible solution.

We acquire the sequential skeletal data of the subject using a single 3D sensor. The skeletal stream is further processed to decouple the complex spatio-temporal information. Finally we generate a representative skeletal sequence which exhibits the subject's most consistent motion pattern. Based on this representation, we extract detailed spatio-temporal objective measurements.

Our main contributions are: 1) A non-invasive home monitoring and evaluation system for patients with musculo-skeletal disorders using remote sensor technologies; 2) A methodology for skeletal data processing to decouple the spatio-temporal information; 3) An accurate and reliable skeletal sequence representation based on which multiple objective measurements are extracted; 4) Experimental evaluation results of PD people and non-PD people.

Sec 2 presents the relevant literature. Sec 3 describes our proposed method in details including overview, data acquisition, segmentation, temporal alignment, spatial summarization and TASS validation. Sec 4 demonstrates the effectiveness of our method by several experiments on the PD and non-PD subjects. Sec 5 ends with the conclusion and future work.

2. Related Work

Many works have been proposed to monitor and evaluate patients with musculo-skeletal disorders. The most prevailing methodology suggests using one or a combination of intrusive sensors, such as accelerometer, gyroscope, magnetometer, pedometer, *etc*. The clinically meaningful indicators are further extracted by analyzing the pattern of time series data generated by these sensors. Zijlstra and Hof [25] attach a triaxial accelerometer to the pelvis of the subject with Parkinson's Disease as he/she walks. They differentiate left and right steps and extract measures including the duration of subsequent stride cycles, step size as well as walking speed. Weiss *et al*. [22] use the same setup and analyze data in the frequency domain instead of the time domain which is robust to noise. They find the dominant frequency, amplitude and slope of PD people is lower than non-PD people indicating a larger gait variability. Others [9, 20] use a wrist-worn activity monitor to examine the

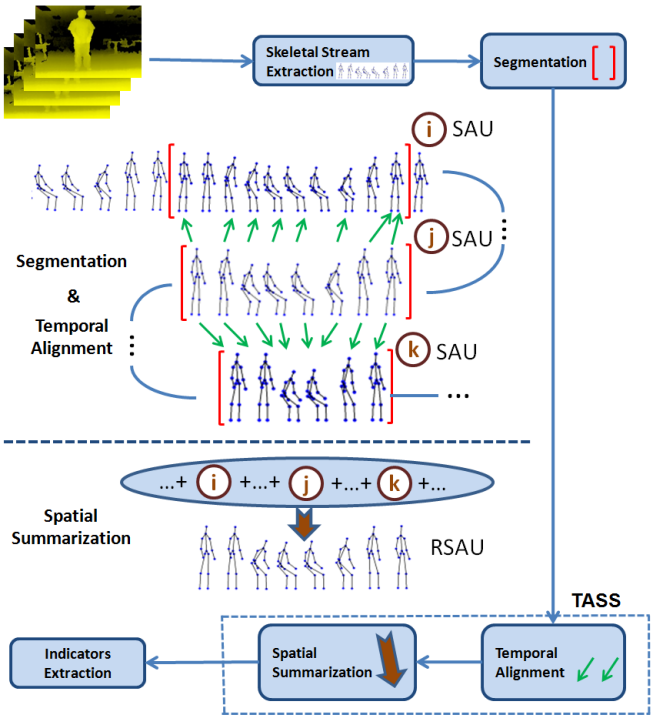


Figure 2. General pipeline of the proposed method

mobility patterns in PD patients. Salarian *et al*. [16, 17] examined activity patterns in PD patients (*e.g*. tremor and bradykinesia) by placing 2 gyroscopes on the forearms and 3 inertial sensors on the shanks and trunk. These methods are cumbersome by asking the subject to put on and put off sensors.

A few vision-based methods have also been proposed. Cho *et al*. [7] differentiate the PD people and the non-PD people by analyzing the corresponding gait videos. They extract the subject's silhouette in scene and further reduce the dimensionality using Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). However their method requires specific setup (*e.g*. pure color background for silhouette extraction) and is not feasible for the home setting. Also their method stays on the stage of differentiation instead of quantification.

To the authors' best knowledge, ours is the first work proposed for clinical evaluation of patients with musculoskeletal disorders using a 3D sensor. Our method relies on several key computer vision subfields: stereo vision, pose estimation and temporal alignment. Relevant works of respective subfields are introduced in corresponding sections.

3. Method

3.1. Overview

The general pipeline of our method is shown in Fig 2. While the outer loop displays the flowchart, the inner square exhibits the data. The corresponding symbols (*i.e*. brackets, arrows) visualize our process and intermediate results. We use the 3D sensor to capture a depth stream of the subject performing a standardized clinical assessment and further extract a skeletal stream (Sec 3.2). While the complete skeletal stream is cyclical in nature, we segment it into multiple repeated Skeletal Action Units (SAU) by detecting the periodicity in a projected feature space (Sec 3.3). Although all SAUs are motion sequences of the same subject in a short time period, they differ in both temporal domain and spatial domain. We propose the key Temporal Alignment Spatial Summarization (TASS) method to decouple the complex spatial-temporal information and generate a single robust representation. First, a SAU is selected as the reference and all other SAUs are temporally aligned with the reference (Sec 3.4) to build the temporal correspondences. Then all SAUs are summarized spatially to generate the Representative Skeletal Action Unit (RSAU) (Sec 3.5) which encodes the most consistent motion pattern observed among all SAUs. We validate the choice of TASS method on the MSR-Action 3D Dataset (Sec 3.6).

3.2. Data Acquisition

The system set-up is easy and convenient. A 3D sensor is horizontally fixed (*e.g*. on table, top of TV) and a skeletal stream is extracted while the subject performs standardized tests in front of it.

3D sensor and skeleton extraction algorithm. We use the Kinect sensor which can provide 640×480 depth images at 30 fps. Skeletons are extracted in real time using [18] which is implemented in the Microsoft Kinect SDK [2].

Standardized tests. A number of standardized instruments for patients with musculo-skeletal disorders have been proposed and clinically validated. We list the ones we use for our experiments.

Walking Test. The subject is asked to walk back and forth multiple times along a line for a one way distance of four meters. The test is adapted from a standardized six meters walking test considering the working range of the 3D sensor.

Walking-with-counting Test. The subject is asked to count from 1 to 100 while walking. A constant number is added each time, *e.g*. 1, 5, 9,...,97. The added cognitive load of counting is an indicator of automatic control of the primary walking task.

Sit-to-stand Test. The subject is asked to sit down and stand up as fast as possible multiple times.

Each test usually consists of five repetitions and can be carried out by a subject in several minutes.

All tests are cyclical in nature and consist of multiple simple action units. *For the walking-based tests, we define an action unit to be two steps. For the Sit-to-stand Test, we define an action unit to be sitting and standing one time.* Al-

though the test is performed by a subject in a short period of time, we often observe variations between repeated action units. We analyze and summarize them into two categories.

Large motion variation. For a person with musculo-skeletal disorders, some action units exhibit extreme pose. For example, in the Walking Test while the patient is doing fine with most steps, he/she shows extremely unbalanced pose during a few steps. On the other hand, for a person without musculo-skeletal disorders, large variation is rarely observed and the motion is more consistent.

Small motion variation. No one performs the same task exactly the same every repetition. For example, each step actually differs slightly in length while people walk. This type of variation is generally small and observed for all people.

The small motion variation is random, and we regard as noise. Our purpose is to eliminate it by averaging multiple repeated action units (Sec 3.5). In other words, we enhance the consistent motion pattern by removing this noise. On the other hand, we use a robust method to detect the few action units with large motion variation. The action units are regarded as outliers and should not be taken into account for averaging. In order to average or reject an action unit, we must decouple the complex spatio-temporal information so that frame-to-frame correspondenes are built and comparing two skeletons from two different SAUs in Euclidean Space is meaningful.

3.3. Segmentation

A key observation is that the skeletal stream consists of multiple repeated Skeletal Action Units (SAU). For example, walking two steps is defined as a SAU for walking-based tests while sitting down and standing up one time is defined as a SAU for the Sit-to-stand Test. We formulate the segmentation problem of the skeletal stream as follows. A skeletal stream is represented as $\mathbf{X} = [\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_N}] \in \mathbb{R}^{3M \times N}$ where N is the total number of frames, M is the number of 3D joints per frame, and $\mathbf{x_i} = [\mathbf{p^t_{i,1}}, \mathbf{p^t_{i,2}}, ..., \mathbf{p^t_{i,M}}]^T \in \mathbb{R}^{3M \times 1}$ with $\mathbf{p_{i,j}} = (x_{i,j}, y_{i,j}, z_{i,j})^t \in \mathbb{R}^{3 \times 1}$. We are trying to find the Segmentation Vector

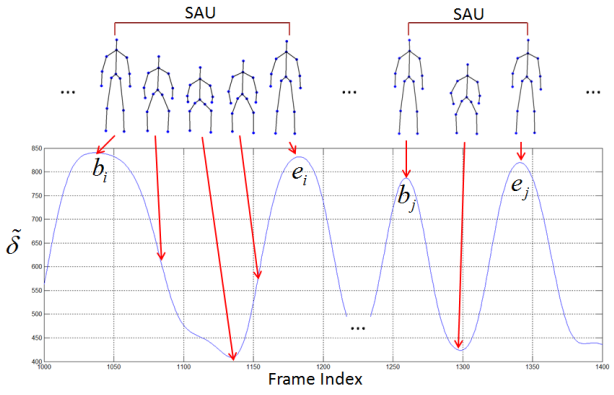


Figure 3. Illustration of segmentation based on periodicity of feature space $\delta(\mathbf{X})$

$$\mathbf{S} = [b_1, e_1, b_2, e_2, ..., b_K, e_K]^t \in \mathbb{R}^{2K \times 1} \quad (1)$$

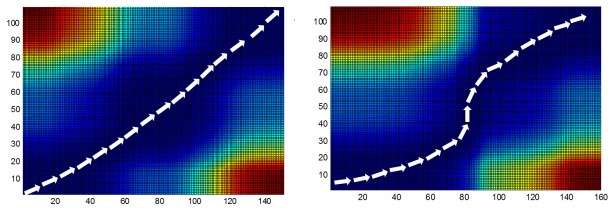
such that the K segmented SAUs can be represented as $\mathbf{X_S} = \{\mathbf{X_{b_i:e_i-1}}, i = 1, 2, ..., K\}$. In other words, (b_i, e_i) are the indices of the start frame and the end sentinel frame of the ith segmented SAU and they must satisfy $1 \leq b_1 < e_1 \leq b_2 < e_2 \leq ... < t_K \leq N$.

We define a feature function $\delta : \mathbb{R}^{3M \times 1} \rightarrow \mathbb{R}$ on $\mathbf{X}$ such that $\delta(\mathbf{X}) = [\delta(\mathbf{x_1}), \delta(\mathbf{x_2}), ..., \delta(\mathbf{x_N})] = [\delta_1, ..., \delta_N] \in \mathbb{R}^{1 \times N}$. While the periodicity of motion is observed in the $\mathbb{R}^{3M \times N}$ space, it is projected to the $\mathbb{R}^{1 \times N}$ subspace by a well designed δ function. To better deal with noise, the feature subspace is further convoluted with a Gaussian filter such that $\tilde{\delta}_i = \sum_{j=i-h}^{i+h} \frac{e^{-\frac{(j-i)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} \delta_j$ and $\mathbf{S}$ is finally determined by detecting the local maximas of $[\tilde{\delta}_1, \tilde{\delta}_2, ..., \tilde{\delta}_N]$. *For the sit-to-stand test, δ is defined to be the height of the head joint and an example is shown in Fig 3. For the walking-based tests, δ is defined to be the signed distance between two feet.*

3.4. Temporal Alignment

Temporal alignment methods. Many temporal alignment algorithms have been initially proposed to solve the video synchronization problem [15], using Dynamic Time Warping (DTW) or its variant. Recently, Canonical Component Analysis (CCA), which is proposed for learning the shared subspace between two high dimensional features, has been extended as Canonical Time Warping (CTW) [24] to address the spatio-temporal alignment between two human motion sequences. Another work proposed by Gong and Medioni [10] address the problem using Dynamic Manifold Warping (DMW). They extend previous works on spatio-temporal alignment by incorporating manifold learning and employing a novel robust similarity metric. In this paper, we use their method and give an intuition and brief introduction. We strongly encourage the interested readers to read the original paper.

Problem formulation. Given multiple segmented SAUs $\mathbf{X_S}$, we pick one SAU as the reference (as explained later) and all other SAUs are aligned with the reference in a pairwise manner. Two simplify the notations, we formulate the problem of temporal alignment between two SAUs. Given two SAUs $\mathbf{X_{1:L_x}} \in \mathbb{R}^{3M \times L_x}$ and $\mathbf{Y_{1:L_y}} \in \mathbb{R}^{3M \times L_y}$, we try to find the Optimal Alignment Path $\mathbf{Q} = [q_1, q_2, ..., q_{L_y}] \in \mathbb{R}^{1 \times L_y}$ that aligns Y_i with X_{q_i}. We do that by minimizing the following loss function ($\|\|_F$ is the



(a) (b)

Figure 4. (a) Linear alignment path between two SAUs (b) Non-linear alignment path between two SAUs

Frobenius norm operator),

$$Ł_{DMTW}(\mathcal{F}(\cdot),\mathcal{F}(\cdot),\mathbf{W}) = \|\mathcal{F}(\mathbf{X_{1:L_x}}) - \mathcal{F}(\mathbf{Y_{1:L_y}})\mathbf{W}^T\|_F^2 \quad (2)$$

where $\mathcal{F}$: $\mathbb{R}^{3M\times 1} \rightarrow \mathbb{R}^{L\times 1}$ maps $\mathbf{X_{1:L_x}}$ and $\mathbf{Y_{1:L_y}}$ to $\mathcal{F}(\mathbf{X_{1:L_x}}) \in \mathbb{R}^{L\times L_x}$ and $\mathcal{F}(\mathbf{Y_{1:L_y}}) \in \mathbb{R}^{L\times L_y}$ in a L dimensional subspace. $\mathbf{W} \in \{0,1\}^{L_x\times L_y}$ encodes $\mathbf{Q}$ such that when $q_i = k$ we have $\mathbf{W_{i,k}} = 1, i = 1,2,...,L_x$.

Completion score. Although each $\mathbf{x_i}$ in $\mathbf{X_{1:L_x}}$ lies in a high dimensional space, the natural property of human pose suggests that $\mathbf{x_i}$ has lower intrinsic number of degrees of freedom. In other words, $\mathbf{X_{1:L_x}}$ can be regarded as traversing along a path $\mathfrak{M}_p$ on a spatial-manifold $\mathfrak{M}$. The geodesic distance $d_{Geo}(\mathbf{x_i},\mathbf{x_{i+1}};\mathfrak{M}_p)$in $\mathfrak{M}$ can be further estimated using Tensor Voting [13] which is a non-parametric framework proposed to estimate the geometric information of manifolds. Knowing the geodesic distance between consecutive frame, we assign a completion score ζ_i to frame $\mathbf{x_i}$ as

$$\zeta_i = \frac{\sum_{s=1}^{i-1} d_{Geo}(\mathbf{x_s},\mathbf{x_{s+1}};\mathfrak{M}_p)}{\sum_{s=1}^{L_x-1} d_{Geo}(\mathbf{x_s},\mathbf{x_{s+1}};\mathfrak{M}_p)}. \quad (3)$$

By defining $\mathcal{F}(\cdot)$ as $\mathcal{F}(\mathbf{x_i}) = \zeta_i^x$ we can rewrite the loss term in (2) as $Ł_{DMTW}(\mathbf{W}) = \|\zeta^{\mathbf{x}} - \zeta^{\mathbf{y}}\mathbf{W}\|_F^2$ where $\zeta^{\mathbf{x}} = [\zeta_1^x,\zeta_2^x,...,\zeta_{L_x}^x]$ and $\zeta^{\mathbf{y}} = [\zeta_1^y,\zeta_2^y,...,\zeta_{L_y}^y]$.

Algorithm. The key for solving Eq 2 lies in the Temporal Aligning Matrix $\mathbb{A} = \{a_{i,j}\}_{L_x\times L_y}$ where $a_{i,j} = \|\mathcal{F}(\mathbf{x_i}) - \mathcal{F}(\mathbf{y_j})\|^2$. And in our case $a_{i,j} = (\zeta_i^x - \zeta_j^y)^2$. Two examples are shown in Fig 4. The Optimal Alignment Path is found by looking for the shortest path traversing from $\mathbb{A}_{1,1}$ to $\mathbb{A}_{L_x,L_y}$ (white arrows in Fig 4). Dynamic programming provides an efficient solution in $O(L_xL_y)$ to obtain $\mathbf{Q}$ after calculating $\mathbb{A}$ following the traditional Dynamic Time Warping (DTW) algorithm.

Temporal Aligning Score η. The Optimal Alignment Path indicates how two time series match each other temporally. Fig 4(a) displays a linear path indicating that while one motion sequence is slower than another the delay is equally distributed. On the contrary, Fig 4(b) exhibits non-linear correspondences between two time series. From a medical perspective, this indicates possible pauses, or even worsely, Freezing of Gait (FOG) (Sec 4) of the subject during specific repetition. While not covered in the original paper, we introduce η to quantify the non-linearity of the Optimal Alignment Path. Given $\mathbf{X_{1:L_x}}$, $\mathbf{Y_{1:L_y}}$ and the optimal alignment path $\mathbf{Q} \in \mathbb{R}^{1\times L_y}$, we define the the Temporal Aligning Score as

$$\eta \propto \frac{\sum_{i=1}^{L_y} |(L_x-1)(i-1) - (L_y-1)(q_i-1)|}{L_y((L_x-1)^2 + (L_y-1)^2)}. \quad (4)$$

η is a normalized scale-invariant score quantifying how much the alignment path $\mathbf{Q}$ deviates from the line $y = \frac{L_x-1}{L_y-1}(x-1)+1$. The larger the score, the bigger the non-linearity among the temporal correspondences between two SAUs. For Fig 4(a), $\eta = 0.0093$ while for Fig 4(b) $\eta = 0.0355$.

Reference selection. The reference SAU must match all other SAUs as much as possible in the temporal domain. So we pick one SAU at a time, align it with all remaining SAUs and calculate the sum of the Temporal Aligning Scores. The sum of all Temporal Aligning Scores indicates how worse the current SAUs conform to other SAUs and we pick the one with the minimum score. In other words, the reference SAU is

$$\arg\min_i \sum_{j\neq i} \eta_{i,j} \quad (5)$$

when defining $\eta_{i,j}$ as the Temporal Aligning Score between the ith and the jth SAUs.

3.5. Spatial Summarization

After segmentation and temporal alignment, we have segmented SAUs $\mathbf{X_S} = \{\mathbf{X_{b_i:e_i-1}}, i = 1,2,...,K\}$ as well as the Optimal Alignment Path for each SAU $\{\mathbf{Q}^i \in \mathbb{R}^{1\times\tilde{L}}, i = 1,2,...,K\}$ where $\mathbf{Q}^i = [q_1^i, q_2^i, ..., q_{\tilde{L}}^i]$ and $\tilde{L}$ is the reference SAU's total number of frames. In other words, we have a set of temporally aligned SAUs $\{\mathbf{X_{Q^i}}, i = 1,2,...,K\}$ where $\mathbf{X_{Q^i}} = [\mathbf{x}_{q_1^i}, \mathbf{x}_{q_2^i}, ..., \mathbf{x}_{q_{\tilde{L}}^i}]$. As mentioned before (Sec 3.2), these SAUs exhibit small and large motion variations. We regard the small motion variation as noise and the large motion variation as outlier. We try to cancel out small variations by averaging over SAUs of consistent motion pattern while detecting the SAUs with large motion variation.

Problem Formulation. The RSAU is represented as $\mathbf{X_R} = [\mathbf{x_1^R}, \mathbf{x_2^R}, ..., \mathbf{x_{\tilde{L}}^R}] \in \mathbb{R}^{3M\times\tilde{L}}$. We find the optimal RSAU by minimizing

$$\sum_{i=1}^{K}\sum_{j=1}^{\tilde{L}} \mathcal{D}(\mathbf{x_j^R}, \mathbf{x_{q_j^i}}) + \lambda \sum_{j=1}^{\tilde{L}-1} \mathcal{D}(\mathbf{x_j^R}, \mathbf{x_{j+1}^R}), \quad (6)$$

where $\mathcal{D}$: $\mathbb{R}^{3M\times 1} \times \mathbb{R}^{3M\times 1} \rightarrow \mathbb{R}$ is interpreted as a distance measure between two skeletal frames. In all

our experiments, we define $\mathcal{D}(\cdot,\cdot)$ to be $\mathcal{D}(\mathbf{x_i},\mathbf{x_j}) = \sum_{k=1}^{M} \|\mathbf{p_{i,k}} - \mathbf{p_{j,k}}\|^2$, *i.e.* the Euclidean Distance between $\mathbf{x_i}$ and $\mathbf{x_j}$. Eq 6 has a data term and a smoothness term and λ is a parameter used to tune the ratio between them.

Data Term. $\sum_{i=1}^{K}\sum_{j=1}^{\tilde{L}} \mathcal{D}(\mathbf{x_j^R}, \mathbf{x_{q_j^i}})$ minimizes the distance between the RSAU with all SAUs. This term forces RSAU to behave like an average result of multiple SAUs and is used to remove small motion variations.

Smoothness Term. $\lambda\sum_{j=1}^{\tilde{L}-1} \mathcal{D}(\mathbf{x_j^R}, \mathbf{x_{j+1}^R})$ minimizes the distance between two consecutive frames of the RSAU. While the 3D sensor extracts skeletons at 30fps, the relative motion between two consecutive frames is small. Hence this term is useful for rejecting large measurement noise (*e.g.* Some joints of the skeleton jumping between consecutive frames). λ tunes the ratio between these two terms. When the input skeleton is very noisy, λ is set larger to enforce smoothness. In all our experiments, we set $\lambda = 0.1$ for the Microsoft Kinect SDK skeletal data (actually this data has already been smoothed).

Algorithm. Eq 6 is rewritten as a Linear Least Square problem which is efficiently solved by Gaussian Elimination. In practice, we combine RANSAC with Eq 6 to help detect the SAUs with large motion variation. At each iteration, we select a subset of SAUs and calculate the corresponding RSAU. Then we compare the distance between the RSAU with all SAUs to find the inliers. The RSAU with the most number of inliers is output as our final result.

3.6. TASS Validation

TASS is used to capture the most consistent motion pattern among multiple SAUs and incorporate such information into a single RSAU. RSAU can be regarded as a rich and robust representation from which many clinically relevant parameters can be extracted. To validate our choice of the TASS approach, we apply it on the 3D activity recognition task and compare its performance on the MSR-Action 3D Dataset [12] with several state-of-the-art 3D activity recognition algorithms.

MSR-Action 3D Dataset contains 20 actions performed 3 times by each of 10 subjects. We removed some extreme outliers and used 547 out of the original 567 SAUs. These SAUs cover various movement of arms, legs and torso.

First we normalized all skeletons to the same size. Then for each action type we trained a single RSAU using the TASS method. In this case, RSAU captures the most consistent motion pattern of a specific action type crossing different subjects. At the testing stage, given a SAU, we temporally aligned all 20 RSAUs with it and calculate the Euclidean Distance. The testing SAU is associated with the action with the minimum Euclidean Distance. We report accuracy on the cross-subject test setting [12] and compare our method with other state-of-the-art methods (Tab 1).

If we randomly select a SAU to represent an action type, we achieve an accuracy of 0.606. After we enhance the representation using the TASS method, we boost the accuracy up to 0.819. Our method reaches competitive performance even comparing with most fine-tuned 3D activity recognition algorithms. The result demonstrates the effectiveness of the TASS method and shows that RSAU is a much richer representation than single SAU.

Table 1. Recognition Accuracy Comparison on MSR-Action 3D Dataset

Method	Accuracy
Dynamic Temporal Warping [14]	0.540
Random Sampled SAU	**0.606**
Action Graph on Bag of 3D Points [12]	0.747
View-invariant Histogram of 3D Joints [23]	0.789
RSAU	**0.819**
Actionlet Ensemble [21]	0.882

4. Experiments

Parkinson's Disease is a degenerative disorder of the central nervous system [11]. The most obvious symptoms of Parkinson's Disease are motion related which include slowness of movement (*i.e.* Bradykinesia), resting tremor, rigidity and postural instability. While some symptoms are consistent, others occur on an episodic basis (*e.g.* Festination and Freezing of Gait) [6]. Detecting and quantifying these symptoms are crucial for deciding exact medicine level, designing rehabilitation plan and preventing falls which are most risky for PD patients [5]. Two experiments were conducted and the data was processed using our proposed TASS method. The RSAU was able to quantify a series of spatio-temporal clinical measurements which are crucial for evaluation of the subject's mobility level.

Walking-based experiment. This experiment was performed by a PD subject and a non-PD subject. These two subjects were similar in size. The non-PD subject performed the Walking Test while the PD subject carried out both the Walking Test and the Walking-with-counting Test. Three RSAUs were generated from our data. A series of spatio-temporal indicators were then automatically extracted.

Step size. We project the joints of two feet to the ground. The step size is calculated based on the difference between the start position of the first frame and the end position of the last frame.

Postural Swing Level. Postural Swing Level (PSL) quantifies how stable the pose is while a person walks. We extract it by projecting the joint of torso (*i.e.* Center of Mass) to the ground. Then we find its Maximum Absolute Deviation along the direction perpendicular to the subject's moving direction. The larger the value, the more unstable the pose is.

Arm Swing Level. Arm Swing Level (ASL) indicates the degree of arm motion. We extract it by projecting the hand joint and the torso joint to the ground and calculating the signed distance between them for each frame. Again we use the Maximum Absolute Deviation. The smaller the ASL, the stiffer the specific arm.

Stepping time. This indicates the most consistent time of the subject walking two steps.

Table 2. Results of PD and non-PD on walking-based tests

Indicator	non-PD W	PD W	PD W&C
L step size	95.2	96.1	56.7
R step size	91.3	90.2	57.3
PSL	3.05	5.13	8.35
L ASL	12.29	5.49	3.76
R ASL	10.68	10.66	3.28
Stepping time	0.89	1.61	2.14

The experimental results are displayed in Tab 2 with spatial indicators measured in centimeters and temporal indicators measured in seconds. We have the following observations.

—While the non-PD subject and the PD subject show similar step sizes during the Walking Test, the PD subject's step size decreases dramatically as soon as the dual counting task is added.

—The PD subject exhibits larger Postural Swing Level than the non-PD subject and the situation gets worse after adding the dual task. The result quantifies the PD subject's Postural Instability.

—The PD subject has a stiff left arm during the walking test and both arms become stiff with the dual task added. The result quantifies the Rigidity of the PD subject's arms.

—It takes more time for the PD subject to walk two steps and even more time to walk two steps while counting. The result quantifies the Bradykinesia of the PD subject.

Sit-to-stand experiment. The same PD subject and non-PD subject carried out the Sit-to-stand Test. Similar spatio-temporal measurements were extracted from the respective RSAU which quantified the motion differences between the PD subject and the non-PD subject. They are not displayed considering redundancy. The PD subject exhibited pauses during a specific repetition, and we temporally aligned the corresponding SAU with the RSAU. Fig 5 displays the corresponding Temporal Aligning Matrix. During pauses, the Optimal Alignment Path shows sparser points and this is well captured by the Kernel Density Estimation [19] method as local minimas. In Fig 5 we picked up two pauses with one representing the subject leaning against the chair and the other representing the subject having difficulty standing up. Detection of pause and/or FOG is important for the clinician to evaluate the subject's postural instability as well as risk of fall.

Applications. Quantifying the PD patients' musculoskeletal disorders has many potential applications.

—Falls prevention and warning. Abnormal pause, occurence of Freezing of Gait, and extreme postural instability are often strong indicators of falls which are very dangerous for the PD patients. Hence quantification of them is crucial for prevention or alerting the patient/clinician of possible falls.

—More precise medication level. The PD patient usually takes medicine everyday to maintain their 'on' state, *i.e.* smooth skeletal muscle movements. As the medicine effect 'wears off', the person becomes very stiff, slow and may even be unable to move in a few minutes. In some cases of the PD patients, the 'on-off' fluctuations are unpredictable. Our system can recommend them to take medicine at the exact time with the exact amount by analyzing several objective measurements, *e.g.* step size, stepping time.

—Long term monitoring to observe disease progression and/or treatment effectiveness. The PD patient gradually lose the motor ability as an inevitable effect of the progressive deterioration of the nervous system. Building a long term quantified profile for each specific patient is important for deterioration evaluation and prediction, rahabilitation plan design and even help the clinicians better understand the underlying mechanism and come up with a more effective solution.

Discussion. The key idea here is not what medical measurements we can extract from this specific experiment, but rather we provide a robust methodology to decouple the complex spatio-temporal information. Instead of providing single scores, we provide the source (*i.e.* RSAU) from which many well-defined medical indicators can be extracted and used in many related applications. It is worth noting that, however, our main focus in this paper is not

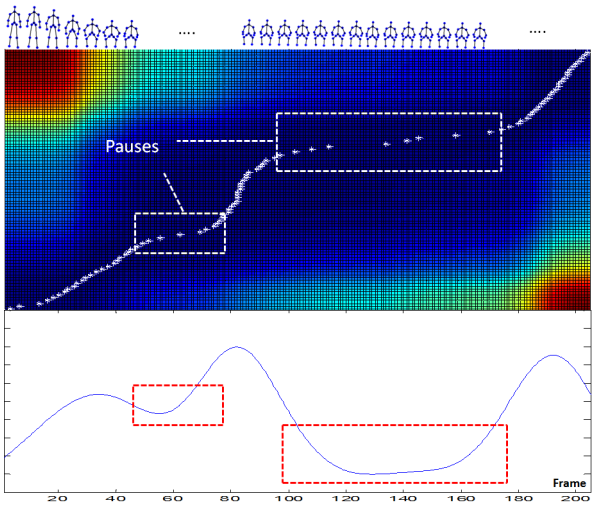


Figure 5. Detection of Pause/FOG (Top to bottom: Skeletal stream; Temporal Aligning Matrix; Estimated density)

clinical validation of these spatio-temporal gait parameters. This process asks for more experiments on the PD subjects and close collaboration with experienced clinicians.

5. Conclusion and Future Work

To summarize, we address the following problems: A non-invasive home monitoring and evaluation system for patients with musculo-skeletal disorders using a 3D sensor; A methodology to decouple the spatio-temporal information between multiple skeletal sequences; A compact and robust skeletal sequence representation based on which multiple medical indicators are extracted; Experimental evaluation results on the PD and non-PD subjects.

For the future work, we plan to focus on studying people with Parkinson's Disease and gather more data to refine our methodology. Also we will work closely with the clinicians and design the most clinically informative spatio-temporal measurements.

References

[1] Kinesia homeview. http://glneurotech.com/kinesia/homeview/. 2

[2] Microsoft kinect sdk. http://www.microsoft.com/en-us/kinectforwindows/. 3

[3] Motus motion monitor. http://www.motusbioengineering.com/index.htm. 2

[4] Objective parkinson's disease measurement. http://kineticsfoundation.org/. 2

[5] B. R. Bloem, Y. A. Grimbergen, M. Cramer, M. Willemsen, and A. H. Zwinderman. Prospective assessment of falls in parkinson's disease. *Journal of neurology*, 248(11):950–958, 2001. 2, 6

[6] B. R. Bloem, J. M. Hausdorff, J. E. Visser, and N. Giladi. Falls and freezing of gait in parkinson's disease: a review of two interconnected, episodic phenomena. *Movement Disorders*, 19(8):871–884, 2004. 6

[7] C.-W. Cho, W.-H. Chao, S.-H. Lin, and Y.-Y. Chen. A vision-based analysis system for gait recognition in patients with parkinsons disease. *Expert Systems with applications*, 36(3):7033–7039, 2009. 3

[8] S. Fahn, R. Elton, U. D. Committee, et al. Unified parkinsons disease rating scale. *Recent developments in Parkinsons disease*, 2:153–163, 1987. 1

[9] P. J. Garcia Ruiz and V. Sanchez Bernardos. Evaluation of actitrac®(ambulatory activity monitor) in parkinson's disease. *Journal of the neurological sciences*, 270(1):67–69, 2008. 2

[10] D. Gong and G. Medioni. Dynamic manifold warping for view invariant action recognition. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 571–578. IEEE, 2011. 4

[11] J. Jankovic. Parkinsons disease: clinical features and diagnosis. *Journal of Neurology, Neurosurgery & Psychiatry*, 79(4):368–376, 2008. 6

[12] W. Li, Z. Zhang, and Z. Liu. Action recognition based on a bag of 3d points. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 9–14. IEEE, 2010. 6

[13] P. Mordohai and G. Medioni. Dimensionality estimation, manifold learning and function approximation using tensor voting. *The Journal of Machine Learning Research*, 11:411–450, 2010. 5

[14] M. Müller and T. Röder. Motion templates for automatic classification and retrieval of motion capture data. In *Proceedings of the 2006 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 137–146. Eurographics Association, 2006. 6

[15] C. Rao, A. Gritai, M. Shah, and T. Syeda-Mahmood. View-invariant alignment and matching of video sequences. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 939–945. IEEE, 2003. 4

[16] A. Salarian, H. Russmann, F. J. Vingerhoets, P. R. Burkhard, and K. Aminian. Ambulatory monitoring of physical activities in patients with parkinson's disease. *Biomedical Engineering, IEEE Transactions on*, 54(12):2296–2299, 2007. 3

[17] A. Salarian, H. Russmann, C. Wider, P. R. Burkhard, F. J. Vingerhoets, and K. Aminian. Quantification of tremor and bradykinesia in parkinson's disease using a novel ambulatory monitoring system. *Biomedical Engineering, IEEE Transactions on*, 54(2):313–322, 2007. 3

[18] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from single depth images. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 1297–1304. IEEE, 2011. 3

[19] B. W. Silverman. *Density estimation for statistics and data analysis*, volume 26. Chapman & Hall/CRC, 1986. 7

[20] B. van Hilten, J. I. Hoff, H. A. Middelkoop, E. A. van der Velde, G. A. Kerkhof, A. Wauquier, H. A. Kamphuisen, and R. A. Roos. Sleep disruption in parkinson's disease: assessment by continuous activity monitoring. *Archives of neurology*, 51(9):922, 1994. 2

[21] J. Wang, Z. Liu, Y. Wu, and J. Yuan. Mining actionlet ensemble for action recognition with depth cameras. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1290–1297. IEEE, 2012. 6

[22] A. Weiss, S. Sharifi, M. Plotnik, J. P. van Vugt, N. Giladi, and J. M. Hausdorff. 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. 2

[23] L. Xia, C.-C. Chen, and J. Aggarwal. View invariant human action recognition using histograms of 3d joints. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*, pages 20–27. IEEE, 2012. 6

[24] F. Zhou and F. De la Torre. Canonical time warping for alignment of human behavior. *Advances in Neural Information Processing Systems (NIPS)*, pages 1–9, 2009. 4

[25] W. Zijlstra and A. L. Hof. Assessment of spatio-temporal gait parameters from trunk accelerations during human walking. *Gait & posture*, 18(2):1–10, 2003. 2