Kinematic and Dynamical Analysis Techniques for Human Movement Analysis from Portable Sensing Devices

by

Vinay Venkataraman

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved February 2016 by the
Graduate Supervisory Committee:

Pavan Turaga, Chair
Antonia Papandreou-Suppappola
Narayanan Krishnamurthi
Baoxin Li

ARIZONA STATE UNIVERSITY

May 2016

ABSTRACT

Today's world is seeing a rapid technological advancement in various fields, having access to faster computers and better sensing devices. With such advancements, the task of recognizing human activities has been acknowledged as an important problem, with a wide range of applications such as surveillance, health monitoring and animation. Traditional approaches to dynamical modeling have included linear and nonlinear methods with their respective drawbacks. An alternative idea I propose is the use of descriptors of the shape of the dynamical attractor as a feature representation for quantification of nature of dynamics. The framework has two main advantages over traditional approaches: a) representation of the dynamical system is derived directly from the observational data, without any inherent assumptions, and b) the proposed features show stability under different time-series lengths where traditional dynamical invariants fail.

Approximately 1% of the total world population are stroke survivors, making it the most common neurological disorder. This increasing demand for rehabilitation facilities has been seen as a significant healthcare problem worldwide. The laborious and expensive process of visual monitoring by physical therapists has motivated my research to invent novel strategies to supplement therapy received in hospital in a home-setting. In this direction, I propose a general framework for tuning component-level kinematic features using therapists overall impressions of movement quality, in the context of a Home-based Adaptive Mixed Reality Rehabilitation (HAMRR) system.

The rapid technological advancements in computing and sensing has resulted in large amounts of data which requires powerful tools to analyze. In the recent past, topological data analysis methods have been investigated in various communities, and the work by Carlsson establishes that persistent homology can be used as a powerful

topological data analysis approach for effectively analyzing large datasets. I have explored suitable topological data analysis methods and propose a framework for human activity analysis utilizing the same for applications such as action recognition.

To My Family

ACKNOWLEDGMENTS

There are many individuals who have contributed towards this work. First, I would like to thank Dr. Pavan Turaga for showing confidence in me and giving me an opportunity to work in his lab. His unconditional support has always kept me going during tough times. I would like to thank Dr. Antonia Papandreou-Suppappola, Dr. Narayanan Krishnamurthi, and Dr. Baoxin Li for serving as my committee members. I am thankful to my fellow lab mates for their support and critical review of my work. Last, but the most important, my family for being there for me when I needed the most. The work in this thesis was partly supported by NSF CAREER grant 1452163 and NIH R24 grant 5R24HD050821-10.

TABLE OF CONTENTS

Page

LIST OF TABLES ix

LIST OF FIGURES x

CHAPTER

1 INTRODUCTION 1

1.1 Signal Acquisition 2

1.2 Action Recognition 2

1.3 Movement Quality Assessment 3

1.4 Research Objectives 4

2 DYNAMICAL SYSTEMS AND CHAOS 5

2.1 Properties of Chaotic Systems 6

2.2 Lorenz Attractor 7

2.3 Dynamical Modeling in Computer Vision 8

2.3.1 Preliminaries 9

2.4 Classical Dynamical Invariants 13

2.4.1 Largest Lyapunov Exponent 13

2.4.2 Correlation Sum 17

2.4.3 Correlation Dimension 17

2.4.4 Drawbacks of Traditional Chaotic Invariants 18

2.5 Applications of Interest 19

2.5.1 Activity Recognition 19

2.5.2 Activity Quality for Stroke Rehabilitation 20

2.5.3 Natural Scene Classification 21

3 DYNAMICAL SHAPE FEATURE EXTRACTION 23

3.0.4 Test on Models 26

CHAPTER Page

3.1 Experiments and Results 26
3.1.1 Motion Capture Dataset 29
3.1.2 Kinect Dataset 30
3.1.3 Activity Quality for Stroke Rehabilitation 35
3.1.4 Dynamic Scene Recognition 39
3.2 Conclusion and Future Work 40
4 KINEMATIC ANALYSIS FOR STROKE REHABILITATION 44
4.1 Related Work 49
4.2 System Design 50
4.3 Data Collection 51
4.3.1 Trajectory Error 52
4.3.2 Speed Profile Deviation 53
4.3.3 Jerkiness 54
4.3.4 Segmentation 55
4.3.5 Estimation of Optimal Weights and Thresholds 57
4.4 Experimental Results 60
4.5 Conclusion and Future Work 62
5 DECISION SUPPORT FOR STROKE REHABILITATION 65
5.1 Methods for Collecting Kinematics and Therapist Ratings 66
5.1.1 Collection of Kinematics 66
5.1.2 Therapist Rating Protocol 67
5.1.3 Data Collection 68
5.2 Definitions of Kinematic Features 71
5.3 Conclusion and Future Work 74

CHAPTER Page

6 DYNAMICAL REGULARITY FOR MOTION ANALYSIS: APPLICATIONS TO ACTION SEGMENTATION, RECOGNITION AND QUALITY ASSESSMENT 78
6.1 Related Work 82
6.2 Approximate Entropy (ApEn) 86
6.2.1 Choice of Parameters 89
6.3 Experimental Evaluation 91
6.3.1 Coupled Rossler Model 91
6.3.2 Segmentation and Action Classification 93
6.3.3 Temporal Segmentation 99
6.3.4 Movement Quality Assessment 100
6.3.5 Action Quality Assessment on Diving Datasets 103
7 MULTIVARIATE EMBEDDING BASED QUALITY ASSESSMENT OF DIVING ACTIONS 105
7.1 Introduction 105
7.2 Framework 107
7.2.1 Phase Space Reconstruction 107
7.2.2 Features from Reconstructed Phase Space 109
7.3 Experimental Evaluation 112
7.3.1 Diving Action Dataset 112
7.4 Conclusion 113
8 PERSISTENT HOMOLOGY OF ATTRACTORS FOR ACTION RECOGNITION 115
8.1 Related Work 116

CHAPTER Page

8.2 Preliminaries 117
8.2.1 Phase Space Reconstruction 119
8.2.2 Persistent Homology 119
8.3 Topological Features from Attractor 121
8.4 Experimental Results 123
8.4.1 Motion Capture Data 123
9 Conclusion and Future Directions 125
REFERENCES 127

LIST OF TABLES

Table Page

3.1 Experiments on Lorenz and Rossler Models 27

3.2 Classification Rates for Motion Capture Dataset 31

3.3 Confusion Table for Motion Capture Dataset 31

3.4 Classification Results for Cross-Subject Test Setting 34

3.5 Classification Resuts for Cross-subject Test Setting 34

3.6 Comparison of Classification Rates on Stroke Rehabilitation Dataset ... 36

3.7 Comparison of Performance for Movement Quality Assessment 39

3.8 Comparison of Classification Rates on UMD Dataset 41

3.9 Comparison of Classification Rates for Yupenn Dataset 42

4.1 Optimized Values for Linear Model for Quality Assessment 63

5.1 Rating Rubric for Quality Assessment 69

5.2 Demographics of Stroke Survivors 71

6.1 Confusion Table for Weizmann Dataset 95

6.2 Automatic Segmentation and Recognition Performance 98

6.3 Automatic Segmentation on UTKinect and Florence3D Dataset 99

6.4 Comparison of Average Temporal Segmentation Accuracy 101

6.5 Mean Rank Correlation for Quality Assessment of Diving Actions 104

7.1 Mean Rank Correlation for Various Methods 113

8.1 Comparison of Classification Rates on the Motion Capture Dataset 124

8.2 Confusion Table for Motion Capture Dataset 124

LIST OF FIGURES

Figure		Page
2.1	Phase Space Reconstruction Of Lorenz Attractor	7
2.2	Estimation of Time Delay	13
2.3	Examples of Phase Space Reconstruction	14
2.4	The Algorithm for Estimation of Largest Lyapunov Exponent	16
3.1	Illustration of The Effect of Time-series Lengths On Reconstructed Phase Space	27
3.2	Illustration of Stability of The Dynamical Shape Distribution	28
3.3	Illustration of Phase Space Reconstruction And Dynamical Shape Feature Extraction	29
3.4	Example Actions From Action Classes	32
3.5	Proposed Framework For Movement Quality Assessment	34
3.6	Block Diagram Representation For Movement Quality Assessment	37
3.7	Comparison Between Impairment Level Given By WMFT and MQS	37
3.8	Dynamic Scene Recognition	41
4.1	Exemplar Visual Feedback Summaries Based On Low-level Kinematic Analysis	46
4.2	The Home-based Adaptive Mixed Reality Rehabilitation System	47
4.3	The Proposed Linear Model of Kinematic Features	58
4.4	Comparison Between The Predicted Cumulative Score And Therapist Rating	59
4.5	Comparison of Cumulative Score and Therapist Rating	60
4.6	Linear Regression Plots for Various Low-level Kinematic Features	60
5.1	A Sample of Video Data Provided to Therapists	70
5.2	The Decision Tree Model For Movement Quality Assessment	75

Figure Page

5.3 Comparison Between Impairment Level Given By Component-level Score for Wrist Trajectory and Decision Tree Predictions 75
6.1 Visual Representation of Our Applications of Interest 80
6.2 Picture Showing Sliding Window To Estimate Approximate Entropy ... 90
6.3 Estimation of Delay Time 91
6.4 Illustration of Utility of Approximate Entropy 93
6.5 Approximate Entropy Features Estimated on Left and Right Hand Trajectories 94
6.6 Typical Video Frames from Weizmann Dataset 95
6.7 Exemplar Recurrence Plots from Action Sequences 97
6.8 Illustration of Utility of Approximate Entropy Feature 101
6.9 Comparison of Temporal Clustering Methods on the CMU Motion Capture Dataset 102
6.10 The Impairment Scores Assigned to Movements 103
7.1 Block Diagram Showing Algorithmic Flow for Quality Assessment of Diving Actions 106
7.2 Exemplar Video Frames Shown From The Diving Action Dataset 109
8.1 Phase Space Reconstruction of Dynamical Attractors 118

1 INTRODUCTION

Computer vision community has been interested in modeling human activity for numerous applications including video surveillance, automatic video annotation and health monitoring [4]. Understanding the underlying dynamics in human motion forms the core idea of such systems. Human activity analysis has attracted the attention of many researchers providing extensive literature on the subject. A detailed review of the approaches in literature for modeling and recognition of human activities are discussed in [4, 51]. Recent advancements in sensing platforms like motion capture systems and Kinect have opened doors to several applications including home-based health monitoring, gaming and entertainment. Take for instance, the task of developing algorithms for understanding the dynamics in human activities. This problem is non-trivial due to the complexity of natural human movement, which is a result of interactions between multiple body joints having high degrees of freedom. In addition, the task of recognizing human actions is challenging due to several factors including inter-class similarities between actions (e.g., running and walking), intra-class variations due to multiple strategies for an action (e.g., dance) and inter-subject variations.

An 'action' is defined as simple motion patterns usually executed by a single person typically lasting for a short duration of time (around 10 sec) [134]. An activity is a complex sequence of actions performed by several individuals interacting with each other. Natural human movements (such as walking, running) are composed of

periodic action sequences in the form of repetitions, with some variability [127]. In our research, we focus our interest towards human activity analysis with two main applications: (a) action recognition, and (b) movement quality assessment for stroke rehabilitation.

1.1 Signal Acquisition

Within the framework of our research, we work with various sensing modalities such as optical motion capture systems, Kinect and RGB cameras. These sensing modalities are classified as "outside-in" systems which use external sensors to collect data from sources placed on the human body. Optical motion capture systems use infrared cameras to track the motion of reflective markers placed on the body. Such systems are highly accurate and can operate at 100 frames/second or higher. These can track a large number of markers, but the experimental data has to be captured in a controlled environment away from reflective noise and without occlusion of markers. Traditional sensing in the vision community has been using RGB cameras which are cheaper and operate at lower frequency of 30 frames/second. A recent technology of Kinect uses depth information along with RGB data to achieve markerless motion capture at 30 frames/sec. Our experimental analysis show results on publicly available datasets which were collected using these sensing modalities.

1.2 Action Recognition

The aim here is to recognize the type of action performed by a subject in the sequence of images using the training examples provided for each class of actions. In a real world scenario, it would require automatic recognition of action sequences from continuous untrimmed videos. Traditionally, the vision community works with

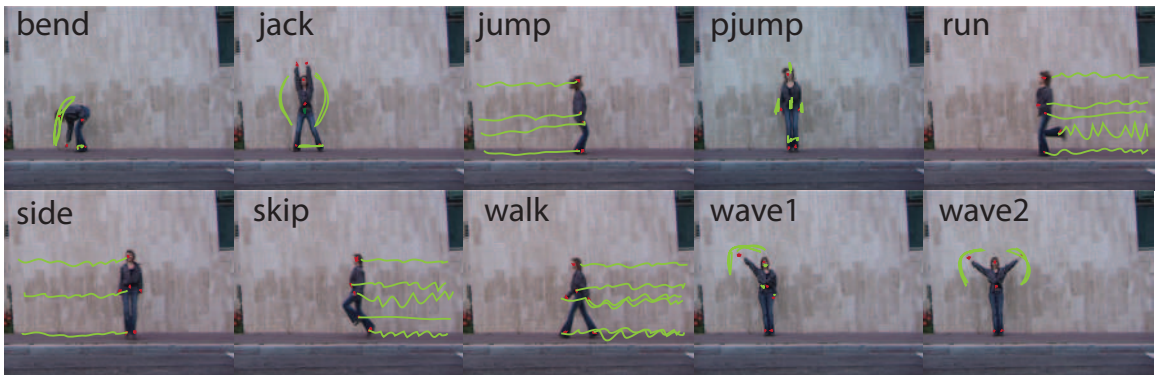


Figure 1.1: Typical video frames of 10 actions performed by a subject from the Weizmann dataset [54]. The trajectories corresponding to six body joints namely head, belly, two hands and two feet were extracted by Ali *et al.* [5].

the simpler, unrealistic assumption that temporal segmentation of videos is a step which has been done beforehand, resulting in pre-segmented videos containing individual action sequences as shown in Figure 6.6. Action recognition has got the industry interested with applications such as gaming systems using Kinect, security and surveillance systems, animated movies and gait analysis.

1.3 Movement Quality Assessment

The application of interest here is to develop a computational framework for movement quality assessment to aid physical therapists in providing supervised rehabilitation therapy for stroke survivors. Stroke is the most common neurological disorder worldwide leaving behind a significant number of survivors every year disabled with chronic impairments such as problems with vision, difficulty to formulate or understand speech, or inability to move limbs. Increasing healthcare costs paired with insufficient coverage by insurance for long-term therapy treatment has often left impairments untreated. Several validated clinical measures which requires visual monitoring by a therapist for movement quality assessment have been proposed, and researchers aim to match these clinical scores using a computational framework. The

existing approaches in literature to quantify movement quality use nonlinear dynamical system theory [127, 142, 95], random forests [93], and SVMs [94]. Chen *et al.* [30] proposed several kinematic attributes which requires access to reach trajectories from unimpaired subjects, thereby limiting the generalizability of the framework to different reach targets. In our research agenda, we aim to propose a framework which is general enough to recognize coarse differences in different actions and as well quantify fine variations (impairments) in a given action.

1.4 Research Objectives

The aim of this research is two-fold:

(a) To propose a novel approach based on nonlinear dynamical analysis and shape analysis to address the drawbacks of traditional measures used in literature for action recognition. In this direction, we propose to use dynamical shape features representative of the shape of the reconstructed phase space as our feature representation in our framework. We also test the generality of the proposed feature representation to other tasks such as movement quality assessment and dynamical scene recognition.

(b) To propose a kinematics-based framework to generate movement quality scores matching therapists' impressions of movement quality in the context of a home-based rehabilitation setting.

2 DYNAMICAL SYSTEMS AND CHAOS

Dynamical systems are mathematical models which are used to simulate a physical phenomenon whose states evolve over time. Chaos theory studies the behavior of nonlinear dynamical systems, that are highly sensitive to initial conditions. Any perturbation to the initial conditions of such systems yields widely diverging dynamics. This behavior is known as deterministic chaos. Convincing evidence for existence of deterministic chaos has been provided from a variety of research experiments [111, 128]. Differential equations have been used to model physical systems to determine how they behave temporally under different experimental conditions and try to predict their future states. Modeling a physical system using differential equations is essentially impossible when the order and degree of the modeled systems are very high. Nonlinear systems with closed form analytical solutions typically settle in a steady state or in a periodic motion. In early sixties, a new kind of motion was observed which was erratic. This type of motion was termed chaos, and the theory developed to explain such systems as chaos theory.

Many natural systems showing chaotic behavior have been comprehensively studied [59, 114], the most famous one being the weather. The initial study on chaos theory was pursued by a meteorologist, Edward Lorenz, while working on weather prediction models on a computer with a set of differential equations to model the weather. When he started the same experiment with a different set of initial conditions, he found that rounding-off errors in initial conditions had a large influence on

the subsequent dynamics of the model equations.

A detailed description of such systems was first described mathematically by Lorenz in his seminal paper in 1963. He presented a system of 3 coupled differential equations which demonstrate chaotic behavior. This led him to his now famous speculation that a butterfly flapping wings in Brazil (which is a small change in the initial conditions in the atmosphere) might cause a tornado in Texas. This dependence of the evolution of a system on its initial conditions makes chaotic motion a complex phenomenon. In this sense, it is intuitive to expect that systems in nature are complex, and the larger the number of systems state variables, the more complex the system is.

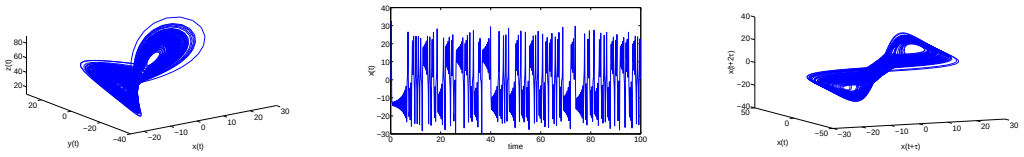
2.1 Properties of Chaotic Systems

1. Determinism: Even though chaotic systems exhibit random behavior, they are classified as deterministic systems. This is because if the initial conditions are known precisely, future behavior of the system can be predicted. However, initial conditions are never known for a real system.

2. Nonlinearity: Nonlinearity is a necessary condition for a system to exhibit chaos. A perfectly linear system can never exhibit chaos.

3. Sensitivity to initial conditions: This is the most important characteristic of chaotic systems. Chaotic systems for any two different initial conditions (however close) always diverge exponentially as they evolve in time. Hence, a small change in the initial conditions takes the system in a completely different trajectory.

4. Boundedness: If the divergent orbits go to infinity, the system is considered not to be chaotic as the system is unbounded and cannot produce steady states.



(a) (b) (c)

Figure 2.1: Phase space reconstruction of Lorenz attractor by delay embedding. (a) shows the 3D view of trajectories of Lorenz attractor with control parameters $\rho = 45.92, \sigma = 16.0$ and $\beta = 4.0$. We can see that trajectories of Lorenz system settle down and are confined within the attractor. The one-dimensional time series (observed) of the Lorenz system is shown in (b). We see that a low-dimensional nonlinear system can generate such complex and chaotic signal. (c) shows the reconstructed phase space from observed time series of the Lorenz system using delay embedding ($\tau = 11$). The above example illustrates that the reconstructed phase space preserves certain topological properties of the original Lorenz attractor.

2.2 Lorenz Attractor

The Lorenz attractor is the steady state of a nonlinear chaotic system of three coupled nonlinear ordinary differential equations [133] as given below:

$$\dot{x} = \sigma(y - x), \tag{2.1a}$$

$$\dot{y} = x(\rho - z) - y, \tag{2.1b}$$

$$\dot{z} = xy - \beta z, \tag{2.1c}$$

where x, y, z are the state variables and σ, ρ and β are non-negative and dimensionless parameters. These equations were defined by Lorenz in 1963 [145] to represent a simplified model of thermal convection in the lower atmosphere. Lorenz showed that this relatively simple-looking set of equations could have highly erratic dynamics for a range of defined control parameters, for which the dynamics are chaotic.

Upon close inspection of the plots shown in Fig. 2.1, the trajectories depicted therein never intersect each another. For any small perturbation of initial conditions, the state-space trajectory will never follow the same path. Furthermore, if one were to plot the trajectories of the solution for one set of initial conditions and then for

another set of initial conditions (infinitesimally close to the first), the two trajectories would diverge from one another exponentially. This means that not only does a small perturbation to initial condition result in a trajectory that will never intersect with that of the original system but it results in a completely different trajectory.

The dynamics of the Lorenz system in the 3-dimensional state space generated from these set of equations is illustrated in Fig. 2.1(a). Lorenz attractor also illustrates that deterministic nonlinear models of low dimension can produce signal with complex dynamics. Furthermore, Fig. 2.1 illustrates that it is possible to recreate an approximate attractor generated by a multidimensional system (such as Lorenz) using only a one-dimensional observed time series.

2.3 Dynamical Modeling in Computer Vision

Dynamical modeling methods for understanding signals from various sensing platforms have been the cornerstone of many applications in the computer vision community, such as human activity analysis [4] and dynamical natural scene recognition [118]. Natural human movements (such as walking, running) are composed of periodic action sequences in the form of repetitions, with some variability [127]. These inherent attributes of human movement (periodicity with variability) descriptive of a complex nonlinear chaotic system has motivated researchers to employ tools from nonlinear dynamical systems theory to model human movement [5, 64, 127, 95, 37, 38, 57, 81]. Dynamical modeling of spatio-temporal evolution of human activities are traditionally accomplished by defining a state space and learning a function that maps the current state to the next state [104, 16]. A recent alternate approach has attempted to derive a representation for the dynamical system directly from the observation data using tools from chaos theory [5]. The main idea here is that, by using a top-down approach of dynamical modeling, one would only approximate the true-dynamics of the system

with attempts to fit a model to the observational data. Whereas, in the bottom-up approach [5], the dynamical system parameters such as the number of independent variables, degrees of freedom and other unknown parameters are estimated from the data. Such an approach can be seen as a generalized representation without any strong assumptions, suitable for analyzing a wide range of dynamical phenomenon.

2.3.1 Preliminaries

In this section, we introduce the background necessary to develop an understanding of nonlinear dynamical system analysis and chaos theory for applications in activity analysis, activity quality assessment and natural scene analysis.

Dynamical System Analysis

Dynamical systems are governed by a set of functions defining the variations in the behavior of the system over time. A dynamical system is termed linear or nonlinear if the function defining the behavior of the system is linear or nonlinear respectively. Dynamical systems can be represented using state variables defining the state of the system at a given time t. A dynamical system is termed deterministic if there exists a unique future state for a given current state and is termed stochastic if the future state is derived from a probability distribution of possible states. Chaos theory is the field of study of such deterministic dynamical systems that show high sensitivity to initial conditions. A chaotic system is a dynamical system with deterministic behavior showing sensitivity to initial conditions.

The states of a chaotic system are generally considered to be in an n-dimensional manifold also called *phase space*. A chaotic system evolves over time in its phase space according to the system variables governing the dynamics. The path traversed by the system over time is called a *trajectory* and the region of the phase space where

the trajectories settle down as time approaches infinity is denoted as an *attractor*.

One would intend to have access to all independent variables of the system and their interactions for a complete understanding of the system. In a real world scenario, the data recorded is of low-dimension and is insufficient to model the dynamics of the system. In addition, model-based (parametric) approaches, such as LDS assume an underlying mapping function f to describe the dynamics of the system. It has been established that such approaches may not be suitable for modeling the dynamics of complex systems such as human movements due to the simplifying assumptions [15]. The theory of chaotic systems allows for determining certain invariants of the dynamical system function f without making any assumptions about the system.

Phase Space Reconstruction

The *phase space* is defined as the space with all possible states of a system [145, 3]. In a deterministic dynamical system that can be mathematically modeled, future states of the system can be determined using present and past state information. However, for applications such as human activity understanding and dynamical scene understanding, the system equations are complex. Furthermore, sensing systems in the real-world do not allow us to observe all variables of the system (e.g., the home-based setting for stroke rehabilitation with single marker on the wrist). To address these problems, we have to employ methods for reconstructing the attractor to obtain a phase space which preserves the important topological properties of the original dynamical system. This process is required to find the mapping function between the one-dimensional observed time series and the m-dimensional attractor, with the assumption that all variables of the system influence one another. The concept of phase space reconstruction was expounded in the embedding theorem proposed by Takens, called Takens' embedding theorem [129] and an example of the procedure is

shown in Fig. 2.1. For a discrete dynamical system with a multidimensional phase space, time-delay vectors (or embedding vectors) are obtained by concatenation of time-delayed samples given by

$$\mathbf{x}_i(n) = [x_i(n), x_i(n+\tau), \cdots, x_i(n+(m-1)\tau)]^T, \tag{2.2}$$

where 'm' is the embedding dimension and 'τ' is the embedding delay. These parameters should be carefully selected in order to facilitate a good phase space reconstruction. For a sufficiently large 'm', the important topological properties of the unknown multidimensional system are reproduced in the reconstructed phase space [3]. The embedding method has proven to be useful, particularly for time series generated from low-dimensional deterministic dynamical systems, by providing a way to apply theoretical concepts of nonlinear dynamical systems onto observed time series. The embedding theorem does not suggest methods to estimate the optimal values for 'm' and 'τ'. We use false nearest neighbors [68] approach to estimate m and the first zero crossing of the autocorrelation function [122] to estimate τ. Fig. 2.1 shows an example of phase space reconstruction from a one-dimensional observed time-series of a Lorenz system.

Embedding Dimension

The embedding dimension refers to the number of time-delayed samples concatenated to form the time-delay vector. The aim here is to estimate an integer embedding dimension which can *unfold* the attractor thereby removing any self-overlaps due to projection of the attractor onto lower dimensional space. Hence, the embedding dimension can be defined as the minimum dimension required to unfold the attractor completely. The false nearest neighbor approach finds this minimum embedding dimension to remove any *false* nearest neighbors (neighbors due to projection onto

lower dimension) [3]. Consider a vector in reconstructed phase space in dimension m given by

$$\mathbf{x}(k) = [x(k), x(k+\tau), \cdots, x(k+(m-1)\tau)]^T, \tag{2.3a}$$

and a nearest neighbor in the phase space given by

$$\mathbf{x}^{NN}(k) = [x^{NN}(k), x^{NN}(k+\tau), \cdots, x^{NN}(k+(m-1)\tau)]^T. \tag{2.3b}$$

If the vector $\mathbf{x}^{NN}(k)$ is a true neighbor of $\mathbf{x}(k)$, then it should be because of the underlying dynamics. The vector $\mathbf{x}^{NN}(k)$ can be a false neighbor of $\mathbf{x}(k)$ when dimension m is unable to unfold the attractor. Hence, moving to the next dimension $m+1$ may move this false neighbor out of the neighborhood of $\mathbf{x}(k)$. This process of finding false neighbors to every vector $\mathbf{x}_i(k)$ sequentially removes self-overlaps and identifies m where the attractor is completely unfolded. The embedding dimension m suggested by the false nearest neighbor algorithm for exemplar trajectories of human actions was either 3 or 4. We select a constant embedding dimension $m = 3$ to reconstruct all relevant phase space. Even with this fixed value of m, we obtain excellent results as shown in our experiments.

Embedding Delay

Embedding delay refers to the choice of integer time delay used to construct the time-delay vector. Theoretically, the embedding process allows any value of τ if one has access to infinitely accurate data ([3], chap. 3). Since this is practically impossible, we try to find a value τ which makes the components of the vector $[x(k)$, $x(k+\tau)$, $x(k+2\tau)]^T$ in the embedding sufficiently independent. A low value of τ makes adjacent components to be correlated and hence they cannot be considered as independent variables. On the other hand, a high value of τ may make the adjacent components uncorrelated (almost independent) and cannot be considered as part of

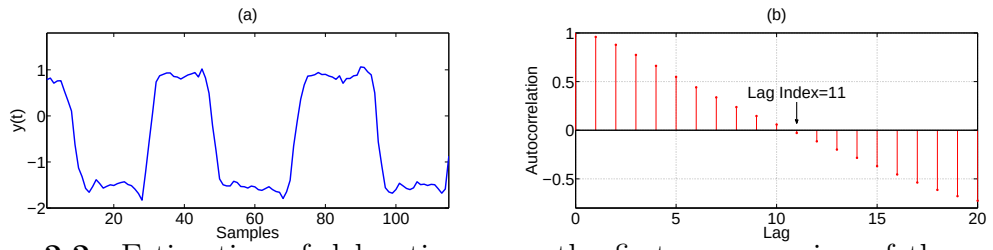


Figure 2.2: Estimation of delay time τ as the first zero-crossing of the autocorrelation function. (b) shows the autocorrelation function of the trajectory data in (a).

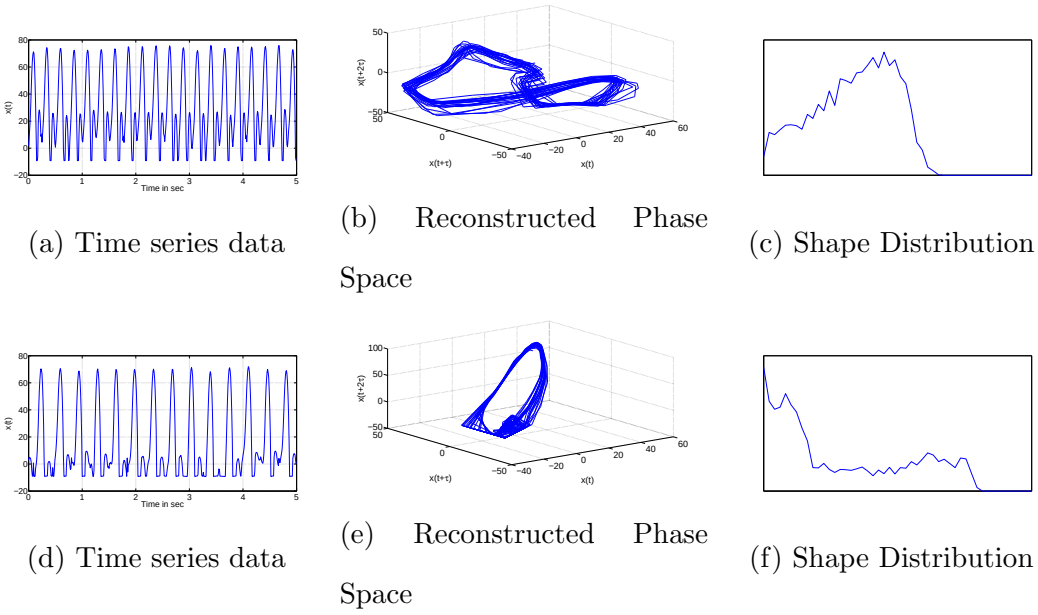
the system that supposedly generated them. The shape of the embedded time series will critically depend on the choice of τ [122]. A good selection of τ should ensure that the data are maximally spread in phase space resulting in smooth phase space reconstruction. We use the first zero-crossing of the autocorrelation function as an estimate of τ as suggested in [122] for strongly periodic data, which is a suitable choice for our experiments.

2.4 Classical Dynamical Invariants

Quantifying divergence of closely spaced trajectories and hence system complexity is a well-studied problem in the field of chaos theory. Correlation dimension [3], largest Lyapunov exponent [148], and correlation sum [3] are a few examples of invariant measures proposed in the literature to quantify complexity of nonlinear dynamical systems. In this section, we study the three commonly used dynamical invariants in the field of chaos theory and computer vision.

2.4.1 Largest Lyapunov Exponent

The Lyapunov exponent is a measure of average rate of divergence (or convergence) of initially closely-spaced trajectories over time [3, 145]. A positive Lyapunov exponent indicates orbital divergence and hence chaos in the system. A negative Lyapunov



(a) Time series data (b) Reconstructed Phase Space (c) Shape Distribution

(d) Time series data (e) Reconstructed Phase Space (f) Shape Distribution

Figure 2.3: Examples of phase space reconstruction of corresponding time series data of a subject performing *Run* and *Walk* action respectively. The embedding parameters were selected as $m = 3$ and τ as described in section 2.3.1. This example illustrates that the *shape* of the reconstructed phase space can be seen as a discriminative feature for classification of actions. We use shape distributions proposed by Osada *et al.* [92] as a representation for shape of phase space. (c) and (f) together support our hypothesis that shape distribution (**D2**) can be used for classification of actions.

exponent indicates orbital convergence and hence a dissipative system.

Chaos theory has found its applications in the analysis of chaotic dynamical systems. In comparison, largest Lyapunov exponent is a widely used measure of chaos in various engineering applications, including computer vision and biomechanics to model human movements and quantify chaos in the reconstructed phase space [37, 5, 95, 127, 132, 118]. It is used to quantify the variability in human movement [127], which is believed to exhibit a chaotic structure. The inherent assumption here is that different action classes possess different levels of chaos and quantification using Lyapunov exponents help in classification of these action classes. While quantification of chaos using the largest Lyapunov exponent have been used to monitor varying chaos levels (level of complexity of the system) for recognition or prediction

purposes [62], experimental studies for modeling human activities have not reported any evidence for different levels of chaos in human activities. Hence, we believe that a representation for level of chaos may not be a suitable approach to model human activities. While previous experiments on action modeling using Lyapunov exponents have reported good results, certain data requirements make it less suitable for action modeling where the number of data samples are less.

Estimation of Largest Lyapunov Exponent (λ_1)

A recent practical method for estimating the largest Lyapunov exponent from a time series proposed by Rosenstein [109] quantifies chaos by monitoring the rate of divergence of closely spaced trajectories over time. The algorithm claims to be fast, easy to implement and robust to changes in embedding dimension, size of dataset, embedding delay and noise level. Rosenstein's algorithm was developed to address the limitations of the Wolf's algorithm [148] and has been shown in [132] that it is more robust to changes in data length than the Wolf's algorithm. The algorithmic flow as proposed by Rosenstein is shown in Figure 2.4.

However, experimental results on Lorenz and Rossler models for different time series lengths (N) with fixed embedding dimension and embedding delay shows that the estimate approaches the true value only after $N = 5000$ and 2000, respectively. Furthermore, both Rosenstein and Wolf suggest that the minimum number of data samples required for accurate estimation of largest Lyapunov exponent is 10^m (where m is the embedding dimension) [132, 55]. Therefore, we believe that the use of largest Lyapunov exponent may not be a suitable approach in modeling short-duration video data.

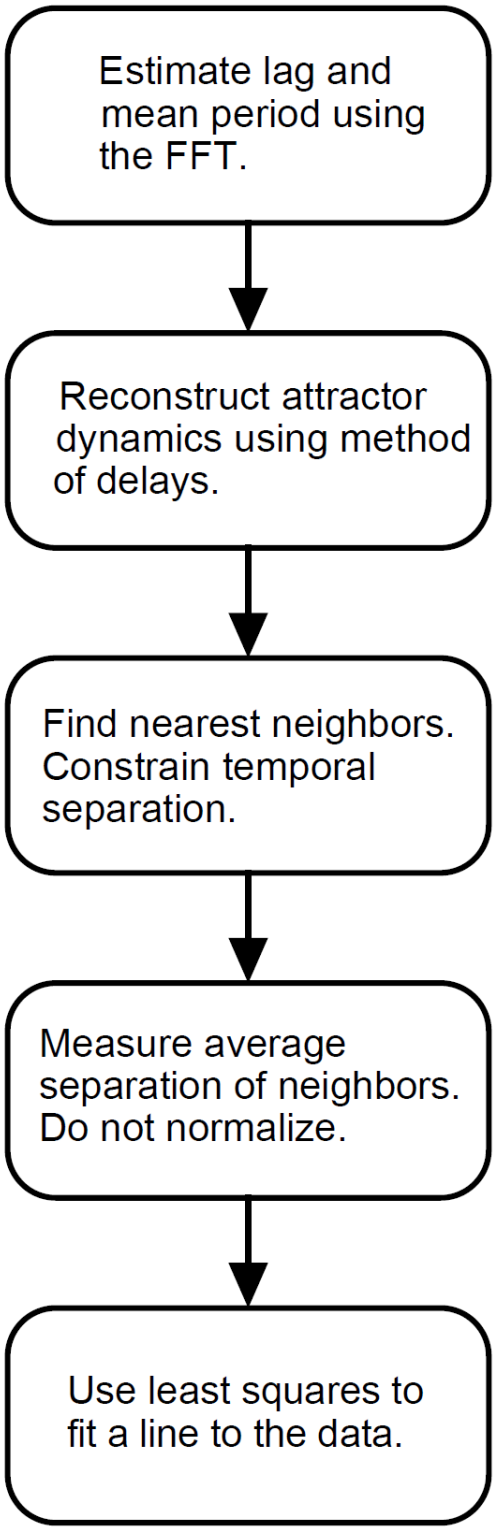


Figure 2.4: The algorithm for estimation of largest Lyapunov exponent from experimental time series data.

2.4.2 Correlation Sum

Correlation sum is a chaotic invariant used to quantify density of points in the reconstructed phase space. For a given point in the reconstructed phase space, draw a circle of radius 'r' around it and count the number of points which fall inside the circle. Repeat the procedure for all points in the reconstructed phase space. This process can be mathematically represented as

$$C(r) = \frac{2}{N(N-1)} \sum_{j=1}^{N} \sum_{i=j+1}^{N} \Theta(r - d(\mathbf{x}(i), \mathbf{x}(j))), \tag{2.4}$$

where:

$$\Theta(a) = \begin{cases} 1, & \text{if } a \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

and

$$d(\mathbf{x}(i), \mathbf{x}(j)) = \sqrt{\sum_{k=0}^{m-1} (X_{i-k} - X_{j-k})^2}$$

Θ is the Heaviside function, $C(r)$ is called the correlation sum which converges to correlation integral when N $\to \infty$. This procedure of estimating correlation sum was proposed by Grassberger *et al.* and is called as the Grassberger-Procaccia algorithm. Correlation sum ($C(r)$) refers to the probability that two randomly chosen vectors will be closer than r in the reconstructed phase space.

2.4.3 Correlation Dimension

One would expect the correlation sum $C(0) = 0$ for a chaotic system, as the points in reconstructed phase space never repeat in a nonperiodic system embedded without false nearest neighbors. A plot of $\log C(r)$ versus $\log r$ should give an approximately straight line whose slope in the limit of small r and large N is called as the correlation dimension given by

$$D_2 = \lim_{r \to 0} \lim_{N \to \infty} \frac{\log C(r)}{\log r} \tag{2.5}$$

It is important to note here that these invariants of the dynamics (largest Lyapunov exponent, correlation dimension and correlation sum) have been extracted directly from the given time series without making specific assumptions about the system.

2.4.4 Drawbacks of Traditional Chaotic Invariants

The proposed algorithms to estimate chaotic invariants suggest that these invariant measures require large number of data samples (of the order of $10^m - 30^m$) [132, 109] for accurate estimation (where m is a parameter used in the estimation procedure called as the embedding dimension), with typical values of $m = 3$ and above, corresponding to a minimum of 1000 data samples. In computer vision applications such as action recognition, the signal acquisition operates at a frequency of 30 frames/sec. Hence, the observation time of any given action should be at least 33 seconds, which is impractical. In general, these traditional chaotic invariants suffer from at least one of these drawbacks: (a) unreliable for small datasets, (b) computationally intensive, (c) relatively difficult to implement [109]. In recent years, these methods have been applied to model various visual dynamical phenomenon such as video-based recognition of human activities [5] as well as recognition of dynamical scenes [118]. However, when one needs to make inferences from short videos, or for instance when the activity of interest lasts only a few seconds, the classical approaches have significant drawbacks.

2.5 Applications of Interest

2.5.1 Activity Recognition

Human activity analysis has attracted the attention of many researchers providing extensive literature on the subject. A detailed review of the approaches in literature for modeling and recognition of human activities are discussed in [4, 51]. Since our present work is related to non-parametric approaches for dynamical system analysis for action modeling, we restrict our discussion to related methods.

Human actions have been modeled using dynamical system theory in computer vision [5, 16] and biomechanics [37, 95, 127]. Differential equations can be used to model such a system, which requires access to all independent variables of the system. This approach would facilitate an understanding of the system behavior and also allow for the prediction of future states using present and past state information. However, this is not realizable in practice, as it is extremely hard to determine the independent variables and the interactions governing the dynamics of human actions.

Dynamical modeling of human actions can be broadly categorized into parametric and nonparametric methods. Furthermore, human actions have been modeled with the assumption that the underlying dynamical system is linear [16] or nonlinear [5, 104]. In parametric modeling approaches, the dynamics of a system is represented by imposing a model and learning the model parameters from training data. Hidden Markov Models (HMMs) [103] and Linear Dynamical Systems (LDSs) [26] are the most popular parametric modeling approaches employed for action recognition [154, 146, 139, 33] and gait analysis [65, 77, 16]. Nonlinear parametric modeling approaches like Switching Linear Dynamical Systems (SLDSs) have been utilized to model complex activities composed of sequences of short segments modeled by LDS [20]. While, nonlinear approaches can provide a more accurate model, it is difficult

to precisely learn the model parameters. In addition, one would only approximate the true-dynamics of the system with attempts to fit a model to the experimental data. An alternative nonparametric action modeling approach is based on tools from chaos theory, with no assumptions on the underlying dynamical system. Traditional chaotic measures, like the largest Lyapunov exponent, correlation dimension and correlation integral, have been extensively used to model human actions [5, 37, 95, 127]. However, [109] and [132] have shown that these nonlinear dynamical measures need large amounts of data to produce stable results (10^m, where m is the embedding dimension). Junejo *et al.* [64] used a self-similarity matrix, a graphical representation of distinct recurrent behavior of nonlinear dynamical systems, to learn an action descriptor. In this work, through illustrative examples and experimental validation, we show that our framework works better than traditional chaotic invariants for action modeling.

2.5.2 Activity Quality for Stroke Rehabilitation

While recognizing human activities is seen as a challenging task in the computer vision community, recently researchers from various backgrounds have shown interest in the development of computational frameworks for quantification of *quality* of movement, for possible applications in health monitoring and rehabilitation [30, 127, 132, 142]. Stroke being the most common neurological disorder, leaves millions disabled every year who are unable to undergo long-term therapy treatment due to insufficient coverage by insurance. Recent directions in rehabilitation research has been towards development of portable systems for therapy treatment. Traditional quantitative scales such as the Fugl Meyer Test [50] and the Wolf Motor Function Test (WMFT) [149], have proven to be effective in evaluating movement quality. However, these approaches involve visual monitoring which would greatly benefit from the devel-

opment of an objective computational framework for movement quality assessment. The aim here is to develop standardized methods to describe the level of impairment across subjects. We show the utility of the proposed action modeling framework for quantifying the quality of reaching tasks using a single marker on the wrist, and obtain comparable results to a heavy marker-based setup (14 markers placed on arm, shoulder and torso [30]).

The focus of existing approaches for movement quality assessment has been towards finding typical patterns in kinematics which differ between healthy and impaired subjects. While these approaches are successful in giving an insight into understanding human movement, they fail to utilize the inherent dynamical nature of the movement. Rehabilitation therapies are composed of repetitive movements (e.g., reach to a target) that are strongly periodic with inherent variability. Traditional methods have assumed that this variability arises from noise in the system. However, it is evident that variability is an integral part of repetitive movements due to the availability of multiple strategies for the movement. Also, it is believed that variability produced in human movement is a result of nonlinear interactions and have deterministic origin [127]. Extensive research has been carried out to model this variability using nonlinear dynamical system theory [37, 95, 127]. In this work, we utilize the action modeling framework for movement quality assessment using a single wrist marker.

2.5.3 Natural Scene Classification

Natural scene classification has been an active area of research in computer vision with applications in automated image and video understanding. Much research has been focused around scene classification using single still images [47, 153], thereby neglecting dynamical motion information available in videos. Recently, the problem

of dynamical modeling of natural scenes was introduced by Shroff *et al.* [118] who utilized tools from chaos theory along with GIST [90, 89] to model the spatio-temporal evolution in natural scenes in an unconstrained setting.

Dynamic texture representation using LDS proposed by Soatto *et al.* have been used to recognize and synthesize dynamic textures such as sea-waves, smoke, traffic [124, 39]. Such low-dimensional models have been used to capture complex natural phenomena. However, experimental results reported in [118] show that these simple models might not be effective for dynamic scene classification in an unconstrained setting. Shroff *et al.* utilized traditional chaotic invariants to model the dynamics and have shown that dynamical attributes augmented with spatial attributes (GIST [89]) can be effectively used for categorization of dynamic scenes [118]. Another recent approach utilized spatio-temporal oriented energy filters for dynamic natural scene classification [36]. In this work, we test the generality of the proposed action modeling framework for dynamic scene classification application.

3 DYNAMICAL SHAPE FEATURE EXTRACTION

In this chapter, we present a framework which combines the strong theoretical concepts of nonlinear dynamical analysis and ideas in shape theory to effectively represent the nature of dynamics. From Fig. 2.3, we see that the '*shape*' of the reconstructed phase space can be seen as a discriminative feature for classification between *Run* and $Walk$ action classes. Hence, our aim will be to extract feature representations for the shape of the reconstructed phase space. It is important to note here that the process of phase space reconstruction preserves certain topological properties and global shape is not a topological invariant, while local shape properties are. However, our goal here is to suggest a shape-based descriptor (both global and local) which possess sufficient discriminatory properties and robustness.

We consider the attractor as having its own characteristic shape in the high-dimensional phase space. Shape analysis of 3D surfaces is a well-studied problem in the computer vision community. In [92], Osada *et al.* present a method for finding a similarity measure between 3D shapes by computing shape distributions of the 3D surface sampled from the shape function by measuring their global geometric properties. We use the shape distribution of the reconstructed phase space as the dynamical feature representation in our experiments. While the shape distributions was originally proposed to measure similarity between 3D shapes, we believe that shape distributions can be used as feature representations for any n-dimensional phase space. In addition, it is said that any function can be used to extract the shape

distribution [92], but we adopt simpler shape functions based on geometric properties (distance and area) which are listed below:

(a) *Global Shape Functions*:

- **D1**: measures the distance between one fixed point and one random point sampled from the reconstructed phase space. The fixed point is selected as the centroid of the attractor.
- **D2**: measures the distance between two random points in the phase space represented as $||\mathbf{x}_i - \mathbf{x}_j||_2$.
- **D3**: measures the square root of the area of the triangle formed by three random points on the attractor.

For example, the **D2** shape function can be represented as

$$\mathbf{D2}_{ij} = ||\mathbf{x}_i - \mathbf{x}_j||_2, \tag{3.1}$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are points (embedding vectors) in the reconstructed phase space. A set of these distances for randomly chosen embedding vector pairs are computed. From this set, we construct a histogram by counting the number of samples which fall into each of B=50 fixed sized bins to obtain the attractor's shape distribution.

These shape functions encode global geometric properties of the phase space, lacking information about local shape and dynamical evolution in the phase space. While previous investigation shows that global geometric shape function (**D2**) performs sufficiently better than the traditional nonlinear dynamical measures (largest Lyapunov exponent, correlation dimension and correlation integral) [142], we hypothesize that a shape function which encodes local geometry and dynamical evolution information of phase space should improve the performance. In this direction, we propose new

shape functions defined as,

(b) *Local Shape Functions*:

- **DT1**: It is similar to **D2**, with an additional constraint that the time separation between two random points in reconstructed phase space is $\leq \delta$, thereby encoding only the local shape information.

- **DT2**: encodes dynamical evolution of the phase space by exponential weighting given by

$$\mathbf{DT2}_{ij} = e^{-\gamma|t_i - t_j|} * ||\mathbf{x}_i - \mathbf{x}_j||_2, \tag{3.2}$$

  where t_i and t_j are the time indexes of the randomly selected pair of embedding vectors in the reconstructed phase space. 'δ' and 'γ' are empirically determined parameters such that $\delta, \gamma \geq 0$.

Local vs Global: The main idea behind proposing these local shape functions is that, a global shape function would consider data samples from independent repetitions (well separated in time) of a movement. Also, repetitive human movements (such as *running* and *walking*) result in trajectories which wraps around itself in reconstructed phase space, creating an artifact of having closely spaced trajectories in phase space. We believe that such an approach would not provide a robust feature representation, and we suggest the use of local shape functions instead which only considers data samples close in time.

Metric on Shape Distributions: Several metrics exist in literature to calculate the distance between histograms including chi-squared statistic (χ^2 distance), Bhattacharyya distance [13], Riemannian analysis [126] and Earth Mover's Distance (EMD) [112]. In our experiments, we provide results using Euclidean distance and chi-squared distance metrics for comparison due to their simplicity.

3.0.4 Test on Models

The framework was tested on the Lorenz and Rossler models to determine whether the shape feature can be effectively used to classify differences in shape of reconstructed phase space of nonlinear dynamical systems. We compare the performance of the proposed framework with that of largest Lyapunov exponent. The effect of time-series length on estimation of largest Lyapunov exponent was revealed by Rosenstein *et al.* [109], by evaluating the performance of the algorithm they proposed for estimation of λ_1 for various time-series lengths. The simulation results on Lorenz and Rossler models are shown in TABLE 3.1. Their findings indicate that the estimation error increases with reduction in time-series length (N). Fig. 3.1 depicts the variations in reconstructed phase space for different time-series length with defined embedding parameters. It is evident from these plots that the *shape* of the reconstructed phase space remain sufficiently similar and can be used as a discriminative feature for classification purposes. Also, from Fig. 3.2, the shape distribution (using **D2** shape function) was found to be stable for different time-series lengths. This striking ability of our feature representations to be robust to changes in data length will be useful in applications related to human activity analysis, where the signal observation time is small/variable.

3.1 Experiments and Results

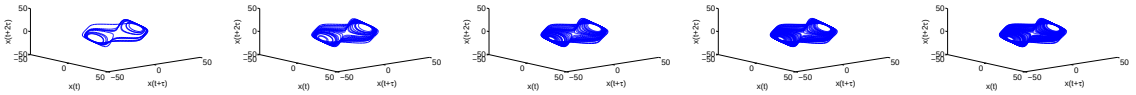
The proposed framework for representation of dynamics was evaluated on the following video-based inference tasks:

(1) Action recognition on a motion capture dataset [5].

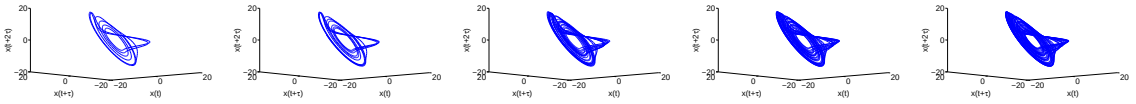
(2) Action recognition on the MSR Action3D dataset released by Microsoft Research [76].

Table 3.1: Experimental results on Lorenz and Rossler models for given embedding parameters ($m_L = 3$, $\tau_L = 11$, $m_R = 3$, $\tau_R = 8$) and different time-series lengths. The true value of λ_1 for Lorenz and Rossler models are 1.50 and 0.09 respectively [148].

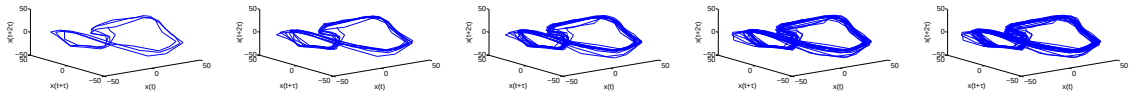
System	N	Calculated λ_1	% error
Lorenz	1000	1.751	16.7
	2000	1.345	-10.3
	3000	1.372	-8.5
	4000	1.392	-7.2
	5000	1.523	1.5
Rossler	400	0.0351	-61.0
	800	0.0655	-27.2
	1200	0.0918	2.0
	1600	0.0984	9.3
	2000	0.0879	-2.3



Reconstructed phase space of Lorenz system for different time-series lengths

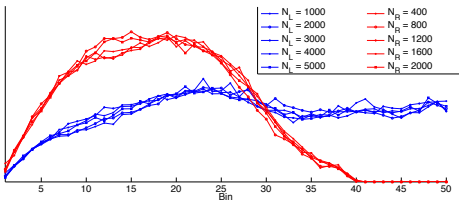


Reconstructed phase space of Rossler system for different time-series lengths

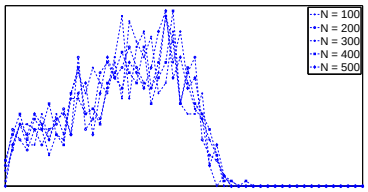


Reconstructed phase space of *Run* action for different time-series lengths

Figure 3.1: Illustration of the effect of time-series lengths on reconstructed phase space for nonlinear dynamical models like Lorenz and Rossler systems, and right-foot trajectory of a subject performing *Run* action. These examples clearly indicate that the *shape* of the reconstructed phase space does not change with time-series length, motivating feature extraction representative of the *shape* of the reconstructed phase space (as reported in Fig. 3.2).



(a) Shape distribution (**D2**) of reconstructed phase space for Lorenz (blue) and Rossler (red) models for different time-series length N (N_L and N_R represent time-series lengths of Lorenz and Rossler systems respectively).



(b) Shape distribution (**D2**) of reconstructed phase space from right-foot trajectory of a subject performing *Run* action for different time-series length.

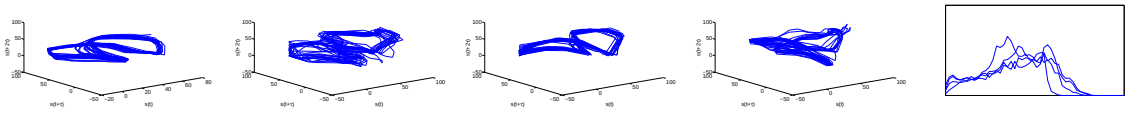
Figure 3.2: Illustration of stability of the dynamical shape distribution (**D2**) extracted from reconstructed phase space for different time-series length. (a) shows the stability of **D2** distribution on Lorenz and Rossler systems while studies have reported significant error in estimation of largest Lyapunov exponent on these models (refer TABLE 3.1). (b) depicts the stability of **D2** distribution for trajectory data collected from right-foot of a subject performing *Run* action.

(3) Action quality estimation on stroke rehabilitation datasets collected in hospital and home based environments [9, 30].

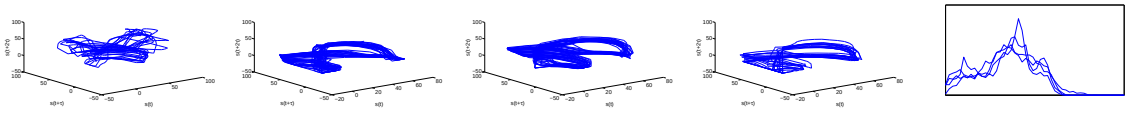
(4) Dynamic scene classification on the Maryland "in-the-wild" natural scene dataset [118] and the Yupenn "stabilized" scene dataset [36].

Baseline: The main contribution of our work is to propose a better way to encode dynamics compared to traditional chaotic invariants. To evaluate the effectiveness of our framework, we provide comparative results in each experiment with a feature vector [1] using traditional chaotic invariants obtained by concatenating largest Lyapunov exponent, correlation dimension and correlation integral (for 8 values of radius) resulting in a 10-dimensional feature vector denoted as ***Chaos***. For a fair comparison,

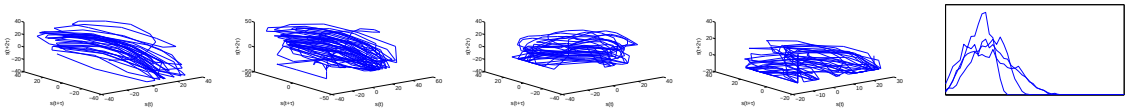
[1] Code available at
http://www.physik3.gwdg.de/tstool/HTML/index.html



Phase space reconstruction of RightLeg X-rotation time-series for 'Run' action Shape Distribution



Phase space reconstruction of RightLeg X-rotation time-series for 'Walk' action Shape Distribution



Phase space reconstruction of RightLeg X-rotation time-series for 'Dance' action Shape Distribution

Figure 3.3: Illustration of the phase space reconstruction and dynamical shape feature extraction (**D2** shape feature) using four examples of *Run*, *Walk* and *Dance* action classes each from the motion capture dataset [5]. As an example, phase space reconstruction of X-rotation time-series from right leg of subjects performing these actions is shown. Embedding parameters, m was selected to be 3 and τ was calculated by method explained in section 2.3.1. It is evident from these examples that the 'shape' of phase space is a representative feature for an action class and can be captured using shape distributions.

the embedding procedure is fixed as mentioned in earlier sections.

3.1.1 Motion Capture Dataset

In the first experiment, we evaluate the performance of the proposed framework using 3-dimensional motion capture sequences of body joints of subjects performing actions released by FutureLight, R&D division of Santa Monica Studios [5]. The dataset is a collection of five actions: *dance, jump, run, sit* and *walk* with $31, 14, 30, 35$ and 48 instances respectively. The classification problem on this dataset is shown to be challenging due to the presence of significant intra-class variations [5]. The data is in the form of trajectories of 3D rotation angles from 18 body joints. We use all body joints except the hip joint, to remove any effects of translational movement of the body. The 3D time-series from these 17 body joints were divided into scalar

time-series resulting in a 51-dimensional vector representation for each action. Phase space reconstruction and dynamical shape feature extraction was performed. The results of the leave-one-out cross-validation approach using a nearest neighbor classifier (using Euclidean and χ^2 distance metrics) are tabulated in TABLE 3.2. The best classification performance we achieved was a mean accuracy of 99.37% using **DT2** dynamical shape feature, in comparison with 89.7% reported by Ali *et al.* in [5] using traditional chaotic invariants. In addition, we see that the classification performance of each dynamical shape feature is significantly better than the results achieved by using traditional chaotic invariants (***Chaos*** with $m = 3$ & $m = 5$). The proposed action modeling framework achieves near-perfect classification accuracy on the motion capture dataset even in the presence of significant intra-class variations indicating its stability. This is also evident from the examples shown in Fig. 3.3, where minor variations in the reconstructed phase space (in the form of intra-class variations) has not produced any significant effect on the dynamical shape feature indicating the stability of the proposed framework. From these results, we see that the dynamical shape features with temporal evolution information (**DT1** and **DT2**) performs better than the shape features **D1**, **D2** and **D3**, hence substantiating our hypothesis that shape functions with dynamical evolution information should only improve the recognition performance.

3.1.2 Kinect Dataset

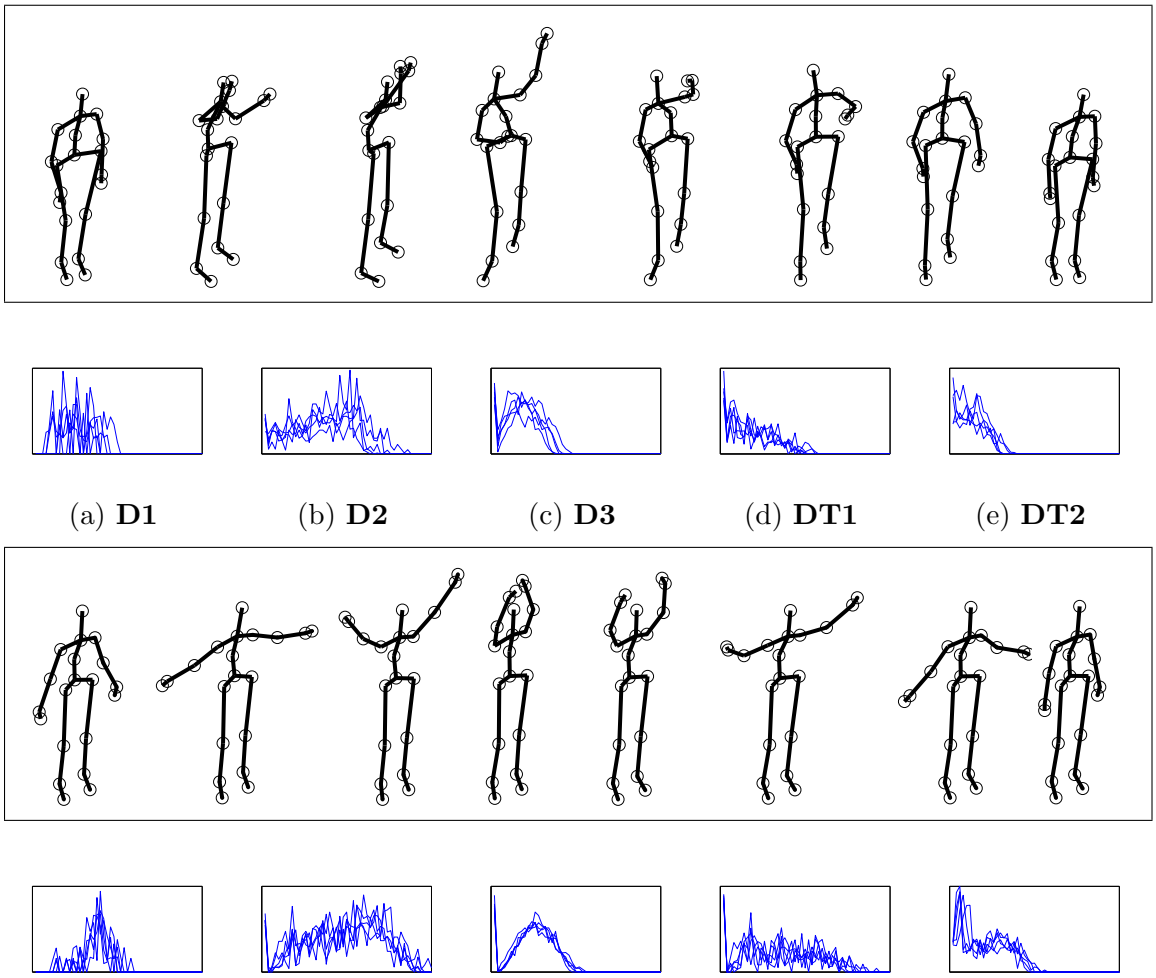
The framework was also evaluated on a more comprehensive dataset released by Microsoft Research called *MSR Action3D* dataset [76] having 20 action classes: *high arm wave, horizontal arm wave, hammer, hand catch, forward punch, high throw, draw x, draw tick, draw circle, hand clap, two hand wave, side boxing, bend, forward kick, side kick, jogging, tennis swing, tennis serve, golf swing, pick up & throw* with 10 subjects

Table 3.2: Classification rates for the various proposed dynamical shape features of phase space on the motion capture dataset. For comparison, we use Euclidean distance and chi-squared distance metrics as a measure of distance between probability distributions. We see that **DT2** achieves highest classification rate of 99.37%. The confusion table of the same is reported in TABLE 3.3.

Dynamical Shape Feature	**Distance Measure**	
	L_2	χ^2
Chaos ($m = 3$)	80.38	83.54
Chaos ($m = 5$)	82.28	85.54
Ali *et al.*	89.70	-
D1 ($m = 3$)	94.30	98.10
D2 ($m = 3$)	96.84	96.84
D3 ($m = 3$)	97.47	97.47
DT1 ($m = 3$)	97.47	98.73
DT2 ($m = 3$)	96.84	99.37

Table 3.3: Confusion table for motion capture dataset using **DT2** as the dynamical shape feature achieving mean classification rate of 99.37% when compared to 89.7% reported by Ali *et al.* in [5].

Action	**Dance**	**Jump**	**Run**	**Sit**	**Walk**
Dance	**30**	1	0	0	0
Jump	0	**14**	0	0	0
Run	0	0	**30**	0	0
Sit	0	0	0	**35**	0
Walk	0	0	0	0	**48**



(a) **D1** (b) **D2** (c) **D3** (d) **DT1** (e) **DT2**

(f) **D1** (g) **D2** (h) **D3** (i) **DT1** (j) **DT2**

Figure 3.4: Example actions from action class *Tennis serve* (a) and *Two hand wave* (b) from the MSR Action3D dataset. Skeleton data of 20 joints provided in the dataset will be used in our action recognition experiment. Shape distributions from reconstructed phase space using the hand trajectory from five instances each of tennis serve and two hand wave actions is shown here to illustrate the insensitivity of the framework to inter-class similarities.

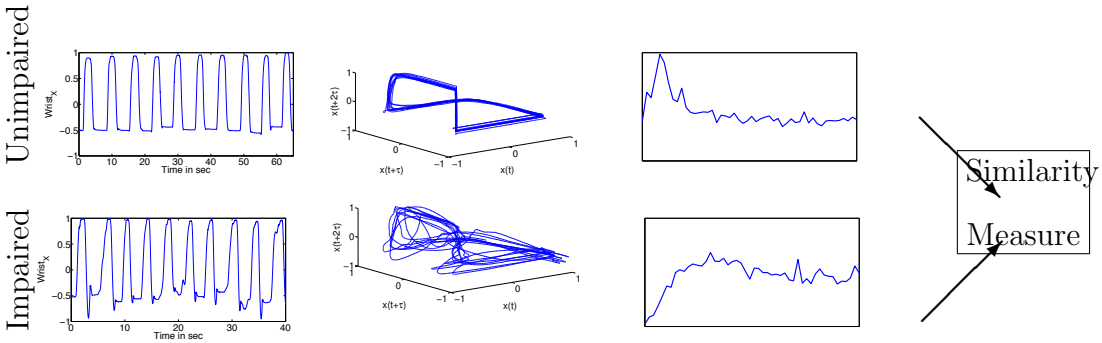
performing each action thrice (see Fig. 3.4 for example actions). The action classes in this dataset were selected to ensure the use of arms, legs and torso by subjects to simulate interaction with gaming consoles. High similarity between classes (e.g., forward punch and hammer, high throw and pickup & throw) makes this a challenging dataset. The 20 action classes were further divided into 3 Action Sets: *AS1*, *AS2* and *AS3* in [76] to account for the large amount of computation involved in classification of these actions. The action sets 1 and 2 were intended to group actions with similar movement and action set 3 to group complex movements. The dataset provides 3D joint positions on which phase space reconstruction and extraction of shape distribution were carried out individually on every dimension (x, y & z). These shape distributions were concatenated to form our feature vector representative of any given action. The classification results on the cross-subject test setting using a linear SVM are tabulated in TABLE 3.4 and as seen, the proposed framework performs better than the traditional chaotic invariants. Examples shown in Fig. 3.4 further support our hypothesis that shape distributions can be used as discriminative feature of reconstructed phase space representative of actions. In order to illustrate the proposed framework's stability to intra-class variations and insensitivity to inter-class similarities, we compare the dynamical shape features of hand trajectory for five instances of *tennis serve* and *two hand wave* action classes. Evident from these examples is that even actions using similar hand movements are represented by dynamical shape features with enough differences to successfully recognize these actions. Furthermore, from results in TABLE 3.4, we see that the dynamical shape feature **DT2** has the highest overall classification accuracy, indicating that the shape distribution based on temporal evolution of phase space is better than traditional global shape representations. We have also provided classification results using a nearest neighbor classifier in TABLE 3.5 for a comprehensive comparison of the proposed shape distributions.

Table 3.4: Classification results for cross-subject test setting where 50% subjects were used for training and the remaining 50% subjects for testing in proposed method using linear SVM.

	Shape Distribution $(m = 3)$					***Chaos***	
Set	**D1**	**D2**	**D3**	**DT1**	**DT2**	$m = 3$	$m = 5$
AS1	88.35	89.32	87.13	88.57	90.48	72.28	74.56
AS2	69.72	72.65	71.43	73.21	74.11	51.85	52.40
AS3	90.74	96.40	98.20	98.25	99.09	76.36	78.86
Avg.	**82.94**	**86.12**	**85.59**	**86.68**	**87.89**	66.83	68.61

Table 3.5: Classification results for cross-subject test setting where 50% subjects were used for training and the remaining 50% subjects for testing in proposed method using nearest-neighbor classifier.

	Shape Distribution $(m = 3)$					***Chaos***	
Set	**D1**	**D2**	**D3**	**DT1**	**DT2**	$m = 3$	$m = 5$
AS1	67.00	74.62	75.73	75.05	78.43	52.30	55.67
AS2	59.63	67.66	65.77	64.47	68.21	42.53	49.23
AS3	87.83	89.96	89.66	88.11	91.13	53.45	60.59
Avg.	**71.49**	**77.41**	**77.05**	**75.87**	**79.25**	49.43	55.16



(a) Time-series data (b) Reconstructed phase space (c) Shape distribution

Figure 3.5: Proposed framework for movement quality assessment and action recognition by extraction of dynamical shape feature from reconstructed phase space. (a) shows the time-series of x-location of wrist marker; its respective reconstructed phase space is shown in (b). These two exemplar trajectories are collected from the stroke rehabilitation dataset [30] and belong to unimpaired and impaired subjects respectively. The corresponding dynamical shape feature represented by shape distribution is shown in (c). Similarity measure (e.g., Euclidean distance) can be used to classify these trajectories.

3.1.3 Activity Quality for Stroke Rehabilitation

Our aim in this experiment is two-fold: a) to classify movements of unimpaired (neurologically normal) and impaired (stroke survivors) subjects, b) to quantitatively assess the quality of movement performed by the impaired subjects during repetitive task therapy. Fig. 3.5 illustrates the differences in shape of reconstructed phase space between unimpaired and impaired subjects using trajectories from the wrist marker (reflective marker placed on the subject's wrist). The experimental data was collected using a heavy marker-based system (14 markers on the right hand, arm and torso) in a hospital setting. Seven unimpaired and 15 impaired subjects perform multiple repetitions of reach and grasp movements, both on-table and elevated (the subject must move against gravity to reach the target). Each subject would perform 4 sets of reach and grasp movements to different target locations, with each set having 10 repetitions. To account for a small number of training examples, we adopt leave-one-reach-out cross validation scheme where one set of reach movement was used as testing example and rest as training examples. The stroke survivors were also evaluated by the Wolf Motor Function Test (WMFT) [149] on the day of recording, which evaluates the subject's functional ability on a scale of $1-5$ (with 5 being least impaired and 1 being most impaired) based on predefined functional tasks. Since our focus is on development of quantitative measures of movement quality for a home-based rehabilitation system that would use a single marker on the wrist, we only use the data corresponding to the single marker on the wrist from the heavy marker-based hospital system.

The focus of traditional methods for quantitative assessment of movement quality has been towards kinematics. Hence, in TABLE 3.6, we compare our results with an approach which uses kinematic analysis on the same dataset [30]. We also compare

Table 3.6: Comparison of classification rates for different methods using leave-one-reach-out cross-validation and nearest neighbor classifier on the stroke rehabilitation dataset.

Method	**Classification Rate (%)**
KIM [30]	85.2
Chaos ($m = 3$)	81.82
Chaos ($m = 5$)	83.43
D1 ($m = 3$)	**84.32**
D2 ($m = 3$)	**88.60**
D3 ($m = 3$)	**86.04**
DT1 ($m = 3$)	**87.65**
DT2 ($m = 3$)	**92.05**

our results with the performance of traditional chaotic invariants. It is evident from these results that our framework performs better than the two promising quantitative measures for movement analysis in the field of stroke rehabilitation.

We also propose a framework for movement quality assessment (shown in Fig. 3.6) for stroke rehabilitation. Using the WMFT scores of impaired subjects, we learn a regression function using SVM to compute a movement quality score from dynamical shape feature (using **D2** shape distribution). The regressor was trained using leave-one-reach-out cross-validation technique. The outputs of the regressor were averaged per subject to get the Movement Quality Score (MQS). Fig. 3.7 shows a comparison between the actual WMFT score and the quality assessment score by the proposed method (MQS). The Pearson correlation coefficient between the MQS and the Function Activity Score (FAS) of the WMFT was found to be 0.8527. When we repeat the same experiment with kinematic attributes on a single wrist marker, the correlation coefficient was found to be 0.6481. In comparison, kinematic

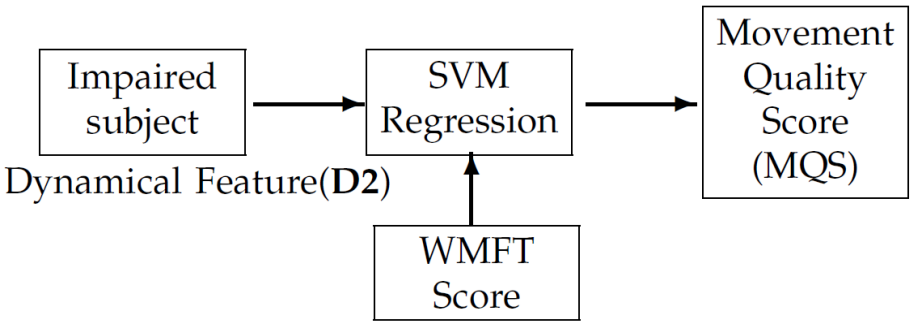


Figure 3.6: Block diagram representation for learning a regressor for movement quality assessment using Functional Activity Score (FAS) from the Wolf Motor Function Test (WMFT).

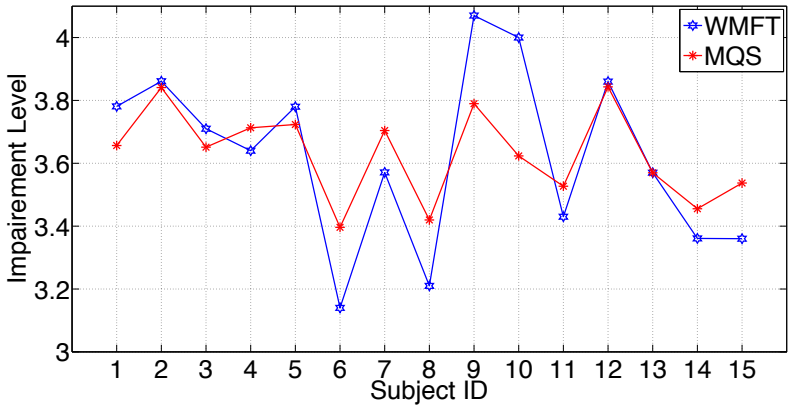


Figure 3.7: Comparison between impairment level (with 5 being least impaired and 1 being most impaired) given by actual WMFT score and MQS for 15 impaired subjects. The Pearson correlation coefficient was found to be 0.8527 with a two-tail P-value of 5.35×10^{-5}, proving its statistical significance.

analysis of data from all 14 markers gave a correlation coefficient of 0.9041. This experiment clearly shows that the proposed framework achieves comparable results obtained by the heavy marker-based system even when using a single wrist marker, which is facilitated by the phase space reconstruction and robust feature extraction from phase space using shape distribution.

The WMFT scores are based on several functional tasks (e.g., folding a towel, picking up a pencil) and not on evaluation of the actual movements during repetitive

therapy treatment (reach and grasp movements). In the above experiment, we utilize these WMFT scores as an approximate high-level quantitative measure for movement quality of impaired subjects performing reach and grasp movements, as both WMFT evaluation and 3D marker data on the wrist were obtained on the same day.

To address this conflict in collection of ground truth (movement quality labels) and trajectory data, we have collected a dataset from eight stroke survivors performing reach and grasp movement tasks and have developed a rating scale for movement quality in collaboration with physical therapists. Within this scale, physical therapists would provide us an overall rating on a scale of $1-5$ based on the therapist's impression of the participant's performance. A score of 1 denotes that the participant could not complete the task (most impaired) and a 5 denotes that the participant performed the task with the same quality of performance as the therapist if he/she were to perform it (least impaired or unimpaired). We have collected both 3D position of the wrist and physical therapist ratings in order to make comparisons among the kinematics, our proposed measure, and the therapist ratings, across the same reach action. Utilizing the expert knowledge of the therapist ratings for these rated actions will also help us better contextualize the data to better shape our framework as a therapy tool. Using the same framework for regression as earlier, we see from TABLE 3.7 that the proposed framework (using **DT2**) performs better than the traditional methods for movement quality assessment in terms of correlation coefficient and mean squared error. It should be noted that the proposed framework does not require data collected from unimpaired subjects for generating MQS, while kinematic methods like KIM [30] does, making the framework more suitable to model complex tasks during therapy treatment.

Table 3.7: Comparison of performance of the proposed dynamical shape features with the performance of traditional methods used for movement quality analysis.

Method	Correlation Coefficient	MSE
KIM [30]	0.4918	0.0066
Chaos ($m = 3$)	0.4717	0.0101
Chaos ($m = 5$)	0.5089	0.0100
D1 ($m = 3$)	0.3877	0.1190
D2 ($m = 3$)	0.5029	0.0078
D3 ($m = 3$)	0.4935	0.0061
DT1 ($m = 3$)	0.4582	0.0100
DT2 ($m = 3$)	0.5510	0.0057

3.1.4 Dynamic Scene Recognition

Natural dynamic scene recognition has been gaining interest in recent years [118, 36]. In an attempt to test the generality of the proposed framework to dynamical modeling for applications in video analysis, we evaluate its performance on dynamical scene classification. In this experiment, we use the Maryland "in-the-wild" dataset [118] which is a collection of 13 classes with 10 examples per class and a larger Yupenn stabilized dynamic dataset [36] which is a collection of 14 classes with 30 examples per class. The former has videos collected from video hosting websites with no control over recording process leading to a dataset with large variations in illumination, view and scale [118]. The latter dataset was recently released to emphasize only the scene-specific temporal information rather than camera-induced ones. In addition, the scene classes in the datasets were selected to illustrate potential failure of static scene representations leading to confusion between classes (e.g., chaotic traffic and smooth traffic).

Recent research on dynamical modeling of scenes have shown that temporal (motion) information can provide better classification performance than traditional feature representations (e.g., GIST [89]) on static scenes [118, 36]. The GIST feature is based on the hypothesis that humans recognize scenes by holistic understanding of a scene [89, 14], thereby providing a global spatial representation of a scene. Shroff *et al.* employed traditional chaotic invariants to model the dynamics in the time-series of the 960-dimensional GIST descriptor extracted from each video and will be treated as our baseline. Similarly, we compare the performance of our proposed shape distribution features estimated on the 960-dimensional GIST descriptor to further support our hypothesis that proposed shape-based features can perform better than traditional chaotic invariants in video-based inference tasks.

The average classification accuracy for all the proposed dynamical shape features in comparison with traditional chaotic invariants using a nearest neighbor classifier are tabulated in TABLE 3.8 and 3.9. It is evident from these results that the proposed dynamical shape features (**D2** and **DT2**) perform better than the traditional chaotic invariants used in literature for dynamical scene classification. Evidently it is possible to improve classification performance further by fusion of dynamical and spatial features as in [118], but here we restrict ourselves to comparison with core dynamical approaches.

3.2 Conclusion and Future Work

In this work, we have proposed a shape theoretic dynamical analysis framework for applications in action and gesture recognition, movement quality assessment for stroke rehabilitation and dynamical scene classification. We address the drawbacks of tradi-

[2]Here "our" refers to our implementation of traditional chaotic invariants using the OpenTSTOOL package.

Figure 3.8: Dynamic scene "in-the-wild" dataset consisting of 13 scene classes with 10 examples per class [118]. Sample video frames from scene classes (left-to-right, top-to-bottom) avalanche, boiling water, chaotic traffic, forest fire, fountain, iceberg collapse, landslide, smooth traffic, tornado, volcanic eruption, waterfall, waves and whirlpool are shown here. This dataset has large intra-class variations with significant changes in illumination and scale.

Table 3.8: Comparison of classification rates for various approaches on the Maryland "in-the-wild" dataset (with $m = 3$).

Class	**Chaos** [118]	**Chaos** (our) [2]	**D1**	**D2**	**D3**	**DT1**	**DT2**
avalanche	30	40	0	0	20	10	0
b. water	30	40	30	40	20	30	30
c. traffic	50	30	80	100	50	60	90
f. fire	30	20	10	30	30	30	30
fountain	20	0	40	30	30	30	40
i. collapse	10	0	10	0	0	10	0
landslide	10	50	0	10	20	10	20
s. traffic	20	20	20	30	30	40	30
tornado	60	10	40	70	60	50	60
v. eruption	70	0	60	70	60	40	70
waterfall	30	20	10	40	20	20	30
waves	80	40	70	80	80	90	80
whirlpool	30	20	40	50	30	70	50
Avg. (%)	36	22.31	31.54	42.31	34.62	37.69	40.77

Table 3.9: Comparison of classification rates for various approaches on the Yupenn "stabilized" dynamic dataset (with $m = 3$).

Class	**Chaos** [118]	**Chaos** (our)[2]	**D1**	**D2**	**D3**	**DT1**	**DT2**
beach	27	17	77	80	77	83	77
c. street	17	70	3	87	90	100	93
elevator	40	17	7	37	10	23	17
f. fire	50	10	40	50	57	40	50
fountain	7	10	0	27	17	47	0
highway	17	17	77	47	53	33	60
l. storm	37	97	97	97	93	97	100
ocean	43	30	60	70	80	87	77
railway	3	17	60	57	23	40	60
r. river	3	87	60	90	83	87	77
sky	33	23	30	47	43	50	57
snowing	10	77	73	80	90	90	93
waterfall	10	17	50	37	30	37	37
w. farm	17	03	30	13	20	10	33
Avg. (%)	22.43	35.14	48.64	58.50	54.71	58.85	59.35

tional measures from chaos theory for modeling the dynamics by proposing a framework combining the concepts of nonlinear time-series analysis and shape theory to extract robust and discriminative features from the reconstructed phase space. Our experiments on nonlinear dynamical models and joint trajectory data from motion capture support our hypothesis that the *shape* of the reconstructed phase space can be used as feature representation for the above discussed applications. Furthermore, the wide range of experimental analysis on publicly available datasets for recognition of actions, gestures and scenes validate our claims. The framework was also tested on movement analysis on a finer scale, where we were interested in quantifying the *movement quality* (level of impairment) for applications in stroke rehabilitation. Our experiments using a single marker indicate that with combination of dynamical features and machine learning tools, we are able to achieve comparable performance levels to a heavy marker-based system in movement quality assessment.

In this work, we perform phase space reconstruction on every dimension independently (univariate phase space reconstruction). Our future directions will be towards

employing techniques for multi-variate phase space reconstruction [23]. It has been shown in [12] that multi-variate phase space reconstruction method provides better modeling than univariate phase space reconstruction, and hence lower error in predictions for human motion. We would also like to explore the use of approximate entropy [99], a dynamical measure quantifying regularity in a time-series. The suggested number of data samples required for computation of approximate entropy is between 50 and 5000 [99], which makes it more a suitable feature representation for applications in video-based inferences.

4 KINEMATIC ANALYSIS FOR STROKE REHABILITATION

Stroke is the most common neurological disorder worldwide [79] leaving behind a significant number of survivors every year disabled with chronic impairments such as problems with vision, difficulty to formulate or understand speech, or inability to move limbs. Even with persistent efforts to lower blood pressure and reduce smoking, the incidence of stroke remain high due to the ageing population, with nearly three-quarters of stroke related events experienced by people over the age of 65 [53, 107]. This increasing demand for rehabilitation facilities has been seen as a significant healthcare problem worldwide [91, 84]. In addition, studies indicate that the increasing healthcare costs paired with insufficient coverage by insurance for long-term therapy treatment has often left impairments untreated [6]. Hence, it is important to have well-thought-out strategies to manage these stroke survivors by providing low-cost long-term rehabilitation therapy for their recovery.

Traditional rehabilitation therapy is usually composed of repetitive movement tasks such as reaching and grasping an object. A participant performs these movement tasks in a hospital under the supervision of a physical therapist, who visually monitors the quality of movement over time to provide personalized rehabilitation therapy. This laborious and expensive process has motivated researchers to invent novel strategies to accelerate hospital discharge without compromising on clinical outcomes.

Challenges in Developing Component-level Kinematic Features: Thera-

pists are trained to assess the overall performance of a task, which can also be achieved through existing validated clinical measures such as the Wolf Motor Function Test (WMFT) [149] and the Fugl-Meyer Assessment (FMA) [49]. Such clinical measures do not provide enough information about the component-level impairments, which will be useful in providing focused rehabilitation. The motivation of our research was to develop a computational framework for component-level tuning of kinematic features such as trajectory error, speed profile deviation, jerkiness, and segmentation using the composite (overall) therapist impressions of movement quality to drive the feedback module in the HAMRR system.

One recurring problem in the stroke rehabilitation community is the general lack of consensus among physical therapists in defining an ontology of component level labels for movement quality, thereby leading to lack of training datasets to develop algorithms for movement quality assessment. In addition, therapists only provide composite assessments indicative of quality of overall movement without any information about components such as deviation in speed profile, leading to a challenging problem to train the component-level kinematic features, which are required to provide personalized rehabilitation and facilitate active learning without therapist supervision. An illustration of the above concept is shown in Fig. 4.1, where the aim is to induce active learning by providing auditory and visual feedback implying the impairments in low-level components such as trajectory inaccuracy, tremor, and segmentation [10]. In Fig. 4.1, (a) and (c) represent the visual feedback seen during an efficient reach (reach trajectory without any impairments) marked by a straight path of rocks or a complete boat, while (b) represents a reach with trajectory error on the right marked by curved path of rocks (in red), and (d) represents a reach with segmentation error marked by a broken boat.

Towards Home-based Rehabilitation Systems: Clinical intervention alone

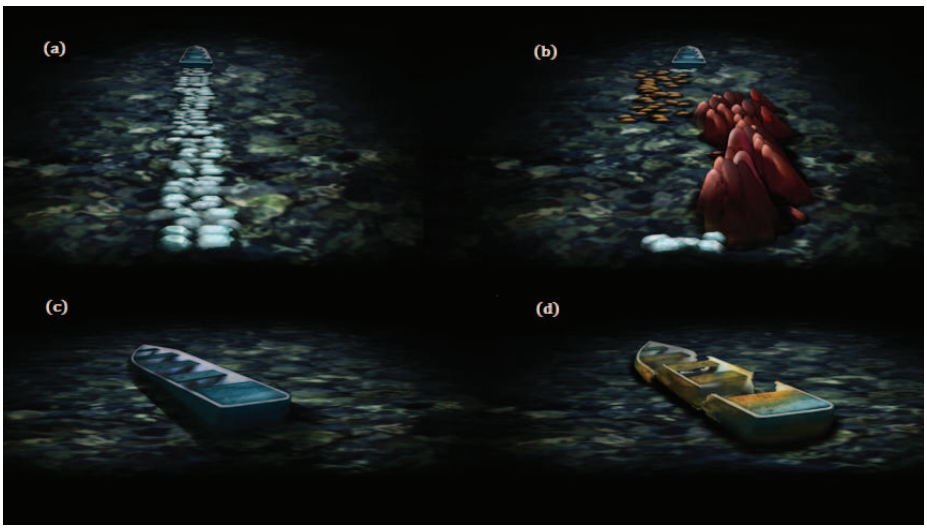


Figure 4.1: Exemplar visual feedback summaries based on low-level kinematic analysis. (a) represents an efficient reach, (b) represents trajectory error to the right. (c) is a representation of an efficient and consistent task completion and (d) represents segmented movement.

is not completely effective for restoring daily activity functionality in a stroke survivor [72, 130, 34, 40]. A comprehensive study involving 1277 stroke survivors has reported that an early hospital discharge and home-based rehabilitation strategy resulted in reduced length of stay by 13 days, and overall mean costs being 15% lower compared to standard care, without any significant effect on mortality or clinical outcomes [7]. A similar long-term study has reported significant reduction in hospital stay without any change in health outcomes in stroke survivors who experienced home-based rehabilitation compared to traditional rehabilitation care [8].

Interactive neurorehabilitation systems which computationally evaluate and deliver feedback based on a subject's movement performance have been utilized to provide home-based rehabilitation care. With advances in 3D motion capture and wearable sensor technology, researchers from various backgrounds have developed objective measures for movement quality assessment during and following rehabilitation [30, 142, 127, 156, 28]. Virtual and mixed reality environments have been employed in novel stroke rehabilitation strategies [113, 82, 73, 74]. In this direction, Adaptive

Figure 4.2: The Home-based Adaptive Mixed Reality Rehabilitation (HAMRR) system designed for stroke survivors. The system uses four OptiTrack cameras to track the wrist movements as well as a computer and speakers to provide audio and visual feedback during therapy treatment. The table is designed to accommodate custom touch and grasp objects for training reaches in different orientations. In the inset, we see the placement of a wrist marker on a participant performing reaching tasks to a cone. The system design is discussed in detail in [10].

Mixed Reality Rehabilitation (AMRR) system which integrates rehabilitation and motor learning theories with motion capture, activity analysis, and multimedia feedback [31, 29], has been shown as an effective rehabilitation system in helping improve the kinematic and functional performance of a stroke survivor's upper extremity in a hospital setting. Examples of visual feedback for active learning using the home system are shown in Fig. 4.1. In addition, accommodating heavy marker-based systems in a home-based setting is unrealistic, as inaccurate placement of markers can negatively affect the movement quality assessment framework and place a heavy burden on the stroke survivor and/or caregiver. In recent years, the focus of rehabilitation research has been towards devising multi-modal interventions and accompanying tools to assist home-based therapy [121, 10, 31], thereby supplementing traditional therapy received in the hospital. A solution to this was proposed in [10], where a single reflective marker was placed on the participant's wrist to track the movement (see Fig. 4.2). A recent study has shown that a single marker-based system (marker on the wrist) can achieve comparable performance levels of movement quality assessment to a heavy marker-based system [142].

In this work, our aim is to use the composite labels provided by therapists' impressions to learn the underlying movement components. We propose several kinematic features and learn the associated thresholds and weights using composite labels for reach data. This research facilitates better understanding of the underlying components defining movement quality and also the generation of a '*cumulative score*' for movement quality, which can aid physical therapists in visual monitoring during supervised rehabilitation therapy.

Contributions: Our aim is to decompose the movement quality score (given by therapists) into its constituent kinematic components. We assume a linear relation between kinematic features and composite movement quality score. This work has

two main contributions: 1) propose component-level kinematic features for movement quality assessment of wrist movement, 2) propose a generic framework for tuning the thresholds and weights associated with each of these kinematic features using movement quality labels provided by therapists.

4.1 Related Work

Quantifying movement quality is useful for physical therapists to provide improved and personalized rehabilitation therapy. Several quantitative scales for movement quality assessment have been proposed, including the FMA [49] and the WMFT [149]. For example, the WMFT has been used to quantify the upper extremity motor ability through timed and functional tasks [83]. However, these methods rely on visual monitoring of movements by experienced and trained physical therapists. Hence, these methods can be subjective, as a therapist will apply their individual training and impressions when evaluating a participant's movement quality. Developing an objective computational framework for movement quality assessment will be beneficial, thereby minimizing the influence of a therapist.

The focus of existing approaches for movement quality assessment has been towards finding typical patterns in kinematic attributes which differ between healthy and impaired participants. Kinematic Impairment Measure (KIM) proposed by Chen *et al.* [30] employs 33 kinematic attributes derived from a heavy-marker based system in a hospital setting to quantitatively evaluate the movement quality. This study showed that the weighted average of individual kinematic attributes was strongly correlated with the WMFT scores. Similar work using kinematics to model the smoothness of the movement have also been explored [48, 61]. In a similar study, it was shown that features derived from wearable sensor data can be used to estimate the FMA score [35].

Rehabilitation robotics has gained a lot of attention in quantification of motor functionality due to its ability to offer objective and repeatable therapy treatment [11, 137, 44, 32, 106, 105, 71]. Linear regression model-based kinematic scales were developed using the MIT-Manus robot to achieve highly a repeatable and high resolution framework for quantification of motor performance [18]. Another robotics-based rehabilitation technique proposed four measures showing correlation with clinical measures such as FMA, MAL, Action Research Arm Test, and Jebsen-Taylor Hand Function Test [27]. A recent work using movement time, trajectory length, directness, smoothness, and mean and maximum velocity claims that such kinematic features can be effectively used to assess upper limb motor recovery and is linked to FMA score [138].

Nonlinear dynamical analysis methods have been employed to model the variability in repetitive movements, which are an integral part of rehabilitation therapy [95, 127]. To address the drawbacks of traditional nonlinear dynamical measures, a shape theory based dynamical analysis framework for movement quality assessment was proposed [142]. This study also demonstrated that the information contained in a single marker on the wrist is sufficient to achieve comparable performance levels to a heavy marker-based system in movement quality assessment.

4.2 System Design

The HAMRR system has four Natural Point Opti-Track cameras facing down on a table to track a single reflective marker placed on the participant's wrist (wrist marker). The selection of the wrist marker was motivated by previous investigations indicating that the wrist trajectory is the most informative joint about the reach trajectory [142, 30, 73, 102, 151]. The system also tracks torso movement using four reflective markers attached to a badge worn on the left side of the participant's chest.

Effective upper extremity rehabilitation requires monitoring of such aspects of the body movement to evaluate the extent participant's compensation while performing a task. In this study, we focus solely on the data collected from the wrist marker.

The table houses a contact switch rest position pad and can accommodate a target location of the cone object based on the participant's reaching ability. While we only consider reaching tasks for the cone object located on the left of a participant, the system was designed to accommodate custom touch and grasp objects for training reaches in different orientations. The system is shown in Fig. 4.2 and detailed information of the system design can be found in [10]. The main objective of this design was to be able to install the system in a participant's home for long term therapy treatment, which prohibits the use of a heavy marker-based system.

4.3 Data Collection

Therapists undergo training to assess both the overall performance of a task and monitor some individual coarse aspects of movement for a set of reaches. While validated clinical measures exist for assessing overall task performance, no such measures currently relate these to performance of component-level kinematic attributes for an individual reach. Therefore, in this study we have collected therapist ratings for quality of wrist trajectory for each reach in an attempt to build a computationally generated component-level assessment that correlates with therapist impressions.

The dataset consists of reaching tasks performed by a total of ten participants (refer to Table 5.2 for demographics) to an on-table cone left of the participant's rest position. Each participant performs five reaches in each of four sessions. An iPad application was developed to assist therapists in administering the system experience questionnaire, recording videos of reaching tasks, and providing movement quality labels. These videos were later segmented to contain individual reaches, which were

randomized across participants and provided to two physical therapists (each therapist would rate a reach movement which was not repeated by the other therapist) to rate each reach in terms of overall performance of the task. Overall reaching performance was rated on a scale from 1-5 based on the therapist's impression of the participant's performance, where a 1 denotes that the participant could not complete the task and a 5 denotes that the participant performed the task with the same quality of performance as the therapist if he/she were to perform it. This rating scale was adapted from the WMFT Functional Assessment Score [149] by rehabilitation experts who collectively created a rubric for the purposes of this study.

4.3.1 Trajectory Error

Trajectory error is a measure of spatial deviation of the wrist trajectory from the reference trajectory. For every point in the reach trajectory, horizontal error (E_{hor}) and vertical error (E_{vert}) were defined as

$$E_{hor}(i) = \mathbf{x}(i) - \mathbf{x}_{ref}(i), \quad i = 0, \ldots, N_s - 1 \tag{4.1a}$$

$$E_{vert}(i) = \mathbf{y}(i) - \mathbf{y}_{ref}(i), \quad i = 0, \ldots, N_s - 1 \tag{4.1b}$$

where N_s is the number of points in the reach trajectory. A thresholded error function was calculated as

$$\hat{E}_{hor}(i) = \begin{cases} E_{hor}(i) & \text{if } E_{hor}(i) > T_1 \\ 0 & \text{otherwise.} \end{cases} \tag{4.1c}$$

Similarly,

$$\hat{E}_{vert}(i) = \begin{cases} E_{vert}(i) & \text{if } E_{vert}(i) > T_1 \\ 0 & \text{otherwise.} \end{cases} \tag{4.1d}$$

Confidence values for the movement being curved were estimated as

$$C_{\mathbf{x}}^{curved} = \frac{\sum\limits_{<i>} |\hat{E}_{hor}(i)|}{\sum\limits_{<i>} |E_{hor}(i)|} \tag{4.1e}$$

$$C_{\mathbf{y}}^{curved} = \frac{\sum\limits_{<i>} |\hat{E}_{vert}(i)|}{\sum\limits_{<i>} |E_{vert}(i)|} \tag{4.1f}$$

The final confidence of curved movement was a combination of the above two confidences,

$$C_{T_1}^{curved} = \begin{cases} \lambda_1 & \text{if } \lambda_1 > 2\lambda_2 \\ min(1.5\lambda_1, 1) & \text{otherwise} \end{cases} \tag{4.1g}$$

$$\text{where } \lambda_1 = 1 - max(C_{\mathbf{x}}^{curved}, C_{\mathbf{y}}^{curved}),$$
$$\lambda_2 = 1 - min(C_{\mathbf{x}}^{curved}, C_{\mathbf{y}}^{curved}).$$

4.3.2 Speed Profile Deviation

It is a measure of deviation of the speed profile from the reference speed profile (speed profiles collected from 10 unimpaired participants to generate a reference). For a given reach trajectory, a point-to-point comparison of speeds with the reference speed profile was calculated. The speed vector for the reference and test data are denoted as $v_{ref}(i)$ and $v(i)$ respectively and was calculated as the first derivative of the position vector. The thresholded speed vector for *fastness* feature was calculated as

$$\hat{v}_f(i) = \begin{cases} v(i) & \text{if } v(i) - v_{ref}(i) > T_2 \\ 0 & \text{otherwise} \end{cases} \tag{4.2a}$$

The confidence score for movement being too-fast was computed as C^{fast} given by

$$C_{T_2}^{fast} = 1 - \frac{\sum\limits_{<i>} \hat{v}_f(i)}{\sum\limits_{<i>} v(i)} \tag{4.2b}$$

Similarly, the thresholded speed vector for *slowness* feature is given by

$$\hat{v}_s(i) = \begin{cases} v(i) & \text{if } v(i) - v_{ref}(i) < T_3 \\ 0 & \text{otherwise} \end{cases} \tag{4.2c}$$

The confidence score for movement being too-slow was calculated as C^{slow} given by

$$C_{T_3}^{slow} = 1 - \frac{\sum\limits_{<i>} \hat{v}_s(i)}{\sum\limits_{<i>} v(i)} \tag{4.2d}$$

4.3.3 Jerkiness

The jerkiness (or smoothness) feature is a measure of variations in the velocity profile. An '*efficient*' reach movement should have a smooth velocity profile with an accelerating pattern followed by a decelerating pattern without any jerks. Jerkiness of a movement was calculated using the method described in [30] (similar to [48]) and is given by

$$J = \int_{t_{som}}^{t_{eom}} \sqrt{\left(\frac{d^3\mathbf{x}}{dt^3}\right)^2 + \left(\frac{d^3\mathbf{y}}{dt^3}\right)^2 + \left(\frac{d^3\mathbf{z}}{dt^3}\right)^2}\, dt \tag{4.3a}$$

where $\mathbf{x}, \mathbf{y}$ and $\mathbf{z}$ are 3-D coordinates of the position of participant's wrist. t_{som} is the time index corresponding to the start of the movement and t_{eom} is the time index of the end of the movement. The thresholded jerkiness function was calculated as

$$\hat{J}(i) = \begin{cases} J(i) & \text{if } J(i) > T_4 \\ 0 & \text{otherwise} \end{cases} \tag{4.3b}$$

The confidence score for movement being jerky was calculated as

$$C_{T_4}^{jerk} = 1 - \frac{\sum\limits_{<i>} \hat{J}(i)}{\sum\limits_{<i>} J(i)} \tag{4.3c}$$

4.3.4 Segmentation

A movement is termed as '*segmented*' if the elbow does not open in synchrony with the shoulder moving forward. Instead, the forward movement of the shoulder and the opening of the elbow happens in sequence, resulting in a disjointed movement (or presence of submovements). Rohrer *et al.* [108] have shown how paretic movement can be represented by submovements using MIT-MANUS and InMotion2 robots, which allows motion within a horizontal plane. An accurate analysis of this phenomenon (presence of submovements) requires tracking of both shoulder and elbow in addition to the wrist. In the proposed home-based rehabilitation system, this was not possible with the one marker sensing solution, and we wanted to learn if such movements can be described computationally using only the wrist marker.

After consultation with domain experts, it was found that segmented movements give rise to notches (sudden change in direction) in the wrist trajectory. These notches can be quite subtle and often occur towards the end of the movement. We quantify segmented movements by calculating the following:

1. The number of times the movement changes its turning direction
2. The magnitude of direction change
3. The ratio of the magnitude of direction change

We project the 3D trajectory onto the **X**-**Z** and **Y**-**Z** planes to detect the direction changes (notches). In the projection onto the **X**-**Z** plane, we first compute displacement vectors from the spatial locations. The direction change was quantified as the

signed angle ($\alpha_{\mathbf{xz}}(i)$) between successive displacement vectors. The sign of the angle is positive if the displacement is clockwise from the previous displacement vector and negative if it is counter-clockwise. Using this, the number of significant changes in turning direction of the movement is calculated (N_C), and the corresponding confidence is calculated as

$$C_{seg1,\mathbf{xz}} = \begin{cases} 1 - e^{-(a \cdot N_C)^b} & \text{if } N_C > N_{ref} \\ 0 & \text{otherwise} \end{cases} \tag{4.4a}$$

The magnitude of direction change is computed as $S = \sum_{<i>} |\alpha_{\mathbf{xz}}(i)|$, and the corresponding confidence score was given by

$$C_{seg2,\mathbf{xz}} = 1 - e^{-(a \cdot \lambda_S)^b} \tag{4.4b}$$

$$\lambda_S = \begin{cases} 1 - S/ref_{\mathbf{xz}} & \text{if } S < ref_{\mathbf{xz}} \\ 0 & \text{otherwise} \end{cases} \tag{4.4c}$$

The ratio of magnitude of direction change is defined as $\gamma = \frac{|\sum \alpha_{\mathbf{xz}}(i)|}{\sum |\alpha_{\mathbf{xz}}(i)|}$, and the corresponding confidence score was computed as

$$C_{seg3,\mathbf{xz}} = \begin{cases} 1 & \text{if } \gamma < \gamma_{ref} \\ 1.47 * (1 - \gamma) & \text{otherwise} \end{cases} \tag{4.4d}$$

The final confidence for segmentation of the projected movement on $\mathbf{X} - \mathbf{Z}$ plane is computed as

$$C_{\mathbf{xz}} = C_{seg1,\mathbf{xz}} \cdot C_{seg2,\mathbf{xz}} \cdot C_{seg3,\mathbf{xz}} \tag{4.4e}$$

Similarly, we can compute $C_{\mathbf{yz}}$ in the $\mathbf{Y}$-$\mathbf{Z}$ plane. Let $\beta_1 = 1 - max(C_{\mathbf{xz}}, C_{\mathbf{yz}}), \beta_2 =$

$1 - min(C_{\mathbf{xz}}, C_{\mathbf{yz}})$. The final confidence of segmented movement is given by

$$C_{T_5}^{seg} = \begin{cases} \beta_1 & \text{if } \beta_1/\beta_2 > T_5 \\ min(1.5\beta_1, 1) & \text{otherwise} \end{cases} \tag{4.4f}$$

The thresholds $T_1, \ldots, T_5$ were difficult to define and hence optimal values for these thresholds was estimated using movement quality label provided by therapist. Thresholds such as $N_{ref}, ref_{\mathbf{xz}}, \gamma_{ref}$ were determined from the data collected from unimpaired participants. The constants a and b were selected through empirical observation. The confidence scores range from 0 to 1, with 0 indicating maximum impairment and 1 indicating movement being similar to an unimpaired participant's reach.

4.3.5 Estimation of Optimal Weights and Thresholds

A physical therapist rating the quality of reach trajectory will pay careful attention to many kinematic attributes, including speed, trajectory and jerkiness. We believe that a linear combination model of the non-linear kinematic features will be correlated with the therapist rating. In this work, we propose a linear model of kinematic features for movement quality assessment by posing an optimization problem to determine the thresholds and weights associated with each kinematic feature in the linear combination model. Hence, the equation for the linear model for movement quality assessment for the wrist trajectory can be written as

$$\Big\{ w_1 C_{T_1}^{curved} + w_2 C_{T_2}^{fast} + w_3 C_{T_3}^{slow} + w_4 C_{T_4}^{jerk} + w_5 C_{T_5}^{seg} \Big\} \approx R_j^w \tag{4.5}$$

where, $w_1, \ldots, w_5$ are weights for each of the confidence scores of kinematic attributes *curvedness, fastness, slowness, jerkiness and segmentation*, respectively. R_j^w

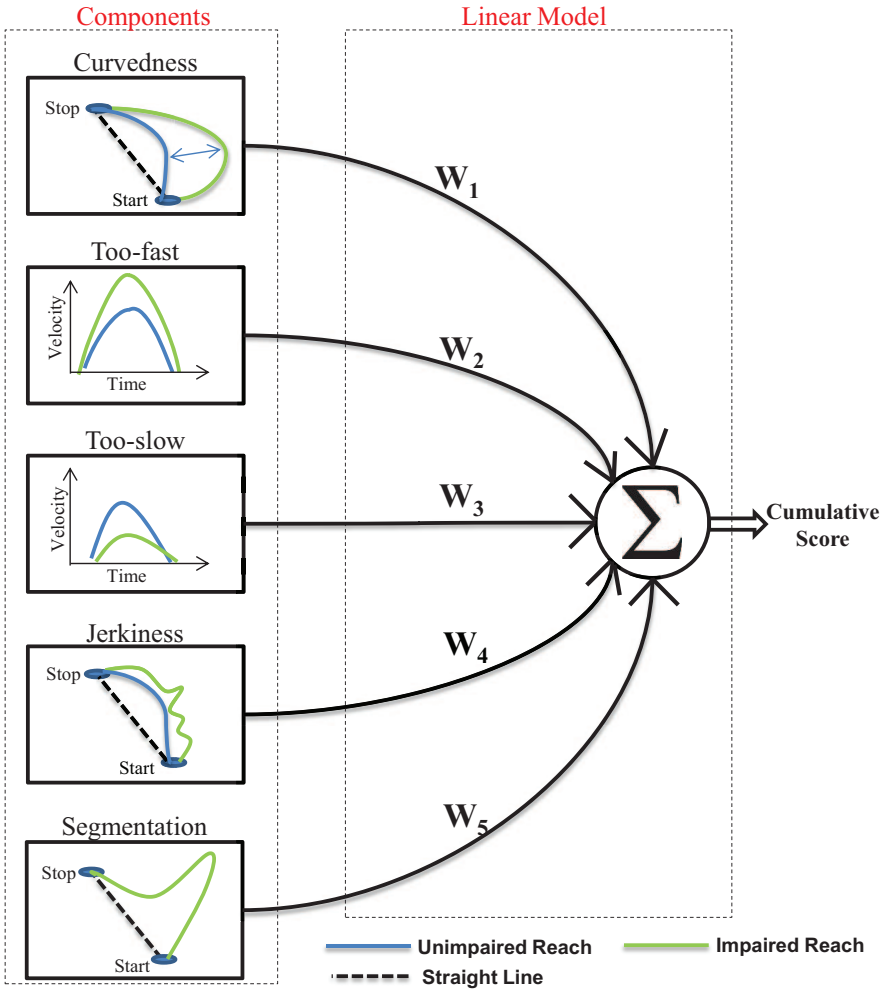


Figure 4.3: The proposed linear model of kinematic features extracted from the wrist marker. The weights (W_1, W_2, ..., W_5) and a unique threshold associated with each kinematic feature (T_1, T_2, ..., T_5) were estimated by minimizing the L_1 norm between *cumulative score* and *therapist rating* (R_j^w).

is the therapist rating for quality of wrist trajectory. The thresholds $T_1, \ldots, T_5$ bound a region called 'zero-zone' where the attribute value is termed '*efficient*' (indicating a reach movement without any impairments). For example, eq. (4.2a) has a threshold T_2 which represents a 'zone' of ideal speed profiles. Eq. 4.5 is pictorially depicted in Fig. 4.3. The aim here is to minimize the error between cumulative score and therapist rating in L_1 sense to estimate thresholds and weights associated with each kinematic feature. The cost function can be written as

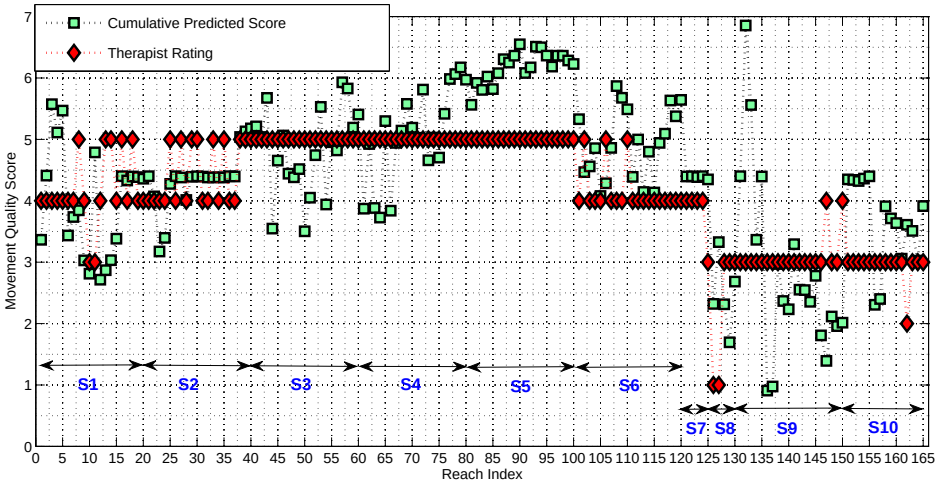
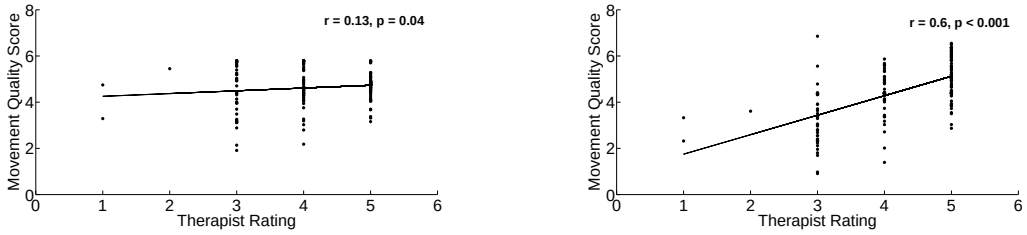


Figure 4.4: Comparison between the predicted cumulative score and therapist rating for movement quality. Each of 10 participants performed 20 reach and grasp to cone tasks except subjects S7, S8, and S10. The demographics and FMA score for each subject is tabulated in Table 5.2. A correlation of 0.6 exists between the predicted cumulative score and therapist rating for movement quality.

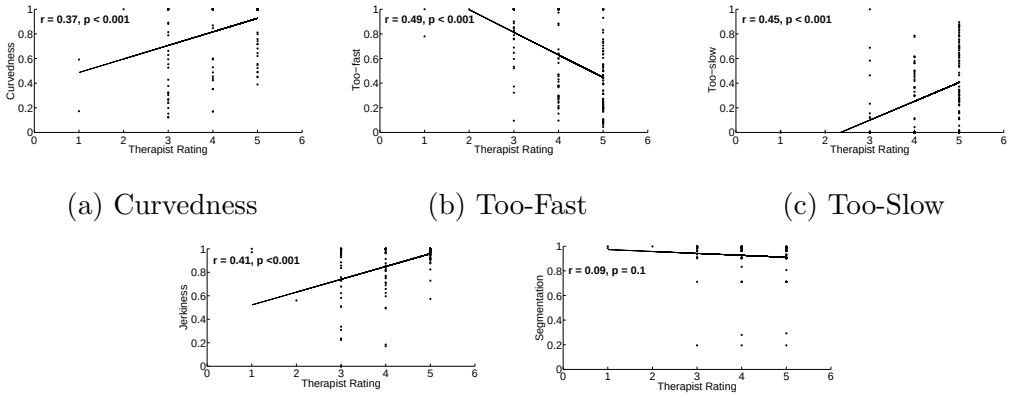
$$
\begin{aligned}
P1 : \{w_1, \ldots, w_5, T_1, \ldots, T_5\}^{opt} = & \\
& \arg\min_{w_1,\ldots,w_5,T_1,\ldots,T_5} \sum_{<j>} | \sum_{i=1}^{5} w_i C^i_{(T_i)} - R^w_j | \\
\text{subject to} \quad & w_i \geq 0, \\
& 0 \leq T_i \leq 10.
\end{aligned}
\tag{4.6}
$$

This cost-function is difficult to optimize, and is non-convex. In order to solve this optimization problem, we use the active-set method [87], because of its reduced complexity of the search, as the algorithm uses a subset of inequalities while searching the solution. We use the implementation of the active-set method available in Matlab.



(a) Before Optimization (b) After Optimization

Figure 4.5: Comparison of cumulative score and therapist rating before and after optimization procedure. A linear regression plot between cumulative score and therapist rating indicates that correlation coefficient increases from 0.13 to 0.6 with a significant p-value and increased slope.



(a) Curvedness (b) Too-Fast (c) Too-Slow

(d) Jerkiness (e) Segmentation

Figure 4.6: Linear regression plots for various low-level kinematic features used in our linear model for movement quality assessment with estimated thresholds. (a) Curvedness, (c) Too-slow and (d) Jerkiness show positive and significant correlation with therapist rating. (b) Too-fast shows a negative and significant correlation. (e) Segmentation shows a weak correlation with therapist rating.

4.4 Experimental Results

The aim of our experiments is two-fold: a) generate a cumulative score indicative of the overall quality of the reach movement, and b) detect the anomalies in the low-level kinematics to provide accurate feedback.

Due to the lack of low-level kinematic labels, we have collected a dataset of reach movements simulating impairments in individual low-level kinematics. These movements were performed by unimpaired participants with hands-on experience with

stroke survivors. A database of 62 reach movements were collected and a classification experiment using nearest neighbor classifier with 10-fold crossvalidation scheme shows that the anomalies in speed, curvedness and jerkiness were easy to detect with 100% classification accuracy. The classification results for segmentation was 98.38%, as accurate analysis of segmentation requires tracking of elbow and shoulder joints, which was not possible due to the design constraints. Similar analysis on torso leaning and twisting movements gave 100% classification accuracy.

In order to measure the efficacy of the proposed optimization procedure, we look at the output (cumulative score) generated by the forward-model in eq. 5. The results of our analysis using the linear combination of kinematic features for movement quality assessment of the wrist trajectory are shown in Fig. 4.4. The information about participants who experienced our system is tabulated in Table 5.2. Each participant performed 20 repetitions of reach and grasp to a cone target, except participants $S7, S8$, and $S10$ who performed 5, 5, and 15 repetitions, respectively. Fig. 4.4 shows the comparison between the movement quality scores provided by a trained physical therapist against the cumulative score predicted by our proposed framework. If the feature thresholds and combination weights were tuned, we expect the cumulative predicted scores to be correlated with the therapist ratings. The Pearson correlation coefficient between the cumulative scores and the therapist ratings was found to be 0.6 with a significant p-value ($p < 0.001$). The results of our analysis using the linear combination of kinematic features for quality assessment of wrist trajectory before and after optimization is shown in Fig. 4.5. We see that before optimization, the predicted movement quality scores of all classes (therapist ratings from 1 to 5) are overlapping (Fig. 4.5a). The use of optimized weights and thresholds resulted in an increased correlation between cumulative predicted score and therapist rating from 0.13 to 0.6. The contribution of each of the low-level kinematic features with

optimized threshold towards movement quality assessment is shown in Fig 4.6. A linear regression analysis between each kinematic feature and therapist rating shows that curvedness, too-slow and jerkiness show a significant positive correlation, while too-fast and segmentation respectively show negative and weak correlation. The weak correlation between segmentation and therapist rating could be due to the fact that the segmentation feature needs data from elbow and shoulder joints, which is not available in our single marker-based system. The obtained values for thresholds and weights after solving the optimization problem $P1$ are listed in Table 4.1. Kinematic features curvedness and too-fast have the highest weight of 2.5 in our linear model, with jerkiness and segmentation having lowest weight. It is evident from these results that the estimation of weights and thresholds of linear model using the proposed framework provides a novel methodology to combine low-level kinematic features to generate a cumulative score for movement quality of wrist trajectories. Furthermore, the cumulative score aligns with the ratings given by a therapist, which makes it a suitable tool to assist physical therapists in assessing the movement quality during supervised rehabilitation, leading to better evaluation and adaptation of therapy. The estimation of thresholds for low-level kinematic features facilitates better evaluation of components of movement (e.g., curvature, segmentation), thereby improving the efficacy of audio and visual feedback in our home-based rehabilitation system.

4.5 Conclusion and Future Work

In this work, we have introduced the problem of developing a computational framework for movement quality assessment suitable for home-based rehabilitation systems using kinematic analysis. We have proposed and evaluated a linear model of component-level kinematic features for movement quality assessment of the wrist. We propose a framework to learn these component-level kinematic features indicating im-

Parameter	Optimized Value
T_1	$0.13m$
T_2	$0.2m/s$
T_3	$0.1m/s$
T_4	$2.5m/s^3$
T_5	0.99
w_1	2.5
w_2	2.5
w_3	1.8
w_4	0.05
w_5	0.05

Table 4.1: The optimized values for thresholds and weights in the proposed linear model for movement quality assessment.

pairments in underlying movement components using composite therapist impressions of movement quality. Our results indicate that the proposed framework can be used to provide improved and efficient audio and visual feedback indicative of the impairments in component-level kinematics of a participant's reach. Further, this framework can be used to generate a cumulative score indicative of overall reach quality, which can be used to aid therapists during supervised rehabilitation. It should be noted that kinematic analysis of movement has an inherent requirement of "reference" trajectory data, which is difficult to define for complex movements (e.g., lift and transport an object) due to variability. Since we are interested in analyzing such complex movements of stroke survivors, our future directions will be focused towards developing suitable quantitative frameworks for modeling such complex movements.

Monitoring body movement during upper extremity tasks is necessary to determine the extent to which the stroke survivor is using body compensation. Prelim-

inary work using the data collected from the marker plate worn by the participant (not presented here due to scope) is promising for applying similar methods to aspects of movement beyond wrist trajectory performance. However, the consistent marker placement on the torso requires assistance from a caregiver, and we believe markerless solutions for monitoring the torso movements, such as using the Kinect, could provide a robust alternative. This points to several interesting directions of future work. From a sensor fusion perspective, one can explore the utility of multiple Kinect sensors and study the effects on obtaining high fidelity tracking results. Such efforts are already underway, with early commercial systems that are limited to a few gestures [2]. Accuracies of such multi-Kinect systems and its efficacy for rehabilitation systems are still unknown. We are currently working on pilot experiments with Kinect and mono-vision systems.

For the computer vision and machine learning communities, this application area opens up several interesting questions related to the design of robust features for movement quality analysis. Significant research in computer vision has been focused on activity and gesture recognition and not much on measures for '*quality*' of the movement. While this problem is traditionally addressed in the bio-mechanics community, the tools developed there are based on precise clinical measurements of biomechanics. These tools have limited applicability in home-based deployments, where data is of significantly lower quality. Thus, one needs to rely on larger datasets and advanced feature selection and machine learning tools to design movement quality measures. This can form the basis of several interesting research questions in the future.

5 DECISION SUPPORT FOR STROKE REHABILITATION

Researchers have been motivated to develop frameworks for quantification of movement quality [30, 142, 127, 156, 28] given its potential impact on disseminating interactive rehabilitation training to unsupervised contexts such as the home. Several automated approaches exist in literature to quantify movement quality based on complex models including nonlinear dynamical system theory [127, 142, 95], random forests [93], and SVMs [94]. While these approaches provide a computational framework for movement quality assessment showing high correlation with the clinical assessment scores, it would be beneficial to have an interpretable framework which can be used as a decision support tool by physical therapists during rehabilitation treatment.

To assess the level of functional ability of a stroke survivor, therapists can employ validated rating rubrics such as the Wolf Motor Function Test [149], to systematically assign a movement quality assessment score after observing a participant perform a predefined set of functional tasks. Such a rubric imposes a hierarchical set of rules for a therapist to consider, in order to help evaluate a participant's performance. Given this method of translating visual observation of movement to a quantitative score, we were motivated to investigate if a computational framework based on kinematic features can also be structured in a hierarchical form that can be easily understood by a therapist. We believe that such a framework would be useful in providing recommendations to physical therapists especially in the context of telerehabilitation, where a therapist reviews large amounts of movement performance data produced by

a participant performing rehabilitation exercises without supervision (e.g., in the home). Large scale movement quality evaluation would greatly benefit from such systems by providing recommendations to therapists and also allowing them to check the reason for recommended movement quality score using describable attributes indicative of the impairments.

We propose a hierarchical model using decision trees to simulate the results of the rating rubric created by rehabilitation experts to rate reach to grasp tasks across stroke survivors of various deficit. This is a step towards development of generalized models for knowledge representation of movement quality assessment of reach and grasp action based on previous work [73, 75, 43]. Within this experimental framework, we assume a simplified kinematic representation of reach and grasp action which focuses on a few specific elements of reach movement suitable for real-time monitoring and quantification of movement quality. The elements of the reaching movement chosen in our experiments include hand trajectory error in the horizontal and vertical planes, peak speed, jerkiness [30], velocity bellness [30] and torso rotation along XYZ axes. The main goal of this work is to learn a model that can simulate the resultant ratings of therapists using a rating rubric for movement quality assessment based on low-level kinematics indicative of the participant's impairment which can be used as a decision support system to aid the therapist during supervised rehabilitation therapy.

5.1 Methods for Collecting Kinematics and Therapist Ratings

5.1.1 Collection of Kinematics

The HAMRR system was designed to provide rehabilitation therapy to stroke survivors in a home-setting with reduced supervision by a physical therapist. This system

was used as an apparatus to collect kinematics when participants perform movement tasks without any assistance of feedback. The HAMRR system has four Natural Point Opti-Track cameras facing down on a table to track a single reflective marker placed on the participant's wrist (wrist marker) and four markers on the corners of a rectangular rigid plate placed on the participant's left side of chest (Fig. 4.2 inset). The selection of the wrist marker was motivated by previous investigations indicating that the wrist trajectory as the most informative joint with respect to analyzing reach trajectory performance [142, 30]. In addition, we believe that it is important to monitor the torso compensatory strategies for efficient movement analysis.

The selection of the plate was motivated by efforts to capture body compensation. To compensate for the lack of extension during a reach, many stroke survivors use excessive shoulder movement (elevation and/or protraction) and excessive torso movement (flexion and/or rotation). Therefore, a system for rehabilitation training should monitor movement of the body to determine the extent to which a participant is utilizing pre-stroke movement strategies to advance his/her hand towards the target. The HAMRR system was designed for home-based use, and the sensing apparatus worn by the participant must be simple and easy to wear. Therefore, we are only using a single plate worn on the chest of the participant, which captures coarse torso movement as opposed to both shoulder and torso movement separately. The system is shown in Fig. 4.2 and detailed information of the system design can be found in [10].

5.1.2 Therapist Rating Protocol

Stroke rehabilitation experts have standardized means for systematically rating overall functional performance of a defined set of tasks (relevant to activities of daily living) included within the WMFT protocol. However, within the stroke rehabili-

tation community there lacks a consensus among physical therapists in defining an ontology of component-level labels for movement quality (i.e., methods for rating the movement components that contribute to completion of a functional task), thereby leading to lack of training datasets to develop algorithms for movement quality assessment. In other words, while kinematics can capture the component-level aspects of movement (trajectory, compensation) which are important for evaluating movement quality, there is not yet a corresponding rating system in the stroke rehabilitation community for these components. Therefore, our team has collaborated with rehabilitation experts to introduce a new rubric for physical therapists to rate movement quality for specific tasks trained by the HAMRR system. Movement quality is assessed in terms of trajectory, compensation, manipulation, transport of an object, and release. However, we limit our focus on trajectory and compensation in the context of reaching to grasp a stationary cone, as these movement components have established corresponding methods for quantifying performance using kinematics derived from 3D positions of reflective markers described in the section 5.1.1. The rating rubric used by therapists to rate trajectory and compensation is provided in Table 5.1. One should note that this rubric was designed given the constraints of the therapist viewing a single camera video of the participant while performing a task from the right side (as shown in Fig. 5.1).

5.1.3 Data Collection

The dataset used in our experiments consists of reaching tasks performed by a total of eight participants (refer Table 5.2 for demographics) to a cone on-table located at the participant's midline. Each participant performed five reaches in each of four sessions (one session per week). These reaches were performed without any feedback from the system or therapist unless the participant was unclear on how to perform

Table 5.1: The Rating Rubric for Movement Quality Assessment Provided to Therapists

Score	Trajectory	Compensation
1	Does not ever reach the target	Demonstrates compensatory shoulder movement with compensatory torso movement in more than one plane
2	Demonstrates profound deviation from a direct path during the reaching phase, which may be affected by but is not limited to one or more of the following secondary factors: Synergy, Ataxia and Spasticity	Demonstrates compensatory shoulder movement with trunk compensatory movement mainly in one plane
3	Demonstrates slight deviation (relative to how the rater would perform the task) from a direct path during the reaching phase	Demonstrates noticeable compensatory shoulder or trunk movement
4	The trajectory appears to be similar to that of the rater if he/she were performing the task	The shoulder and trunk are positioned in a manner similar to the rater if he/she were performing the task

Figure 5.1: A sample of video data provided to therapists to evaluate movement quality of stroke survivors interacting with the HAMRR system.

the task. During the task, each participant was seated at the HAMRR system and his/her movement was captured by the Opti-Track system. A custom designed iPad application was also concurrently used to capture video footage of a participant performing these tasks. These videos were randomized across participants and sessions before they were provided to therapists for evaluation. Therapists could only view one video at a time and were allowed to watch the videos as many times as they needed to form a decision on the ratings. However, therapists were not allowed to see or change responses to previous videos once they were submitted.

Trajectory performance was rated on a scale from $1-4$ based on the therapist's impression of the participant's performance, where a 1 denotes that the participant could not complete the task and a 4 denotes that the participant performed the task with the same quality of performance as the therapist if he/she were to perform it. Compensation was rated on a scale from $1-4$ based on the participant's excessive use of the shoulder and/or torso and if compensation was used in single or multiple planes of movement. A 1 denotes that the participant used both excessive shoulder and torso movement in multiple planes of movement, while a 4 denotes that the shoulder and trunk are positioned in a manner similar to the therapist if he/she was performing the task.

Table 5.2: The Demographics of Stroke Survivors Who Participated in Our Study

Name	**Age**	**Gender**	**Time since stroke** (in months)	**# of strokes**
1	63	Male	14	1
2	69	Male	44	1
3	65	Male	31	1
4	47	Male	26	1
5	56	Male	28	1
6	49	Male	18	1
7	64	Female	6	1
8	27	Male	12	1

5.2 Definitions of Kinematic Features

The following kinematic features were extracted to quantify the impairments of a participant while performing a reach to grasp a cone task.

Kinematic Features from Wrist Trajectory

Trajectory Error Trajectory error is a measure of spatial deviation of the wrist trajectory from the reference trajectory. The three-dimensional positions of the wrist marker $p(t) = [\mathbf{x}(t), \mathbf{y}(t), \mathbf{z}(t)], \quad t = 0, \ldots, \tau$ were recorded from the start of the movement to the target grasp state. The coordinate system was rotated such that $p(0)$ was the origin, $\mathbf{X} - \mathbf{Z}$ plane was the horizontal plane and the straight line connecting $p(0)$ and $p(\tau)$ lies along the new $\mathbf{Z}$-axis. This in effect re-parameterizes (after normalization) the trajectory $[\mathbf{x}(t), \mathbf{y}(t), \mathbf{z}(t)], \quad t = 0, \ldots, \tau$ to $[\mathbf{x}'(\mathbf{z}), \mathbf{y}'(\mathbf{z})], \quad \mathbf{z} = 0, \ldots, 1$. This re-parameterization works without introducing significant ambiguity in our experiments due to the strong directionality of the reach action. The $\mathbf{Z}$-axis was further quantized into $N = 50$ bins, thereby transforming the trajectory

to $[\mathbf{x}'(n), \mathbf{y}'(n)], \quad n = 0, \ldots, N-1$. We now have a vectorial representation of the trajectory suitable for real-time comparisons. For every point in the reach trajectory, horizontal error (E_{hor}) and vertical error (E_{vert}) were defined as

$$E_{hor}(i) = \mathbf{x}(i) - \mathbf{x}_{ref}(i), \quad i = 0, \ldots, N-1 \tag{5.1a}$$

$$E_{vert}(i) = \mathbf{y}(i) - \mathbf{y}_{ref}(i), \quad i = 0, \ldots, N-1 \tag{5.1b}$$

The horizontal trajectory error ($\hat{E}_{hor}$) and vertical trajectory error ($\hat{E}_{ver}$) were defined as (units in mm)

$$\hat{E}_{hor} = \max_{0<i<N-1}(E_{hor}) \tag{5.1c}$$

$$\hat{E}_{ver} = \max_{0<i<N-1}(E_{ver}) \tag{5.1d}$$

Jerkiness The jerkiness (or smoothness) feature is a measure of variations in the velocity profile. An '*efficient*' reach movement should have a smooth velocity profile with an accelerating followed by a decelerating pattern without any jerks. Jerkiness (in m/s^3) of a movement was computed using the definition given in [30] as

$$J = \int_{t_{som}}^{t_{eom}} \sqrt{\left(\frac{d^3\mathbf{x}}{dt^3}\right)^2 + \left(\frac{d^3\mathbf{y}}{dt^3}\right)^2 + \left(\frac{d^3\mathbf{z}}{dt^3}\right)^2}\, dt \tag{5.2}$$

where $\mathbf{x}, \mathbf{y}$ and $\mathbf{z}$ are 3-D coordinates of the participant's wrist trajectory. t_{som} is the time index corresponding to start of the movement and t_{eom} is the time index of end of the movement.

Velocity Bellness Ideally, the velocity profile of a reaching task should be a bell curve. Typically, stroke survivors throw their arm towards the target and then make fine adjustments to grasp the object. These adjustments show up as additional phases

in the speed profile. It is believed that these occur during the deceleration phase and we use normalized area to evaluate velocity bellness (B_{NA}) given by

$$B_{NA} = \frac{\int_{t_{1st}}^{t_{eom}} v(t)dt}{\int_{t_{v_{max}}}^{t_{eom}} v(t)dt} \tag{5.3}$$

where $v(t)$ is the instantaneous velocity, $t_{v_{max}}$ is the time index corresponding to maximum velocity, t_{1st} is the end of the first phase.

Peak Speed An efficient reach movement is typically accomplished by a hand velocity between 0.4m/s and 0.6m/s. We use peak speed (in m/s) as a measure of deviation from this ideal range defined as the maximum velocity of each trial given by

$$V_{max} = \max_{t_{som}<t<t_{eom}} [v(t)] \tag{5.4}$$

Our results indicate that the kinematic components we chose (hand trajectory error in the horizontal and vertical planes, peak speed, jerkiness, velocity bellness and torso rotation along XYZ axes) combined with a decision tree model are capable of simulating the results of an imposed hierarchical structure used by trained therapists. The selected low-level kinematic attributes are representative of the impairments in reach and grasp action and can collectively be used to generate a movement 'component score' showing high correlation with the therapist rating. These results also indicate that the proposed framework can be used as an assistive tool to therapists during supervised rehabilitation to reduce the time spent on movement quality assessment.

To more specifically qualify our findings: the rehabilitation experts were able to create an imposed hierarchy based on expert knowledge (presented in Table 5.1). Given this hierarchy developed by expert knowledge and its careful implementations

by highly trained therapists, we are able to replicate the results of their ratings through a decision tree approach. Our initial results support that these decision trees can help with semi-automated ratings when the therapist is absent, and assist therapists to provide ratings faster when they log-in remotely to fine tune a home-based training system for a participant. Since we achieved favorable results using this decision tree approach given a particular imposed hierarchy, when the hierarchy needs to be switched for different types of training, we propose that similar trees can be estimated based on different hierarchies across tasks, stages of therapy, and participants. Thus, our process is dependent on clear declarations of hierarchies by therapists and their consistent implementation.

Defining an ontology of component-level labels for movement quality assessment is seen as a difficult problem in the stroke rehabilitation community. While the current research was directed towards learning a simple decision tree model for knowledge representation of given physical therapists, our future goal is to extract a generalized knowledge representation for movement quality assessment using evaluations from multiple therapists. Similar problems have been discussed in the machine learning community [56]. We are currently collecting evaluation ratings from multiple therapists as different knowledge representations for movement quality assessment and will be used to estimate a generalized knowledge model using existing approaches for matching of knowledge structures.

5.3 Conclusion and Future Work

In this work, we present a computational framework capable of simulating the component-level movement quality assessment rubric with imposed hierarchical structure on physical therapists. This automatic assessment of movement quality framework can provide suggestions to physical therapists during supervised rehabilitation reducing

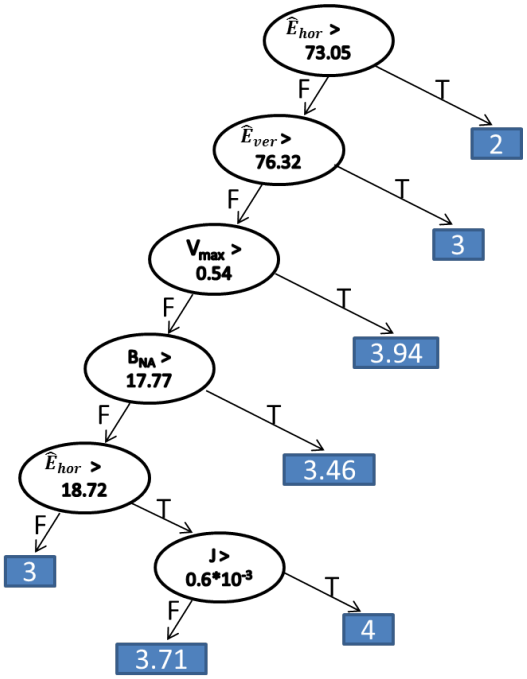


Figure 5.2: The decision tree model for movement quality assessment of wrist trajectory. The low-level kinematic features used were horizontal trajectory error ($\hat{E}_{hor}$), vertical trajectory error ($\hat{E}_{ver}$), peak speed (V_{max}), velocity bellness (B_{NA}) and jerkiness (J). The scores highlighted in blue are the decision tree outputs for wrist trajectory analysis.

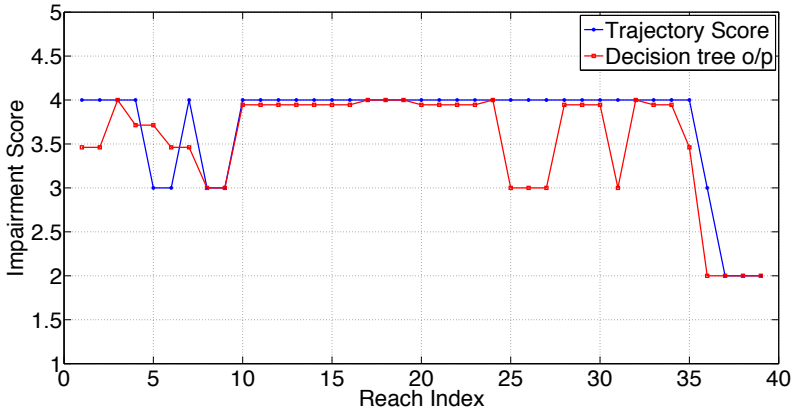


Figure 5.3: Comparison between impairment level (with 4 being least impaired and 1 being most impaired) given by component-level score for wrist trajectory and decision tree predictions. The Pearson correlation coefficient was found to be 0.8049.

the time spent on evaluating the quality of movements, thereby reducing the cost of long-term rehabilitation treatment.

Our results indicate that the kinematic components we chose (hand trajectory error in the horizontal and vertical planes, peak speed, jerkiness, velocity bellness and torso rotation along XYZ axes) combined with a decision tree model are capable of simulating the results of an imposed hierarchical structure used by trained therapists. The selected low-level kinematic attributes are representative of the impairments in reach and grasp action and can collectively be used to generate a movement 'component score' showing high correlation with the therapist rating. These results also indicate that the proposed framework can be used as an assistive tool to therapists during supervised rehabilitation to reduce the time spent on movement quality assessment.

To more specifically qualify our findings: the rehabilitation experts were able to create an imposed hierarchy based on expert knowledge (presented in Table 5.1). Given this hierarchy developed by expert knowledge and its careful implementations by highly trained therapists, we are able to replicate the results of their ratings through a decision tree approach. Our initial results support that these decision trees can help with semi-automated ratings when the therapist is absent, and assist therapists to provide ratings faster when they log-in remotely to fine tune a home-based training system for a participant. Since we achieved favorable results using this decision tree approach given a particular imposed hierarchy, when the hierarchy needs to be switched for different types of training, we propose that similar trees can be estimated based on different hierarchies across tasks, stages of therapy, and participants. Thus, our process is dependent on clear declarations of hierarchies by therapists and their consistent implementation.

Defining an ontology of component-level labels for movement quality assessment is

seen as a difficult problem in the stroke rehabilitation community. While the current research was directed towards learning a simple decision tree model for knowledge representation of given physical therapists, our future goal is to extract a generalized knowledge representation for movement quality assessment using evaluations from multiple therapists. Similar problems have been discussed in the machine learning community [56]. We are currently collecting evaluation ratings from multiple therapists as different knowledge representations for movement quality assessment and will be used to estimate a generalized knowledge model using existing approaches for matching of knowledge structures.

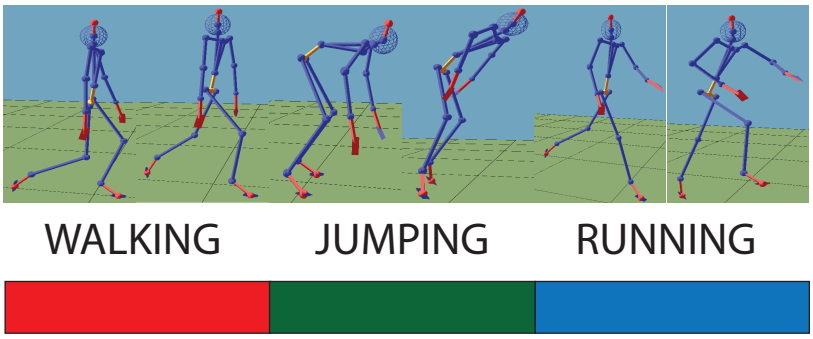
6 DYNAMICAL REGULARITY FOR MOTION ANALYSIS: APPLICATIONS TO ACTION SEGMENTATION, RECOGNITION AND QUALITY ASSESSMENT

Human motion recognition from untrimmed videos is a challenging problem in the vision community [60, 125]. In a real world scenario, these applications require automatic recognition of action sequences from continuous untrimmed videos. Traditionally, the vision community works with the simpler, unrealistic assumption that temporal segmentation of videos is a step which has been done beforehand, resulting in pre-segmented videos containing individual action sequences. In literature, most of the proposed frameworks for action recognition assume that each clip contains just one action sequence. Temporal segmentation of human motion from untrimmed videos into its constituent action sequences is a challenging problem due to large variations in temporal scale of actions and extremely large number of possible movement combinations. In this work, we focus our interest towards developing a framework to simultaneously achieve both temporal segmentation of untrimmed videos and action classification.

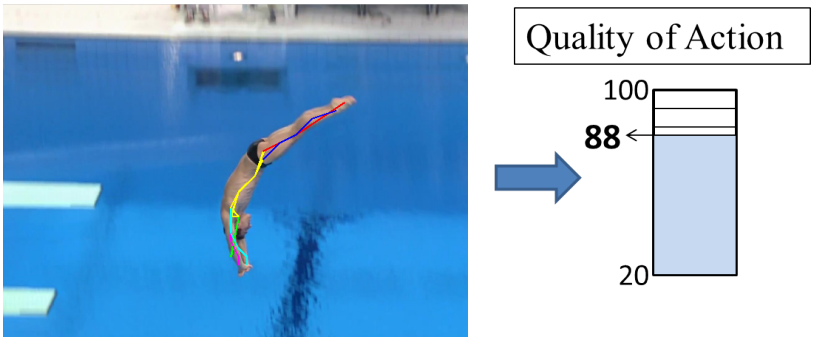
The computer vision community has been interested in modeling human activities for many applications including video surveillance, automatic video annotation and health monitoring [4]. Modeling the underlying dynamics in an activity forms the core idea in many systems. An activity can be seen as a resultant of coordinated movement of body joints and their respective interdependencies to achieve a goal-

directed task. This idea is further supported by Johansson's demonstrations that visual perception of the entire human body motion can be represented by a few bright spots which holistically describe the motion of important joints [63]. Traditional dynamical modeling approaches usually operate on the level of individual joints of the human body, lacking any information about the interdependencies between joints [5]. Only recently, researchers have started exploring relationships between body joints, using rotations and translations in 3D space [140], which lacks dynamical information. In this paper, we propose a novel approach for dynamical modeling by extending conventional ideas to quantify the interdependencies between body joints. Towards this end, we propose a new approach – approximate entropy-based feature representation to model the dynamics in human movement by quantifying dynamical *regularity*.

Our use of the term *regularity* represents the frequency of repetition of typical patterns in the data. The main principle in our work is that different actions correspond to different levels of regularity, and quantification of regularity can be used for human activity analysis. For instance, *walking* is inherently periodic and hence corresponds to a higher level of regularity when compared to *dancing*, which is more towards randomness due to multiple movement strategies. From the system complexity perspective, *walking* can be represented by simple dynamical systems, while more complex systems with a large number of variables may be required to represent *dancing*. Quantifying regularity and system complexity is a well-studied problem in the field of signal processing. Correlation dimension [3] and largest Lyapunov exponent [147] are examples of invariant measures proposed in the literature to quantify complexity of dynamical systems. It was found that robust estimation of these invariant measures requires large number of data samples (of the order of 10^d), where d is related to the dimension of the dynamical system's state space used in the estimation



(a) Temporal segmentation of actions using motion capture data.



(b) Quality assessment of diving actions using videos.

Figure 6.1: A visual representation of our applications of interest in this work. In (a), our aim is to achieve temporal segmentation of actions from continuous untrimmed motion capture data in an unsupervised manner. In (b), we use a supervised learning framework to assess the quality of diving actions from videos.

procedure, with typical values of 3 and above. Later, a probabilistic measure called approximate entropy was proposed to overcome the drawbacks of the above traditional measures for quantification of system complexity [99]. Approximate entropy assigns lower values for ordered time series and higher values for time series towards randomness. In this paper, we utilize the algorithmic framework of [99] for estimating approximate entropy from time series data and extend it to model the dynamics in human activities for applications such as temporal segmentation and fine-grained quality assessment of actions.

Much work in the domain of action recognition over the past few decades is carried out using RGB videos [4], which are sensitive to factors like background clutter and

illumination changes. In addition, it is difficult to capture complex articulated human motion using monocular video sensors. Recent advances in sensing platforms, such as the Kinect, provide access to 3D locations of body joints in real time [117], thereby providing a better representation of human body motion in 3D space compared to monocular video sensors. One can also get access to 3D locations of body joints using large and expensive motion capture systems which require the participant to wear reflective markers on his/her body. Kinect offers a cheaper and more user-friendly joint tracking solution compared to other motion capture systems. This motivates us to design a framework for automatic segmentation and action classification using trajectories of body joints derived from Kinect sensors.

One popular approach for temporal segmentation of human motion is to detect the presence of an action sequence (or event) by evaluating a classifier function over a sliding window [42, 67, 86, 119]. Approaches for change point detection of time series data such as in [58] temporally monitors the test statistic for change-point analysis in a sliding window. Based on this idea, in our framework, we monitor a measure for *regularity* of human motion patterns in a sliding window over time for automatic segmentation and classification of actions. The term regularity represents the frequency of repetition of typical patterns in the data. The main principle in our work is that different actions correspond to different levels of regularity, and quantification of regularity of human actions can be used for simultaneous segmentation and action classification. For instance, *walking* action is inherently periodic and hence corresponds to a higher level of regularity when compared to *dancing* action, which is more towards randomness due to multiple available movement strategies. From the system complexity perspective, *walking* action can be represented by simple dynamical systems, while more complex systems with large number of variables may be required to represent *dance* action.

Quantifying regularity and system complexity is a well-studied problem in the field of statistics and signal processing. Correlation dimension [3], largest Lyapunov exponent [147], and Kolmogorov-Sinai entropy [70] are a few examples of invariant measures proposed in the literature to quantify complexity of dynamical systems. It was found that robust estimation of these invariant measures requires large number of data samples (of the order of 30^d), where d is a parameter used in the estimation procedure with typical values of 3 and above. A more recent work by Pincus proposed a measure called approximate entropy to overcome the drawbacks of the above traditional measures for quantification of system complexity [99]. Approximate entropy is a probabilistic measure which assigns lower values for ordered time series and higher values for time series towards randomness. In this paper, we utilize the algorithmic contributions by Pincus for estimating approximate entropy to quantify regularity in time series of human actions.

Researchers in the vision community have shown growing interest towards fine-grained analysis of human activities by developing frameworks for quantification of movement quality [30, 142]. Stroke being the most common neurological disorder, has motivated us to develop a computational framework to assist physical therapists during supervised rehabilitation therapy, and potential unsupervised contexts such as the home. We use the approximate entropy based feature representation and show its utility to quantify impairment in a stroke survivor's movement trajectory.

6.1 Related Work

Most of the contributions in the domain of human activity analysis are carried out on pre-segmented action sequences. Tremendous amount of research has been conducted on action recognition using RGB videos [4, 51]. The advent of Kinect sensors has brought recent interest in skeletal-based action recognition. Existing skeletal-

based approaches for action recognition can be categorized as joint-based approaches and body part-based approaches. The relevant works in the literature on these two approaches have been explained well in [140]. Since our current work is on dynamical modeling of trajectories of human actions, we restrict our discussion to related methods focused on applications of interest.

Segmentation and Action Classification: In the literature, many approaches have been proposed for temporal segmentation of human actions based on hidden Markov models (HMMs). Bregler *et al.* [20] utilized HMMs to model complex human gestures as successive phases of simple movements. Brand *et al.* [19] applied coupled HMMs demonstrating superiority to conventional HMMs towards classifying two-handed human motion. Spriggs *et al.* [125] used HMMs for temporal segmentation of activities in a kitchen environment using wearable camera and inertial measurement units. Sminchisescu *et al.* [123] introduced conditional models as complimentary tools based on conditional random fields and maximum entropy Markov models. Hoai *et al.* [60] proposed a framework based on multi-class SVMs for joint temporal segmentation and action classification. Zhou *et al.* [158] proposed aligned cluster analysis for temporal segmentation by extending standard kernel k-means clustering combined with dynamic time warping for temporal invariance. Niebles *et al.* [86] utilized probabilistic Latent Semantic Analysis and Latent Dirichlet Allocation for unsupervised learning of human actions.

Some of the early approaches for temporal segmentation of actions include learning representations for motion primitives using the theory of linear dynamical systems [78, 135, 136], thereby segmenting the human motion into its constituent action sequences. Oh *et al.* [88] utilized switching linear dynamical system to learn and infer motion patterns. Such parametric approaches may approximate the true dynamics

of human actions and fit experimental data to the model. Ali *et al.* in [5] claims through validated experiments that by constraining the dynamical system to be of a particular type (linear or nonlinear), one would only approximate the true dynamics of human motion. They proposed a novel feature representation based on the tools from chaos theory namely largest Lyapunov exponent, correlation integral and correlation dimension to characterize nonlinear dynamics of human actions in trajectory data extracted from videos and motion capture systems. While it is typical for human actions to last for 10 sec or less corresponding to 300 samples (at 30 frames per second), it is not advisable to use such feature representations as in [5], as the suggested number of data points required for robust estimation is large (of the order of 30^d, where d is the dimension of embedded phase space) [110]. This is evident from the findings of Wu *et al.* [150] who showed that the estimation of largest Lyapunov exponent in [5] produced negative values, which is incorrect for chaotic systems. It has been shown that approximate entropy can quantify system complexity with as low as 50 data samples [99], which makes it a suitable feature representation for modeling human actions.

Quality Assessment: The application of interest here is to develop a computational framework for movement quality assessment to aid physical therapists in providing supervised rehabilitation therapy for stroke survivors. Several validated clinical measures which requires visual monitoring by a therapist for movement quality assessment have been proposed [52, 149], and researchers aim to match these clinical scores using a computational framework. The existing approaches in literature to quantify movement quality use nonlinear dynamical system theory [127, 142, 95], random forests [93], and SVMs [94]. Chen *et al.* [30] proposed several kinematic attributes which requires access to reach trajectories from unimpaired subjects, thereby

limiting the generalizability of the framework to different reach targets. We evaluate the performance of approximate entropy-based feature representation for movement quality assessment on a dataset collected from stroke survivors.

Even though researchers have been working towards automatic recognition of human actions for decades, the task of automatically quantifying the quality of a given action has remained unexplored until recently. Such automated frameworks for quality assessment of actions will find real-world applications in sports and healthcare. Hamed *et al.* [101] used a regression model to predict the scores given by human expert judges on diving actions using spatio-temporal pose features. A similar approach using a regression model learned from shape-based dynamical features to quantify the quality of movement has been proposed for stroke rehabilitation [142]. In [96], authors quantified team performance in a multi-player basketball activity context using Bayesian networks. In this paper, we utilize the approximate entropy-based feature to quantify the quality of diving actions and show that using a dynamical measure performs better than the previously used frequency domain representation using discrete cosine transform (DCT).

Contributions: We propose a feature representation for modeling human motion by quantification of regularity using approximate entropy measure. The proposed feature representation encodes both the dynamics of individual joints and the cross-coupling information between joints by respectively using an univariate and bivariate form of approximate entropy. We show the its utility for simultaneous segmentation and action classification, and movement quality assessment.

6.2 Approximate Entropy (ApEn)

Approximate entropy is a statistical tool proposed by Pincus [99, 100] for quantification of regularity of time series data and system complexity. It is a probabilistic measure based on the log-likelihood of repetitions of patterns of length m being close within a defined tolerance window that will exhibit similar characteristics as patterns of length $(m + 1)$ [98, 99]. It assigns a non-negative number to time series data, with lower values for predictable (ordered) signals and higher values for signals with increased irregularity (or randomness). Ideally, a pure sine wave should have a zero value of approximate entropy. It has an advantage over Shannon's entropy [116] in that it takes into account the temporal order, which makes it more suitable to represent the dynamical evolution of time series data. The development of approximate entropy was motivated to address the drawbacks of traditional measures to quantify system complexity, thereby having a measure to successfully handle noise and address the limitations of data length requirements and other model constraints [100].

It is defined using three parameters: embedding dimension (m), radius (r), and time delay (τ). Here, m represents the length of pattern (also called as embedding vector) in the data which is checked for repeatability, τ is selected so that the components of the embedding vector are sufficiently independent, and r is used for the estimation of local probabilities. Given N data samples $\{x_1, x_2, x_3, \ldots, x_N\}$, we can define embedding vector $\mathbf{x}(i)$ as,

$$\mathbf{x}(i) = [x_i, x_{i+\tau}, x_{i+2\tau}, \ldots, x_{i+(m-1)\tau}]^T; \quad for \quad 1 \leq i \leq N - (m-1)\tau. \tag{6.1a}$$

The frequency of repeatable patterns of the embedding vector within a tolerance r is given by $\mathbf{C}_i^m(r)$ as

$$\mathbf{C}_i^m(r) = \frac{1}{N-(m-1)\tau} \sum_{<j>} \Theta(r - d(\mathbf{x}(i), \mathbf{x}(j))). \tag{6.1b}$$

where:

$$\Theta(a) = \begin{cases} 1, & \text{if } a \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

$$d(\mathbf{x}(i), \mathbf{x}(j)) = \max_{k=1,2,..,m} (|x(i+(k-1)\tau) - x(j+(k-1)\tau)|).$$

Approximate Entropy is given by

$$ApEn(m, r, \tau) = \Phi^m(r) - \Phi^{m+1}(r). \tag{6.1c}$$

where:

$$\Phi^m(r) = \frac{1}{N-(m-1)\tau} \sum_{i=1}^{N-(m-1)\tau} \ln \mathbf{C}_i^m(r). \tag{6.1d}$$

In the above equations, $\mathbf{C}_i^m(r)$ represents the frequency of repeatable patterns (local probabilities) in the embedding vector $\mathbf{x}(i)$, $\Theta(a)$ is the Heaviside step function, and $\Phi^m(r)$ represents the conditional frequency estimates. Evident from the above algorithm, the estimation procedure requires parameters m, τ, and r to be specified. In an ideal case, where one has access to an infinite amount of data of infinite accuracy, any set of parameters which can result in smooth embedding would give similar results ([3], chap. 3). With real world data, the choice of these parameters should ensure smooth embedding with components of the embedding vectors being sufficiently independent.

Multivariate Approximate Entropy: Motion capture sensing allows us to observe 3-dimensional time series data per body joint. A trivial solution to model the dynamics would be to consider each dimension of a body joint independently to create the embedding vector (eq. 6.1a) as in [5, 142]. Recent theoretical and empirical

findings have demonstrated that multivariate embedding of time series data by simple concatenation of individual univariate embedding vectors achieves good state space reconstruction as evaluated by the shape and dynamics distortion measures [144]. In this work, we propose to use the multivariate embedding procedure as described by Cao *et al.* [23] per body joint and estimate the approximate entropy feature representation.

Natural human movement involves multiple body joints interacting with each other to together accomplish a particular action task. Hence, it would be beneficial to utilize the cross-coupling information between these joint trajectories. Research carried out by Kavanagh *et al.* [66] using cross approximate entropy to model trunk motion during walking supports our hypothesis that adding information about cross-coupling offers better feature representation to model human motion and will be validated by our experiments.

Cross Approximate Entropy (XApEn): Cross approximate entropy is defined as the amount of asynchrony between two time series data [98, 97]. Let $\mathbf{u} = [u_1, u_2, \ldots, u_N]^T$ and $\mathbf{v} = [v_1, v_2, \ldots, v_N]^T$ denote two time series data of length N. The embedding vectors for given parameters m, τ, and r are defined as

$$\mathbf{x}_1(i) = [u_i, u_{i+\tau}, \ldots, u_{i+(m-1)\tau}]^T; \quad \mathbf{x}_2(i) = [v_i, v_{i+\tau}, \ldots, v_{i+(m-1)\tau}]^T. \tag{6.2a}$$

The frequency of repeatable patterns within the embedding vectors $\mathbf{x}_1(i)$ and $\mathbf{x}_2(i)$ for a tolerance r is given by $\mathbf{C}_i^m(r)(v||u)$ as

$$\mathbf{C}_i^m(r)(v||u) = \frac{1}{N-(m-1)\tau} \sum_{<j>} \Theta(r - d(\mathbf{x}_1(i), \mathbf{x}_2(j))). \tag{6.2b}$$

The cross approximate entropy is then given by

$$XApEn(m, r, \tau) = \Phi^m(r)(v||u) - \Phi^{m+1}(r)(v||u). \tag{6.2c}$$

where:

$$\Phi^m(r) = \frac{1}{N-(m-1)\tau} \sum_{i=1}^{N-(m-1)\tau} \ln \mathbf{C}_i^m(r)(v||u). \tag{6.2d}$$

We estimate the XApEn feature across all pairs of body joints (after performing multivariate embedding using data available from each body joint). It is evident from the above equations that XApEn is an asymmetric measure. We note here that our initial analysis on exemplar human action data did not show a significant difference in the values of XApEn for forward and backward directions. Hence, we use only one of these values in our feature representation. We then concatenate ApEn and XApEn values to form our final approximate entropy-based feature vector to model actions denoted by ***ApEnFT***.

Framework: For any given time series data, we calculate univariate approximate entropy on every individual dimension and bivariate cross approximate entropy across pair of dimensions over a sliding window as shown in Fig. 6.2. Here, the feature value estimated from the samples in the sliding window is assigned to the sample at the center of the window. This window is moved by one frame and the process is repeated till the end of action data.

In our supervised training protocol, we use the frame-level features to train Partial Least Squares (PLS) regressor. With a set of binary PLS regressors with thresholded outputs, a majority voting scheme is used to determine the action class to be assigned to a particular frame.

6.2.1 Choice of Parameters

Data Length (N): The suggested value for N was typically between 50 and 5000. This constraint was imposed by Pincus in [100] to ensure a homogeneous segment of data under certain experimental conditions, and this range for N was not an

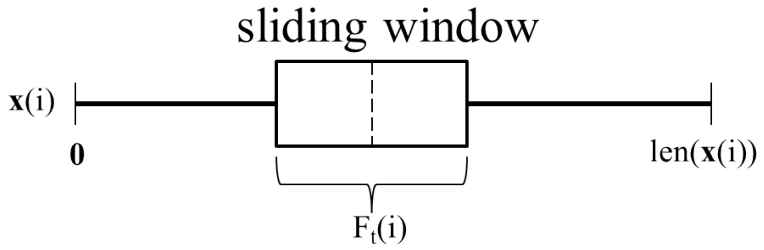


Figure 6.2: A picture showing sliding window run over a given time series $\mathbf{x}(i)$ to estimate the approximate entropy based features per frame.

algorithmic limitation. Our choice of N depends on the dataset used, and typically ranges between 30 and 50.

Embedding Dimension (m): Through theoretical analysis and extensive experimental validation, it has been shown that both $m = 1$ and $m = 2$ can distinguish data on the basis of regularity [100]. In our application, we use a fixed value of $m = 2$.

Delay Time (τ): The purpose of delay time τ is to ensure that the components in the embedding vectors are sufficiently independent. A low value of delay time will make adjacent components in the embedding vector to be correlated and hence cannot be considered as independent. On the other hand, a high value of delay time will make adjacent components to become uncorrelated (almost independent). Suggested methods in the literature to estimate an optimum delay time has been first minimum of the lagged auto-mutual information, and the time lag when the autocorrelation drops to 1/e of its initial value or the first zero of the autocorrelation function [3]. In our experiments, we use the lag when autocorrelation function drops below zero (refer Fig. 6.3). This estimate for delay time is suggested for use with strongly periodic data, which is a suitable choice to work with human actions.

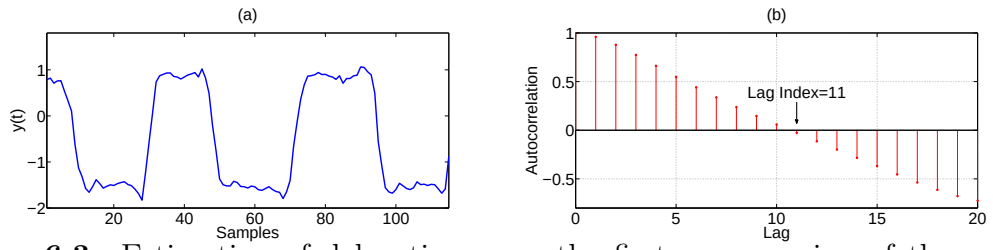


Figure 6.3: Estimation of delay time τ as the first zero-crossing of the autocorrelation function. (b) shows the autocorrelation function of the trajectory data in (a).

Radius (r): The value of r could range anywhere between 0.1 to 0.25 times the standard deviation of the data . A good choice of r should ensure that the conditional frequencies defined in Eq. 6.1c are reasonably estimated. Smaller values of r results in poor conditional frequency estimates, while large values of r cannot capture enough information of the system.

6.3 Experimental Evaluation

In this section, we evaluate the performance of our feature representation on (1) synthetic data generated from coupled Rossler oscillators, (2) action datasets from Kinect sensor, and (3) stroke rehabilitation dataset.

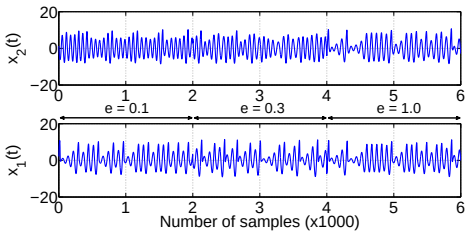
6.3.1 Coupled Rossler Model

In order to demonstrate the utility of the proposed feature representation for quantifying regularity and cross-coupling in time series data, we use two coupled Rossler oscillators given by the equations shown below. The main motive behind this experiment is to provide an analogy to human actions as coupled systems with changing coupling strengths to accomplish a particular type of action.

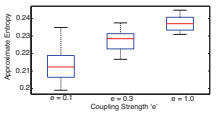
$$\begin{aligned} \dot{x_1} &= -w_1 y_1 - z_1 \\ \dot{y_1} &= w_1 x_1 + \alpha y_1 \\ \dot{z_1} &= \beta + z_1(x_1 - \gamma) \end{aligned} \tag{6.3a}$$

$$\begin{aligned} \dot{x_2} &= -w_2 y_2 - z_2 + e(x_1 - x_2) \\ \dot{y_2} &= w_2 x_2 + \alpha y_2 \\ \dot{z_2} &= \beta + z_2(x_2 - \gamma) \end{aligned} \tag{6.3b}$$

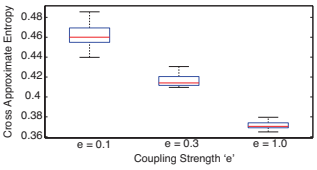
Here, the Rossler system in Eq. 6.3a *drives* the Rossler system in Eq. 6.3b. 'e' is the coupling strength between the two Rossler oscillators. As the coupling strength is increased, the two oscillators become synchronized. For this configuration of Rossler oscillators, the parameters were chosen as $\alpha = 0.2$, $\beta = 0.2$, $\gamma = 5.7$, $w_1 = 1$, and $w_2 = 0.2$. We choose three values of coupling strength, $e = 0.1, 0.3$, and 1.0 to demonstrate the sensitivity of cross approximate entropy measure to coupling strength. For each value of e, we generate 20 data segments from the coupled Rossler system, with each segment having 2000 samples. Fig. 6.4 shows exemplar time series of $x_1(t)$ and $x_2(t)$ for different coupling strengths. From Fig. 6.4a, we see that as e approaches 1.0, $x_2(t)$ becomes more synchronized with $x_1(t)$. In a coupled Rossler system where one oscillator drives the other, the dynamics of the receiver oscillator depends on the coupling strength and becomes more synchronized with the driver as coupling strength increases. From Fig. 6.4b, we see the changes in distribution of ApEn values for different e, showing that univariate ApEn can capture the change in dynamics (or regularity). Similarly, Fig. 6.4c shows the changes in distribution of XApEn values for different e, indicating that as the two oscillators become more synchronized, the cross approximate entropy value decreases, thereby capturing the amount of asynchrony between two time series data.



(a) Time series data



(b) ApEn of $x_2(t)$



(c) XApEn of $x_1(t)$ and $x_2(t)$

Figure 6.4: Illustration of utility of approximate entropy feature representation for quantifying regularity and cross-coupling on coupled Rossler model. (a) shows exemplar time series data synthesized from the coupled Rossler model for three different coupling strength $e = 0.1, 0.3, 1.0$. (b) and (c) respectively show the distribution of ApEn values of $x_2(t)$ and the distribution of XApEn values of $x_1(t)$ and $x_2(t)$ for 20 trials each for different values of e.

The dynamics in human motion can be considered as analogous to the dynamics of such coupled systems in that different coupling strength between body joints corresponds to different actions. For instance, actions two-hand wave and one-hand wave (as shown in Fig. 6.5) can be considered as cases with different coupling strengths between the two hand joints. Fig. 6.5 also illustrates that the proposed approximate entropy-based features can successfully differentiate the individual action sequences. This experimental analysis support our idea of using approximate entropy measures for quantifying regularity and cross-coupling in dynamics of human motion.

6.3.2 Segmentation and Action Classification

We evaluate the performance of using approximate entropy based feature for joint segmentation and recognition on publicly available action databases: Weizmann [54], MSR Action 3D [76], UTKinect-Action [152], and Florence3D-Action [115].

Evaluation Criterion: In our experiments, we follow the evaluation criterion introduced by Hoai *et al.* [60], where we calculate the overall frame-level accuracy as the ratio of the number of agreements (a match between the predicted label by our

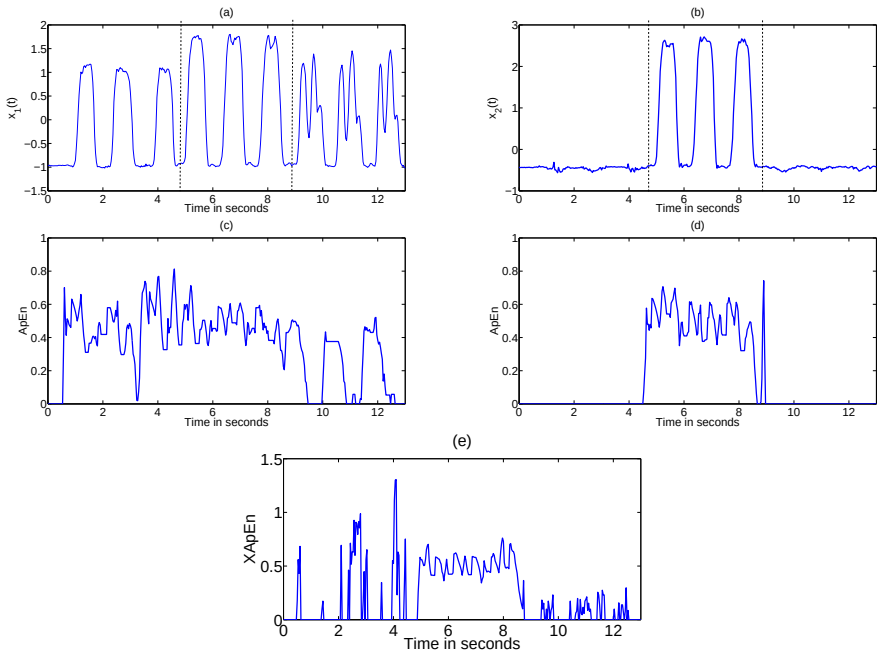


Figure 6.5: Approximate entropy features estimated on left and right hand trajectories of a subject performing horizontal arm wave, two-hand wave, and draw circle actions (in order) as shown in (a) and (b). The vertical lines in (a) and (b) denotes temporal action segments. (c) and (d) shows the univariate ApEn values estimated over a sliding window of 30 samples with parameters $m = 2$ and $r = 0.2$*std(x). The XApEn between left and right hand trajectories is shown in (e), where $m = 2$, and $r = 0.2$.

framework and the ground truth label) to the total number of frames. It should be noted that the evaluation criterion used in our framework is different from the traditional recognition accuracy, and hence our numbers cannot be compared with recognition accuracies on pre-segmented action data.

Weizmann Dataset

The Weizmann dataset [17] is a collection of 90 videos with 10 actions performed by 9 participants. The action classes are: bend, jack, jump, jump on two legs (pjump), run, gallop sideways (side), skip, walk, one handed wave (wave1), and two handed wave (wave2) (see Fig. 6.6). Since we are interested in the automatic segmentation-recognition problem, we have created long video sequences by concatenating the ac-

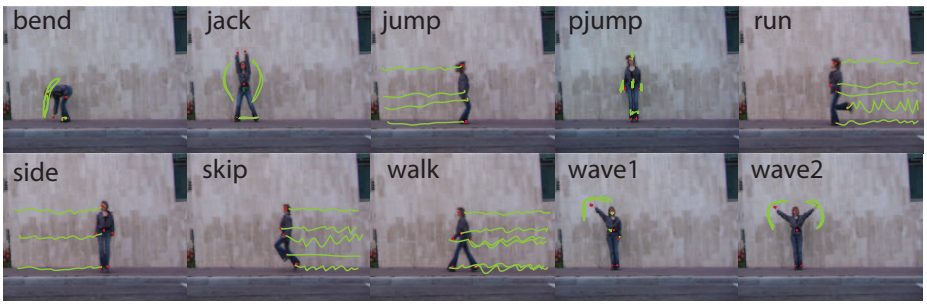


Figure 6.6: Typical video frames of 10 actions performed by subject-1 from the Weizmann dataset [54]. The trajectories corresponding to six body joints namely head, belly, two hands and two feet were extracted by Ali *et al.* [5], which will be used as input to our framework.

	bend	jack	jump	pjump	run	side	skip	walk	wave1	wave2
bend	**.975**	.017	0.002	0	0.002	0	0.002	0	0.003	0
jack	.019	**.981**	0	0	0	0	0	0	0	0
jump	0	.088	**.779**	.119	0	.013	0	0	0	0
pjump	0	0	.017	**.983**	0	0	0	0	0	0
run	0	0	.025	.179	**.592**	.063	.14	0	0	0
side	0	0	0	0	.014	**.938**	.039	.009	0	0
skip	0	0	0	0	.002	.144	**.667**	.187	0	0
walk	0	0	0	0	0	0	.054	**.858**	.088	0
wave1	0	0	.025	0	.016	0	0	.049	**.889**	.021
wave2	0	0	0	0	0	0	0	0	.039	**.961**

Table 6.1: Confusion table for Weizmann dataset for frame-level segmentation and action classification achieving a mean accuracy of 88.18%.

tion sequences performed by every participant to get 9 videos each containing 10 action sequences. The joint locations of two hands, two feet, head and belly were extracted by Ali *et al.* [5] through a semi-supervised joint detection and tracking. We extract the approximate entropy based features as described in section 6.2, and train a PLS regressor using leave-one-video-out cross validation scheme. Table 6.1 shows confusion matrix for simultaneous segmentation and action classification, and we achieve a mean accuracy of 88.18% in comparison with 87.7% reported by Hoai *et al.* [60].

Baselines for Kinect Datasets: The baselines used for comparison of evaluation results were:

(1) **Joint positions (JP):** We concatenate 3D coordinates of all the body joints, as in [140], to form our feature representation.

(2) **Recurrence plots (RP-3D):** These are matrices obtained by calculating distances between observation vectors of every pair of frames, such that the $(i, j)^{th}$ entry in the recurrence plot is the distance between observation vectors at i^{th} frame and j^{th} frame. The observation vector used is the vector of (x, y, z) co-ordinates of all body joints, and the distance metric used is Euclidean distance.

(3) **Linear Dynamical System (RP-LDS):** For each frame, we consider a window of 32 frames centered at the frame concerned. Using the vector of (x, y, z) coordinates of all body joints as feature vector for individual frames, we fit LDS [136] and extract the system parameters, A (transition matrix) and C (measurement matrix). Using (A_i, C_i) as the observation vector for the i^{th} frame, and the Martin distance as the distance metric, we construct recurrence plots. The recurrence plots for one sequence in MSR Action 3D dataset is shown in Fig 6.7, where the temporal segments of action sequences are clearly visible. Treating recurrence plots as textures, we extract texture-based per-frame feature in the following way: For each frame i, we extract a band of 32 rows, centered at the i^{th} row in the recurrence plot. We compute local binary pattern (LBP) feature [80] and use it as the feature vector for the i^{th} frame.

MSR Action 3D Dataset

The MSR Action 3D dataset consists of 20 actions performed by 10 subjects, with each subject performing every action twice or thrice. The dataset is comprehensive with many different actions and consists of a total of 557 action sequences. The dataset provides 3D joint positions and will be used as input to our framework. These 20 action classes were further divided into 3 Action Sets: AS1, AS2 and AS3 by Li *et al.* in [76] to account for large computation involved in classification of these actions.

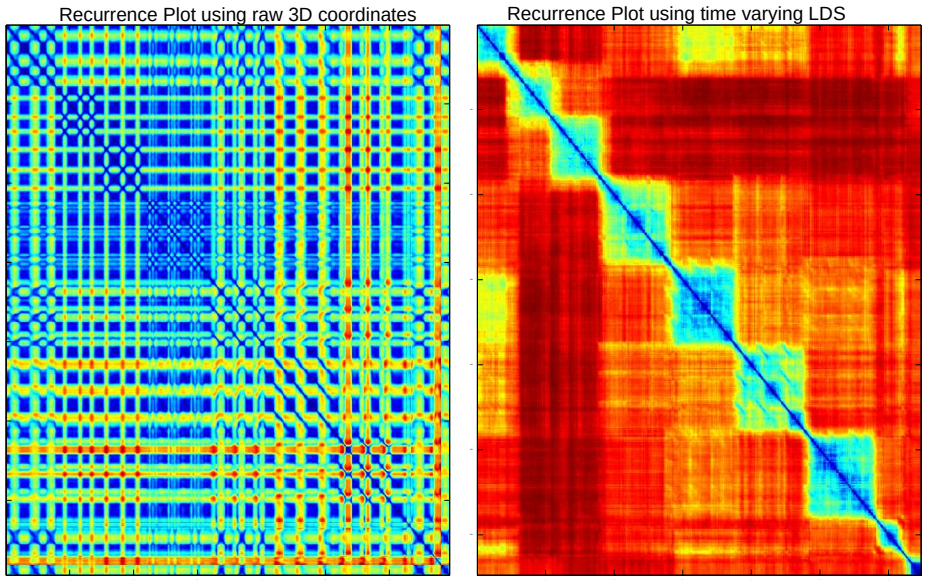


Figure 6.7: Exemplar recurrence plots generated from action sequences performed by a single subject in MSR Action 3D dataset. A distinctive structure is evident in the recurrence plot marking temporal segments of action sequences.

To generate a dataset suitable for the evaluation protocol in our framework, we concatenate the existing action sequences containing all actions performed by a given subject. This process of synthetically generating untrimmed action data from pre-segmented action sequences is repeated for all three action sets. We follow the cross-subject test setting as described in [76] using subjects $1, 3, 5, 7$, and 9 for training and the rest for testing. The evaluation results of frame-level accuracy are tabulated in Table 6.2, with a mean accuracy of 82.62% across all three action sets. The state-of-the-art recognition results reported on this dataset is 92.46% [140], but cannot be compared with our numbers as their framework requires access to pre-segmented action sequences and hence does not provide any segmentation accuracy. The mean accuracy achieved by the proposed framework is much higher than the three baseline representations indicating its usefulness.

Dataset	JP	RP-3D	RP-LDS	Proposed
AS1	27.51	55.12	52.22	**80.39**
AS2	30.66	43.98	45.77	**74.49**
AS3	37.72	57.86	53.04	**92.98**
Average	31.96	52.32	50.34	**82.62**

Table 6.2: Automatic segmentation and recognition performance on the MSR Action 3D dataset following the cross-subject evaluation protocol of [76] in that subjects 1,3,5,7, and 9 were used for training and the rest were used for testing.

UTKinect-Action Dataset

The UTKinect-Action dataset [152] consists of a total of 199 action sequences with 10 action classes performed twice by each of 10 subjects. In addition, the subject performs all actions in one go during recording, which makes it a suitable choice for our experimental protocol of simultaneous segmentation and action classification. The dataset provides 3D locations of 20 body joints, and is considered as a challenging dataset due to variations in view-point. The frame-level action segmentation and classification accuracy achieved using proposed feature representation was 80.3%, when using 50% subjects as training and rest as testing data. In comparison, baseline **JP** achieved highest accuracy of 66.6% across other baseline measures.

Florence3D-Action Dataset

The Florence3D-Action dataset [115] consists of 9 actions performed by 10 subjects twice or thrice resulting in a total of 215 action sequences. The dataset provides 3D locations of 15 body joints. We achieve an accuracy of 61.9% on the cross-subject test setting.

All the above experimental results clearly indicates that the proposed feature representation is a good choice for simultaneous segmentation and action classification.

Dataset	JP	RP-3D	RP-LDS	Proposed
UTKinect	66.6	56.0	37.6	**80.3**
Florence3D	46.9	54.3	48.5	**61.9**

Table 6.3: Automatic segmentation and recognition performance on the UTKinect and Florence3D action datasets following the cross-subject evaluation protocol of [160] in that 50% of the subjects were used for training and the rest were used for testing.

6.3.3 Temporal Segmentation

In this experiment, we use the publicly available Carnegie Mellon University motion capture database [1]. As in [157], we use the data collected from subject 86 with 14 markers placed on the most informative body joints with the motion capture system recording at 120 Hz. The dataset is a collection of 14 action sequences, each sequence containing multiple natural actions such as walking, punching, drinking, running. The main idea in [157, 158] is that such natural actions are inherently periodic, and this periodicity can be observed in the recurrence matrix showing block structures. Clustering methods such as spectral clustering can be used to cluster (segment) these blocks to achieve temporal segmentation of actions, and hence the clustering accuracy will greatly depend on the *quality* of the recurrence matrix. In this work, we demonstrate that quantifying regularity in actions using approximate entropy-based features can be used to improve the quality of recurrence matrix. We calculate the approximate entropy features as explained in section 6.2 over a sliding window and the estimated feature values are indexed to the center of the sliding window. The recurrence matrix is now calculated on the approximate entropy feature values instead of the time series data collected from the mo-cap system. Figure 6.8 shows an illustration of our proposed idea using one-dimensional time series data, where we clearly see that the recurrence matrix in (d) calculated from approximate entropy fea-

ture values looks more suitable to segment the three actions than recurrence matrix in (b) calculated directly from mo-cap raw time series data. We follow the evaluation protocol as in [157] using the Hungarian algorithm to find the optimum cluster correspondence and to compute clustering accuracy [22]. We compute the confusion matrix between the segmentation provided by the algorithm and the ground truth such that each entry C_{c_1,c_2} in the confusion matrix represents the total number of frames that belong to the cluster segment c_1 that are shared by the cluster segment c_2 in the ground truth. The accuracy is then given by the equation

$$accuracy = \mathbf{max}\frac{tr(C\mathbf{P})}{tr(C1_{k\times k})} \tag{6.4}$$

where $\mathbf{P} \in \{0,1\}^{k\times k}$ is a permutation matrix.

Figure 6.9 shows exemplar segmentation results obtained using the approximate entropy-based features along with Spectral Clustering (SC) and HACA on two action sequences. Different colors mark different actions and the ground truth segmentation was obtained from human observers. In both these examples we see that using approximate entropy features provides better segmentation than just using SC or HACA on mo-cap time series data. Due to space constraints, we only show the segmentation results on two sequences. We report the average segmentation accuracy using various features in Table 6.4, which further supports our claim that using the proposed approximate entropy-based features along with a clustering approach will provide better segmentation accuracy compared to using a clustering approach on mo-cap time series data.

6.3.4 Movement Quality Assessment

The main objective in this experiment is to quantitatively assess the quality of movements performed by stroke survivors. We use the approximate entropy based feature

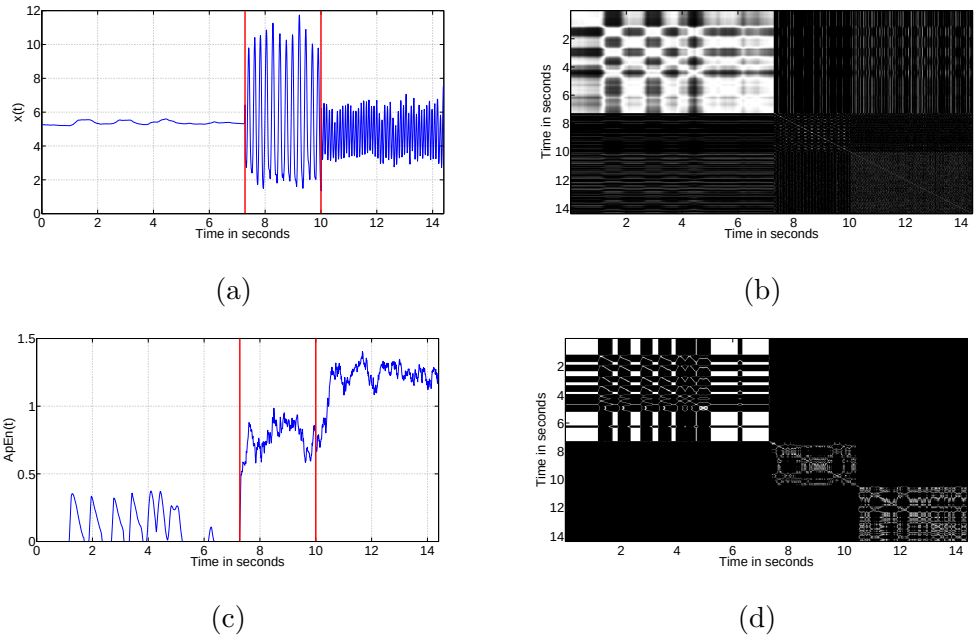


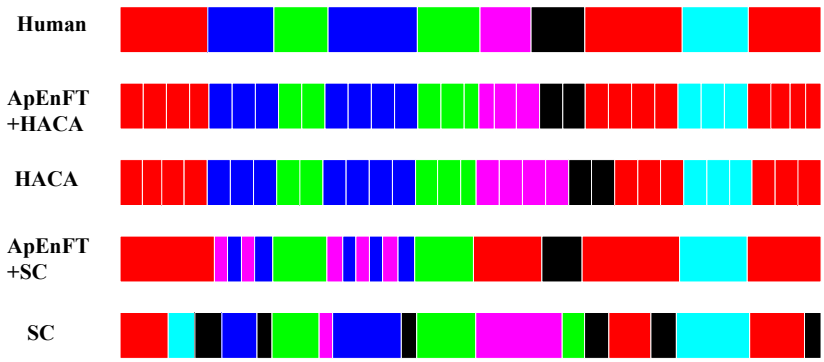
Figure 6.8: Illustration of utility of approximate entropy feature for quantifying regularity and improving quality of recurrence matrix. (a) shows exemplar time series data collected from hip joint of a subject performing *DANCE, JUMP* and *RUN* actions, (c) shows the corresponding ApEn feature values, (b) and (d) respectively show the recurrence matrix estimated on raw time series data in (a) and ApEn feature values in (c).

Method	**Avg. Accuracy**
ApEnFT+HACA	**0.93**
HACA	0.91
ApEnFT+SC	**0.86**
SC	0.75

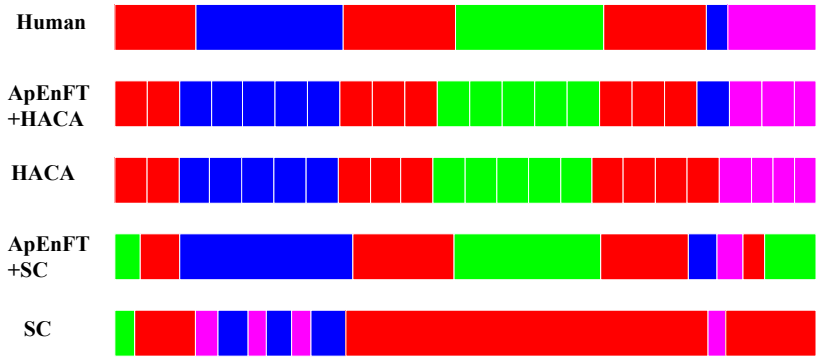
Baseline	**Avg. Accuracy**
UniAp+HACA	0.67
UniAp+SC	0.56
Dynamics+HACA	0.65
Dynamics+SC	0.63

Table 6.4: Comparison of average temporal segmentation accuracy for various methods.

representations to computationally estimate the impairment scores as assigned by traditional clinical measures such as the Wolf Motor Function Test [149]. The data was collected using a motion capture system using 14 markers on the right-hand, arm and torso. A total of 15 impaired subjects perform reach and grasp movements



(a)



(b)

Figure 6.9: Comparison of temporal clustering methods on the CMU motion capture dataset. Different colors indicate different actions. Ground truth motion segmentation was provided by human observers.

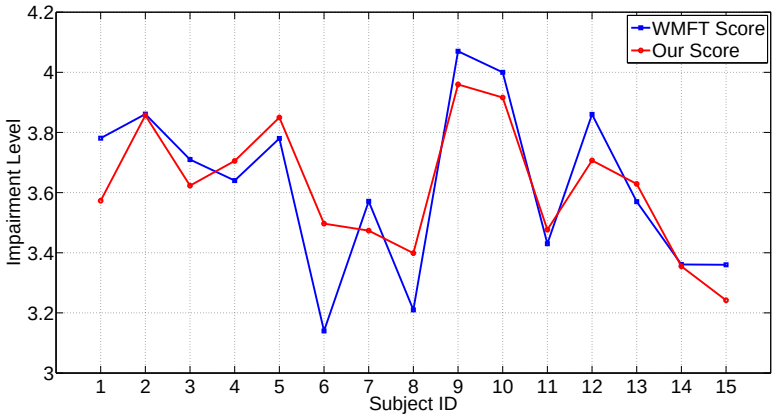


Figure 6.10: The impairment scores assigned to movements performed by stroke survivors by the WMFT (in blue) and our proposed framework (in red). The pearson correlation coefficient was found to be 0.8603.

to a target. The experimental protocol defined in [142] allows us to use only the data corresponding to single marker on the wrist. Using the approximate entropy based feature representation as explained in section 6.2 along with a PLS regressor, we achieve a correlation coefficient of 0.8603 with the scores assigned by the WMFT protocol as compared to correlation coefficient of 0.8527 reported in [142].

6.3.5 Action Quality Assessment on Diving Datasets

In the next experiment, we show that the proposed feature can also be used to quantify the quality of diving actions. For this experiment, we use the diving dataset released by Pirsiavash *et al.* [101] which is a collection of videos downloaded from YouTube. The diving dataset consists of 159 videos of diving actions performed by multiple subjects with their respective quality scores given by expert judges. The dataset also provides estimated pose for each frame of the video which is used as input to our framework. The problem of quantifying the quality of diving actions on this dataset is shown to be challenging by the experimental analysis done by Pirsiavash *et al.* in [101], where the best performance achieved was of mean rank correlation of

Method	STIP	Hierarchical	Pose+DFT	Pose+DCT	UniAp	Dynamics	Proposed
SVR	0.07	0.19	0.27	0.41	0.05	0.17	**0.45**

Table 6.5: Mean rank correlation for various methods. Our proposed feature achieves 10% improvement in the correlation coefficient compared to the state-of-the-art. [101] reported correlation coefficient using STIP, hierarchical and pose+DCT features.

0.41 between predicted scores and ground truth scores given by judges. We use the same evaluation protocol of generating random training and testing example splits 200 times as introduced in [101] with 100 instances as training examples and the rest as testing examples. Using the estimated pose for each frame, we calculate the approximate entropy features as explained in section 6.2 for different values of radius ($r = 0.1, 0.12, 0.14, 0.18$) and concatenate to get a high-dimensional feature vector. Using PCA to achieve dimensionality reduction and an SVM regressor to generate real-valued scores indicative of the quality of diving actions, we show that our approximate entropy-based feature performs better than the traditional DCT-based feature. We believe that this is achieved due to the fact that our feature encodes the dynamical information in the time series of poses while DCT does not. In addition, traditional approaches consider each joint independently, while the proposed framework incorporates the interdependency between the joints. The results are tabulated in Table 6.5 and we achieve a rank correlation of 0.45 in comparison with 0.41 reported in [101].

7 MULTIVARIATE EMBEDDING BASED QUALITY ASSESSMENT OF DIVING ACTIONS

7.1 Introduction

The vision community has been interested in modeling human motion for numerous applications including video surveillance, automatic video annotation, and health monitoring [4]. Many methods have been proposed in the literature to model the underlying dynamics in human motion, and forms the core idea for activity analysis. An *activity* can be seen as a resultant of coordinated movement of body joints and their respective interdependencies to achieve a goal-directed task. Traditional approaches to model the dynamics operate on the level of individual dimensions of body joints of the human body [5]. Only recently, researchers have started exploring relationships between body joints, using rotations and translations in 3D space [140]; however these approaches lack dynamical information. In this paper, we use the multivariate embedding approach to model the dynamics of individual body joints, and show improved performance on fine-grained quality assessment of actions.

After achieving adequate success in recognizing actions from videos, researchers in the vision community have become interested in fine-grained analysis of human activities. Frameworks for quantification of movement quality for applications in stroke rehabilitation and sports have been developed [101, 142]. In this paper, we focus our interest on the quantification of quality of diving actions from RGB videos

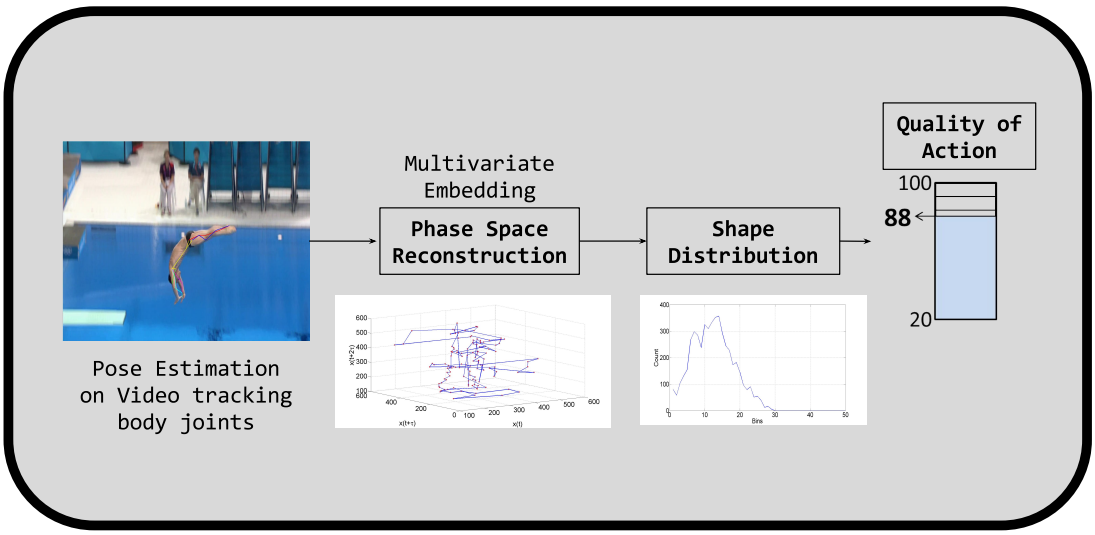


Figure 7.1: Block diagram showing the algorithmic flow of our framework for assessing quality of diving actions using videos. The dataset provides access to high-level pose features for each frame.

as shown in Fig. 7.1. We propose the use of global descriptors of the shape of the attractor of the dynamical system as a feature as in [142, 141] to quantify the quality of diving actions.

A tremendous amount of research has been conducted on activity analysis using RGB videos [4, 51]. Traditional dynamical modeling approaches for action recognition include parametric methods such as Hidden Markov Models (HMMs) and Linear Dynamical Systems (LDSs), which have been used for computer vision applications like action recognition [146, 154] and gait analysis [16]. Such parametric approaches assume a model and impose it on the data trying to fit the observed data to the assumed model. Recent work by Ali *et al.* proposed the use of a nonparametric modeling approach using ideas from chaos theory to model the dynamics in human actions [4]. The authors use largest Lyapunov exponent, correlation dimension and correlation integral from trajectories of action data as part of their feature representation. These traditional chaotic measures have been extensively used to model human actions [5, 37, 95, 127]. However, [109] and [132] have shown that these non-

linear dynamical measures need large amounts of data to have good estimates using the existing algorithms (10^m, where m is the embedding dimension). In [142, 141], the authors propose a shape-theoretic framework for dynamical analysis of human movement from 3D data. They use global descriptors of the shape of the attractor of the dynamical system as a feature for modeling actions. The shape distribution descriptor operated on an individual dimension of body joints using a univariate embedding technique to reconstruct the attractor. Here we use multivariate embedding to reconstruct the attractor and we show improved results on the application of interest.

7.2 Framework

In this section, we introduce the necessary background and present each block in the pipeline of our framework.

7.2.1 Phase Space Reconstruction

Sensing systems such as RGB cameras allow us to sense a series of 2D images of human movement which is a result of projection of high dimensional data onto 2D space without allowing us to observe all the variables of the system. One would prefer to have access to all independent variables of the system and their interactions for a complete understanding of the system. Traditional parametric approaches assume an underlying mapping function f to describe the dynamics of the system. The theory of chaotic systems allows for determining certain invariants of the dynamical system function f without making any assumptions about the system using a method called phase space reconstruction.

The *phase space* is defined as the space with all possible states of a system [145, 3]. Given one-dimensional time series data, we can reconstruct the important topo-

logical properties of the original dynamical system using time delay embedding as proposed by Takens [**?**]. This process finds the mapping function f between the one-dimensional observed time series and the m-dimensional attractor, with the assumption that all variables of the system influence one another.

For a discrete dynamical system with a multidimensional phase space, time-delay vectors (or embedding vectors) are obtained by concatenation of time-delayed samples given by

$$\mathbf{x}_i(n) = [x_i(n), x_i(n+\tau), \cdots, x_i(n+(m-1)\tau)]^T. \tag{7.1}$$

where m is the embedding dimension and τ is the embedding delay. This method of embedding is called as *Univariate Embedding*, as the method uses one-dimensional observed time series data to recover the system dynamics. The embedding theorem by Takens does not suggest optimal values for parameters m and τ, but there are approaches in literature to estimate these values such as the false nearest neighbors [68] for m and the first zero crossing of the autocorrelation function [122] for τ.

Recent theoretical and empirical findings have demonstrated that multivariate embedding of time series data by simple concatenation of individual univariate embedding vectors achieves good phase space reconstruction as evaluated by the shape and dynamics distortion measures [144] and significantly improves the attractor reconstruction when compared to univariate embedding. In this work, we propose to use the multivariate embedding procedure as described by Cao *et al.* [23] per body joint to reconstruct the attractor.

Multivariate Embedding – This simple yet powerful extension of univariate embedding as proposed by Cao et al. [23] has proven to be useful in computer vision applications such as action synthesis and dynamic texture synthesis [12]. The findings in [144] indicate that in scenarios with access to more than one-dimensional observed time series data, one can reconstruct the phase space better using the multivariate

Figure 7.2: Exemplar video frames shown from the diving action dataset collected by [101].

embedding approach compared to univariate embedding. Given multivariate time series data $[x_i(n)]_{n=1}^{T}$, $i = 1, \ldots, p$, where p is the dimension of time series data, the reconstructed phase space vector is of the form

$$\begin{aligned} \mathbf{z}_t = [&x_1(n), x_1(n+\tau_1), \ldots, x_1(n+(m_1-1)\tau_1), \\ &x_2(n), x_2(n+\tau_2), \ldots, x_2(n+(m_2-1)\tau_2), \\ &\ldots, \\ &x_p(n), x_p(n+\tau_p), \ldots, x_p(n+(m_p-1)\tau_p)]^T. \end{aligned} \tag{7.2}$$

where m_i and τ_i are respectively the embedding dimension and time delay for each of the p-dimension in the multivariate time series data. The authors in [144] define evaluation metrics shape distortion and dynamics distortion to quantify the quality of reconstructed phase space and discover that multivariate embedding outperforms univariate embedding in reconstructing the dynamics of standard nonlinear dynamical models such as the Lorenz, Rossler, and coupled Rossler and Lorenz. We use the multivariate embedding technique in our framework and show improvement over other methods in the literature for quality assessment of diving actions.

7.2.2 Features from Reconstructed Phase Space

In this section, we present various feature representations extracted from the reconstructed phase space.

Traditional Dynamical Invariants

Quantifying divergence of closely spaced trajectories in the reconstructed phase space is a well-studied problem in the field of chaos theory. Largest Lyapunov exponent [148], correlation sum [3], and correlation dimension [3] are a few examples of invariant measures proposed in the literature to quantify complexity of nonlinear dynamical systems. The largest Lyapunov exponent is a measure of the average rate of divergence (or convergence) of initially closely-spaced trajectories over time [3, 145]. Correlation sum is an invariant used to quantify density of points in the reconstructed phase space. Correlation dimension or the fractal dimension defines the dimensionality of the reconstructed phase space, and is given by the slope of the line in the plot of correlation sum for different values of radius (r) versus radius. The main contribution of our work is to propose a better way to encode dynamics compared to traditional chaotic invariants. To evaluate the effectiveness of our framework, we provide comparative results with a feature vector [1] of traditional chaotic invariants obtained by concatenating the largest Lyapunov exponent, correlation dimension and correlation integral (for 8 values of radius) resulting in a 10-dimensional feature vector denoted as ***Chaos***.

Attractor Shape Distributions

In this section, we discuss the feature representation used in [142, 141] for action recognition and quality assessment by quantifying the shape of the reconstructed phase space. Using the idea proposed by Osada *et al.* in [92], we compute shape distribution of the reconstructed phase space sampled from the shape function by measuring the global geometric properties. It is said that any function can be used

[1] Code available at
http://www.physik3.gwdg.de/tstool/HTML/index.html

Algorithm 1 Multivariate Attractor Shape Distribution

1: Input: $\mathbf{x}(n) \in \mathbb{R}^D, n = 1, \ldots T$

2: **for** $dim = 1 \rightarrow D$ **do**

3: Reconstruct attractor using method of delays [**?**].

$$\mathbf{z}_t = [x_1(n), x_1(n+\tau_1), \ldots, x_1(n+(m_1-1)\tau_1),$$
$$x_2(n), x_2(n+\tau_2), \ldots, x_2(n+(m_2-1)\tau_2),]^T.$$

4: **for** $iter = 1 \rightarrow N$ **do**

5: $\mathbf{DT2}_{ij} = e^{-\gamma|t_i - t_j|} * ||\mathbf{z}_i - \mathbf{z}_j||_2$.

6: **end for**

7: Calculate histogram with 50 bins on $\mathbf{DT2}_{ij}$.

8: **end for**

to extract the shape distribution [92]; we use the shape function which encodes information about dynamical evolution in the phase space as proposed by [141] given by,

$$\mathbf{DT2}_{ij} = e^{-\gamma|t_i - t_j|} * ||\mathbf{z}_i - \mathbf{z}_j||_2, \tag{7.3}$$

where t_i and t_j are the time indexes of the randomly selected pair of embedding vectors in the reconstructed phase space. δ and γ are empirically determined parameters such that δ, $\gamma \geq 0$. $\mathbf{z}_i$ and $\mathbf{z}_j$ are points (embedding vectors) in the reconstructed phase space. A set (of size N) of these distances for randomly chosen embedding vector pairs are computed. From this set, we construct a histogram by counting the number of samples which fall into each of $B = 50$ fixed sized bins to obtain the attractor's shape distribution. The procedure to extract the multivariate attractor shape distribution from a given multivariate time series data is outlined in algorithm 1.

7.3 Experimental Evaluation

The proposed framework for representation of dynamics was evaluated on the diving action video dataset.

7.3.1 Diving Action Dataset

The diving action dataset was collected by a research group at MIT [101] consists of 159 videos of diving actions performed by athletes participating in the Olympics. Each video has approximately 150 frames. The ground truth labels were collected from judges whose scores varied between 20 (worst) and 100 (best). An exemplar diving action is shown in Fig. 7.2 showing the transition from left to right. The dataset provides access to high-level pose features using a pose estimation algorithm [155] for each frame independently and links the poses using a dynamic programming algorithm to find the best track of poses in the entire video. We use the evaluation protocol of generating random training and testing example splits 200 times as introduced in [101] with 100 instances as training examples and the rest as testing examples. Using the estimated pose for each frame, we perform multivariate embedding on each body joint ($p = 2$) and concatenate the calculated attractor shape distribution feature for all body joints to form our feature representation. Using an SVM regressor, we generate real-valued scores indicative of the quality of diving actions.

The problem of quantifying the quality of diving actions on this dataset is shown to be challenging by the experimental analysis done by Pirsiavash *et al.* in [101], where the best performance achieved was of mean rank correlation of 0.41 between predicted scores and ground truth scores given by judges using Discrete Cosine Transform (DCT) on the estimated poses. The authors in [101] use DCT to reject noise due to pose estimation errors by keeping only the low frequency components to cre-

Table 7.1: Mean rank correlation for various methods. Our proposed feature achieves 10% improvement in the correlation coefficient compared to the state-of-the-art. [101] reported correlation coefficient using STIP, Hierarchical, and Pose+DCT features.

Method	**Mean Rank Correlation**
STIP [101]	0.07
Hierarchical [101]	0.19
Pose+DFT [101]	0.27
Pose+DCT [101]	0.41
Pose+Chaos [5]	0.17
Pose+Univariate **DT2** [141]	0.24
Proposed	**0.45**

ate the feature vector. We achieve a mean rank correlation of 0.45 in comparison with 0.41 reported in [101] using Pose+DCT. We also show comparative results with STIP (Space Time Interest Points), low-level hierarchical features and pose-based features with Discrete Fourier Transform (DFT) which performs significantly worse compared to our method. Use of the traditional chaotic invariants achieves only mean correlation coefficient of 0.17. These results indicate that the proposed feature representation provides a better way to encode the temporal information in diving action and is also robust to noise due to pose estimation errors.

7.4 Conclusion

In this paper, we are interested in the problem of fine-grained assessment of the quality of actions with real-world applications in sports. The proposed framework is an extension of the shape theory based dynamical analysis framework for movement quality assessment and action recognition introduced in [142, 141]. In this work, we

use the multivariate embedding approach to reconstruct the dynamical attractor per body joint as opposed to the traditional way of operating on individual dimensions of time series data. In addition, the proposed framework addresses the drawbacks of traditional measures from chaos theory by combining the concepts of nonlinear time series analysis and shape theory to extract robust and discriminative features from reconstructed phase space.

8 PERSISTENT HOMOLOGY OF ATTRACTORS FOR ACTION RECOGNITION

The rapid technological advancements in sensing and computing has resulted in large amounts of data warranting the development of new methods for their analysis. In the past decade, topological data analysis (TDA) has shown to be a promising new paradigm for analyzing and deriving inferences [24]. In this paper, we explore the suitability of TDA for analyzing human actions by modeling each action as a dynamical system and extracting the topological features of the attractor. These features are then used in a demonstrative application of classifying actions.

The task of recognizing human activities has a wide range of applications such as surveillance, health monitoring and animation. Modeling the spatio-temporal evolution of human body joints is traditionally accomplished by defining a state space and learning a function that maps the current state to the next state [16, 104]. An alternate approach proposed derives a representation for the dynamical system directly from the observation data using tools from chaos theory [5, 142, 141, 143], thereby learning a generalized model representation suitable for analyzing a wide range of dynamical phenomenon. In this paper, we use the framework proposed in [5, 142] to extract a reconstructed phase-space from the available time series data, which preserves the topological properties of the underlying dynamical system of a given action. We treat the reconstructed attractor as a point cloud and we extract topological features from the point cloud based on persistent homology [45, 25].

8.1 Related Work

Human activity analysis is a well-studied problem in the vision community with extensive literature on the subject. We suggest the readers to refer [4, 51] for a detailed review of the approaches for modeling and recognition of human activities. Since our contribution in this paper is related to topological data analysis and non-parametric approaches for dynamical system analysis for action modeling, we restrict our discussion to related methods.

Activity Analysis using Dynamical Invariants: Traditional methods for action recognition by parametric modeling approaches impose a model and learn the associated parameters from the training data. Hidden Markov Models (HMMs) [103] and Linear Dynamical Systems (LDSs) [26] are the most popular parametric modeling approaches employed for action recognition [154, 146, 139, 33] and gait analysis [65, 77, 16]. Nonlinear parametric modeling approaches like Switching Linear Dynamical Systems (SLDSs) have been utilized to model complex activities composed of sequences of short segments modeled by LDS [20]. While, nonlinear approaches can provide a more accurate model, it is difficult to precisely learn the model parameters. In addition, one would only approximate the true-dynamics of the system with attempts to fit a model to the experimental data. An alternative nonparametric action modeling approach based on tools from chaos theory, with no assumptions on the underlying dynamical system like the largest Lyapunov exponent, correlation dimension and correlation integral, have been extensively used to model human actions [5, 37, 95, 127].

Topological Data Analysis: Topological data analysis has gained its importance in analyzing point cloud data [25], and is seen as a tool to obtain the *shape* of high-dimensional data as opposed to geometric approaches that try to understand the *size*

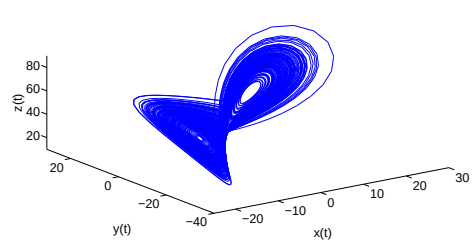
of the data. Such tools are also very useful in visualization applications [120, 41]. The representations of persistent homology such as persistence diagrams and barcodes have several applications, such as speech signal analysis [21], wheeze detection [46], document structure representation [159], detection of cancer [85], characterizing decision surfaces in classifiers [**?**] to name a few. There are also a number of freely available software for computing persistent homology from point clouds [131**?**].

Contributions: Our work has the following contributions: (1) We treat the reconstructed phase-space of the dynamical system as a point cloud and derive features based on homological persistence. (2) We incorporate links between adjacent time points when building simplicial complexes from the point cloud. (3) We demonstrate the value of the proposed framework in an action recognition task on a publicly available motion capture dataset, using a nearest neighbor classifier with the the persistence-based features.

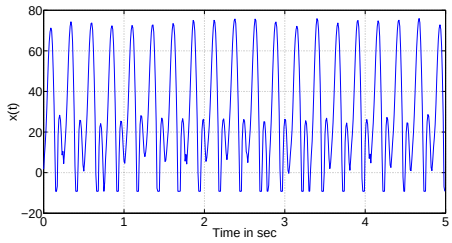
Outline: In section 8.2, we introduce the theoretical concepts of phase-space reconstruction and persistent homology. The feature which encodes the temporal evolution information in the persistence diagrams will be introduced in section 8.3. In section 8.4, we present our experimental results on the motion capture dataset [5].

8.2 Preliminaries

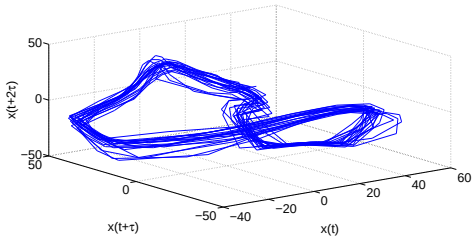
In this section, we introduce the background necessary to develop an understanding of nonlinear dynamical system analysis using tools from chaos theory and persistent homology.



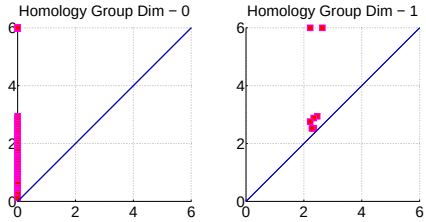
(a) Lorenz Attractor



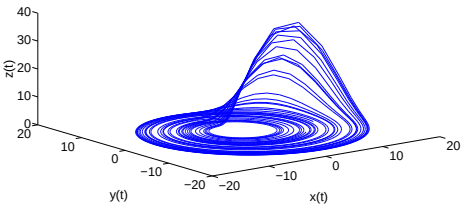
(b) Time series data



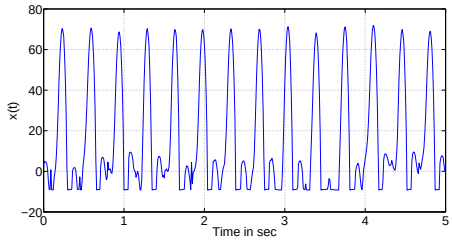
(c) Reconstructed Phase Space



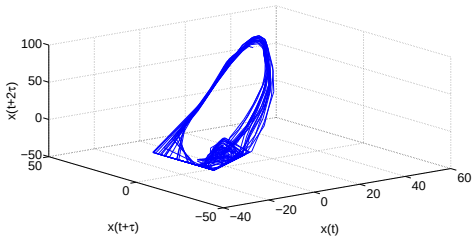
(d) Persistence Diagram



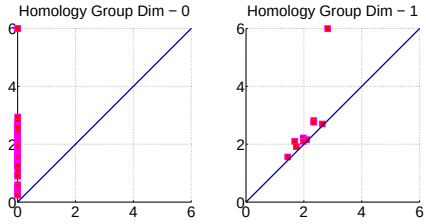
(e) Rossler Attractor



(f) Time series data



(g) Reconstructed Phase Space



(h) Persistence Diagram

Figure 8.1: Phase space reconstruction of dynamical attractors by delay embedding. (a), (e) shows the 3D view of trajectories of Lorenz and Rossler attractors. The one-dimensional time series (observed) of the Lorenz and Rossler systems are shown in (b), (f). (c), (g) shows the reconstructed phase-space from observed time series using delay embedding. The above example illustrates that the reconstructed phase-space preserves certain topological properties of the original attractor.

8.2.1 Phase Space Reconstruction

The data that we obtain from sensors is usually a projection of the original dynamical system to a lower dimensional space, and hence do not represent all the variables in the system. Hence, the available data is insufficient to model the dynamics of the system. To address this, we have to employ methods for reconstructing the attractor to obtain a phase-space which preserves the important topological properties of the original dynamical system. This process is required to find the mapping function between the one-dimensional observed time series data and the m-dimensional attractor, with the assumption that all variables of the system influence one another. The concept of phase-space reconstruction was proposed in the embedding theorem proposed by Takens, called Takens' embedding theorem [129]. For a discrete dynamical system with a multidimensional phase-space, time-delay vectors (or embedding vectors) are obtained by concatenation of time-delayed samples given by

$$\mathbf{x}_i(n) = [x_i(n), x_i(n+\tau), \cdots, x_i(n+(m-1)\tau)]^T. \tag{8.1}$$

where m is the embedding dimension and τ is the embedding delay. The idea here is that for a sufficiently large m, the important topological properties of the unknown multidimensional system are reproduced in the reconstructed phase-space [3]. The process of phase-space reconstruction from a one-dimensional observed time-series of a Lorenz and Rossler system is shown in Fig 8.1, where the reconstructed phase-space and the original attractor are topologically equivalent.

8.2.2 Persistent Homology

Consider a point cloud of T data samples in $\mathbb{R}^D$: $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T]^T$. The point cloud data can be viewed as samples from a unknown shape. Our aim is to estimate the topological properties of the underlying shape by constructing a simplicial

complex $\mathcal{S}$ using the point cloud $\mathbf{X}$ and examining the topology of the complex. A simplicial complex is a set of simplices constructed from $\mathbf{X}$ glued together in a particular way. It is denoted by $\mathcal{S} = (\mathbf{X}, \Sigma)$, where Σ is a family of non-empty subsets of $\mathbf{X}$, with each element $\sigma \in \Sigma$ being a simplex. The other necessary condition is that $\sigma \in \Sigma$ and $k \subseteq \sigma$ implies that $k \in \Sigma$. The simplices are usually constructed using some neighborhood rule, such as the $\epsilon-$neighborhood, where ϵ is the scale parameter.

We are interested in computing the rank of homology groups of a given dimension, aka, Betti numbers (β), since they are one of the simple but informative characterizations of topology of the point cloud. Betti-0 or β_0 denotes the number of connected components, β_1, the number of holes of dimension-1, β_2, the number of holes of dimension-2 and so on. Betti numbers depend on the scale (which is same as the scale used with $\epsilon-$nearest neighbors) at which the complex is constructed. Homology groups that are stable across a wide range of scale values, i.e., *persistent homology groups*, are the ones that provide the most information about the underlying shape. Homology that do not persist are considered to be noise. The Betti numbers of a given dimension can be compactly encoded in a $2-$dimensional plot, which provides the birth versus death times of each homology group, also known as the persistence diagram. Persistence diagrams are multi-sets of points, with infinite number of points on the diagonal where birth time equals death time. They admit several metrics and hence distances between two diagrams can be estimated numerically [69].

Various approaches exist for constructing simplicial complexes from $\mathbf{X}$ at a given scale ϵ. In our work, we use the Vietoris-Rips (VR) complex, $\text{VR}(\mathbf{X}, \epsilon)$, where a simplex is created if and only if the Euclidean distance between every pair of points is less than ϵ [161]. Efficient construction of the VR complex can proceed by creating an ϵ-neighborhood graph, also referred to as the one-skeleton of $\mathcal{S}$. Then inductively, triplets of edges that form a triangle are taken as two-dimensional simplices, sets of

Algorithm 2 Persistence diagrams from phase-space

1: **Input**: $\mathbf{x}_i(n) \in \mathbb{R}^D, n = 1, \dots T$

2: **Output**: Persistence diagram for homology group dimensions 0 & 1.

3: **for** $i = 1 \rightarrow D$ **do**

4: Reconstruct attractor using method of delays [3]

$\mathbf{x}_i(n) = [x_i(n), x_i(n+\tau), \cdots, x_i(n+(m-1)\tau)]^T$.

5: Construct metric space encoding temporal evolution

Temporal link between $[\mathbf{x}_i(n-1), \mathbf{x}_i(n), \mathbf{x}_i(n+1)]$.

6: Build Vietoris-Rips complexes [131, 161]

7: **end for**

four two-dimensional simplices that form a tetrahedron are taken as three-dimensional simplices, and so on. This is repeated for increasing values of scale, known as filtration, and the persistence diagrams are estimated. Although several types of topological features can be extracted from point clouds, in our work, we will use it to refer exclusively to persistence diagrams.

8.3 Topological Features from Attractor

Although VR complexes can successfully retrieve the topological features of a general point cloud, topological features that incorporate the dynamical evolution in phase-space can model actions better. In this section, we present a method to encode temporal information in persistence diagrams which in turn can be used as a representative topological feature for the reconstructed phase-space.

Methods to build simplicial complexes from the point cloud data, such as the VR filtration approach, only takes into consideration the adjacency in space, but not in time. An activity is a resultant of coordinated movement of body joints and their re-

spective interdependencies to achieve a goal-directed task with temporal information in trajectories of body joints. Modeling the underlying dynamics in the trajectories forms the core idea in designing action recognition systems. Therefore, we explicitly we create temporal links between $\mathbf{x}_i(n-1)$, $\mathbf{x}_i(n)$, and $\mathbf{x}_i(n+1)$ in the one-skeleton of $\mathcal{S}$, thereby creating a metric space which encodes adjacency in both space and time. The persistence diagrams for homology groups of dimensions 0 and 1 are then estimated. The pseudo code for our framework is outlined in algorithm 2.

As a demonstrative example, we use this approach to estimate the persistence diagrams of Lorenz and Rossler attractors. From Fig. 8.1, we see that for the Lorenz attractor, the ranks of homology groups that persist are, $\beta_0 = 1$ and $\beta_1 = 1$, whereas for the Rossler attractor, $\beta_0 = 1$ and, $\beta_1 = 2$. Clearly they indicate the connected components and $1-$dimensional holes in each of the cases. Note that the points close to the diagonal are considered to be noise with their birth and death times being close to each other. Therefore these points represent homology groups that die in a short time after they are born.

Distance Between Persistence Diagrams: For any two persistence diagrams X and Y, the distance between the diagrams are usually quantified using the bottleneck distance or the $q-$Wasserstein distance [69]. In our experiments, we use the 1-Wasserstein distance given by,

$$W_1(X,Y) = \inf_{\eta: X \to Y} \sum_{x \in X} ||x - \eta(x)||_1 \tag{8.2}$$

Since each diagram contains an infinite number of points in the diagonal, this distance is computed by pairing each point in one diagram uniquely to another non-diagonal or diagonal point in the other diagram, and then computing the distance. This can be efficiently obtained with the Hungarian algorithm or using a more efficient variant [69].

8.4 Experimental Results

The proposed framework for topological data analysis for action representation was evaluated on the motion capture dataset [5].

Baseline: To evaluate the effectiveness of our framework, we provide comparative results using $10-$dimensional feature vectors [1] of traditional chaotic invariants obtained by concatenating the largest Lyapunov exponent, correlation dimension and correlation integral (for 8 values of radius). The results with this approach are denoted with *Chaos* in Table 8.1. We also tabulate the results using persistence diagrams obtained from VR filtrations with no additional temporal encoding (*VR Complex*), and a recent shape-theoretic framework **D2** and **DT2** [141]. The evaluation with VR complexes follow the same protocol as our proposed approach described below.

8.4.1 Motion Capture Data

We evaluate the performance of the proposed framework using 3-dimensional motion capture sequences of body joints used in the [5]. The dataset is a collection of five actions: *dance, jump, run, sit* and *walk* with $31, 14, 30, 35$ and 48 instances respectively. The dataset provides $3-$dimensional time-series from 17 body joints which were further divided into scalar time-series resulting in a 51-dimensional vector representation for each action. We generate 100 random splits having 5 testing examples from each action class and use a nearest neighbor classifier with the $1-$Wasserstein distance measure. The mean recognition rates for the different methods are given in Table 8.1. Traditional chaotic invariants (*Chaos*) only achieves a mean recognition rate of 52.44%. The best classification performance reported on the dataset uses **DT2** dynamical shape feature achieves a mean recognition rate of 93.92% which en-

[1] Code available at
http://www.physik3.gwdg.de/tstool/HTML/index.html

Table 8.1: Comparison of classification rates for different methods using nearest neighbor classifier on the motion capture dataset.

Method	**Mean Accuracy (%)**	**Std. dev**
Chaos [5]	52.44	0.0081
VR Complex [131]	93.68	0.0054
D2 [142]	91.96	0.0036
DT2 [141]	93.92	0.0051
Proposed	**96.48**	**0.0053**

Table 8.2: Confusion table for motion capture dataset using our proposed framework which achieves mean classification rate of 96.48%.

Action	**Dance**	**Jump**	**Run**	**Sit**	**Walk**
Dance	**0.98**	0	0	0.02	0
Jump	0.08	**0.92**	0	0	0
Run	0	0	**0.96**	0	0.04
Sit	0.03	0	0	**0.97**	0
Walk	0	0	0.01	0	**0.99**

codes temporal information. In comparison, our proposed method achieves 96.48% which is significantly better than the results achieved by any of the previous methods. Clearly, topological persistence features are informative, since they summarize the feature evolution over a range of scale values when compared to chaotic invariants such as largest Lyapunov exponents. The standard deviation of classification accuracy over the different random splits are also tabulated. The class confusion matrix for the proposed framework is shown in Table 8.2.

9 Conclusion and Future Directions

In this work, we have proposed a shape theoretic dynamical analysis framework for applications in action and gesture recognition, movement quality assessment for stroke rehabilitation and dynamical scene classification. We address the drawbacks of traditional measures from chaos theory for modeling the dynamics by proposing a framework combining the concepts of nonlinear time-series analysis and shape theory to extract robust and discriminative features from the reconstructed phase space. Our experiments on nonlinear dynamical models and joint trajectory data from motion capture support our hypothesis that the *shape* of the reconstructed phase space can be used as feature representation for the above discussed applications. Furthermore, the wide range of experimental analysis on publicly available datasets for recognition of actions, gestures and scenes validate our claims. The framework was also tested on movement analysis on a finer scale, where we were interested in quantifying the *movement quality* (level of impairment) for applications in stroke rehabilitation. Our experiments using a single marker indicate that with combination of dynamical features and machine learning tools, we are able to achieve comparable performance levels to a heavy marker-based system in movement quality assessment.

We have also proposed the use of an approximate entropy-based feature representation to quantify dynamical regularity in time series of action data for applications in (a) temporal segmentation of actions and (b) quantification of quality of diving actions. The novelty in the proposed feature is in the use of the multivariate embedding

approach for approximate entropy to model dynamics in individual body joints and cross approximate entropy to model interaction between body joints. Using nonlinear dynamical models such as the coupled Rossler system, we showed that the proposed feature is sensitive to changes in coupling factor, analogous to interactions between body joints in different actions. Extensive experimental evaluation was presented on two publicly available databases showing better results than the state-of-the-art and the traditional approaches used as baseline measures.

Another idea we have proposed is a novel topological feature representation for persistent homology which encodes temporal information in any given point cloud suitable for applications in action recognition. The proposed framework addresses the drawbacks of conventional methods, by combining the principles from nonlinear time-series analysis and topological data analysis, to extract robust and discriminative features from the reconstructed phase-space.

Since computing distances between persistence diagrams is similar to obtaining Wasserstein distance between two probability mass functions, a well-designed multi-resolution approach can be used to reduce complexity, particularly in applications where we only need approximate distances. Further, using recently proposed persistence kernels can significantly widen the scope of applications of topological persistence features.

REFERENCES

[1] Carnegie mellon university motion capture database. 2012.

[2] Finger-precise hand tracking.

[3] H. D. Abarbanel. *Analysis of observed chaotic data*. New York: Springer-Verlag, 1996.

[4] J. Aggarwal and M. S. Ryoo. Human activity analysis: A review. *ACM Computing Surveys (CSUR)*, 43(3):16, 2011.

[5] S. Ali, A. Basharat, and M. Shah. Chaotic invariants for human action recognition. In *IEEE International Conference on Computer Vision*, pages 1–8, Oct. 2007.

[6] C. Anderson, K. Jamrozik, and E. Stewart-Wynne. Patterns of acute hospital care, rehabilitation, and discharge disposition after acute stroke: the perth community stroke study 1989–1990. *Cerebrovascular Diseases*, 4(5):344–353, 1994.

[7] C. Anderson, C. N. Mhurchu, P. M. Brown, and K. Carter. Stroke rehabilitation services to accelerate hospital discharge and provide home-based care. *Pharmacoeconomics*, 20(8):537–552, 2002.

[8] C. Anderson, S. Rubenach, C. N. Mhurchu, M. Clark, C. Spencer, and A. Winsor. Home or hospital for stroke rehabilitation? results of a randomized controlled trial i: Health outcomes at 6 months. *Stroke*, 31(5):1024–1031, 2000.

[9] M. Baran, N. Lehrer, D. Siwiak, Y. Chen, M. Duff, T. Ingalls, and T. Rikakis. Design of a home-based adaptive mixed reality rehabilitation system for stroke survivors. In *IEEE Conference on Engineering in Medicine and Biological Society*, pages 7602–7605, Aug. 2011.

[10] M. Baran, N. Lehrer, D. Siwiak, Y. Chen, M. Duff, T. Ingalls, and T. Rikakis. Design of a home-based adaptive mixed reality rehabilitation system for stroke survivors. In *Engineering in Medicine and Biology Society*, pages 7602–7605. IEEE, 2011.

[11] M. P. Barnes, B. H. Dobkin, and J. Bogousslavsky. *Recovery after stroke*. Cambridge University Press, 2005.

[12] A. Basharat and M. Shah. Time series prediction by chaotic modeling of nonlinear dynamical systems. In *IEEE International Conference on Computer Vision*, pages 1941–1948, 2009.

[13] A. Bhattacharyya. On a measure of divergence between two statistical populations defined by their probability distributions. *Indian Journal of Statistics*, 35(99-109):4, 1943.

[14] I. Biederman. Recognition-by-components: a theory of human image understanding. *Psychological review*, 94(2):115–147, 1987.

[15] A. Bissacco. Modeling and learning contact dynamics in human motion. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 421–428, June 2005.

[16] A. Bissacco, A. Chiuso, Y. Ma, and S. Soatto. Recognition of human gaits. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 52–57, 2001.

[17] M. Blank, L. Gorelick, E. Shechtman, M. Irani, and R. Basri. Actions as space-time shapes. In *International Conference on Computer Vision*, volume 2, pages 1395–1402. IEEE, 2005.

[18] C. Bosecker, L. Dipietro, B. Volpe, and H. I. Krebs. Kinematic robot-based evaluation scales and clinical counterparts to measure upper limb motor performance in patients with chronic stroke. *Neurorehabilitation and neural repair*, 2009.

[19] M. Brand, N. Oliver, and A. Pentland. Coupled hidden markov models for complex action recognition. In *Conference on Computer Vision and Pattern Recognition*, pages 994–999. IEEE, 1997.

[20] C. Bregler. Learning and recognizing human dynamics in video sequences. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 568–574, June 1997.

[21] K. A. Brown and K. P. Knudson. Nonlinear statistics of human speech data. *International Journal of Bifurcation and Chaos*, 19(07):2307–2319, 2009.

[22] R. E. Burkard, M. Dell'Amico, and S. Martello. *Assignment Problems, Revised Reprint*. Siam, 2009.

[23] L. Cao, A. Mees, and K. Judd. Dynamics from multivariate time series. *Physica D: Nonlinear Phenomena*, 121(1):75–88, 1998.

[24] G. Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009.

[25] G. Carlsson. Topological pattern recognition for point cloud data. *Acta Numerica*, 23:289–368, 2014.

[26] J. L. Casti. *Linear Dynamical Systems*. Academic Press Professional, Inc., 1986.

[27] O. Celik, M. K. O'Malley, C. Boake, H. S. Levin, N. Yozbatiran, and T. A. Reistetter. Normalized movement quality measures for therapeutic robots strongly correlate with clinical motor impairment measures. *IEEE Transactions on Neural Systems and Rehabilitation Engineering,*, 18(4):433–444, 2010.

[28] C.-Y. Chang, B. Lange, M. Zhang, S. Koenig, P. Requejo, N. Somboon, A. A. Sawchuk, and A. A. Rizzo. Towards pervasive physical rehabilitation using microsoft kinect. In *Pervasive Computing Technologies for Healthcare*, pages 159–162. IEEE, 2012.

[29] Y. Chen, M. Duff, N. Lehrer, S.-M. Liu, P. Blake, S. L. Wolf, H. Sundaram, and T. Rikakis. A novel adaptive mixed reality system for stroke rehabilitation: principles, proof of concept, and preliminary application in 2 patients. *Topics in stroke rehabilitation*, 18(3):212–230, 2011.

[30] Y. Chen, M. Duff, N. Lehrer, H. Sundaram, J. He, S. L. Wolf, and T. Rikakis. A computational framework for quantitative evaluation of movement during rehabilitation. In *AIP Conference Proceedings-American Institute of Physics*, volume 1371, pages 317–326, 2011.

[31] Y. Chen, W. Xu, R. I. Wallis, H. Sundaram, T. Rikakis, T. Ingalls, L. Olson, and J. He. A real-time, multimodal biofeedback system for stroke patient rehabilitation. In *Proceedings of the 14th annual ACM international conference on Multimedia*, pages 501–502. ACM, 2006.

[32] A. M. Coderre, A. A. Zeid, S. P. Dukelow, M. J. Demmer, K. D. Moore, M. J. Demers, H. Bretzke, T. M. Herter, J. I. Glasgow, K. E. Norman, et al. Assessment of upper-limb sensorimotor function of subacute stroke patients using visually guided reaching. *Neurorehabilitation and neural repair*, 24(6):528–541, 2010.

[33] N. P. Cuntoor and R. Chellappa. Epitomic representation of human activities. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2007.

[34] M. de Niet, J. B. Bussmann, G. M. Ribbers, and H. J. Stam. The stroke upper-limb activity monitor: its sensitivity to measure hemiplegic upper-limb activity during daily life. *Archives of physical medicine and rehabilitation*, 88(9):1121–1126, 2007.

[35] S. Del Din, S. Patel, C. Cobelli, and P. Bonato. Estimating fugl-meyer clinical scores in stroke survivors using wearable sensors. In *Engineering in Medicine and Biology Society*, pages 5839–5842. IEEE, 2011.

[36] K. G. Derpanis, M. Lecce, K. Daniilidis, and R. P. Wildes. Dynamic scene understanding: The role of orientation features in space and time in scene classification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1306–1313, June 2012.

[37] J. B. Dingwell and J. P. Cusumano. Nonlinear time series analysis of normal and pathological human walking. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 10(4):848–863, 2000.

[38] J. B. Dingwell and H. G. Kang. Differences between local and orbital dynamic stability during human walking. *Journal of Biomechanical Engineering*, 129(4):586–593, 2007.

[39] G. Doretto, A. Chiuso, Y. N. Wu, and S. Soatto. Dynamic textures. *International Journal of Computer Vision*, 51(2):91–109, 2003.

[40] A. W. Dromerick, C. E. Lang, R. Birkenmeier, M. G. Hahn, S. A. Sahrmann, and D. F. Edwards. Relationships between upper-limb functional limitation and self-reported disability 3 months after stroke. *Journal of rehabilitation research and development*, 43(3):401, 2006.

[41] Q. Du, V. Faber, and M. Gunzburger. Centroidal voronoi tessellations: applications and algorithms. *SIAM review*, 41(4):637–676, 1999.

[42] O. Duchenne, I. Laptev, J. Sivic, F. Bach, and J. Ponce. Automatic annotation of human actions in video. In *International Conference on Computer Vision*, pages 1491–1498. IEEE, 2009.

[43] M. Duff, Y. Chen, L. Cheng, S.-M. Liu, P. Blake, S. L. Wolf, and T. Rikakis. Adaptive mixed reality rehabilitation improves quality of reaching movements more than traditional reaching therapy following stroke. *Neurorehabilitation and neural repair*, 27(4):306–315, 2013.

[44] S. P. Dukelow, T. M. Herter, K. D. Moore, M. J. Demers, J. I. Glasgow, S. D. Bagg, K. E. Norman, and S. H. Scott. Quantitative assessment of limb position sense following stroke. *Neurorehabilitation and neural repair*, 24(2):178–187, 2010.

[45] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Discrete and Computational Geometry*, 28(4):511–533, 2002.

[46] S. Emrani, T. Gentimis, and H. Krim. Persistent homology of delay embeddings and its application to wheeze detection. *Signal Processing Letters, IEEE*, 21(4):459–463, 2014.

[47] L. Fei-Fei and P. Perona. A bayesian hierarchical model for learning natural scene categories. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 524–531, June 2005.

[48] T. Flash and N. Hogan. The coordination of arm movements: an experimentally confirmed mathematical model. *The journal of Neuroscience*, 5(7):1688–1703, 1985.

[49] A. Fugl-Meyer, L. Jääskö, I. Leyman, S. Olsson, and S. Steglind. The post-stroke hemiplegic patient. 1. a method for evaluation of physical performance. *Scandinavian journal of rehabilitation medicine*, 7(1):13–31, 1974.

[50] A. Fugl-Meyer, L. Jääskö, I. Leyman, S. Olsson, S. Steglind, et al. The post-stroke hemiplegic patient. 1. a method for evaluation of physical performance. *Scandinavian journal of rehabilitation medicine*, 7(1):13–31, 1975.

[51] D. M. Gavrila. The visual analysis of human movement: A survey. *Computer Vision and Image Understanding*, 73(1):82–98, 1999.

[52] D. J. Gladstone, C. J. Danells, and S. E. Black. The fugl-meyer assessment of motor recovery after stroke: a critical review of its measurement properties. *Neurorehabilitation and Neural Repair*, 16(3):232–240, 2002.

[53] A. Go, D. Mozaffarian, V. Roger, E. Benjamin, J. Berry, W. Borden, D. Bravata, S. Dai, E. Ford, C. Fox, et al. on behalf of the american heart association statistics committee and stroke statistics subcommittee. *Heart disease and stroke statistics2013 update: a report from the American Heart Association. Circulation*, 127(1):e1–e240, 2013.

[54] L. Gorelick, M. Blank, E. Shechtman, M. Irani, and R. Basri. Actions as space-time shapes. *Transactions on Pattern Analysis and Machine Intelligence*, 29(12):2247–2253, 2007.

[55] P. Grassberger and I. Procaccia. Characterization of strange attractors. *Physical review letters*, 50(5):346–349, 1983.

[56] F. Hadzic and T. S. Dillon. Application of tree mining to matching of knowledge structures of decision tree type. In *On the Move to Meaningful Internet Systems 2007: OTM 2007 Workshops*, pages 1319–1328. Springer, 2007.

[57] R. T. Harbourne and N. Stergiou. Movement variability and the use of nonlinear tools: principles to guide physical therapist practice. *Physical Therapy*, 89(3):267–282, 2009.

[58] Z. Harchaoui, E. Moulines, and F. R. Bach. Kernel change-point analysis. In *Advances in Neural Information Processing Systems*, pages 609–616, 2009.

[59] A. Hastings and T. Powell. Chaos in a three-species food chain. *Ecology*, pages 896–903, 1991.

[60] M. Hoai, Z.-Z. Lan, and F. De la Torre. Joint segmentation and classification of human actions in video. In *Conference on Computer Vision and Pattern Recognition*, pages 3265–3272. IEEE, 2011.

[61] N. Hogan and D. Sternad. Sensitivity of smoothness measures to movement duration, amplitude, and arrests. *Journal of motor behavior*, 41(6):529–534, 2009.

[62] L. D. Iasemidis, D.-S. Shiau, W. Chaovalitwongse, J. C. Sackellares, P. M. Pardalos, J. C. Principe, P. R. Carney, A. Prasad, B. Veeramani, and K. Tsakalis. Adaptive epileptic seizure prediction system. *IEEE Transactions on Biomedical Engineering*, 50(5):616–627, 2003.

[63] G. Johansson. Visual perception of biological motion and a model for its analysis. *Perception & psychophysics*, 14(2):201–211, 1973.

[64] I. N. Junejo, E. Dexter, I. Laptev, and P. Pérez. View-independent action recognition from temporal self-similarities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(1):172–185, 2011.

[65] A. Kale, A. Sundaresan, A. Rajagopalan, N. P. Cuntoor, A. K. Roy-Chowdhury, V. Kruger, and R. Chellappa. Identification of humans using gait. *IEEE Transactions on Image Processing*, 13(9):1163–1173, 2004.

[66] J. J. Kavanagh. Lower trunk motion and speed-dependence during walking. *Journal of neuroengineering and rehabilitation*, 6(1):9, 2009.

[67] Y. Ke, R. Sukthankar, and M. Hebert. Event detection in crowded videos. In *International Conference on Computer Vision*, pages 1–8. IEEE, 2007.

[68] M. B. Kennel, R. Brown, and H. D. Abarbanel. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical review A*, 45(6):3403, 1992.

[69] M. Kerber, D. Morozov, and A. Nigmetov. Geometry helps to compare persistence diagrams.

[70] A. N. Kolmogorov. A new metric invariant of transient dynamical systems and automorphisms in lebesgue spaces. In *Dokl. Akad. Nauk SSSR (NS)*, volume 119, pages 861–864, 1958.

[71] T. Krabben, G. Prange, B. Molier, J. Rietman, and J. Buurke. Objective measurement of synergistic movement patterns of the upper extremity following stroke: an explorative study. In *IEEE International Conference on Rehabilitation Robotics (ICORR),*, pages 1–5. IEEE, 2011.

[72] G. Kwakkel, B. Kollen, and E. Lindeman. Understanding the pattern of functional recovery after stroke: facts and theories. *Restorative neurology and neuroscience*, 22(3):281–299, 2004.

[73] N. Lehrer, S. Attygalle, S. L. Wolf, and T. Rikakis. Exploring the bases for a mixed reality stroke rehabilitation system, part i: A unified approach for representing action, quantitative evaluation, and interactive feedback. *Journal of neuroengineering and rehabilitation*, 8(1):51, 2011.

[74] N. Lehrer, Y. Chen, M. Duff, S. L. Wolf, and T. Rikakis. Exploring the bases for a mixed reality stroke rehabilitation system, part ii: Design of interactive feedback for upper limb rehabilitation. *Journal of neuroengineering and rehabilitation*, 8(1):54, 2011.

[75] N. Lehrer, Y. Chen, M. Duff, S. L. Wolf, and T. Rikakis. Exploring the bases for a mixed reality stroke rehabilitation system, part ii: Design of interactive feedback for upper limb rehabilitation. *Journal of neuroengineering and rehabilitation*, 8(1):54, 2011.

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

[77] Z. Liu and S. Sarkar. Improved gait recognition by gait dynamics normalization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(6):863–876, 2006.

[78] C. Lu and N. J. Ferrier. Repetitive motion analysis: segmentation and event classification. *Transactions on Pattern Analysis and Machine Intelligence*, 26(2):258–263, 2004.

[79] J. Mackay, G. A. Mensah, and K. Greenlund. *The atlas of heart disease and stroke*. World Health Organization, 2004.

[80] T. Mäenpää. *The Local binary pattern approach to texture analysis: Extenxions and applications*. Oulun yliopisto, 2003.

[81] D. J. Miller, N. Stergiou, and M. J. Kurz. An improved surrogate method for detecting the presence of chaos in gait. *Journal of biomechanics*, 39(15):2873–2876, 2006.

[82] A. Mirelman, B. L. Patritti, P. Bonato, and J. E. Deutsch. Effects of virtual reality training on gait biomechanics of individuals post-stroke. *Gait & posture*, 31(4):433–437, 2010.

[83] D. M. Morris, G. Uswatte, J. E. Crago, E. W. Cook, E. Taub, et al. The reliability of the wolf motor function test for assessing upper extremity function after stroke. *Archives of physical medicine and rehabilitation*, 82(6):750–755, 2001.

[84] C. J. Murray and A. D. Lopez. Global mortality, disability, and the contribution of risk factors: Global burden of disease study. *The Lancet*, 349(9063):1436–1442, 1997.

[85] V. Nanda and R. Sazdanović. Simplicial models and topological inference in biological systems. In *Discrete and Topological Models in Molecular Biology*, pages 109–141. Springer, 2014.

[86] J. C. Niebles, H. Wang, and L. Fei-Fei. Unsupervised learning of human action categories using spatial-temporal words. *International journal of computer vision*, 79(3):299–318, 2008.

[87] J. Nocedal and S. Wright. Numerical optimization, series in operations research and financial engineering. *Springer, New York*, 2006.

[88] S. M. Oh, J. M. Rehg, T. Balch, and F. Dellaert. Learning and inferring motion patterns using parametric segmental switching linear dynamic systems. *International Journal of Computer Vision*, 77(1-3):103–124, 2008.

[89] A. Oliva and Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3):145–175, 2001.

[90] A. Oliva and A. Torralba. Building the gist of a scene: The role of global image features in recognition. *Progress in brain research*, 155:23–36, 2006.

[91] W. H. Organization et al. The world health report: 2002: Reducing the risks, promoting healthy life. 2002.

[92] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin. Shape distributions. *ACM Transactions on Graphics*, 21(4):807–832, 2002.

[93] S. Patel, R. Hughes, T. Hester, J. Stein, M. Akay, J. G. Dy, and P. Bonato. A novel approach to monitor rehabilitation outcomes in stroke survivors using wearable technology. *Proceedings of the IEEE*, 98(3):450–461, 2010.

[94] S. Patel, K. Lorincz, R. Hughes, N. Huggins, J. Growdon, D. Standaert, M. Akay, J. Dy, M. Welsh, and P. Bonato. Monitoring motor fluctuations in patients with parkinson's disease using wearable sensors. *TITB*, 13(6):864–873, 2009.

[95] M. Perc. The dynamics of human gait. *European journal of physics*, 26(3):525–534, 2005.

[96] M. Perše, M. Kristan, J. Perš, and S. Kovačič. *Automatic Evaluation of Organized Basketball Activity using Bayesian Networks*. Citeseer, 2007.

[97] S. Pincus and R. E. Kalman. Not all (possibly) "random" sequences are created equal. *Proceedings of the National Academy of Sciences*, 94(8):3513–3518, 1997.

[98] S. Pincus and B. H. Singer. Randomness and degrees of irregularity. *Proceedings of the National Academy of Sciences*, 93(5):2083–2088, 1996.

[99] S. M. Pincus. Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences*, 88(6):2297–2301, 1991.

[100] S. M. Pincus. Irregularity and asynchrony in biologic network signals. *Methods in enzymology*, 321:149–182, 2000.

[101] H. Pirsiavash, C. Vondrick, and A. Torralba. Assessing the quality of actions. In *European Conference on Computer Vision*, pages 556–571. Springer, 2014.

[102] W. Prinz. Perception and action planning. *European journal of cognitive psychology*, 9(2):129–154, 1997.

[103] L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[104] L. Ralaivola, F. d'Alché Buc, et al. Dynamical modeling with kernels for nonlinear time series prediction. In *Neural Information Processing Systems*, volume 4, pages 129–136, 2003.

[105] D. Reinkensmeyer, B. Schmit, and W. Rymer. Mechatronic assessment of arm impairment after chronic brain injury. *Technology and Health Care*, 7(6):431–435, 1999.

[106] D. J. Reinkensmeyer, L. E. Kahn, M. Averbuch, A. McKenna-Cole, B. D. Schmit, and W. Z. Rymer. Understanding and treating arm movement impairment after chronic brain injury: progress with the arm guide. *Journal of rehabilitation research and development*, 37(6):653–662, 2000.

[107] V. L. Roger, A. S. Go, D. M. Lloyd-Jones, E. J. Benjamin, J. D. Berry, W. B. Borden, D. M. Bravata, S. Dai, E. S. Ford, C. S. Fox, et al. Heart disease and stroke statistics–2012 update: a report from the american heart association. *Circulation*, 125(1):e2, 2012.

[108] B. Rohrer, S. Fasoli, H. I. Krebs, B. Volpe, W. R. Frontera, J. Stein, N. Hogan, et al. Submovements grow larger, fewer, and more blended during stroke recovery. *Motor Control*, 8:472–483, 2004.

[109] M. Rosenstein, J. Collins, and C. De Luca. A practical method for calculating largest lyapunov exponents from small data sets. *Physica D: Nonlinear Phenomena*, 65(1):117–134, 1993.

[110] M. T. Rosenstein, J. J. Collins, and C. J. De Luca. A practical method for calculating largest lyapunov exponents from small data sets. *Physica D: Nonlinear Phenomena*, 65(1):117–134, 1993.

[111] J.-C. Roux, R. H. Simoyi, and H. L. Swinney. Observation of a strange attractor. *Physica D: Nonlinear Phenomena*, 8(1):257–266, 1983.

[112] Y. Rubner, C. Tomasi, and L. J. Guibas. A metric for distributions with applications to image databases. In *IEEE International Conference on Computer Vision*, pages 59–66, Jan. 1998.

[113] G. Saposnik, R. Teasell, M. Mamdani, J. Hall, W. McIlroy, D. Cheung, K. E. Thorpe, L. G. Cohen, M. Bayley, et al. Effectiveness of virtual reality using wii gaming technology in stroke rehabilitation a pilot randomized clinical trial and proof of principle. *Stroke*, 41(7):1477–1484, 2010.

[114] W. M. Schaffer. Order and chaos in ecological systems. *Ecology*, pages 93–106, 1985.

[115] L. Seidenari, V. Varano, S. Berretti, A. Del Bimbo, and P. Pala. Recognizing actions from depth cameras as weakly aligned multi-part bag-of-poses. In *Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 479–485. IEEE, 2013.

[116] C. E. Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55, 2001.

[117] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, and R. Moore. Real-time human pose recognition in parts from single depth images. *Communications of the ACM*, 56(1):116–124, 2013.

[118] N. Shroff, P. Turaga, and R. Chellappa. Moving vistas: Exploiting motion for describing scenes. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1911–1918, June 2010.

[119] T. Simon, M. H. Nguyen, F. De La Torre, and J. F. Cohn. Action unit detection with segment-based svms. In *Conference on Computer Vision and Pattern Recognition*, pages 2737–2744. IEEE, 2010.

[120] G. Singh, F. Mémoli, and G. E. Carlsson. Topological methods for the analysis of high dimensional data sets and 3d object recognition. In *SPBG*, pages 91–100. Citeseer, 2007.

[121] D. Siwiak, N. Lehrer, M. Baran, Y. Chen, M. Duff, T. Ingalls, and T. Rikakis. A home-based adaptive mixed reality rehabilitation system. In *Proceedings of the 19th ACM international conference on Multimedia*, pages 785–786. ACM, 2011.

[122] M. Small. *Applied nonlinear time series analysis: applications in physics, physiology and finance*, volume 52. World Scientific Publishing Company Incorporated, 2005.

[123] C. Sminchisescu, A. Kanaujia, and D. Metaxas. Conditional models for contextual human motion recognition. *Computer Vision and Image Understanding*, 104(2):210–220, 2006.

[124] S. Soatto, G. Doretto, and Y. N. Wu. Dynamic textures. In *IEEE International Conference on Computer Vision*, volume 2, pages 439–446, 2001.

[125] E. H. Spriggs, F. De La Torre, and M. Hebert. Temporal segmentation and activity classification from first-person sensing. In *Conference On Computer Vision and Pattern Recognition Workshops*, pages 17–24. IEEE, 2009.

[126] A. Srivastava, I. Jermyn, and S. Joshi. Riemannian analysis of probability density functions with applications in vision. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2007.

[127] N. Stergiou and L. M. Decker. Human movement variability, nonlinear dynamics, and pathology: is there a connection? *Human Movement Science*, 30(5):869–888, 2011.

[128] H. L. Swinney. Observations of order and chaos in nonlinear systems. *Physica D: Nonlinear Phenomena*, 7(1):3–15, 1983.

[129] F. Takens. Detecting strange attractors in turbulence. *Dynamical Systems and Turbulence*, 898:366–381, 1981.

[130] E. Taub, G. Uswatte, R. Pidikiti, et al. Constraint-induced movement therapy: a new family of techniques with broad application to physical rehabilitation-a clinical review. *Journal of rehabilitation research and development*, 36(3):237–251, 1999.

[131] A. Tausz, M. Vejdemo-Johansson, and H. Adams. Javaplex: A research software package for persistent (co) homology. *Software available at http://code. google. com/javaplex*, 2011.

[132] T. TenBroek, R. Van Emmerik, C. Hasson, and J. Hamill. Lyapunov exponent estimation for human gait acceleration signals. *Journal of Biomechanics*, 40(2):210, 2007.

[133] W. Tucker. The lorenz attractor exists. *Comptes Rendus de l'Académie des Sciences-Series I-Mathematics*, 328(12):1197–1202, 1999.

[134] P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea. Machine recognition of human activities: A survey. *IEEE Transactions on Circuits and Systems for Video Technology,*, 18(11):1473–1488, 2008.

[135] P. Turaga, A. Veeraraghavan, and R. Chellappa. Unsupervised view and rate invariant clustering of video sequences. *Computer Vision and Image Understanding*, 113(3):353–371, 2009.

[136] P. K. Turaga, A. Veeraraghavan, and R. Chellappa. From videos to verbs: Mining videos for activities using a cascade of dynamical systems. In *Conference On Computer Vision and Pattern Recognition*, 2007.

[137] K. Tyryshkin, A. M. Coderre, J. I. Glasgow, T. M. Herter, S. D. Bagg, S. P. Dukelow, and S. H. Scott. A robotic object hitting task to quantify sensorimotor impairments in participants with stroke. *Journal of neuroengineering and rehabilitation*, 11(1):47, 2014.

[138] L. van Dokkum, I. Hauret, D. Mottet, J. Froger, J. Métrot, and I. Laffont. The contribution of kinematics in the assessment of upper limb motor recovery early after stroke. *Neurorehabilitation and neural repair*, page 1545968313498514, 2013.

[139] N. Vaswani, A. K. Roy-Chowdhury, and R. Chellappa. Shape activity: a continuous-state hmm for moving/deforming shapes with application to abnormal activity detection. *IEEE Transactions on Image Processing*, 14(10):1603–1616, 2005.

[140] R. Vemulapalli, F. Arrate, and R. Chellappa. Human action recognition by representing 3d skeletons as points in a lie group. In *Conference on Computer Vision and Pattern Recognition*, pages 588–595. IEEE, 2014.

[141] V. Venkataraman and P. Turaga. Shape distributions of nonlinear dynamical systems for video-based inference. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016.

[142] V. Venkataraman, P. Turaga, N. Lehrer, M. Baran, T. Rikakis, and S. L. Wolf. Attractor-shape for dynamical analysis of human movement: Applications in stroke rehabilitation and action recognition. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 514–520, June 2013.

[143] V. Venkataraman, I. Vlachos, and P. Turaga. Dynamical regularity for action analysis. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 67.1–67.12, September 2015.

[144] I. Vlachos and D. Kugiumtzis. State space reconstruction from multiple time series. In *Topics on Chaotic Systems: Selected Papers from Chaos 2008 International Conference*, page 378. World Scientific, 2009.

[145] G. P. Williams. *Chaos theory tamed*. Joseph Henry Press, 1997.

[146] A. D. Wilson and A. F. Bobick. Learning visual behavior for gesture analysis. In *IEEE International Symposium on Computer Vision*, pages 229–234, Nov. 1995.

[147] A. Wolf. Quantifying chaos with lyapunov exponents. *Chaos*, pages 273–290, 1986.

[148] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano. Determining lyapunov exponents from a time series. *Physica D: Nonlinear Phenomena*, 16(3):285–317, 1985.

[149] S. L. Wolf, P. A. Catlin, M. Ellis, A. L. Archer, B. Morgan, and A. Piacentino. Assessing wolf motor function test as outcome measure for research in patients after stroke. *Stroke*, 32(7):1635–1639, 2001.

[150] S. Wu, B. E. Moore, and M. Shah. Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scenes. In *Conference on Computer Vision and Pattern Recognition*, pages 2054–2060. IEEE, 2010.

[151] G. Wulf and C. H. Shea. Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychonomic bulletin & review*, 9(2):185–211, 2002.

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

[153] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba. Sun database: Large-scale scene recognition from abbey to zoo. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3485–3492, June 2010.

[154] J. Yamato, J. Ohya, and K. Ishii. Recognizing human action in time-sequential images using hidden markov model. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 379–385, June 1992.

[155] Y. Yang and D. Ramanan. Articulated pose estimation with flexible mixtures-of-parts. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 1385–1392. IEEE, 2011.

[156] M. Zhang, B. Lange, C.-Y. Chang, A. A. Sawchuk, and A. A. Rizzo. Beyond the standard clinical rating scales: Fine-grained assessment of post-stroke motor functionality using wearable inertial sensors. In *Engineering in Medicine and Biology Society*, pages 6111–6115. IEEE, 2012.

[157] F. Zhou, F. De la Torre, and J. K. Hodgins. Hierarchical aligned cluster analysis for temporal clustering of human motion. *Transactions on Pattern Analysis and Machine Intelligence*, 35(3):582–596, 2013.

[158] F. Zhou, F. Torre, and J. K. Hodgins. Aligned cluster analysis for temporal segmentation of human motion. In *Conference on Automatic Face & Gesture Recognition*, pages 1–7. IEEE, 2008.

[159] X. Zhu. Persistent homology: An introduction and a new text representation for natural language processing. In *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*, pages 1953–1959. AAAI Press, 2013.

[160] Y. Zhu, W. Chen, and G. Guo. Fusing spatiotemporal features and joints for 3d action recognition. In *Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 486–491. IEEE, 2013.

[161] A. Zomorodian. Fast construction of the vietoris-rips complex. *Computers & Graphics*, 34(3):263–271, 2010.