

Classification-based Segmentation for Rehabilitation Exercise Monitoring

Jonathan Feng-Shun Lin , Vladimir Joukov and Dana Kulić

Abstract

Introduction: Exercise segmentation, the process of isolating individual repetitions from continuous time series measurement of human motion, is key to providing online feedback to patients during rehabilitation and enables the computation of useful metrics such as joint velocity and range of motion that are otherwise difficult to measure in the clinical setting.

Methods: This paper proposes a classifier-based approach, where the motion segmentation problem is formulated as a two-class classification problem, classifying between segment and non-segment points. The proposed approach does not require domain knowledge of the exercises and generalizes to groups of participants and exercises that were not part of the training set, allowing for more robustness in clinical applications.

Results: Using only data from healthy participants for training, the proposed algorithm achieves an average segmentation accuracy of 92% on a 30-participant healthy dataset and 87% on a 44-patient rehabilitation dataset.

Conclusion: A real-time approach for segmentation of rehabilitation exercises is proposed, based on two-class classification approach. The method is validated on both healthy and rehabilitation motion datasets and generalizes to a variety of demographics and exercises not part of the training set.

Keywords

Motion segmentation, physiotherapy, machine learning

Date received: 26 October 2016; accepted: 4 January 2018

Introduction

Temporal segmentation, the process of dividing a continuous sequence of motion data into individual repetitions or *motion primitives*, allows an automated rehabilitation system to provide real-time feedback on the performance of each exercise repetition. The segmented data can then be used by clinicians to perform diagnosis and progress monitoring with automatically collected quantitative data.¹

To meet the requirements of the rehabilitation clinic, a suitable segmentation algorithm must be able to handle intra- and inter-participant variability. *A priori* movement data for a new patient are not likely to be available and thus any supervised algorithm must rely on healthy or other patient data for training. In comparison to healthy participants, patients may exhibit smaller range of motion, inconsistent velocity and *compensatory movements*, where, as a result of redundancy within the neuromusculoskeletal system, patients may utilize alternate muscles and joints to perform

a prescribed motion they find difficult² (Figure 1). However, few of the existing segmentation algorithms have been tested on patient data.³

A segmentation algorithm should be able to handle both *known exercises* (motions in the training set) as well as generalize to *novel exercises*, since exercise regimens are often customized for each patient.

Current solutions to the motion segmentation problem³ can be divided into two classes: (1) approaches that model the exercises using a template library and (2) approaches that model the segment edge

Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Canada

Corresponding author:

Jonathan Feng-Shun Lin, Department of Electrical and Computer Engineering, University of Waterloo, 200 University Ave, Waterloo, N2L 3G1, Canada.
Email: jf2lin@uwaterloo.ca



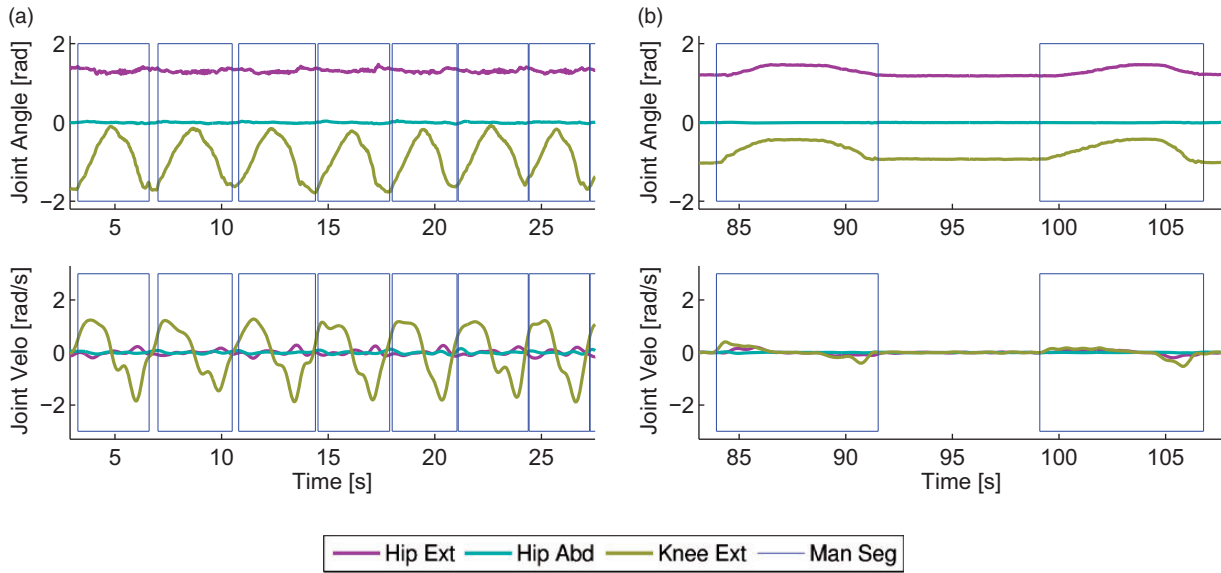


Figure 1. Joint angle and velocity data of healthy (left) and rehabilitation (right) participants performing a knee extension exercise, illustrating hip extension (hip ext), hip abduction (hip abd), knee extension (knee ext) and the manual segments (man seg). These figures, with both vertical and horizontal axes at equal scaling, illustrate motions performed by a rehabilitation patient that are slower, have smaller velocities, and smaller range of motion. Some evidence of compensatory movement can also be seen here, where the patient motion in the knee is accompanied by hip extension. These variations mean techniques that are not validated with patient data may not necessarily generalize. Note that the manual segment definitions also differ slightly. In the healthy data, the manual segments (blue boxes) tend to lag velocity crossings due to human labelling reaction speed, while for the rehabilitation data, the manual segments tend to be on the velocity crossing due to the slower pace of motion. (a) Healthy participant data. (b) Rehabilitation patient data.

points by assuming segment boundaries share common characteristics.

Among template-based methods, hidden Markov models (HMMs)⁴⁻⁶ are commonly used for motion segmentation, where each segment is considered to be a hidden state and the state sequence is recovered using the Viterbi algorithm. Regression-based techniques, where a piecewise linear fit is applied to the observation data and a segment declared when the error between the regression line and the data exceeds a threshold,^{7,8} have also been used. A third category of approaches uses dynamic time warping (DTW), where temporal and spatial variations between the template and the observed data are minimized by selectively altering the time scale to minimize the error, and segment on threshold.^{9,10} Lastly, classifier-based approaches have been applied, by extracting sequences of discretized data as segments.¹¹

Approaches that use template libraries can provide accurate segments. Some methods account for temporal and spatial variability by either modelling them stochastically or manipulate the time scale so that two sets of data can be compared, allowing these methods to handle human motion variability. However, these methods tend to be offline^{4,9,10} and cannot account for novel primitive types.

Segment edge point modelling methods learn or assume that data at segment boundaries will have certain characteristics, which are utilized to perform segmentation. Kinematic zero crossings (KZC) methods declare segments on velocity,^{12,13} acceleration,¹⁴ jerk¹⁵ or curvature¹⁶ zero crossings as they denote turning points in the motion. These methods serve as a simple way to perform segmentation. Derived metrics, such as Euclidean^{17,18} and Mahalanobis distance,¹⁹ signal variance,^{20,21} and likelihood tests¹⁶ can also be used, where segments are declared when the metric distance between two subsequent data windows exceeds a given threshold.

Segment edge models tend to be defined based on heuristics, and rely on specific domain knowledge of the underlying data. There is often no mechanism to reject false positives, necessitating further post-processing. These algorithms can be employed online but may require offline training.¹²

Proposed approach

This paper proposes a classifier approach for segmentation, where machine learning techniques are used to automatically extract segment edge characteristics. Normalization techniques are applied to improve

similarity and generalizability between healthy training data and patient testing data and improve the segmentation accuracy.¹

All data points are classified as either a *segment point* (p_1) or a *non-segment point* (p_0). This converts the task of temporal segmentation into a data classification task. If segment edge points share common features which are equivalent between different exercises, such as changes in velocity directions or changes in contact conditions, classifying all data points into either a p_1 or p_0 point allows the algorithm to handle motions and participants that have not been observed before. Rather than specifying features similar to the KZC-based methods,²² the characteristics of a segment point are learned from the training data, the training data are annotated as either p_1 and p_0 points (Figure 2).

The proposed algorithm is illustrated in Figure 3. After pre-processing, principal component analysis (PCA)²³ is used to reduce the dimensionality and the computational cost. The dimensionally reduced data are used to train a support vector machine (SVM). The trained classifier is then used to classify the observation data.

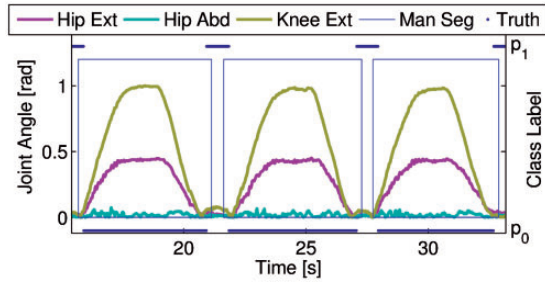


Figure 2. A time-series waveform of joint angles during a squatting exercise. This figure denotes the hip extension (hip ext), hip abduction (hip abd), knee extension (knee ext), the manual segments (man seg) and p_1/p_0 ground truth data generated from the manual segments (truth). The segments provided by a human observer (blue boxes) are expanded into multiple segment edge points (top blue lines) and non-segment points (bottom blue lines).

Pre-processing

Normalizing. To reduce sensitivity to the starting pose, the initial value of each time-series can be removed such that the resting pose is always at zero, by applying an offset calculated from the mean value of the first 10 samples of each trajectory.

The data can also be rectified. Taking the absolute value of the joint angle data would allow motions that involve the same joints but moving in opposite directions to be recognized as one exercise, improving generalization.

Magnitude scaling, by dividing by the mean of the absolute peak values over all the segments in the time-series data, compensates for the varying joint angle ranges between healthy and patient data, as well as enables exercise generalization. However, the determination of this value requires the full observation set, which is impractical in clinical applications because the full observation sequence is not available *a priori*. For training, the scaling factor is calculated from the full dataset, while for testing, the scaling factor is calculated from only the first segment.

Temporal. Data points are interpolated to ensure even temporal spacing between time points; then the data points that are n_{stack} around the time point being examined (t_n) are concatenated into the current data point to incorporate temporal information before and after t_n into the classifier.

Balancing. Imbalanced datasets can skew the classifier towards the class that has the most samples. To minimize this problem, data points n_{exp} around manual segment points, as well as the data points between segments, are converted into p_1 points, as they are similar to the manual segment point. Downsampling is also applied to p_0 to further balance the dataset. A Gaussian sampling method was employed, where p_0 points closer to p_1 points have a higher chance of being sampled than p_0 points that are not close to any p_1 points. This is to include more data points that are close to the segmentation boundaries and thus are

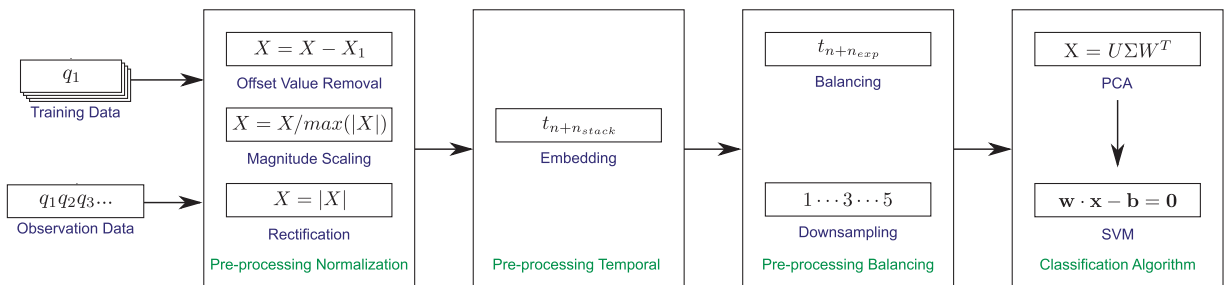


Figure 3. Flow diagram of the algorithm.

more likely to be misclassified (Figure 4). During testing, no downsampling is applied.

Feature extraction

After the pre-processing steps, a feature extraction technique can be applied, to reduce the feature space dimensionality and decrease computational time. To achieve this, PCA is used. PCA is an orthogonal transformation algorithm that finds a projection to a new feature space, consisting of principal components (PCs) which are constructed to capture the maximum amount of variance and constrained such that all subsequent PCs are orthogonal to the previous ones. PCA is commonly used in conjunction with the scree plot method,²³ which is a plot that denotes the fraction of total variance in the data represented by each PC. By selecting a threshold in the amount of variance to be captured by the resulting projection and using only the PCs that explain the required level of variance, dimensional reduction is achieved.

Discriminative classifier

The data can now be used for training and classification. For the purposes of this paper, SVM²⁴ is used. SVM is a two-class classification algorithm where the training data are mapped to a higher dimensional kernel space, and a decision hyperplane is generated between the classes. SVM forms its decision hyperplane by selecting key training points, or *support vectors*, that form the hyperplane, such that the margins between the two classes are maximized. The SVM parameters are the slack variable C and the selection of the kernel. For this paper, the radial kernel is used.

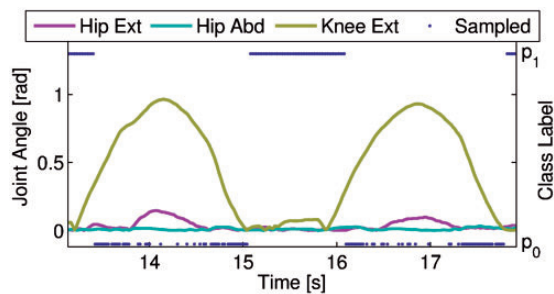


Figure 4. Knee extension training data, denoting the hip extension (hip ext), hip abduction (hip abd), knee extension (knee ext) and the sampled data for training (sampled). The blue points at $y=0$ denote the training data selected for p_0 , sampled by Gaussian resampling, favouring non-segment points closer to the boundary. The blue points at $y=1$ denote the training data selected for p_1 , which has no downsampling. The Gaussian covariance coefficient is reduced to show this effect more clearly.

Experimental validation

While many algorithms were designed with clinical applications in mind, few to date have been tested against patient populations or report specific segmentation results.³ Three different datasets are used to validate the proposed approach: One collected from healthy participants and two from rehabilitation participants undergoing knee or hip total-joint replacement surgery. Two types of validations were performed: (1) cross-validation, to validate inter-participant generalization and (2) validation with known and novel exercises, to test for inter-exercise generalization.

Datasets

All three datasets were collected using three Shimmer accelerometer and gyroscope sensors²⁵ worn on the hip, knee and ankle. Using the measured accelerometer and gyroscope data, joint angles were estimated via the extended Kalman filter,²⁶ using a five-joint kinematic model, where the hip is modelled as a three-joint system with flexion/extension, abduction and internal rotation, and the knee is modelled as a two-joint system with flexion/extension and internal rotation. Poorly reconstructed data were removed from consideration.

The first dataset is a 30-healthy participant dataset (17 M, 13 F, $\mu_{age}=24.4$)¹² collected from the University of Waterloo (UW) student population, with 10 to 20 repetitions per exercise type. All participants were healthy and had no lower-body injuries for the six months prior to the data collection. Prior to data collection, participants were instructed on how to perform each motion. Motion capture data were collected simultaneously. The supine exercises collected were hip abduction, hip flexion, knee extension and knee/hip flexion. The seated exercises collected were knee extension and sit to stand. The standing exercises collected were gait, hip abduction, hip extension, marching, lunges and squats.

The second dataset is an 18-patient dataset (4 M, 14 F, $\mu_{age}=73.8$)¹² collected from the Toronto Rehabilitation Institute (TRI) musculoskeletal inpatient ward. All patients had undergone a hip or knee replacement surgery and were in TRI for post-surgical rehabilitation. These patients were tracked from the first day of admission until discharge, and had daily rehabilitation sessions, with an average stay length of two weeks. All the exercises collected were part of their prescribed routine; no exercises were modified or added for the data collection. The supine exercises collected were hip abduction, hip flexion, knee extension and knee/hip flexion. The seated exercises collected were knee extension and sit to stand.

The standing exercises collected were gait, hip abduction, hip extension, knee flexion, marching, lunges and squats.

The third dataset is a 26-patient dataset (10 M, 16 F, $\mu_{age} = 66.7$)¹ collected from the St. Joseph's Health Centre Guelph (SJHC) rehabilitation out-patient ward. All patients had undergone hip or knee replacement surgery and were in SJHC for rehabilitation. Patients were in SJHC for a six-week rehabilitation program. Patients were approached in their third to sixth week, and data were collected for a single exercise session for each patient. All the exercises collected were part of their prescribed routine, but each patient performed half of the exercises while using an exercise feedback monitor.¹ The supine exercises collected were knee extension and knee/hip flexion. The seated exercises collected were stationary bicycling, knee extension and sit to stand. The standing exercises collected were gait, hip abduction, hip extension, knee flexion, marching, lunges and squats.

The healthy data were collected with the approval of the University of Waterloo Research Ethics Board (REB). The TRI dataset was approved by both the UW REB and the University Health Network REB. The SJHC dataset was approved by both the UW REB and the SJHC REB. Signed consent was obtained from all participants.

Many of the exercises performed by the SJHC and TRI patients were modified in some way: springs, slings, weights and other physical devices were added to modify the exercise difficulty, providing test cases for

of inter-exercise generalization. The main difference between the two patient datasets was the health status of the patients. In-patients were being treated in hospital due to additional health concerns or co-morbidities that prevented discharge, while out-patients were healthy enough to recover at home.

For the UW dataset, manual segment points were generated by using the motion capture data. The motion capture marker positions were used to produce a video of the Cartesian data and the segments were denoted based on the observer's perception of when a repetition had begun or ended. For the TRI and SJHC dataset, manual segmentations were annotated on the recovered joint angle data, as motion capture data were not available. The accelerometer and joint angle data were plotted as a time-series plot and the segments were denoted based on the observer's perception.

Implementation

Only the healthy data were used for training, while both healthy and rehabilitation data were used for testing. The use of healthy data only for training is preferred, to avoid the need for manual labelling of clinical data. In all experiments, no training data overlap with the testing data. The exercises used for training were selected by totalling the amount of rehabilitation data available and selecting the top five exercises that had exercise examples from all three datasets. A list of these exercises is labelled as known exercises in Table 1.

Table 1. Known (exercises used in algorithm training) and novel (exercises not used in algorithm training) exercises with their corresponding active joints, direction of movement, typical initial postures and the total amount of data collected.

Names	Characteristics			Data collected (s)			
	Short	Joints	Direction	Posture	UW	SJHC	TRI
Long ²⁷							
Known							
Standing hip abduction	HAAO-STD	Hip	Abd	Std	832	976	3535
Standing hip extension	HEFO-STD	Hip	Ext	Std	799	1945	7734
Seated knee extension	KEFO-SIT	Knee	Ext	Sit	1207	1327	8119
Marching	KHEF-STD	Knee/hip	Ext/Flex	Std	836	1934	6892
Supine heel slides	KHEF-SUP	Knee/hip	Ext/Flex	Sup	952	515	21277
					4625	6697	47558
Novel							
Standing hip flexion	HFEO-STD	Hip	Flex	Std	799	—	3817
Quad over roll	KEFO-SUP	Knee	Ext	Sup	—	238	10954
Standing hamstring curls	KFEO-STD	Knee	Flex	Std	—	2921	7118
Resisted knee flexion over roll	KFEO-SUP	Knee	Flex	Sup	733	—	6102
Squats	SQUA-STD	Knee/hip	Ext/Flex	Std	1370	—	6002
					2902	3159	33993

The directions of motion include extension (ext), flexion (flex) and abduction (abd). The initial postures include supine (sup), sitting (sit) and standing (std).

All data were resampled to 50 Hz. Both joint angle and joint velocity data were used as classifier input. All processing and algorithm implementation were done in MATLAB 8.0, along with the LIBSVM toolbox²⁸ and the Toolbox for Dimensionality Reduction.²⁹ n_{stack} was set to 15, while n_{exp} was set to 10, based on previous testing.³⁰ For PCA, a scree plot threshold of 80% was used. For SVM, two-class soft margin radial SVM was used. C was set to 1. For the radial kernel function, γ was set to the inverse of the number of degrees of freedom (DOFs) post-PCA. The same settings as reported in Lin et al.³⁰ are used, where a detailed analysis of parameter sensitivity is provided.

A four-fold cross validation was utilized to test inter-participant generalization. For training, 10 healthy participants from UW were randomly selected, another set of 10 UW participants were used for testing, as well as all of the TRI and SJHC patient data. All reported accuracy values were averaged over the four folds.

Validation metric

The segmentation accuracy was assessed by comparing the algorithmic p_1 and p_0 labels against the manually generated ground truth. The total number of correctly identified points, the p_1 true positives (TP) and p_0 true negatives (TN), as well as the false positives (FP) and false negatives (FN) are reported together as the balanced accuracy Acc_{Bal}

$$Acc_{Bal} = \frac{1}{2} \cdot \frac{TP}{TP + FN} + \frac{1}{2} \cdot \frac{TN}{TN + FP}$$

This measure combines both the sensitivity and the specificity scores and limits inflated accuracy scores in imbalanced dataset cases.

Results

While the proposed method is designed to automatically extract segment edge features, machine learning approaches can be sensitive to input feature scaling. The results show that the normalization allows the healthy and the rehabilitation data to appear more similar, and the temporal stacking and PCA selected a suitable linear combination of features that provide good separability for high segmentation accuracy.

Known exercises

First, this paper considers known exercises, where the same exercises are in the training and test sets to evaluate inter-participant generalization (Table 2). The same classifiers, trained only on healthy data, were used to test both the healthy and rehabilitation datasets, using different pre-processing methods.

For healthy participants, removing the offset significantly improves the accuracy, moving from $74\% \pm 26\%$ to $91\% \pm 6\%$, an increase of 15% in accuracy and a drop of factor of 4 in the standard deviation. The offset removal allows motions that start at different postures to appear more similar, thus improving accuracy. Normalizing without using the scaling factor provided the best performance.

The preprocessing techniques also resulted in a similar outcome for the SJHC dataset, where the offset normalization significantly improves the segmentation accuracy. Applying the rectification alone led to performance degradation, but when combined with the

Table 2. Classifier performance for known and novel exercises.

Pre-processing			UW		STJC		TRI	
Off	Sca	Abs	Known	Novel	Known	Novel	Known	Novel
N	N	N	0.744 ± 0.260	0.691 ± 0.164	0.760 ± 0.164	0.836 ± 0.118	0.607 ± 0.223	0.512 ± 0.213
N	N	Y	0.751 ± 0.260	0.638 ± 0.160	0.719 ± 0.170	0.695 ± 0.121	0.580 ± 0.208	0.456 ± 0.159
N	Y	N	0.774 ± 0.255	0.697 ± 0.149	0.781 ± 0.168	0.709 ± 0.089	0.621 ± 0.210	0.569 ± 0.109
N	Y	Y	0.776 ± 0.255	0.669 ± 0.139	0.784 ± 0.161	0.692 ± 0.085	0.614 ± 0.198	0.545 ± 0.083
Y	N	N	0.909 ± 0.061	0.838 ± 0.115	0.856 ± 0.124	0.879 ± 0.064	0.647 ± 0.163	0.598 ± 0.153
Y	N	Y	0.916 ± 0.054	0.858 ± 0.101	0.830 ± 0.133	0.889 ± 0.060	0.670 ± 0.161	0.583 ± 0.136
Y	Y	N	0.889 ± 0.103	0.869 ± 0.082	0.870 ± 0.122	0.864 ± 0.068	0.813 ± 0.147	0.830 ± 0.152
Y	Y	Y	0.906 ± 0.072	0.870 ± 0.080	0.886 ± 0.107	0.884 ± 0.050	0.853 ± 0.118	0.828 ± 0.144

The reported results are the mean and standard deviation of the four-fold cross-validation in the training data. Three types of normalizations were examined: offset value removal (off), magnitude scaling (sca) and rectification of the joint angle (abs). Bolded value denote maximum in each section. The results are separated into known and novel exercises, as denoted by Table 1.

other normalization methods, up to 17% improvement can be observed.

The preprocessing techniques have the strongest impact on the TRI dataset. Rectifying the data improved performance, but not as significantly as magnitude scaling the data. Magnitude scaling allows the patient datasets to look more similar to the training set, allowing for improvements of over 17% in accuracy for some configurations. Combining all three preprocessing steps improves performance on the TRI dataset to 85%. Offset removal alone improves accuracy from 60% to 65%.

Novel exercises

This paper next considers novel exercises, where the test exercises were not included in the training set to test inter-exercise generalization (Table 2). This test is important for clinical application because if the algorithm can successfully generalize to exercises that were not explicitly included in the training dataset, it reduces the amount of training data that must be provided. Table 2 shows that, without pre-processing, the novel exercises do not perform as well as the known ones.

For healthy participants, the offset value removal is the most influential factor, improving the accuracy by 13%. However, scaling and rectification alone did not seem to impact the accuracy significantly. However, with all three normalizations active, the accuracy is improved by 18%.

For patient data, without normalization, the segmentation performance declines significantly, but with offset value removal, scaling and rectification, the TRI novel exercise approaches performance with the known exercises, at 85% to 83% for the known and novel datasets, respectively. The most influential normalization is the offset and the scaling, which allowed the TRI data to appear more similar to the healthy training data. Similar

results can be observed for the SJHC novel exercises, where the reported accuracy was 89% and 88% for the known and novel motions, respectively.

PCA mapping

The PCA mapping of the data shows that the top PCs are composed of the knee sagittal joint angle, the knee sagittal joint velocity, the hip sagittal joint angle, the hip sagittal joint velocity, the hip abduction joint angle and finally the hip abduction joint velocity, which are the main features that change over time over all the datasets. These PCs, under different configurations, can be seen in Figure 5.

For $n_{stack} = 0$, the first PC contains the knee sagittal and hip sagittal joint velocity. The second PC contains the knee sagittal and hip sagittal joint angles. These two PCs effectively form a phase plot of these two different joints, emphasizing the cyclic characteristics of the exercise movements.

For $n_{stack} = 15$, PCA generates a similar looking projection, even though the input feature has been embedded with temporally offset data. This indicates that PCA is constructing linear combinations of the different joint angles and velocities, generating a space analogous to the phase plot, but automatically incorporating multiple joints and short-term temporal information.

Table 3 shows the classification accuracy of different n_{stack} values. Smaller n_{stack} values, corresponding to less temporal information embedded at each timestep, led to a decrease in accuracy. The accuracy difference between $n_{stack} = 15$ used in the Known and Novel Exercises Results sections and larger n_{stack} values does not vary significantly, due to the wider windows covering components of the trajectory that are both moving and not moving. These suggest that for the dataset examined, the window of 0.6s provided by $n_{stack} = 15$ is a suitable choice.

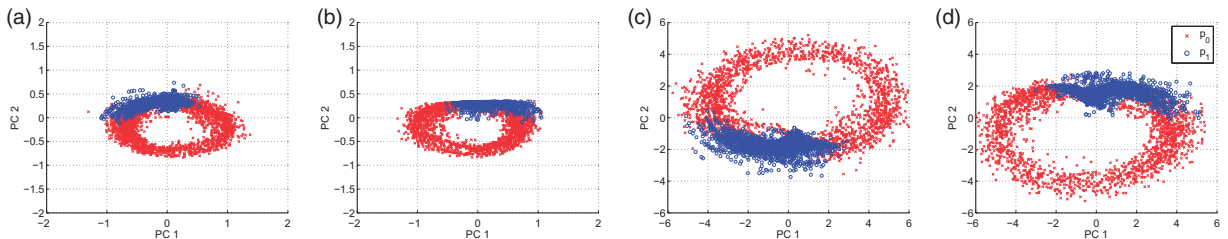


Figure 5. Four plots of the training data, projected onto a 2D plane via PCA using the projection matrix generated by the offset value removal and scaled training data. These plots show the different PCs at different n_{stack} values and rectification state. Note that the scale between the left two figures and the right two figures are different. In all configurations, the first four PCs heavily weigh the knee sagittal and hip sagittal joint angles and velocities, due to their high variance in the trajectories examined. Although not shown here, the subsequent PCs weigh the hip abduction joint angle and velocity more heavily. (a) Unrectified data, $n_{stack} = 0$. (b) Rectified data, $n_{stack} = 0$. (c) Unrectified data, $n_{stack} = 15$. (d) Rectified data, $n_{stack} = 15$.

Table 3. Classification accuracy of testing data using classifiers that have been normalized by using offset value removal, magnitude scaling and rectification.

n_{stack}	UW		SJHC		TRI	
	Known	Novel	Known	Novel	Known	Novel
0	0.862 ± 0.077	0.860 ± 0.083	0.865 ± 0.107	0.858 ± 0.048	0.830 ± 0.130	0.817 ± 0.156
5	0.890 ± 0.066	0.869 ± 0.079	0.870 ± 0.110	0.865 ± 0.046	0.844 ± 0.120	0.823 ± 0.151
10	0.902 ± 0.066	0.870 ± 0.078	0.877 ± 0.111	0.871 ± 0.048	0.849 ± 0.120	0.827 ± 0.144
15	0.907 ± 0.071	0.870 ± 0.078	0.887 ± 0.105	0.884 ± 0.050	0.853 ± 0.118	0.828 ± 0.144
30	0.895 ± 0.108	0.875 ± 0.082	0.861 ± 0.127	0.863 ± 0.085	0.848 ± 0.124	0.833 ± 0.144

The baseline is bolded and denoted as $n_{stack} = 15$. Different n_{stack} parameters are examined. The results are separated into known and novel exercises, as denoted by Table 1.

Discussion

The contributions of this paper are three-fold: (1) the segmentation problem is recast as a two-class classification problem, allowing standard machine learning methods to learn the optimal classification boundary without the need for hand selected features; (2) the differences between healthy and rehabilitation movement patterns and the impact of feature normalization techniques on algorithm performance and generalization are investigated and (3) the algorithm is validated with one healthy and two patient datasets, demonstrating that the algorithm can generalize to both motions and individuals that were not seen during training and that normalization can be used to allow the algorithm to be trained using only healthy data, which can be more easily collected and labelled. The proposed method can be applied to the clinical environment and perform online motion segmentation.

Comparison to existing work

The velocity feature HMM (vfHMM)¹² was implemented as a comparison work against the proposed algorithm. It uses velocity crossing and peaks to identify potential segments, then selects them based on HMM identification. In addition, a zero-velocity crossing (ZVC) algorithm²² was also implemented, where velocity crossings indicated direction change and thus segment points. The HMM and ZVC methods were selected due to their wide-spread use for motion segmentation. DTW^{9,10} is another common method but has heavy computation requirements, making it unsuitable for online clinical deployment.

Joint angle data that had discontinuities, integration drift or noisy recovery were removed, as they caused the training algorithm¹² to diverge. For the ZVC algorithm,²² every second identified ZVC was used to create the segments. The single point segments declared by the

Table 4. Classification accuracy of the proposed algorithm, compared against two other methods, the velocity feature HMM (vfHMM) and the zero-velocity crossing (ZVC), using known exercises.

Algorithm	UW	SJHC	TRI
ZVC ²²	0.707 ± 0.113	0.731 ± 0.146	0.711 ± 0.192
vfHMM ¹²	0.710 ± 0.118	0.709 ± 0.148	0.730 ± 0.180
PCA SVM	0.906 ± 0.072	0.886 ± 0.107	0.853 ± 0.118

The proposed method is bolded.

comparison algorithms were converted into p_0 and p_1 points such that the balanced accuracy metric can be used for all algorithms. For this section, the training and testing data of all three algorithms were normalized by offset removal and magnitude normalization.

Table 4 indicates that the proposed algorithm outperforms both of the comparison algorithms. Both vfHMM and ZVC methods assume segment points only occur on velocity crossings, rather than learning the segment point locations from expert-labelled data. Figure 1 shows that human observers may define segments to lead or lag the velocity crossing. These variations in the manual segments highlight the importance of segment definition flexibility, where the proposed algorithm outperforms the two comparison works. In addition, the proposed approach can generalize to unseen exercises, while vfHMM¹² requires a template for each motion.

With suitable pre-processing, the classifier can generalize both to patient data and to novel exercises, based on only a healthy participant training dataset.

Differences between healthy and rehabilitation data

The results show that normalization plays a key role in enabling generalization to novel participants and that validation on patient populations are important. In Figure 6, intra-participant variance is shown to be

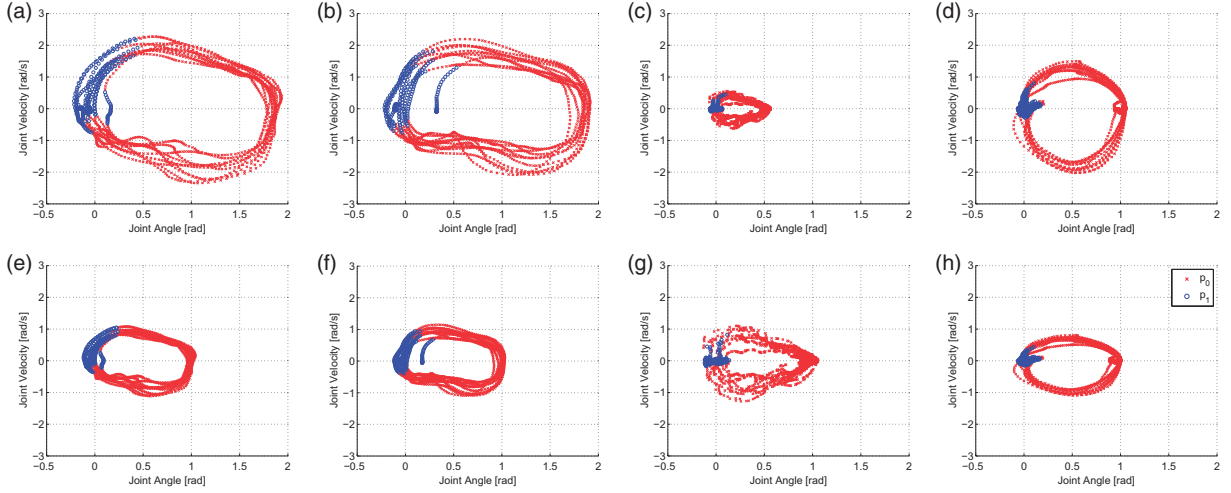


Figure 6. Eight phase plots of seated knee extension (KEFO-SIT), where the x axis denotes knee sagittal joint angle and the y axis denotes knee sagittal joint velocity. The left figures denote the same UW participant performing the exercise at different times within the same hour, illustrating intra-participant variation. The right figures show the same TRI patient performing the exercise at different times over a two-week span, illustrating significant magnitude differences as the patient improves over time. The top figures show the data without magnitude scaling, while the bottom figures show the data after scaling. (a) UW 5, session 1, no scaling. (b) UW 5, session 2, no scaling. (c) TRI 2, session 1, no scaling. (d) TRI 2, session 10, no scaling. (e) UW 5, session 1, with scaling. (f) UW 5, session 2, with scaling. (g) TRI 2, session 1, with scaling. (h) TRI 2, session 10, with scaling.

significant, even within the same healthy participant when the data are collected within a short time period (Figure 6(a) and (b)). In the datasets evaluated here, this difference is magnified between healthy and orthopaedic patient data as the exercises were performed under different conditions. For accelerometers, joint angle recovery with the Kalman filter²⁶ assumes that the sensors are rigidly attached to the limb. While this is generally not a problem for healthy participants, the sensors cannot be attached as tightly to joint replacement patients due to concerns about pain. Sensor placements are also less exact due to the need to avoid surgical sites, bandages, weights and slings, and may be displaced during movement. Many of the total joint replacement patients who participated in the experiments also suffer from complications that impact movement and stability, such as osteoarthritis in the non-operated limbs, obesity or post-surgical swelling.² Compensatory movement due to pain or soreness is another source of variability. Patient data also vary widely due to prescribed patient-specific exercise modifications such as support slings or weights. These differences emphasize the need for normalization to be robust to patient motion variability, as well as highlight the need to perform data validation against both healthy and rehabilitation data.

Reformulating segmentation as a classification problem simplifies the task and allows for machine learning techniques to be used to automatically extract segmentation boundaries without hand crafted features. The various normalization techniques

investigated allowed for the healthy template to generalize to both participants and exercises unseen during training, even though the healthy and rehabilitation participants were observed to have very different movement patterns.

Clinical considerations

In the clinic, segmentation misclassification can lead to a misrepresentation of patient health. Under-segmenting may lead the patient perform more repetitions than prescribed, as well as lead the physiotherapist to believe that the patient is less capable. Over-segmenting may lead the patient to perform fewer repetitions and lead the physiotherapist to overestimate the patient's capability.

Lam et al.¹ used vfHMM¹² as the segmentation algorithm in a clinical study where online computer-based feedback was given to patients on their movements. In Lam et al.'s study,¹ the patients' peak velocity and exercise duration improved with feedback, as well as reducing compensatory movement. This suggests that vfHMM¹² is sufficiently accurate to improve patient metrics. The F_1 accuracy for vfHMM¹² for patient knee extension data was 73%, while the F_1 for the proposed approach on the same exercise is 80%. Since the proposed method outperforms vfHMM,¹² the proposed algorithm should also be suitable for clinical usage.

While the segmentation accuracy outperforms prior work, several limitations of the algorithm should be

considered. The current method assumes that the first manual segment of all observed data is known to perform normalization, requiring physiotherapist input to specify. Although the algorithm would be more flexible if it did not require explicit physiotherapist input, it is reasonable to expect a physiotherapist to observe their patients to ensure that the first repetition of any prescribed exercise is completed properly. However, this supervision process may be cumbersome in group exercise sessions.

Another limitation is if the range of motion changes dramatically over the span of a given observation sequence, the classifier may not be able to sufficiently segment the data since the normalization factors are calculated from the first segment. That is, if the patient improves or degrades significantly from the first segment, the algorithm may not perform as well. The proposed algorithm also relies joint angles reported from body-worn sensors, which may be difficult for all patients to put on unassisted due to limited mobility.¹ Only orthopaedic rehabilitation exercises of the lower body are considered in this paper. To verify clinical feasibility, different population types and wider exercise sets should be considered.

In this paper, only kinematic data of the lower body were examined, which may limit clinical applications. However, earlier versions of this technique have been applied to complex full body kinematic data of healthy participants,³¹ as well as to joint torque³² and electromyogram³³ data, demonstrating that the technique suitable for a variety of input features.

Another limitation of any supervised learning algorithm is the need for manually labelled training data. Manual segments in this paper were obtained by video playback of motion capture marker locations or by annotation on a time-series graph, and may suffer from inter-rater variability. A possible way to alleviate this issue is to collect manual segments from several labellers.

Conclusion

In this paper, an algorithm for motion segmentation was developed and validated with both healthy and patient data. The proposed algorithm converts time-series motion segmentation into a two-class classification problem. The algorithm was tested for both inter-participant generalizability and inter-exercise generalizability and was shown to segment with 92%, 87% and 85% accuracy on healthy, out-patient rehabilitation and in-patient rehabilitation data, respectively, while using only healthy data for training.

For future work, a wider variety of exercises will be examined, including gait and functional tasks.

Acknowledgements

The authors would like to thank the University of Waterloo community, and the staff and patients at the Toronto Rehabilitation Institute and the St. Joseph Health Centre Guelph for providing their time and data. The authors thank Joanna Wong for her insights into physiotherapy practice and Agnes Lam for contributing to the data collection.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Natural Sciences and Engineering Research Council (grant number CRDPJ 433802-12).

Guarantor

JL.

Contributorship

JL, VJ and DK conceived the study. JL and VJ developed the proposed algorithm and conducted the evaluation. JL conducted a literature review, developed the experimental protocol, secured ethical approval, conducted the data analysis and wrote the manuscript. All authors reviewed and edited the manuscript and approved the final version of the manuscript.

Note

I. A preliminary version of this work was presented in conference publications,^{30,31} where the proposed approach was applied to healthy participant datasets. This paper expands upon the earlier work with additional feature processing techniques, validates the proposed approach on two rehabilitation patient datasets and provides an expanded discussion of the strengths and limitations of the proposed approach.

ORCID iD

Jonathan Feng-Shun Lin  <http://orcid.org/0000-0002-0008-7199>.

References

1. Lam AWK, Verona-Marin D, Li Y, et al. Automated rehabilitation system: movement measurement and feedback for patients and physiotherapists in the rehabilitation clinic. *Hum Comput Interact* 2016; 31: 294–334.
2. DeLisa JA, Gans BM, Walsh NE, et al. *Physical medicine and rehabilitation: principles and practice*, 4th ed. Philadelphia, PA, United States: Lippincott Williams and Wilkins, 2005.

3. Lin JFS, Karg ME and Kulić D. Movement primitive segmentation for human motion modeling: a framework for analysis. *IEEE Trans Hum Mach Syst* 2016; 46: 325–339.
4. Varadarajan B, Reiley C, Lin H, et al. Data-derived models for segmentation with application to surgical assessment and training. In: Yang GZ, Hawkes D, Rueckert D, et al. (eds) *Medical image computing and computer-assisted intervention (Lecture Notes in Computer Science)*. Berlin: Springer, 2009, pp.426–434.
5. Castellani A, Botturi D, Bicego M, et al. Hybrid HMM/SVM model for the analysis and segmentation of teleoperation tasks. In: *Proceedings of the IEEE international conference on robotics and automation*. Vol. 3, pp. 2918–2923. IEEE. DOI: 10.1109/ROBOT.2004.1307504.
6. Li M and Okamura AM. Recognition of operator motions for real-time assistance using virtual fixtures. In: *Proceedings of the symposium on haptic interfaces for virtual environment and teleoperator systems*, pp. 125–131.
7. Keogh EJ, Chu S, Hart D, et al. Segmenting time series: a survey and novel approach. In: Last M, Kandel A and Bunke H (eds) *Data mining in time series databases*. Vol 57, Singapore: World Scientific Publishing, 2004, pp.1–22.
8. Amft O, Junker H and Tröster G. Detection of eating and drinking arm gestures using inertial body-worn sensors. In: *Proceedings of the IEEE international symposium on wearable computers*, pp. 160–163. IEEE. DOI: 10.1109/ISWC.2005.17.
9. Ilg W, Bakir GH, Mezger J, et al. On the representation, learning and transfer of spatio-temporal movement characteristics. *Int J HR* 2004; 1: 613–636.
10. Zhou F, de la Torre F and Hodgins JK. Hierarchical aligned cluster analysis for temporal clustering of human motion. *IEEE Trans Pattern Anal Mach Intell* 2013; 35: 582–596.
11. Berlin E and Van Laerhoven K. Detecting leisure activities with dense motif discovery. In: *Proceedings of the ACM conference on ubiquitous computing*. ISBN 978-1-4503-1224-0, pp. 250–259. ACM. DOI: 10.1145/2370216.2370257.
12. Lin JFS and Kulić D. On-line segmentation of human motion for automated rehabilitation exercise analysis. *IEEE Trans Neural Syst Rehabil Eng* 2014; 22: 168–180.
13. Zhang T, Karg ME, Lin JFS et al. Imu based single stride identification of humans. In: *Proceedings of the IEEE international symposium on robot and human interactive communication*, pp. 220–225. IEEE.
14. Ricci L, Formica D, Tamilia E, et al. An experimental protocol for the definition of upper limb anatomical frames on children using magneto-inertial sensors. In: *Proceedings of the IEEE international conference of engineering in medicine and biology society*. IEEE.
15. Yamamoto M, Mitomi H, Fujiwara F, et al. Bayesian classification of task-oriented actions based on stochastic context-free grammar. In: *International conference on automatic face and gesture recognition*, pp. 317–322. DOI:10.1109/FGR.2006.28.
16. Bashir F, Qu W, Khokhar A et al. HMM-based motion recognition system using segmented PCA. In: *Proceedings of the IEEE international conference on image processing*, pp. 1288–1291. DOI:10.1109/ICIP.2005.1530635.
17. Mueen A, Keogh E, Zhu Q, et al. Exact discovery of time series motifs. In: *SIAM international conference on data mining*, pp. 473–484. DOI:10.1137/1.9781611972795.41.
18. Valtazanios A, Arvind DK and Ramamoorthy S. Comparative study of segmentation of periodic motion data for mobile gait analysis. In: *Wireless health*. ISBN 978-1-60558-989-3, pp. 145–154. DOI: 10.1145/1921081.1921099.
19. Barbič J, Safonova A, Pan JY, et al. Segmenting motion capture data into distinct behaviors. In: *Proceedings of Graphics Interface*, ACM, pp. 185–194.
20. Koenig N and Mataric MJ. Behaviour-based segmentation of demonstrated tasks. In: *Proceedings of the international conference on development and learning*.
21. Kohlmorgen J and Lemm S. A dynamic hmm for on-line segmentation of sequential data. In: *Proceedings of the advances in neural information processing systems*. Vol. 14, pp. 793–800.
22. Fod A, Mataric MJ and Jenkins OC. Automated derivation of primitives for movement classification. *Auton Robots* 2002; 12: 39–54.
23. Fodor IK. *A survey of dimension reduction techniques*. Technical report, Lawrence Livermore National Laboratory, 2002.
24. Burges CJC. A tutorial on support vector machines for pattern recognition. *Data Min Knowl Discov* 1998; 2: 121–167.
25. Burns A, Greene BR, McGrath MJ, et al. SHIMMER: a wireless sensor platform for noninvasive biomedical research. *IEEE Sens J* 2010; 10: 1527–1534.
26. Lin JFS and Kulić D. Human pose recovery using wireless inertial measurement units. *Physiological Measurement* 2012; 33: 2099–2115. DOI:10.1088/0967-3334/33/12/2099.
27. Kisner C and Colby LA. *Therapeutic exercise: foundations and techniques*. 5th Edition. Philadelphia, PA, United States: F. A. Davis Company, 2007.
28. Chang CC and Lin CJ. LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol* 2011; 2: 1–27.
29. Van der Maaten LJP, Postma EO and Van Den Herik HJ. Dimensionality reduction: a comparative review. *J Mach Learn Res* 2009; 10: 1–41.
30. Lin JFS, Joukov V and Kulić D. Human motion segmentation by data point classification. In: *Proceedings of the IEEE international conference of engineering in medicine and biology society*, pp. 9–13. IEEE.
31. Lin JFS, Joukov V and Kulić D. Full-body multi-primitive segmentation using classifiers. In: *Proceedings of the IEEE/RAS international conference on humanoid robots*, pp. 874–880.

32. Lin JFS, Bonnet V, Joukov V, et al. Comparison of kinematic and dynamic sensor modalities and derived features for human motion segmentation. In: *Proceedings of the IEEE/NIH Strategic Conference on Healthcare Innovations and Point of Care Technologies*. pp. 109–112. IEEE. DOI: 10.1109/HIC.2016.7797709.
33. Lin JFS, Samadani A and Kulić D. Segmentation by data point classification applied to forearm surface emg. In: *Proceedings of the EAI International Summit of Smart City 360*, volume 166, pp. 153–165. Springer. DOI:10.1007/978-3-319-33681-713.