Movement Analysis of Rehabilitation Exercises: Distance Metrics for Measuring Patient Progress

Roshanak Houmanfar, Michelle Karg, and Dana Kulić

Abstract—**Mobility improvement for patients is one of the primary concerns of physiotherapy rehabilitation. Providing the physiotherapist and the patient with a quantified and objective measure of progress can be beneficial for monitoring the patient's performance. In this paper, two approaches are introduced for quantifying patient performance. Both approaches formulate a distance between patient data and the healthy population as the measure of performance. Distance measures are defined to capture the performance of one repetition of an exercise or multiple repetitions of the same exercise. To capture patient progress across multiple exercises, a quality measure and overall score are defined based on the distance measures and are used to quantify the overall performance for each session. The effectiveness of these measures in detecting patient progress is evaluated on rehabilitation data recorded from patients recovering from knee or hip replacement surgery. The results show that the proposed measures are able to capture the trend of patient improvement over the course of rehabilitation. The trend of improvement is not monotonic and differs between patients.**

Index Terms—**Biomedical monitoring, biomedical signal processing, computer aided diagnosis, human motion analysis, motion measurement, motion quality assessment, rehabilitation robotics.**

I. Introduction

The application of machine learning techniques to human motion analysis has grown rapidly over the past few years. Measurement and analysis of physiotherapy data have the potential to provide an objective and quantitative measure of patient progress over the course of physiotherapy treatment.

During a typical physiotherapy session, the physiotherapist instructs the patient to perform a number of exercises, each with several repetitions. The set of exercises chosen and the number of repetitions may be customized for each patient. In current clinical practice, the patient's performance is typically assessed using visual observation of the patient's motions and questionnaires, e.g., the Community Balance and Mobility Scale [1], the Falls Efficiency Scale [2]. Goniometry, a technique of measuring joint angles which isolates a single body joint in order to evaluate range of motion [3], can also be used, but is not accurate when the subject is moving e.g., during exercises and functional rehabilitation.

The current measurement and assessment techniques require additional physiotherapist effort and monitoring, and are not capable of measuring during movement. An automated system could provide the therapist with numerical metrics to assess the patient's recovery process and potentially allow physiotherapists to assess the effectiveness of various treatment protocols over a population of patients.

Patient data analysis for progress monitoring is a challenging task because of the complexity of human motion. Human movement consists of synchronous recruitment of multiple degrees of freedom (DoF), making single DoF comparisons incomplete and possibly unreliable. Human motion exhibits significant temporal and spatial variability for different repetitions of the same exercise. Since humans differ in characteristics such as age, gender, height, and weight, variability between different subjects is also observed. When recovering from an illness or surgery there are variabilities caused by progress and improvement through rehabilitation, differing levels of pain during the course of treatment, as well as differing levels of fatigue over the course of a session. During the course of rehabilitation, patients frequently are observed to exhibit *compensation*. *Compensation* refers to the recruitment of additional or different DoFs [4] while performing a certain exercise. The correct form of the exercise and the DoFs recruited are prescribed by the physiotherapist. Other sources of variability are due to the measurement system and the algorithms used for deriving the joint angles.

The goal of progress monitoring is to identify the variability caused by recovery and improvement. The presence of multiple other sources of variability makes this task challenging. Furthermore, the exercises are performed based on a specific regimen instructed by the physiotherapist for each patient. Therefore, the proposed approach should be flexible to detect patient improvement for any set of exercises.

In previous work, we have developed a body-worn sensor system and associated algorithms for measuring human movement during rehabilitation. The overall system is illustrated in Fig. 1. The data is collected from body worn inertial measurement unit (IMU) sensors attached to the patient and the joint angle positions, velocities, and accelerations are derived [5]. The data are then segmented such that each segment begins with the start of an exercise repetition and ends when the exercise repetition is finished [6]. In this paper, we propose an approach for progress estimation based on the segmented motion data. Descriptive features are either extracted from joint angle positions, velocities, and accelerations or from a statistical model of these data. The former is a common approach in the biomechanics literature [7]–[9] whereas the latter provides a model of the timeseries based on multiple repetitions and is more common in the machine learning literature [10], [11]. In addition to collecting patient data, we also collect data from healthy participants performing the same exercises, and use the healthy population data as a reference. Distance

Manuscript received October 31, 2013; revised February 28, 2014 and May 20, 2014; accepted May 27, 2014. Date of publication July 1, 2014; date of current version August 23, 2016.

The authors are with the University of Waterloo, Department of Electrical and Computer Engineering, Waterloo, ON N2L 3G1, Canada.

Digital Object Identifier 10.1109/JSYST.2014.2327792

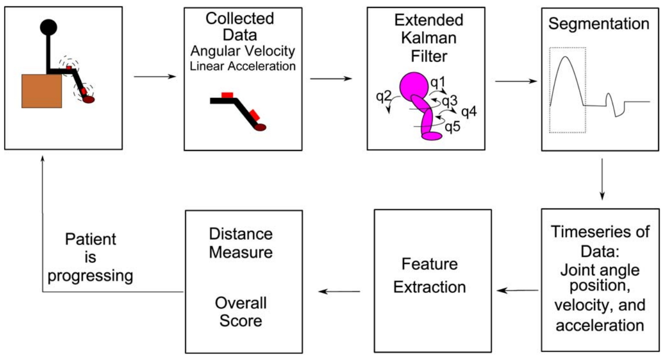


Fig. 1. Overall system is depicted. The IMU sensors are mounted on the patient's knee and ankle. Angular velocity and linear acceleration are collected from the sensors and the five joint angles are estimated using the extended Kalman filter [12] (q_1: extension/flexion of the hip, q_2: internal/external rotation of the hip, q_3: abduction/adduction of the hip, q_4: extension/flexion of the knee, q_5: internal/external rotation of the lower limb). The data is then segmented [6] so that each segment starts with the beginning of one *repetition timeseries* and ends when the repetition is completed. Features are extracted from the joint angles' segmented timeseries and are used to obtain the measures of progress.

measures are proposed to quantify the performance quality of a single repetition of an exercise, a set of repetitions of the same exercise, and a set of different exercises. The distances are calculated based on kinematic data and compare movements of one subject to the average healthy exercise performance.

This paper is organized as follows: Section II overviews the related work and motivates the application of continuous measures in physiotherapy rehabilitation. In Section III, two approaches are proposed to formulate distance measures for each specific exercise and the overall score for a session of multiple exercises. The proposed approaches are evaluated on a synthetic data set in Section IV. Clinical data set collection and the experimental evaluation of the proposed approaches are detailed in Section V. A discussion of the results is presented in Section VI. Section VII outlines conclusions and directions for future work.

II. Related Work

Human movement analysis is an active area of research with a wide field of applications including action recognition [13], gait identification [14], gesture recognition [15], motion imitation in robotics [16], affective human computer interaction [17], sport science [18], medical diagnosis [19], and rehabilitation [20]. The goal of these applications is either to recognize what movement is performed, e.g., [21] or how a movement is performed, e.g., [15]. Automatic human movement analysis for rehabilitation exercises targets the latter to discriminate between movements performed by healthy and patient populations [22] and perform illness diagnosis [23].

Typically, for a set of movements, key elements of human movement performance are extracted concatenating position, velocity, and acceleration information in a common feature vector [10], [11], [24]. These important features are used to separate the unhealthy population from the healthy population. Most studies base their methods on classifiers that can discriminate between the healthy and unhealthy populations e.g., [19], [25]–[27]. These studies rely on a patient database for training such a classifier [26], [27]. There are also studies that focus on monitoring features that change when a certain medication or treatment is applied to a group of patients [28], [29] or focus on detecting features that are specific to the patient population [30]. Unlike classification methods which distinguish only between two classes (healthy versus patient), we focus on patient monitoring and the detection of gradual changes in patient performance due to rehabilitation. To date, only a few studies focus on assessing the correctness of exercises performed [10] and analyzing continuous changes in the movement performed [11], [24].

Upper body functionality post-stroke is considered in [24]. The data from 77 healthy control subjects and 46 stroke patients performing a single exercise is collected using a robotic exoskeleton; both data sets are used in feature selection and classifier training. A multi-layered neural network is used to select features and distinguish between healthy and patient populations. The summation of outputs in the last layer is used to estimate the continuous measure of progress for patients. The analysis is performed on sessions up to 50 days apart.

Taylor *et al.* [10] consider three typical multiple knee osteoarthritis rehabilitation exercises and record the movements with wearable accelerometers. Descriptive features such as mean, minimum, and maximum, are calculated from the sensor readings and directly used in a multi-label classifier to distinguish between correct performance and several common compensation strategies. The Adaboost algorithm with linear classifiers for each feature is used for classification. Only data collected from a healthy population is used in the analysis. The healthy population data is labeled using expert opinion and analysis is performed on motions that have recognizable differences.

Zhang *et al.* [11] focus on post stroke rehabilitation. Motion data is collected with IMUs and raw sensor output is used for feature extraction after basic filtering. The timeseries data for each sensor is partitioned for different exercises. Partitions that correlate least with corresponding partitions of other exercises are considered as motion templates. The patient data is then cross-correlated with the templates and the peak values of the cross-correlation are considered as the features. Data is collected from rehabilitation professionals and a single patient is used for testing. K- Nearest Neighbours is used to classify patient's motions and the distance from the center of cluster is the estimate of continuous progress.

The current state of the art develops models for healthy and patient populations [10], [11], [24] and therefore is capable of assigning the class labels "healthy" or "unhealthy" to captured movement sequences. The disadvantage of classification techniques is that they cannot explicitly model continuous progress and therefore are not suitable for continuous monitoring purposes. Some classifiers, e.g., neural network classifiers [24] often need fine tuning, are hard to replicate or extend because their structure makes clinical interpretation difficult.

Many of the works in the state of the art focus on one specific exercise [11], [24]. This is a limitation for monitoring patients over the course of rehabilitation because the exercise regimen consists of more than one exercise. Furthermore, many of the current works [10], [24] validate their methods based on synthetic and simulated data due to lack of patient data. Validation of studies considering continuous labeling (e.g., [24]) is difficult because an objective quantized ground truth of continuous progress is rarely available. Quantitative assessment

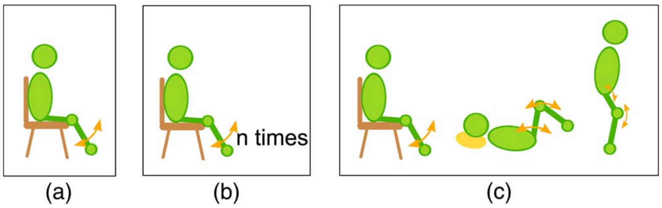


Fig. 2. *Repetition timeseries* and the *repetition set* provide the performance measures δ, and Δ. The overall score S assesses an *exercise set*. During the knee extension exercise, the subject performs a full knee extension/flexion without moving any other joints. During the knee hip extension exercise, the subject lies on the ground and performs a knee hip extension/flexion simultaneously. During the squat, the subject bends his knees and hip while standing. (a) Repetition timeseries. (b) Repetition set. (c) Exercise set.

scores are often not collected for each physiotherapy session of a patient due to the limited time in each session. Therefore, visual graphs and cross validations (e.g. comparison to other classifiers' performance) are common methods for validation.

An objective and quantified measure of patient improvement can be beneficial for monitoring patient progress. In this paper, we propose a technique that estimates the continuous measure of patient improvement, capable of handling a variety of exercises. We validate our proposed approaches based on both synthetic and clinical data. Of the challenges summarized above, we address capturing the variability caused by improvement in human motion, validating the proposed approaches based on clinical data, and handling different exercise regimens for each patients and each session. We do not address factors such as pain and fatigue that affect human motion, and we do not address how the causes for improvement or degradation in performance can be identified.

III. Methodology

Analysing patient progress during physiotherapy requires answering the following questions: 1) How to assess one repetition of one exercise?, 2) How to assess multiple repetitions of one exercise?, and 3) How to combine the evaluations from different exercises and obtain a score that denotes the overall performance for a session?

In answering the questions above, we assume that the motion data is available in the form of joint angle positions, velocities, and accelerations. We also assume that motion data is available from a healthy population performing the same set of exercises. We make the assumption that, at the time of the analysis, we know which exercise is being performed, that the data is segmented such that one single repetition of a certain exercise is a *repetition timeseries* $\omega = [\boldsymbol{\gamma}(1)\ \boldsymbol{\gamma}(2)\ \ldots\ \boldsymbol{\gamma}(T)]$, where T is the duration of the repetition for that exercise, and $\boldsymbol{\gamma}$ is a vector of joint kinematics $\boldsymbol{\gamma} = [q_1\ q_2\ \ldots\ \dot{q_1}\ \dot{q_2}\ \ldots\ \ddot{q_1}\ \ddot{q_2}\ \ldots]$. Multiple repetitions of the same exercise performed in the same session are the *repetition set* for that exercise $\Omega = \{\omega_1, \ldots \omega_n\}$ where n is the number of repetitions. The set of multiple exercises performed in the same session are the *exercise set* of that session $\Gamma = \{\Omega_1, \ldots, \Omega_m\}$ where m is the number of different exercises performed in the session. (See Fig. 2).

We propose two approaches to extract variation due to progress through rehabilitation. In the *feature-based method*, descriptive measures are extracted from the joint angle timeseries. The *HMM-based method* relies on features extracted from a generative model for the joint angle timeseries. For both approaches, we use the healthy population data as the reference for assessing performance. Measures δ and Δ for assessing one and multiple repetitions of one exercise are introduced based on a comparison between the healthy population and the patient. The overall score S is calculated as a function of these measures for multiple exercises in one session.

A. Feature-Based Approach

In the feature-based method, the mean, minimum, maximum, skew and range of motion of the joint angle positions, velocities, and accelerations plus the duration of each *repetition timeseries* are considered as the feature vector

$$\mathbf{v} = \left[\mu_{q_1}\ \min_{q_1}\ \max_{q_1}\ \text{skew}_{q_1}\ \text{rom}_{q_1}\ \mu_{q_2}\ \ldots\ \text{duration}\right] \tag{1}$$

$$\text{skew}_{q_i} = \frac{\frac{1}{T}\sum_{j=1}^{T}\left(q_{i_j} - \mu_{q_i}\right)^3}{\left(\sqrt{\frac{1}{T}\sum_{j=1}^{T}\left(q_{i_j} - \mu_{q_i}\right)^2}\right)^3} \tag{2}$$

$$\text{rom}_{q_i} = \max_{q_i} - \min_{q_i}. \tag{3}$$

This definition of the feature vector is desirable because it allows modeling the timeseries of the data using statistical features. This method is fast to compute and can capture the attributes of the timeseries from one example. However, since the features are defined directly from the timeseries, the approach is more affected by unwanted variabilities such as noise.

1) Feature Selection: To assess the performance, it is essential to extract the informative features from the feature vector and exclude those that are uninformative or redundant. The informative features extracted from the feature vector are the top features. Different features may be informative for different exercises, therefore the top features are selected automatically from the data by looking for those features which show the most variation over the course of treatment and are most different from the healthy population. Features that reflect what changes most throughout the rehabilitation are chosen using Least Absolute Shrinkage and Selection Operator (LASSO) [31]. LASSO is a regression tool which can also be used for selecting features.

For a set of inputs $f_1, f_2, \ldots, f_k$, an output y and the following linear model:

$$\hat{y} = w_0 + w_1 f_1 + w_2 f_2 + w_3 f_3 + \cdots + w_k f_k. \tag{4}$$

LASSO adjusts the weights $w_0, \ldots, w_k$ such that $\sum(\hat{y} - y)^2$ is minimized and $\sum_{i=0}^{k} w_i < t$ where $t \geq 0$ is a tuning parameter [31]. The parameter t is selected so that the weights w are larger than zero for only five features. Preliminary experiments showed that the algorithm is not sensitive to this value. Any value of t resulting in a range of 5–25 features results in the same performance measures. When w_i becomes zero, the input f_i does not contribute to minimizing $\sum(\hat{y} - y)^2$, i.e., f_i is either uninformative or its information is redundant. These remaining 5 features are considered as the top features in the subsequent analysis.

The inputs $f_1, f_2, \ldots, f_k$ are the features of the *repetition timeseries* and the output y is the corresponding session number. The session numbers are normalized between 0 and 1, such that 0 corresponds to a patient's first session, and 1 corresponds to a patient's last session. We select features that change with every session and among these features the ones that most likely correspond to a linear relationship. This selection allows us to find the features that are changing as patients progress through the sessions. For the purposes of feature selection, we also consider the healthy population data in this regression. For the healthy population, the label y is set to be 100 times larger than the patients' last session. Introducing this outlier forces the regression to be in the direction of the healthy population data and helps to detect the features that not only change with the progress of the patients but also separate the healthy population from patients. The value of y for the healthy population directly affects the value of the weights, but the chosen features are not changed as long as y is sufficiently large. We do not use the values of the weights in our analysis and only use the features selected by this method.

As there are multiple sources of variation in human motion, we cannot assume that there is a linear relationship between the number of days in treatment and motion features. We use a linear model only for feature selection, i.e., for identifying which features change during the course of treatment and discriminate between the healthy population and patient data.

2) Measure of Performance for Repetition Timeseries: To obtain a measurement for the performance of one exercise, the top feature vectors are extracted from the patient (V_P) and healthy population (V_H) data as explained in Section III-A1

$$V'_H = V_H(\text{top}_{features}) \tag{5}$$
$$V'_P = V_P(\text{top}_{features}). \tag{6}$$

Based on our observations, to a smaller degree, healthy individuals employ the same compensation strategies that patients use when performing an exercise, i.e., healthy subjects show the same compensation strategies due to mental and physical fatigue, lack of physical readiness, and misunderstanding the exercise instructions. For example, in the knee extension exercise, the correct form of the exercise is to perform a full range of knee extension while keeping the other joints still. Based on our observations, the healthy population often compensates with additional hip extension. We assume that among the features chosen by LASSO, the ones with higher variance in the healthy population are more informative, because the highly variant features are either features of the moving joint or are the features describing the compensation strategy. Therefore, more weight is given to the more variant features in defining the distance measure. When comparing these features, standard normalization (normalizing with a mean zero and standard deviation of one) is performed on the data to make the data unitless. The distance δ between the patient repetition and the healthy population data evaluates each repetition

$$\mu_H = \text{mean}\left(V'_H\right) \tag{7}$$
$$\Sigma_H = \text{diag}\left(\text{std}\left(V'_H\right)\right) \tag{8}$$
$$\delta_i = \left(V'_{P_i} - \mu_H\right)^T \Sigma_H \left(V'_{P_i} - \mu_H\right) \tag{9}$$

where V'_H is the healthy population top feature vector, V'_{P_i} is the patient top feature vector for the ith *repetition timeseries*, μ_H is the mean of the healthy population top feature vectors, Σ_H is the diagonal matrix of standard deviations for the healthy population top feature vectors, and δ_i is the distance between one repetition of the exercise performed and the healthy group's performance. We assume that as patients improve they get closer to the healthy data and therefore a decrease in the value of δ over the course of rehabilitation indicates improvement.

3) Measure of Performance for Multiple Repetitions of the Same Exercise: For each patient, δ represents the measure of performance for one repetition of an exercise. The median of the distance measures (δ) calculated for one exercise over the session is considered as the overall distance measure for the *repetition set* of that exercise

$$\Delta_\Omega = \text{median}(\boldsymbol{\delta}_\Omega) \tag{10}$$

where Δ_Ω is the overall performance of one exercise in one session and $\boldsymbol{\delta}_\Omega$ is the vector of distance measures calculated for every *repetition timeseries data* ω_i in the *repetition set* Ω. The median is used to lessen the sensitivity to outliers.

4) Measure of Performance for a Combination of Exercises: The distance Δ_Ω describes the patient performance for one exercise (i.e., Ω_j) in each session. There are multiple exercises performed in each physiotherapy session (i.e., Γ) that need to be considered together for overall patient progress assessment. Quality and quantity are the two factors that affect scoring an exercise. Based on our observations, the features computed from exercises performed by the healthy population have larger variances when the exercise is more difficult. We therefore assume that the distance measures of the healthy population have larger variance for more difficult exercises.

The distance measures (δ) are calculated for every *repetition timeseries* of the healthy population data according to (9) and are considered as the comparison reference. The healthy population distance measure vector $\boldsymbol{\delta}_{H_j}$ is the vector of the distance measures calculated for every *repetition timeseries* of exercise Ω_j in the healthy population data. The patient distance measures (Δ_{P_j}) are calculated for the *repetition set* of every exercise Ω_j in the *exercise set* Γ. The mean and standard deviation of $\boldsymbol{\delta}_{H_j}$ are considered as the measure of exercise difficulty

$$\mu_{\boldsymbol{\delta}_{H_j}} = \text{mean}\left(\boldsymbol{\delta}_{H_j}\right) \tag{11}$$
$$\sigma_{\boldsymbol{\delta}_{H_j}} = \text{std}\left(\boldsymbol{\delta}_{H_j}\right) \tag{12}$$

where $\mu_{\boldsymbol{\delta}_{H_j}}$ is the mean of the healthy population distance measure vector $\boldsymbol{\delta}_{H_j}$ and $\sigma_{\boldsymbol{\delta}_{H_j}}$ is the standard deviation of the healthy population distance measure vector $\boldsymbol{\delta}_{H_j}$.

We define the measure of quality for a *repetition set* of an exercise j performed by the patient as

$$Q_j = \frac{\left(\Delta_{P_j} - \mu_{\boldsymbol{\delta}_{H_j}}\right)}{\sigma^a_{\boldsymbol{\delta}_{H_j}}} \tag{13}$$

where Δ_{P_j} is the patient's distance measure for the *repetition set* of exercise j, and a is the index that penalizes Q based on the exercise difficulty, i.e., larger a increases the importance of exercise difficulty. We observed that the best value for a is 2, which can be interpreted as the inverted dispersion index [32]. A perfect performance over any *repetition set* Ω results in a value of zero for the overall distance measure $(\Delta_\Omega = 0)$. The overall score for the patient in a specific session is calculated as the difference between the norm of the score resulting from a perfect performance and the norm of the weighted quality measures. The quality measure Q_j of an exercise j is weighted by its number of repetitions. The score of the patient for a given session is calculated using the following equation:

$$S = \sqrt{\sum_{\Omega\in\Gamma}\left(\frac{n_\Omega}{\sum\limits_{\Omega\in\Gamma} n_\Omega}\frac{\mu_{d_{H_\Omega}}}{\sigma^a_{d_{H_\Omega}}}\right)^2} - \sqrt{\sum_{\Omega\in\Gamma}\left(\frac{n_\Omega}{\sum\limits_{\Omega\in\Gamma} n_\Omega}Q_\Omega\right)^2} \tag{14}$$

where Γ is the *exercise set*, Ω is an exercise in the Γ, n_Ω is the number of repetitions for exercise Ω, and Q_Ω is the quality measure calculated for exercise Ω using (13). The score S is formulated such that performing a difficult exercise in a session would improve a patient's score. Furthermore, we assume that exercises with more repetitions in one session are the main focus of that session and therefore, the quality measures Q are weighted by the number of repetitions for each exercise. The score S is defined as the difference between a perfect weighted quality measure and the patient's weighted quality measure hence progress is assumed to result in smaller values for this measure.

The healthy population's distance values are often small and have a small variance compared to patient data. To avoid dividing the quality measure with a value less than 1 when normalizing by the healthy population's distance measure variance $\sigma_{\boldsymbol{\delta}_{H_j}}$, all δ values are scaled uniformly such that all variance values of the healthy population's distance measures become greater than 1. The algorithm flexibility in defining any *exercise set* allows us to calculate the overall score for any arbitrary set of exercises.

B. HMM-Based Approach

The Hidden Markov Model (HMM) [33] based approach relies on features extracted from HMMs modeling the joint angle timeseries. HMMs are trained on the *repetition set* of each exercise for the healthy and patient populations.

Individual HMMs are learned for each member of the healthy population and for each session of each patient. Each *repetition set* is modeled using a 3 state, left-to-right model. State 1 corresponds to moving to the desired posture, state 2 corresponds to holding the desired posture, and state 3 corresponds to returning to the starting posture. The observations of the HMMs are the position, velocity, and acceleration of the joint angles. The mean and variance of the observation distributions in each state are considered as the feature vector

$$\mathbf{v} = \left[\mu_{state1_{q_1}}\ \sigma_{state1_{q_1}}\ \mu_{state2_{q_1}}\ \cdots\ \mu_{state3_{\ddot{q}_5}}\ \sigma_{state3_{\ddot{q}_5}}\right]. \tag{15}$$

LASSO feature selection is used to choose the ten most informative features following Section III-A1. We use the same procedure as Section III-A2 in calculating distance measures: the distance measure, Δ, is calculated using (9), the quality measure Q is calculated using (13), and the overall performance score S is calculated using (14).

The HMM is capable of capturing the statistical essence of a dynamic timeseries. Such a definition of the feature vector is beneficial, because it models the most likely timeseries and the probability of variations. Furthermore, the feature-based approach requires expertise in predefining the features whereas the HMM captures the features that describe the pattern of the data automatically. However, the HMM is computationally more expensive than the feature-based approach and requires multiple samples of the timeseries data, which may not be available. This also means that an individual repetition cannot be evaluated.

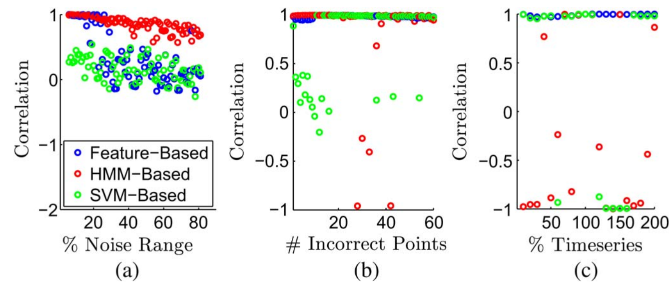


Fig. 3. Effect of each source of variability on the correlation index between the average score of each method and the ground truth average score. (a) Noise. (b) Late segmentation. (c) Early segmentation.

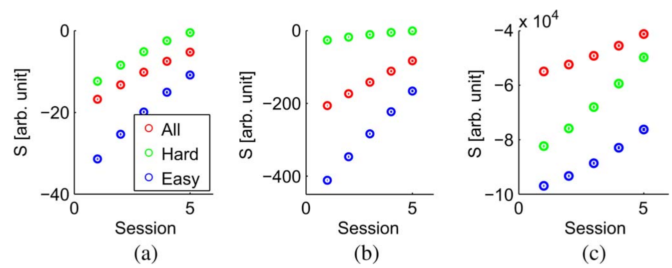


Fig. 4. The results of removing each set of exercise on the average overall score for each of the approaches. (a) Feature-based. (b) HMM-based. (c) SVM.

IV. Synthetic Data Experiments

The synthetic data test is designed to evaluate the proposed approaches with a known ground truth signal. The synthetic data is generated for a scalar position, velocity, and acceleration timeseries based on two simulated exercises, an upward bell curve and a downward bell curve. One of the exercises is considered as a "hard" exercise, while the other an "easy" exercise. We generate a simulated healthy and patient data sets, both including temporal and spatial variability, where the patient range of motion and execution time improve as they advance through sessions. A detailed description of the synthetic data generation procedure is provided in the supplementary material.

The HMM-based approach, feature-based approach, and the Support Vector Machine (SVM)[1] are applied on the synthetic data. When temporal and spatial variability is removed from the synthetic data and all the data is available to all methods, the correlation in average patient scores between the three methods is over 97%. We consider the variability-free, full data case as the ground truth and investigate how correlations between this ground truth and the results from the three approaches are affected under the following conditions: 1) noisy data, 2) poor segmentation, 3) temporal variability, and 4) incomplete data.

The effect of noise is depicted in Fig. 3(a). The HMM-based approach is least affected while the feature-based approach is most affected. This is because the HMM-based approach generates a model based on a *repetition set*. The feature-based

[1]The SVM is chosen for comparison with the proposed approaches because of the results obtained with the clinical data, as described in Section V.

approach and the SVM have the poorest performance because the features are directly affected by the noise.

In the second set of tests, the effect of poor segmentation is investigated. Two types of segmentation error are possible: late segmentation and early segmentation. Late segmentation is modeled by adding points with a constant value to the end of the joint angle position in the patient data set. The results are shown in Fig. 3(b). The feature-based approach is the least affected, the HMM-based approach is affected when the number of points are significant enough to alter the states. The SVM is also affected in many cases. Early segmentation results in an incomplete timeseries. The results of this test are depicted in Fig. 3(c). The feature-based and SVM based approach are not much affected by this variability. This is a limitation of the predefined features since they do not consider the timing and do not include enough information to capture the difference between a complete timeseries and an incomplete timeseries. However, the HMM constructs a statistical model of the timeseries and is significantly affected by the incomplete data. In the third set of tests, the effect of scaling the length of the timeseries in the patient data set is analyzed. None of the approaches are significantly affected by this variability.

The fourth test is designed to investigate the effects of set sparsity when data of some exercises are not available in one session. Fig. 4 shows the values of the average overall score for each exercise when the hard or the easy exercises are not available. It can be seen from the results that set sparsity results in jumps and inconsistencies in the overall score.

V. Clinical Data Experiments

The proposed approaches are evaluated on a patient data set from patients recovering from knee or hip replacement surgery. In this type of rehabilitation, the following exercises are commonly performed: knee extension/flexion while seated, knee and hip extension/flexion while supine, and squat.[2] Data of these exercises was also recorded for a healthy population. The feature-based and the HMM-based approaches are evaluated for two cases: 1) healthy population and a subset of patient data is available for training, and 2) only healthy population data is available for training. The healthy population data is only used to learn a reference model, and results are presented for the patient data.

A. Data Collection and Pre-Processing

Motion data is collected using Shimmer sensors [34] mounted at the knee and ankle providing angular velocity and linear acceleration data (128 Hz). Position q, velocity $\dot{q}$, and acceleration $\ddot{q}$ of five joint angles consisting of knee flexion, knee rotation, hip flexion, hip abduction, and hip rotation are estimated from the sensor databased on a kinematic model and an Extended Kalman Filter [5], see Fig. 1.

Patient data was collected from eighteen inpatients during their rehabilitation at the Toronto Rehabilitation Institute. Each patient performs one 45–60 minute session per day. The number of days a patient stays in the hospital varies from 4–12 days and depends on the patient's needs and health status. The set of exercises specified by the therapist in each session differs between patients and sessions. Therefore, the *repetition set* of one specific exercise is not available for every session. The healthy population data consists of 10 subjects (age: 23 ± 4.5) performing each exercise 10–20 times. The patient population tends to be elderly. Therefore, the healthy population performed the exercises slowly to minimize the speed difference between the two populations.

TABLE I
Patient Information[3]

Patient ID (Age)	Surgery Site	Discharge Session	Special Conditions
2 (80)	Hip Joint	11	Discharged and Readmitted
8 (59)	Hip Joint	9	None
18 (86)	Knee Joint	3	Pain

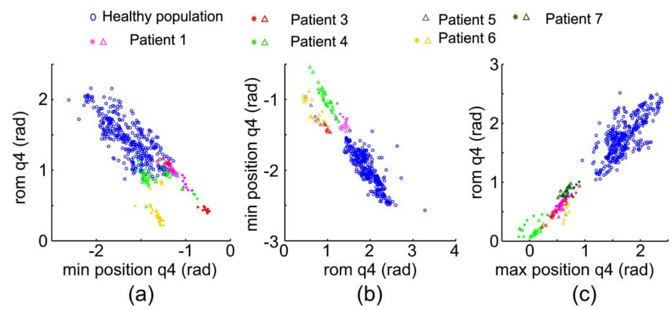


Fig. 5. Patient data and healthy population data are separable for the most informative features. The star indicates the first day the patients have performed the exercises and the triangle indicates the last day the patients have performed the exercises (only 1 session available for patient 5). q_1 is the joint angle corresponding to hip extension and q_4 is the joint angle corresponding to knee extension. (a) Knee extension/flexion. (b) Knee hip extension/flexion. (c) Squats.

A subset of the patient data is used for feature selection (patient 1, 2, 3, 5, 6, 7), all patient data is used for testing. We selected patients 2, 8, and 18 to graphically illustrate the results within this paper. Patient 2 is one of the patients included in the feature selection, and shows gradual improvement during the rehabilitation process. Patients 8 and 18 are examples of patients whose data is not used for feature selection and plots of subjects 8 and 18 illustrate the performance of estimating progress for subjects whose data is unseen during training. Patient 18 is an example of a patient who shows a rapid progress in the course of their rehabilitation, while patient 8 is an example of a patient with a common duration of recovery. General information for patients 2, 8, and 18 is summarized in Table I including any unique circumstances. Plots and information for all 18 patients can be found in the supplementary material.

B. Feature-Based Approach

The LASSO technique described in Section III-A1 is used for feature selection. Fig. 5 shows the distribution of the *repetition timeseries* of the healthy population and the training subset of the patient data over the two features selected by LASSO that have the largest variance in the healthy population. The

[2]The patients do not perform the full squat but lower their torso only slightly (i.e., knee bend of 15 degs).

[3]Patient information for all 18 patients is provided in the supplementary material.

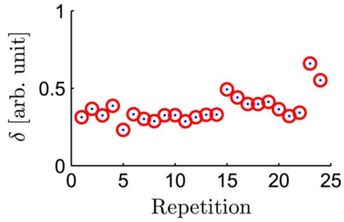


Fig. 6. Measure of performance for each *repetition timeseries* δ is illustrated for session 2 of patient 2 performing the knee extension/flexion exercise.

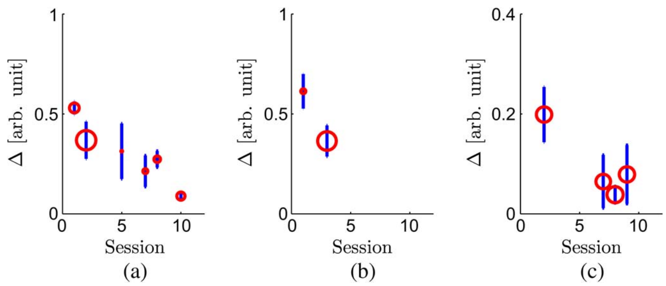


Fig. 7. Results for the distance measure δ calculated using the feature-based approach are shown for three exercises. The red circle illustrates the median of the distance measures (i.e., Δ) in each session and the blue bar depicts the variance of the distance measures δ in one session. The size of the circle indicates the number of repetitions available in each *repetition set*. For knee extension the top features are $\min_{q_4}$, mean_{q_1}, $\text{mean}_{\ddot{q}_4}$, rom_{q_4}, time. (a) P2 knee extension. (b) P18 knee extension. (c) P8 knee extension.

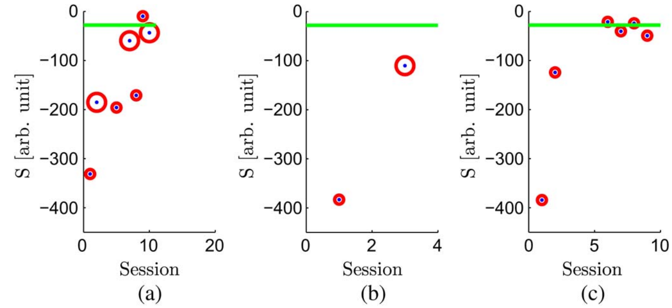


Fig. 8. An overall score S ( ) is calculated for a *exercise set*, combining individual distance measures Δ of knee extension, knee-hip extension, and squat. The size of the marker indicates the number of exercises available in each session. The green line shows the best score of the patients in their last physiotherapy session. (a) Patient 2. (b) Patient 18. (c) Patient 8.

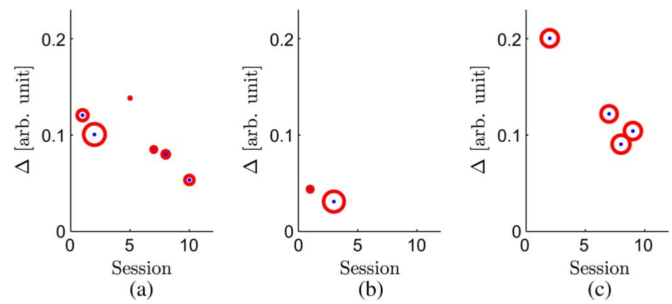


Fig. 9. HMM-based distance measure Δ_{HMM} ( ) shows the trend in progress over the sessions, here illustrated for patients 2, 8, and 18. The marker size indicates the number of repetitions available in the *repetition set*. (a) P2 knee extension; (b) P18 knee extension; (c) P8 knee extension.

clusters of the healthy population data and the patient data are separable. Furthermore, Fig. 5 shows that a patient's progress is in the direction of the variance of the healthy population data and moves toward the mean of the healthy population over the course of rehabilitation.

Fig. 6 illustrates the values of δ for the second session of patient 2. As can be seen from this figure, the approach captures the variation in exercise performance over the course of multiple repetitions. Fig. 7 shows the calculated distance measure Δ and the distribution of δ for the 3 example patients. The exercise regimen is specific to each patient. The exercises are performed in a subset of the sessions, e.g., patient 2 performs knee extensions in session 1, 2, 7, 8, and 10. Furthermore, factors such as pain, fatigue, psychological status, and environmental conditions contribute to patients' performance and it cannot be expected that the patient progress increases monotonically. For all three patients an overall improvement over the course of the physiotherapy treatment can be observed. Some patients show rapid progress and are discharged early, e.g., patient 18 [in Fig. 7(b)]. The distance measure for a *repetition set* is more reliable when the number of repetitions available for that exercise is larger. The feature-based approach generalizes to unseen patient data, e.g., the data of patients 8 and 18 was not used for the feature selection.

The quality measure Q and the overall score S for each session are obtained according to (13) and (14) using the overall distance measure Δ calculated for every *repetition set* of each session. We assume that as patients improve, the overall score increases from negative values toward zero.

Fig. 8 shows the score measures for each patient. It can be seen from the figures that the trend of the score shows progress but there are some inconsistencies in patient 2 session 8 [in Fig. 11(a)]. These inconsistencies are caused by a small number of performed exercise repetitions.

Due to differences in health status, the exercise regimen of each session is different from one patient to the other. Among the three exercises chosen for analysis in this study, there are sessions where only one of these exercises is performed and therefore the score is based solely on the performance quality of that single exercise. This results in inconsistencies in the improvement trend of the score measure since a poor performance in one exercise is not an accurate measure of the patient's overall status. The score measure estimates the patient's overall status more accurately when more exercise data from each session is available.

Visual analysis of the distribution of the most relevant features (see Fig. 5), where it can be seen that patient variation as they proceed through rehabilitation is in the direction of healthy data variability, motivates to investigate whether only healthy population data is sufficient to select the most relevant features. Such an approach is beneficial when a physiotherapist may include a new exercise into the exercise regimen and patient data is not yet available for this exercise. Healthy population data can be easily collected by the physiotherapist him/herself performing the new exercise. We investigate this extension by using only healthy population data for feature selection. Variabilities caused by initial posture and sensor positioning are highly variant in the healthy population. Because we are only considering the variation in the healthy population, such features could get selected using our current approach. To avoid this, features obtained from joint angle position were removed

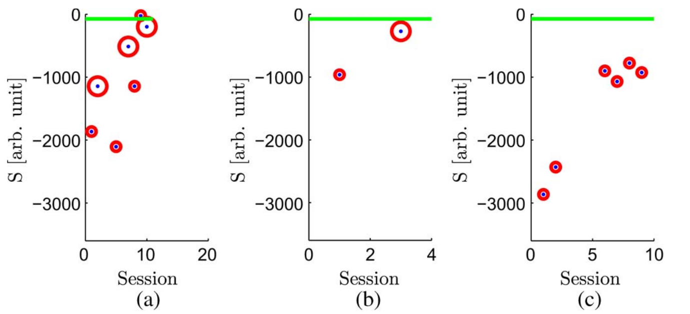


Fig. 10. Overall score S_{HMM} (○) shows the trend of progress during rehabilitation. The marker size indicates the number of exercises available for each session. ▬ shows the best score of the patients in their last session of performing the three exercises. (a) Patient 2; (b) Patient 18; (c) Patient 8.

from the feature vector. The top features are chosen from the fifteen most variant features that correlate less than .5 with each other. Equation (9) is used to calculate the distance measure for each *repetition timeseries*, (10) is used to calculate the distance measure for the *repetition set*, and (13) and (14) are used to calculate the overall score for each session.

We analyze the correlation between the overall score from the feature-based approach when using healthy and patient population data for feature selection and when using only healthy population data. The results correlate highly (over .65) for most patients. Even though feature selection based only on healthy population data does not take compensation strategies specific to the patient population into account, the extension using only healthy population data for feature selection can detect patient progress. When the overall performance of a patient is constantly high (patient 7 and 15) or low (patient 5) over the course of the rehabilitation, changes in the scores are small. In these cases, the correlation index can be low, because the two techniques differ when assessing small changes in performance.

C. HMM-Based Approach

Fig. 9 shows the overall distance measure Δ (see Section III-B) calculated for the sessions when patient 2, 8, and 18 performed the knee extension exercise. The features chosen by LASSO indicate the progress in unseen data by decreasing δ values over the course of the rehabilitation sessions, e.g., shown for patient 8 in Fig. 9(c).

Based on the quality measure Q [see (13)] of each exercise in one session the overall score S for an entire *exercise set* is calculated using (14). Fig. 10 shows the scores S_{HMM} for each patient. The scores show an overall trend of improvement for most patients. As before, the reliability of the score measure depends on the number of available exercises, i.e., outliers are usually observed when only one exercise is available to calculate the score. The features chosen by LASSO generalize well to the unseen data e.g., patient 8 whose trend of improvement is captured by the approach [in Fig. 10(c)]. Furthermore, the method captures the rapid improvement of patient 18 [in Fig. 10(b)].

Moreover, we investigate whether healthy population data is sufficient for feature selection for the HMM-based approach. Feature extraction for the healthy population is described in Section III-B. Among the first fifteen most variant features in the healthy population, those that correlate least with each other (less than .5) are chosen as the top features. The distance measure Δ_{HMM} and the overall score S_{HMM} for each patient and each session are calculated following Section III-B. The correlation between the overall score using healthy and patient population data for feature selection and using only healthy population data for feature selection is above .65 for most patients. Negative correlation indices are observed when the changes in a patient's progress are small.

If only healthy population data is available, an intuitive distance measure δ for the HMM-based approach is the loglikelihood, a common approach in gesture and motion recognition literature [8], [9], [35]. To test this approach, we compute the likelihood that a *repetition timeseries* of a patient is generated by an HMM trained only on healthy population data. The median of the log likelihoods is considered as the distance measure Δ for a *repetition set*. As patients improve and their performance becomes more similar to the healthy population the log likelihood should increase. However, this method does not capture the trend of progress for 80% of patients since distance measures Δ 2–3 times larger than their average range are observed for many repetition sets.

D. Validation

The patient's physical status is visually assessed by the physiotherapist in each rehabilitation session. The physiotherapist uses this assessment to formulate the patient's regimen and decide his or her treatment duration. This evaluation is subjective and does not have a quantified interpretation. Quantified measurements may also be taken (e.g., range of motion score, Berg scale), but these are typically not recorded for every session.

While direct quantified expert evaluation is not available for session-by-session comparison, exercise difficulty and duration can be used as an indirect measure of PT assessment. In the first sessions of rehabilitation, exercises recommended by the physiotherapist are mostly composed of supine and sitting exercises with very few repetitions. As patients improve, the recommended exercise regimen becomes harder (addition of standing exercises) and includes more repetitions. Therefore, the exercise regimen can be utilized to obtain an estimate of the clinical assessment of the patient's overall performance, i.e., overall score. To validate our approaches, we compare the score measures with an estimate of patient progress calculated based on their exercise regimen and we provide a qualitative comparison of the score measures with the physiotherapist notes in the patient health charts. Finally, we compare the proposed approaches with classifier-based approaches.

1) Comparison With Estimate of Patient Progress From Exercise Regimen: To estimate a measure of patient progress from the exercise regimen, we use the complete information of all *exercise set*s available from all patients. We consider the exercises performed in the last session for patients with fewer than 4 sessions, and the exercises performed in the last two sessions for patients with more than 4 sessions as the hard exercise set.[4] The first session exercises are considered as the

[4]We consider the last two sessions because for most patients the last session is a last check up and contains very few exercises.

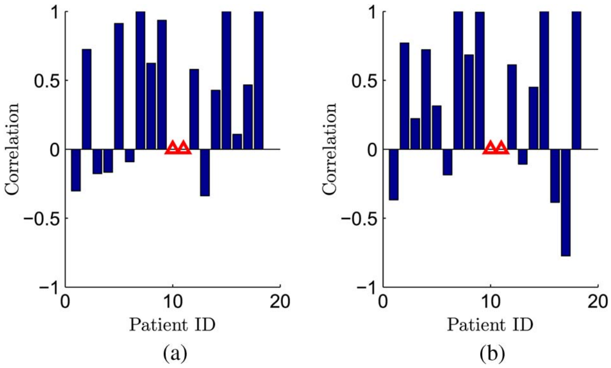


Fig. 11. Correlation between the overall score calculated for each method and the ground truth for each patient. The data of patients 10 and 11 has only 1 session available and therefore correlation cannot be calculated. (a) Feature-based approach. (b) HMM-based approach.

easy exercise set. For each exercise, the number of patients who performed the exercise on their last two sessions are counted and this number is divided by the total number of patients to obtain the probability that the exercise belongs to the hard exercise set, $p_H(\Omega)$. We eliminate the exercises performed by fewer than three patients, i.e., exercises with probability less than .15, from the hard exercise set. The same procedure is performed to determine the probability of belonging to the easy exercise set, $p_E(\Omega)$. If an exercise is not in the hard exercise set the probability that this exercise belongs to the hard exercise set, $p(H|\Omega)$, is assigned a value of .01. The same approach is used for generating the probability that an exercise belongs to the easy exercise set, $p(E|\Omega)$.

The overall measure of progress for each session and each patient is calculated as

$$S_{i_{GT}} = \frac{\sum_{\Omega\in\Gamma}\log\left(p(H|\Omega)\right)}{\sum_{\Omega\in\Gamma}\log\left(p(E|\Omega)\right)}$$
$$S_{GT} = [S_{1_{GT}}, S_{2_{GT}}, \ldots] \tag{16}$$

where i is the number of sessions, $S_{i_{GT}}$ is the ground truth overall score for each session, and S_{GT} is the overall score for all the sessions.

Fig. 11 shows the correlation index comparing each method's overall score, S, for each patient with the overall score obtained from (16). The two approaches correlate moderately in most cases (over 62% for the feature-based approach and over 55% for the HMM-based approach). Low correlations occur mostly in cases where for many of the sessions few exercises are available for evaluation, e.g., patient 4 and 3. As mentioned in Sec. IV the inconsistency in the number of available exercises between different sessions can cause jumps in the overall score, which in turn results in a poor correlation with the ground truth. The cases where the clinical assessment does not correlate well with our proposed approaches are either caused by patients who do not show a visible change in their overall score or are caused by patients who have gaps in the number of available exercises in more than half their sessions. When these patients are excluded (9 patients out of 16 remain) the mean of the correlation becomes 67% for the feature-based approach and 72% for the HMM-based approach.

2) Qualitative Comparison With Patient Health Charts: Physiotherapists assess and record patient performance and condition on admission. Even though these assessment forms often contain unfilled sections and are mostly qualitative, they include information about the initial status of the patients. Furthermore, patient performance during rehabilitation is sometimes recorded by the physiotherapists in the daily charts. In this section, we provide the physiotherapist assessments for the exemplar patients and compare these evaluations to the proposed overall scores.

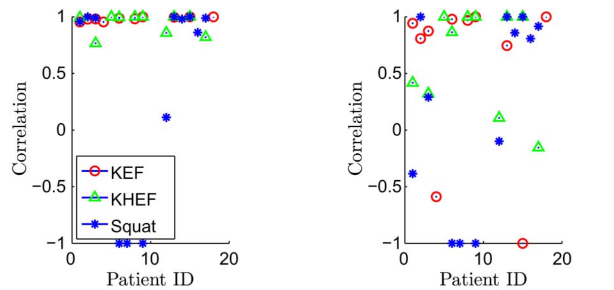
Patient 2 was admitted to the hospital after a hip joint replacement surgery. Based on the first day assessment, she was forbidden to perform hip abduction due to hip precautions. She was capable of bearing her weight, but needed assistance for rolling in bed and transferring from bed to wheelchair. She used a walker and was capable of walking for 2 meters only. The range of motion score in the recovering leg was 8/18 and the patient had a high risk of fall according to her stability test results. On the night before session 5, the patient fell causing pain in her lower extremity joints and therefore affecting her performance in session 5. This information matches the overall scores calculated for both of the proposed approaches in Figs. 8 and 10. The patient was sent back to the ER in session 8 due to complications unrelated to her surgery; the effects of this incident on her performance are captured by both of the proposed approaches in Figs. 8 and 10. The patient comes back a week later to continue her physical rehabilitation. In session 9, the patient is able to walk 70 meters with a walker with supervision and her range of motion score for the rehabilitated leg becomes 18/18. This progress is captured in the 9th and 10th session by both approaches in Figs. 8 and 10.

Patient 8 was admitted to the hospital after a hip joint replacement surgery. Based on her first day assessment, she was capable of bearing her own weight but was feeling severe pain. She had high bed mobility but needed assistance in transitioning from bed to wheelchair. She had a high fall risk according to her stability test and was capable of walking for 10 meters only. Her range of motion score was 12/18 and she could perform the exercises with assistance. In her second session she was capable of performing all her transfers independently and was capable of walking 50 meters independently using a walker. The proposed approaches both capture the progress for this patient between sessions 1 and 3 in Figs. 8 and 10. In her 9th session, she performed 20 repetitions of bilateral exercises which indicates improvement in her performance. The physiotherapist did not record the range of motion score at discharge.

Patient 18 was admitted to the hospital due to knee replacement surgery. He had a high risk of falls and had normal bed mobility. He needed supervision for bed to wheelchair transfer. He could walk 30 meters with supervision. In session 3, he had two physiotherapy sessions where he walked 40 meters supervised using a walker in the morning and 70 meters in the afternoon. Our scores capture this rapid progress for this patient between the first and third sessions in Figs. 8 and 10. For Patient 18, the physiotherapist did not record the range of motion score, either on admission or at discharge.

E. Comparison With Classifier Based Approaches

We provide a comparison of the proposed measures to classifier-based approaches trained using both healthy and patient data. Popular classifiers in the human motion literature are Naive Bayes (NB) [36], Kullback-Leibler (KL) divergence



(a) Correlation between SVM and feature-based approach for different exercises

(b) Correlation between SVM and HMM-based approach for different exercises

Fig. 12. The correlation index between (a) SVM and feature-based approach and (b) SVM and HMM-based approach for the three exercises: knee extension/flexion (KEF), knee hip extension/flexion (KHEF), and squat is for most cases above.6. (a) Correlation between SVM and feature-based approach for different exercises. (b) Correlation between SVM and HMM-based approach for different exercises.

[18], and Support Vector Machines (SVM) [37]. For NB classification, we use the top features selected by LASSO in Section III-A2. The probability of belonging to the healthy population class normalized by the summation of probabilities of belonging to healthy or patient class is considered as the distance measure δ_{NB} for each *repetition timeseries*. To compute the KL divergence, an HMM is trained for the entire healthy population and on every *repetition set* of every exercise and each patient. The symmetric KL divergence [33] for each patient is calculated for each *repetition timeseries* and the average of these values is considered as the measure of progress for one *repetition timeseries*. Both the NB and the KL approaches were unable to capture any trend of progress for the patients. The SVM provided the best results, therefore we base our comparison on the results obtained from an SVM.

The SVM formulation used is as given in [38], [39]. For the purposes of this study, we use a soft margin SVM with a linear kernel. The SVM is trained using the top features chosen by LASSO that have the highest variances in the healthy population. We consider the distance to the SVM decision hyperplane (the magnitude of the SVM-margin) as the distance measure δ_{SVM} for a *repetition timeseries* of an exercise in each session. The median of these values is considered as the overall distance measure Δ_{SVM} for a *repetition set*. The correlation between the distance measures obtained from the SVM and the feature-based approach are shown in Fig. 12(a), while those between the SVM and the HMM-based approach are depicted in Fig. 12(b). Results are reported for those patients who performed knee extension, knee hip extension, and squat exercises during rehabilitation. In 94% of the investigated cases for both approaches, the correlation is high (above .75). Only for a few cases, e.g., for patient 9, small or negative correlations are observed. For these cases the patient's status remains either constantly poor or constantly good over all the sessions, and the methods disagree on the small trends of improvement resulting in correlations less than .2. Such small differences in individual exercise trends do not affect the informative value of the overall score which provides a quantitative assessment of progress status, e.g., the score of patient 2 and 8 is above the green line in both Fig. 8 and 10. Figs. 8 and 10 also capture the rapid recovery of patient 18.

VI. Discussion

Monitoring exercise performance during physiotherapy can provide an objective measure of patient progress. Movement performance shows temporal and spatial variability caused by multiple sources including the stochastic nature of muscle recruitment, as well as individual differences in height, age, pain, fatigue, and progress. The objective of the proposed distance measures and the overall score are to capture the variability caused by improvement and progress over the course of the physiotherapy treatment.

In this paper, we estimate a continuous measure of patient performance to capture their progress through rehabilitation, whereas most existing works [10], [11] can only separate healthy from patient data using classification. We formulate a measure of performance for an *exercise set*, whereas most current works [11], [24] consider only a single exercise. Moreover, we evaluate our approach both on synthetic and patient data, whereas many of the current works focus on synthetic analysis and simulated data only. Furthermore, our proposed approach can be used when patient data for a motion is not available whereas the classification techniques require both healthy and patient data for training. The proposed approaches achieve generalization to new patients by including healthy population data as reference. Furthermore, the score measure formulations can be applied to any set of exercises as long as the corresponding healthy population data is available. This flexibility enables the physiotherapist to include patient specific or novel exercises requiring only a healthy reference set. The score measure is formulated in a way to handle individual exercise regimens and a variable number and type of exercises.

To enable feature selection when little or no patient data is available, we assumed that the healthy population exhibits the same compensation strategies as the patients to a smaller degree. This hypothesis is formulated on the basis that difficult motions result in compensatory strategies in human motion. This assumption is supported for the three exercises discussed in this paper as shown in Fig. 5. In the absence of patient data, we considered the most variant features in the healthy population as the top features. However, feature selection using both healthy and patient population data is more accurate because it allows the method to detect the compensatory strategies which are specific to the patient population.

For both the feature-based and the HMM-based approach, the distance measure Δ for a *repetition set* and the overall score S for a *exercise set* assess patient progress. The feature-based approach is faster to compute whereas the HMM-based approach provides details about each stage of the motion.

We also compared the proposed approach to estimating patient progress based on the magnitude of the SVM-margin between the healthy and patient population data. Our proposed approach has a high degree of correlation with the SVM-based approach, while requiring less training data. SVM requires feature selection on top of our LASSO feature selection to identify the most variant features, and requires training data from both the healthy and patient population. In its current form, the SVM is not capable of capturing the progress based on different exercises. We combine the SVM approach for generating distances with our approach to generate the overall performance score for multiple exercises using the SVM. The results obtained with synthetic data illustrate that the proposed approaches are superior to this classification method in the presence of noise, inaccurate segmentation, and incomplete timeseries.

We also compared the proposed approach to physiotherapist evaluations by computing the correlation between progress estimate and advancement of the exercise regimen, and qualitatively. The method correlates well with physiotherapist evaluations, but it is not yet possible to determine from the current data set whether the proposed approach enhances the ability of the physiotherapist to perform diagnosis and assessment. This will be the focus of our future work. Even if the proposed approach does not provide additional useful information over what a physiotherapist can observe visually, it can be used when a physiotherapist is not available to observe a patients motion (e.g., when a physiotherapist is observing multiple patients in the same session or when the patient is performing rehabilitation at home).

The proposed measures consider the improvement due to exercise performance; other factors such as pain and psychological status are not included in our analysis. Different pain treatments can affect motion performance, e.g., reducing pain killer medication may lead to a decrease in observed exercise performance even though the overall health status improves.

The order of exercises performed in obtaining the overall score is not considered in the proposed formulation and effects of fatigue on movement performance are not included. Exercises vary in their difficulty and the variance of the healthy population's performance is considered as an estimate of exercise difficulty. Considering patients, exercise difficulty may further depend on the type of surgery. Variance in the healthy population depends further on fitness level and familiarity with an exercise.

The proposed approaches can be used both to provide information about how well a patient performs a specific task and repetitions of that task, and also to identify what is different between the ideal motion and the patient's motion. However, since the proposed approaches calculate the performance measures based on a set of features, the information about the contribution of each feature and the reasons for the observed difference between the patient's performance and the healthy performance is not captured. To determine the cause of the difference in performance between the patient and the healthy population, either the features need to be further investigated or the hypothesized causes of the difference should be explicitly modeled, e.g. for fatigue [40].

VII. Conclusion and Future Work

Quantified and continuous measure of performance can be beneficial for monitoring patient progress during the course of physiotherapy rehabilitation. This work introduces two approaches, feature-based and HMM-based, for capturing the continuous change in patient data. A distance measure is introduced as a measure of performance for a *repetition timeseries* and *repetition set*. The overall score is then calculated for the *exercise set* in each session and captures the overall performance of the patient. The proposed approaches are evaluated on data of exercises commonly performed after hip or knee replacement surgery. The results show that the proposed approach is able to track patient progress over the course of treatment. Future work will consider evaluating with an age matched healthy population or physiotherapist demonstrations and evaluating on a larger set of exercises. The proposed approach could also be enhanced by considering the order of the exercises in the formulation, subject independent measures of exercise difficulty, and fatigue and pain. Future directions may also investigate the smallest clinical important difference of the proposed score and whether physiotherapists using the new distance metrics gain additional clinically relevant information not available through visual observation alone.

Acknowledgment

The authors would like to thank the patients and the physiotherapists of the Toronto Rehabilitation Institute.

References

[1] J. A. Howe, E. L. Inness, A. Venturini, J. I. Williams, and M. C. Verrier, "The community balance and mobility scale-a balance measure for individuals with traumatic brain injury," *Clin. Rehabil.*, vol. 20, no. 10, pp. 885–895, Oct. 2006.

[2] L. Yardley *et al.*, "Development and initial validation of the falls efficacy scale-international," *Age Ageing*, vol. 34, no. 6, pp. 614–619, Nov. 2005.

[3] C. C. Norkin and D. J. White, *Measurement of Joint Motion: A Guide to Goniometry*, 4th ed. Philadelphia, PA, USA: Davis, 2009.

[4] M. F. Levin, J. A. Kleim, and S. L. Wolf, "What do motor 'recovery' and 'compensation' mean in patients following stroke?" *Neurorehabil. Neural Repair*, vol. 23, no. 4, pp. 313–319, May 2009.

[5] J. F. Lin and D. Kulić, "Human pose recovery using wireless inertial measurement units," *Physiol. Meas.*, vol. 33, no. 12, p. 2099, Dec. 2012.

[6] J. Lin and D. Kulić, "On-line segmentation of human motion for automated rehabilitation exercise analysis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 1, pp. 168–180, Jan. 2014.

[7] D. Kulić, D. Lee, C. Ott, and Y. Nakamura, "Incremental learning of full body motion primitives for humanoid robots," in *Proc. Humanoids*, 2008, pp. 326–332.

[8] H.-K. Lee and J. Kim, "An hmm-based threshold model approach for gesture recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 10, pp. 961–973, Oct. 1999.

[9] J. Alon, S. Sclaroff, G. Kollios, and V. Pavlovic, "Discovering clusters in motion time-series data," in *Proc. CVPR*, 2003, pp. I-375–I-381.

[10] P. E. Taylor, G. J. Almeida, T. Kanade, and J. K. Hodgins, "Classifying human motion quality for knee osteoarthritis using accelerometers," in *Proc. EMBC*, 2010, pp. 339–343.

[11] Z. Zhang, Q. Fang, L. Wang, and P. Barrett, "Template matching based motion classification for unsupervised post-stroke rehabilitation," in *Proc. ISBB*, 2011, pp. 199–202.

[12] J. F.-S. Lin and D. Kulić, "Human pose recovery for rehabilitation using ambulatory sensors," in *Proc. EMBC*, 2013, pp. 4799–4802.

[13] J. Aggarwal and M. S. Ryoo, "Human activity analysis: A review," *ACM Comput. Surv.*, vol. 43, no. 3, p. 16, Apr. 2011.

[14] P. J. Phillips, S. Sarkar, I. Robledo, P. Grother, and K. Bowyer, "The gait identification challenge problem: Data sets and baseline algorithm," in *Proc. ICPR*, 2002, pp. 385–388.

[15] D. Glowinski *et al.*, "Toward a minimal representation of affective gestures," *IEEE Trans. Affective Comput.*, vol. 2, no. 2, pp. 106–118, Apr./May 2011.

[16] A. Nakazawa, S. Nakaoka, K. Ikeuchi, and K. Yokoi, "Imitating human dance motions through motion structure analysis," in *Proc. IROS*, 2002, pp. 2539–2544.

[17] A. Jaimes and N. Sebe, "Multimodal human-computer interaction: A survey," *Comput. Vis. Image Understanding*, vol. 108, no. 1/2, pp. 116–134, Nov. 2007.

[18] D. Kulić, G. Venture, and Y. Nakamura, "Detecting changes in motion characteristics during sports training," in *Proc. EMBC*, 2009, pp. 4011–4014.

[19] L. Van Gestel *et al.*, "Probabilistic gait classification in children with cerebral palsy: A bayesian approach," *Res. Develop. Disabil.*, vol. 32, no. 6, pp. 2542–2552, Nov./Dec. 2011.

[20] R. Ali, L. Atallah, B. Lo, and G.-Z. Yang, "Detection and analysis of transitional activity in manifold space," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 1, pp. 119–128, Jan. 2012.
[21] K. Jia and D.-Y. Yeung, "Human action recognition using local spatio-temporal discriminant embedding," in *Proc. CVPR*, 2008, pp. 1–8.
[22] B. Sun, X. Liu, J. Shen, and Q. Zhang, "Joint angle measurements based on omni-directional lower limb rehabilitation platform," in *Proc. MHS*, 2012, pp. 337–341.
[23] B. Toro, C. J. Nester, and P. C. Farren, "Cluster analysis for the extraction of sagittal gait patterns in children with cerebral palsy," *Gait Posture*, vol. 25, no. 2, pp. 157–165, Feb. 2007.
[24] J.-Y. Jung, J. I. Glasgow, and S. H. Scott, "Feature selection and classification for assessment of chronic stroke impairment," in *Proc. BIBE*, 2008, pp. 1–5.
[25] B. Zhang, T. Kanno, W. Chen, G. Wu, and D. Wei, "Walking stability by age a feature analysis based on a fourteen-linkage model," in *Proc. CIT*, 2007, pp. 145–150.
[26] A. Webber, N. Virji-Babul, R. Edwards, and M. Lesperance, "Stiffness and postural stability in adults with down syndrome," *Exp. Brain Res.*, vol. 155, no. 4, pp. 450–458, Apr. 2004.
[27] T. Dao and M. H. B. Tho, "Knowledge-based system for orthopedic pediatric disorders," in *Proc. IFMBE*, pp. 125–128, Springer.
[28] A. Świtonński *et al.*, "The effectiveness of applied treatment in parkinson disease based on feature selection of motion activities," *Przegląd Elektrotechniczny*, vol. 88, pp. 103–106, 2012.
[29] B. Rohrer *et al.*, "Movement smoothness changes during stroke recovery," *J. Neurosci.*, vol. 22, no. 18, pp. 8297–8304, Sep. 2002.
[30] H. I. Krebs, M. L. Aisen, B. T. Volpe, and N. Hogan, "Quantization of continuous arm movements in humans with brain injury," *Proc. Nat. Acad. Sci.*, vol. 96, no. 8, pp. 4645–4649, Apr. 1999.
[31] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc., Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.
[32] P. A. W. Cox and D. R. Lewis, *The Statistical Analysis of Series of Events*. London, U.K.: Methuen, 1966.
[33] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.
[34] A. Burns *et al.*, "Shimmer-a wireless sensor platform for noninvasive biomedical research," *IEEE Sens. J.*, vol. 10, no. 9, pp. 1527–1534, Sep. 2010.
[35] C. Bregler, "Learning and recognizing human dynamics in video sequences," in *Proc. CVPR*, Jun. 1997, pp. 568–574.
[36] N. Kern, B. Schiele, and A. Schmidt, "Multi-sensor activity context detection for wearable computing," in *Ambient Intelligence*, vol. 2875, E. Aarts, R. Collier, E. Loenen, and B. Ruyter, Eds. Berlin, Germany: Springer-Verlag, 2003, ser. Lecture Notes in Computer Science, pp. 220–232.
[37] C. Schuldt, I. Laptev, and B. Caputo, "Recognizing human actions: A local svm approach," in *Proc. ICPR*, Aug. 2004, pp. 32–36.
[38] C. Cortes and V. Vapnik, "Support vector machine," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[39] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Process. Lett.*, vol. 9, no. 3, pp. 293–300, Jun. 1999.
[40] M. Karg, G. Venture, J. Hoey, and D. Kulić, "Human movement analysis as a measure for fatigue: A hidden Markov-based approach," *IEEE Trans. Neur. Syst. Rehabil. Eng.*, vol. 22, no. 3, pp. 470–481, May 2014.

Roshanak Houmanfar received the B.A.Sc. degree from Amirkabir University of Technology (Polytechnic of Tehran), Tehran, Iran, in 2011. She is currently pursuing the M.A.Sc. degree in electrical and computer engineering at the University of Waterloo, Waterloo, ON, Canada, developing automated algorithms to estimate patient progress through rehabilitation.

Her research interests include machine learning, and human motion analysis.

Michelle Karg received the B.A.Sc, Dipl.-Ing. (*summa cum laude*), and Dr.-Ing. (*summa cum laude*) degrees in electrical and computer engineering at Technical University of Munich, München, Germany, in 2005, 2006, and 2012, respectively.

Her studies included research stays at LTH Lund, Lund, Sweden, and University of North Texas, Denton, TX, USA. Since 2012, she is a Post-Doctoral Fellow at the Electrical and Computer Engineering Department at the University of Waterloo, Waterloo, ON, Canada. Her research interests include machine learning, human movement analysis, and affective computing.

Dana Kulić received the combined B.A.Sc. and M.Eng. degrees in electro-mechanical engineering, and the Ph.D. degree in mechanical engineering from the University of British Columbia, Vancouver, BC, Canada, in 1998 and 2005, respectively. From 2002 to 2006, she was a Ph.D. degree student and a post-doctoral researcher at the CARIS Lab at the University of British Columbia, developing human-robot interaction strategies to quantify and maximize safety during interaction.

From 2006 to 2009, she was a JSPS Post-doctoral Fellow and a Project Assistant Professor at the Nakamura Laboratory at the University of Tokyo, Tokyo, Japan, working on algorithms for incremental learning of human motion patterns for humanoid robots. She is currently an Assistant Professor at the Electrical and Computer Engineering Department at the University of Waterloo, Waterloo, ON, Canada. Her research interests include human motion analysis, robot learning, humanoid robots, and human-machine interaction.