Statistical Analysis of Natural Human Motion for Animation

Liu Ren

CMU-CS-06-175

September 2006

Computer Science Department
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Jessica K. Hodgins, Chair
Nancy S. Pollard
Alexei A. Efros
James M. Rehg, Georgia Institute of Technology

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

This research was sponsored by the National Science Foundation under grant nos. CNS-0196217, IIS-0205224, and IIS-0326322 and supported by generous software donations from Alias/Wavefront and Autodesk.

Keywords: Character animation, Human motion, Statistics, Natural motion classifier, Ensemble models, PCA, Joint limits, Synergies, Motion summarization.

To my wife Wei and my son Billy.

Abstract

Generating human motion that appears natural is a long standing problem in character animation. Researchers have explored many different approaches including physics-based simulation, optimization, and data-driven methods such as motion graphs and motion interpolation. One major difficulty in applying most of these approaches is the lack of an implementable definition of what it means for motion to be natural or human-like. In this thesis, we explore two techniques to fill this gap. The first technique creates a naturalness measure for quantifying natural human motion. The second technique involves a statistical analysis of human motion to compute aggregate statistics that are needed to guide animation algorithms for human figures toward natural looking solutions.

A naturalness measure should be useful in verifying that a motion editing operation has not destroyed the naturalness of a motion capture clip or that a synthetic motion transition is within the space of those seen in natural human motion. To develop such a measure, we argue that the evaluation of naturalness is not intrinsically a subjective criterion imposed by a human observer but is, instead, an objective measure that can be computed from a large set of representative motions. We base our approach on a statistical analysis of a large motion database. Using positive training data only, the system learns a set of statistical models that represent the motion of individual joints, limbs, and the whole body. Each model produces a score for the naturalness of the test motion and these scores are then combined into an aggregate score to classify the input motion as natural or unnatural. We present ROC curves of the performance of these techniques on a broad set of test sequences and compare the results to human performance in a user study.

Aggregate statistics about the properties of human motion are needed to guide animation algorithms to generate natural looking motion. We compute and report a variety of statistics for joint angle range of motion, joint velocities, and dimensionality reduction using a large and representative motion capture database. We also develop new techniques for identifying motion synergies and summarizing motion in a visually intuitive way.

Acknowledgments

I am extremely grateful to my advisor Jessica Hodgins. Without her guidance and tremendous support, the thesis would not be possible. She guided me through every aspect of my graduate study. She not only taught me how to do research, but also helped me improve my writing skills, speaking skills, and communication skills.

I would like to thank other members in my thesis committee - Alexei Efros, Nancy Pollard and James Rehg for their valuable suggestions on my research work. Their comments and feedbacks greatly helped improve my thesis.

I would like to thank Paul Heckbert who gave me a lot of support when I started my PhD journey in CMU. I would like to thank Hanspeter Pfister who was always a great mentor when I worked in MERL. I should also thank Sharon Burks who helped me a lot to handle all kinds of issues during my PhD study. I also thank my fellow colleagues in CMU's graphics group, as well as many others in CSD, RI, and LTI, for making my student life in CMU so enjoyable.

Finally, I should express my deepest gratitude to my family. I owe a great deal to my wife Wei. Without her love, encouragement, and support, it is impossible for me complete this long journey. I also thank my dearest son Billy who brought me a lot of happy moments during my PhD years. They supported me to pursue my own goals with confidence, to progress with sweat and hard work, and to face and overcome obstacles with courage.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Human motion generation plays an important role in the game and movie industry, where artists or animators need to produce high quality animation for human characters. However, it is impossible to generate high quality human motion fully by hand because human motion contains many degrees of freedom, which must be coordinated in natural and physically consistent patterns to generate compelling motion. In practice, artists or animators employ semi-automatic methods such as keyframing techniques and inverse kinematics to facilitate this process. However, these methods are time consuming because a significant amount of user input is required and creating a good set of key poses manually is quite challenging. Animators have to evaluate the motion quality by hand, checking whether the motion is "in character" and free of unnatural elements.

Animation researchers have developed automatic motion generation procedures to reduce the human input required to create an appealing animated sequence. Many different approaches have been proposed, but they can be divided roughly into three categories: physics-based simulation, optimization, and data-driven methods. Physics-based simulation relies on a set of physical laws for automatic motion synthesis. Building a complete physical structure of a human model and the corresponding control system is very challenging and the physical structure is usually simplified or overly constrained to allow for

stable solutions. Automatic motion synthesis can also be formulated as an optimization problem, a search problem guided by an objective function. However, the search for a natural solution can easily fail because of local minima in the search space or because the objective function does not correctly characterize the naturalness of human motion. More recently, data-driven approaches have been developed that reuse captured motion data to generate variations. For example, a graph structure called a motion graph can be generated by connecting pairs of similar poses (creating new transitions) in the original motion capture clips. By searching the motion graph, we can generate motions that meet the user's constraints. Interpolating similar poses or motion sequences is another data-driven approach to generate variations of existing clips. These two approaches produce good results as long as the new transitions and the interpolation operation do not the break the "naturalness" embedded in the original motion clips.

Although naturalness is the key to the successful generation of realistic human animation, one major difficulty in applying all of these approaches is the lack of an implementable definition of what it means for motion to be natural or human-like. In this thesis, we explore two techniques to fill this gap. The first technique creates a naturalness measure for quantifying natural human motion. A naturalness measure should be useful in verifying that a motion editing operation has not destroyed the naturalness of a motion capture clip or that a synthetic motion transition is within the space of those seen in natural human motion. To develop such a measure, we argue that the evaluation of naturalness is not intrinsically a subjective criterion imposed by a human observer but is, instead, an objective measure that can be computed from a large set of representative motions. We base our approach on a statistical analysis of a large motion database that contains a variety of natural human behaviors. The second technique is a statistical analysis of human motion. These statistics can be used to guide animation algorithms for human figures toward natural looking solutions. They have not been comprehensively reported previously. We believe the set of statistics and the insight gleaned from them will allow others to create significantly better algorithms for generating natural human motion.

For both approaches, we employ the Carnegie Mellon motion capture database (mo-

cap.cs.cmu.edu), the largest publicly available motion capture database. The data are representative as it contains 4 hours of many different behaviors captured from 34 different subjects. We give an overview of these two approaches in the next two sections.

1.1 Quantifying Natural Human Motion

Data-driven approaches to human motion synthesis have been a focus of attention for the past ten years. Much of the research in this area has been on techniques for adapting captured motion data to new situations. Motion capture data can be reordered in time [3, 32, 33], similar motions can be interpolated [72, 52, 31], motion can be edited [16], and new motions can be generated by combining motions for individual limbs [29]. Models of human motion can also be used to synthesize new motion [9, 36].

Each of these techniques employs heuristics or models that attempt to restrict the output of the algorithms to natural-looking motion, but no single naturalness measure exists to assess the quality of the output. In this work, we explore whether it is possible to provide such a measure.

How can we quantify what it means for a sequence of human motion to appear natural? One approach is to propose a set of heuristic rules that govern the movement of various joints. If a given sequence violates any of the rules, it is judged to be unnatural. For example, a character's motion could be tested for angular momentum conservation in flight or violation of the friction cone when the foot is in contact. This bottom-up approach will likely have difficulty with the more stylistic elements of human motion, because a motion can be physically correct without appearing natural.

A second approach is to develop a set of perceptual metrics that provide guidelines for the flaws that people are likely to notice [50, 43, 22, 48]. For example, Reitsma and Pollard measured the sensitivity of users to changes in horizontal and vertical velocity [50]. Taken together, such studies could provide a set of guidelines to assess whether a given motion will be perceived as natural.

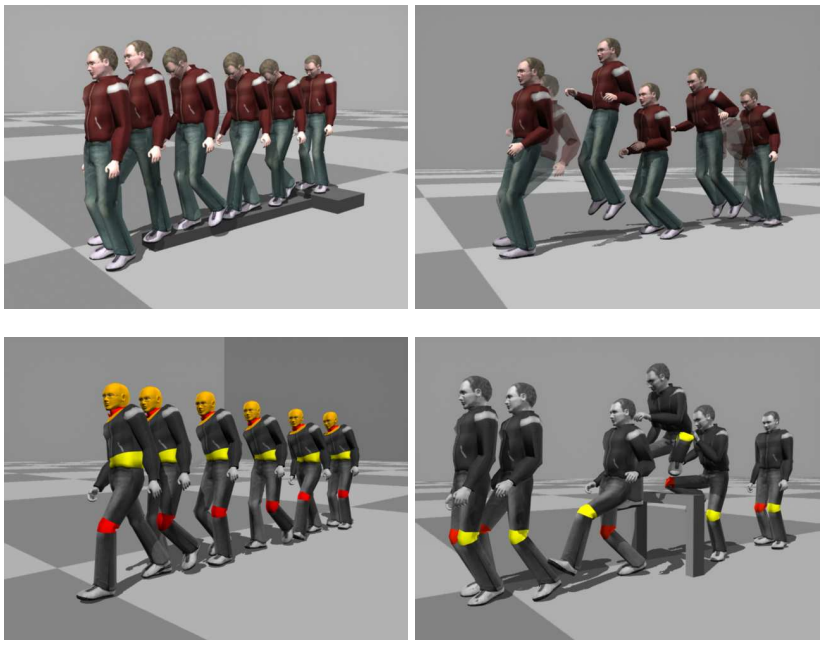


Figure 1.1: Examples from our test set of motions. The upper two images are natural (cleaned and but otherwise unaltered motion capture data). The lower two images are unnatural (badly edited and incompletely cleaned motion). Joints that are marked in red-yellow were detected as having unnatural motion. Frames for these images were selected by the method presented in [4] and discussed in Chapter 3.

A third approach is to train a classifier to distinguish between natural and unnatural movement based on human-labeled, ground-truth data [29, 69]. For example, Wang and Bodenheimer used an optimization approach to find weights for a transition metric that best matched the judgments of sequences by users as to which sequences were natural and which were not. We use an alternative take on this approach and assume that the learning algorithm will be trained only on positive (natural) examples of the motion. We make this assumption because natural motions are readily available from commercial motion capture systems. Negative (unnatural) examples, on the other hand, are precious because each must be hand labeled by a person. As a consequence of this scarcity, the negative examples

that would be required for training do not exist. A further concern is that the characteristics of these negative examples are likely to be specific to the adaptation method that generated them and not representative of unnatural motions in general. We will demonstrate that using our approach, a variety of motions can be assessed using models that have been trained on a large corpus of positive examples (Figure 1.1).

Our approach to this problem is based on the assumption that the evaluation of naturalness is an objective measure imposed by the data as a whole. Simply put, motions that we have seen repeatedly are judged natural, whereas motions that happen very rarely are not. Humans are good at this type of evaluation because they have *seen a lot of data*. The amount of collected motion capture data has grown rapidly over the past few years and we believe that there is now an opportunity for a computer to analyze a lot of data, resulting in a successful method for evaluating naturalness.

We model the naturalness of human motion by training different statistical models *only on positive data — natural human motions*. We explore the performance of mixture of Gaussians (MoG), hidden Markov models (HMM), and switching linear dynamic systems (SLDS) on this problem. We also propose an *ensemble model* to attack this problem. This approach first hierarchically decomposes human motion into its constituent parts (individual joints, limbs, and full body), then builds a statistical model of each one using existing machine learning techniques and finally combines these models into an ensemble model for classification of the motion as natural or unnatural. We also implement a Naive Bayes (NB) model for a baseline comparison. We test these techniques on motion capture data held out from a database, keyframed motions, edited motions, motions with noise added, and synthetic motion transitions. We present the results as receiver operating characteristic (ROC) curves and compare the results to the judgments made by subjects in a user study. This approach is discussed in detail in Chapter 3.

1.2 Exploring the Statistics of Natural Human Motion

The statistics of natural human motion play an important role in understanding and studying human motion in animation, ergonomics, and biomechanics. These statistics can be computed individually from a single subject and behavior or as aggregate statistics computed from many subjects performing a variety of behaviors. In this thesis, we present statistics computed for various classes of behaviors. These statistics include limits and distributions on joint angles, angular velocities and accelerations which characterize the range of human motion. We also present the mean pose and eigenposes computed using principal component analysis (PCA), a statistical analysis of human motion synergies, and motion summary images which expose the characteristics of natural human behaviors in a visually informative way. We believe statistical quantities computed from a large, representative database will allow efficient synthesis of more natural motion by providing better priors and more accurate constraints. These statistics should become a useful reference source for the implementation of existing algorithms and the development of new animation algorithms. We also hope that they will provide insight into natural human motion and verify some common heuristic assumptions.

Some of these statistics serve as important priors or constraints in animation algorithms. For example, joint angle limits have been widely used in such applications as motion editing [76, 14], motion retargeting [15, 60], and motion capture [68]. These parameters are also used to restrict the search space for optimization in inverse kinematics (IK) [76] and trajectory optimization [54]. Joint torque limits are used in IK systems [34] and physics-based motion synthesis algorithms [26, 25, 54, 58, 60]. Principal component analysis has been used as a preprocessing step in many animation algorithms [9, 36, 6, 18, 54, 13] because the resulting mean pose and the eigenposes characterize the distribution of natural human poses and provide important priors for these algorithms.

Although they have often been used, these statistics have not been systematically reported in the animation literature. For example, even though PCA has been widely used

in character animation, general statistics such as the mean pose and eigenposes of natural human motion have not been reported. Although different orientation representations are employed when PCA is used in animation, the impact of these different representations on the performance of PCA compression has not been carefully explored. Some of the statistics used in previous work come from trial-and-error experience [60, 68] or from small motion capture databases [36, 54, 38]. Statistics reported in biomechanics and ergonomics studies are sometimes appropriate although these statistics often apply only to a single behavior such as walking or running [74, 42].

In addition to systematically reporting statistics that are known to be of use in animation algorithms, we also develop new techniques for identifying motion synergies and summarizing motion in a visually intuitive way. Human motion synergies identify joints that exhibit a strong coordination. The topic has not been quantitatively analyzed in either biomechanics or character animation mainly because a clear and computable definition for the coordination of joint movements is not available. In this thesis, we provide a statistical definition to quantify the coordination by computing statistical dependency (measured by mutual information [44]) between joint movements. Summarizing motion in a still image might allow more rapid browsing of motion libraries. Existing techniques can only summarize a single sequence of motion [4] but as motion capture databases become larger, we would like to be able to expose the characteristics of classes of natural human behaviors in a visually informative way. We are inspired by the techniques of artist Jason Salavon [30] and merge sequences of images to create intuitive representations of motion sequences ranging from individual behaviors to a database with four hours of motion(Figure 1.2).

1.3 Organization

In the next chapter, we discuss related work. We describe the details of our data-driven approach to quantifying natural human motion in Chapter 3. In Chapter 4, we present and analyze a number of statistics of natural human motion. Finally, we discuss the contribu-

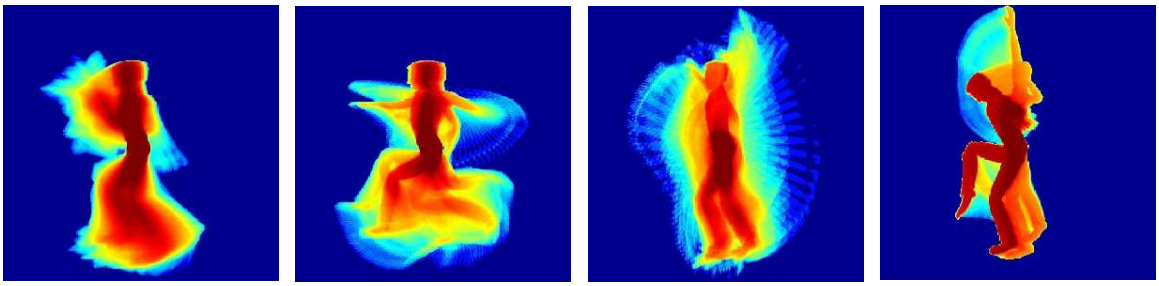


Figure 1.2: Sequences of different behaviors (boxing, run-to-leap, cartwheel, and ballet) are summarized in still images.

tions of this thesis and conclude with a discussion of future work in Chapter 5.

Chapter 2

Background

In this chapter, we give an overview of motion synthesis and analysis work that are related to the naturalness of human motion. These work cover a broad range of research areas such as character animation, biomechanics, computer vision, and art. We first discuss the work most closely related to quantifying natural human motion and then to the statistics of natural human motion.

2.1 Quantifying Natural Human Motion

Even though there is no a clear definition for the naturalness of human motion, many algorithms for synthesizing and editing human motion have been designed with the goal of restricting their output to natural human motion. We use many of these techniques to generate the positive and negative testing examples for the classifier described in Chapter 3. We briefly review this work and then discuss related problems in other disciplines.

One early technique aimed at allowing naive users to create natural looking motion is the use of modulated sine waves and stochastic noise [47]. This approach was based on the hypothesis that many natural human motions can be characterized as sinusoidal waves and stochastic noise. We test our classifier on both positive and negative sequences that

are similar in that sinusoidal noise has been added to motion capture data.

Many motion editing techniques have been proposed, each with a set of optimization criteria intended to ensure that the resulting motion is natural (see, for example [16, 58]). In these works, a natural motion capture clip usually served as the basis of motion editing operations. The edited motion has to satisfy the new constraints (kinematic constraints, for example) provided by the user while maintaining the naturalness. The naturalness is usually measured by how similar the new motion are to the original motion capture clip, which is presented as an objective function in these optimization problem [16, 58]. Some of these techniques have been adapted into commercial software, and we use Maya to perform editing on motion capture data to generate part of our negative test set.

Motion graphs create new animations by resequencing pieces of motion capture data. The naturalness of the resulting motion depends largely on the quality of the motion transitions. Several algorithms have been proposed for creating natural transitions [33, 32, 3]. The quality of motion transitions are usually determined by the similarity of the two pose states. The pose state variable could include joint angles, angular velocities, or angular accelerations. The $L2$ distance metric has been widely used to compute the similarity between two sets of pose state variables. A small distance usually allows for a smooth and natural transitions in these approaches. We use synthetic motion transitions, both good and bad, as part of the test set in our experiments.

Wang and Bodenheimer [69] used optimization to tune the weights of a transition metric based on example transitions classified by a human viewer as good or bad. As it is tedious to label these transitions manually, only a small number of example transitions (16 good transitions and 26 bad transitions) were used in the process. They made several assumptions to make the optimization process tractable. For example, they did not consider how changes in the blending algorithm would affect the naturalness for a given distance metric. They also studied the optimal duration for a transition given a previously learned distance measure [70].

Limb transplant creates new motions by combining limb motions from different mo-

tion clips. It can used to generalize the motion in an available database. However, there is no guarantee that the generated motion looks natural. Ikemoto and Forsyth [29] used an SVM to classify a synthesized motion as "looks human" or "does not look human." Their approach was quite effective for this problem, but it is a supervised learning approach and therefore requires a relatively large number of positive and negative training examples specific to limb transplant. More recently, Ikmoto and colleagues [28] applied logistic regression to carefully chosen features to score motion transitions in the context of multi-way motion blending. They chose features such as footstrike and acceleration information because they observed that these features were related to errors in bad motion transitions. All of these approaches can be used to detect specific error in unnatural human motion as the positive and negative training examples used in these approaches were generated by a specific application. In contrast, our goal is to use unsupervised learning to construct a measure that can be trained only on positive examples and that works for motion produced by a variety of motion editing algorithms. Besides providing a classifier to classify natural and unnatural motion, our approach can also identify specific parts (joint or limb) that demonstrate unnatural motion if the whole motion sequence is classified as unnatural.

The question of how to quantify human motion is also related to research that has been performed in a number of fields other than computer graphics. For example, researchers interested in speaker identification have looked at the problem of deciding whether a particular speaker produced a segment based on a corpus of data for that speaker and for others [11]. Classifying natural vs. unnatural images for fraud detection is similarly related to our problem [12].

Closer to our problem is the work of Troje [64] who was interested in identifying features of a human walk that can be used to label it as male or female. He reduced the dimensionality of the dataset as we do, with PCA, and then fit sinusoids to the resulting components. This approach is specific to a cyclic motion such as walking and would not easily generalize to our very large, heterogeneous database. However, the performance of his classifier was better than that of human subjects on a point light visualization of the walking motion.

Researchers working in activity recognition have looked at detection of unusual activities, which is similar to our problem in that an adequate negative training set would be difficult to collect. As a result, most approaches have focused on unsupervised learning. For example, Zhong and his colleagues [77] used an unsupervised learning approach to detect unusual activity in video streams of human motion. Hara and his colleagues [21] took motion detector data acquired from an intelligent house, performed vector quantization, and estimated the probability of a sequence of sensor data with a HMM. Hamid and his colleagues [20] used clustering of event n-grams to identify and explain anomalous activities. More recently, Boiman and Irani [8] formulated the problem of detecting usual or unusual activities as the problem of composing the new observed visual data using spatio-temporal patches extracted from previous video data. Regions in the observed video which cannot be composed using large continuous chunks of data from a large set of video examples are regarded as suspicious.

2.2 Exploring the Statistics of Natural Human Motion

Although statistics about human motion are widely used in animation algorithms, they have been published more often in the biomechanics literature than in the animation literature. In biomechanics, these statistics are sometimes computed based on a coordinate system that consists of three predefined planes: sagittal, transverse and frontal [74, 42]. Although this coordinate system is commonly used for human gait analysis in biomechanics and has been used for walking motion synthesis in character animation [59], it has not been adapted for more complex and fully three dimensional behaviors (dancing, for example). Moreover, in biomechanics, these statistics are generally computed for specific behaviors or joints and averaged across a group of subjects, which limits their use in animation where we need to synthesize a wide variety of behaviors. For example, Winter [74] analyzed and reported kinematic walking data measured in the sagittal plane.

Researchers have modeled joint angle limits as they are critical for understanding the

human workspace [1]. In these studies, the joint angle of each degree of freedom (DOF) is usually bound to a range between a maximum and a minimum value and each DOF is assumed to be able to take on all values in its range independent of the values of the other DOFs. This simple representation has been widely adopted in computer animation [76, 14, 15, 54, 60]. It is also used in anatomy-based joint models for ergonomic tests, athletic training, surgery planning and surgery simulation. For example, Maciel and colleagues [39] constructed an anatomy-based knee joint by discretizing the sliding curve of the human knee and attaching a local coordinate system at each discrete point to describe a three DOF joint movement at that point. They limited the range of each DOF independently.

Joint angle limits can also be modeled as a joint sinus cone where the apex of the cone is located at the joint center and the cone bounds the movements of two DOFs of a joint, such as abduction/adduction and flexion/extension of the humerus at the shoulder. Joint sinus cones provide more accurate physiological limits on limb circumduction than independent bounds for each joint DOF [40, 73, 41]. An implicit surface can also be used to model joint angle limits. For example, Herda and colleagues [23] used quaternions to represent rotations for 3-DOF joints such as the shoulder. Then using data from an optical motion capture system, they fit an implicit surface to the three imaginary components of these quaternions and used that surface as a joint limit. Each of these different representations determine a configuration space for the joint and the more anatomically accurate representations find a tighter, more accurate description of the space. However, "hard" joint angle limits do not provide information about the likelihood of a joint angle configuration, which is closely related to the naturalness of a joint angle configuration [51] or the "comfort" of joint configurations for task-driven motion generation [34].

PCA has been widely used in character animation [9, 36, 6, 18, 54, 65, 13]. Because there are multiple representations for orientations, researchers have chosen different representations including Euler angles [9, 54, 65], quaternions [6] and exponential maps [36, 18, 13] when PCA was applied to the motion data.

Although the notion of "synergies" has an intuitive meaning in biomechanics and char-

acter animation, it lacks a rigorous and computable definition. The eigenposes found by PCA analysis have been used to represent synergies. For example, they were used for designing human motor control strategies in robotics [19] and for analyzing complex hand manipulation in biomechanics [61]. However, PCA analysis on human poses does not aim to identify the coordination between joint movements but rather represent human poses as a compact linear combination of a few eigenposes (via additions and subtractions). The coordination of joint movements in these eigenposes may not reflect the real coordination observed in the movements of these joints. As a result, PCA analysis cannot explicitly identify groups of joints that exhibit coordination in their motions. Identifying such joint groups has an application in quantifying natural human motion in animation [51] and human motion simulation and control [26].

As observed by Sakamoto and colleagues [55], iconic or image-based motion or posture representation could provide an intuitive user interface for browsing or navigating a large motion database. Assa and colleagues [4] summarized a single motion sequence in one image by compositing images of representative poses. We present a general approach that can visualize motions in a large database in one image. Our technique of motion summarization was inspired by artist Jason Salavon [30]. Each of his works utilizes 100 unique commemorative photographs culled from the internet for averaging. The final compositions are arrived at using both the mean. Based on a similar idea, Torralba and Oliva [63] visualized different categories of natural images by exposing the regularities in the intensity patterns across all the images in the same category. In each category, they used some images constrained to have a particular object at one scale present in the image. In this work, appropriate alignments of objects in the source images is important for exposing regularities. As the alignment process for real images require significant amount of user input, only 100 images were usually used in these work. In our work, we choose to average synthetic silhouettes of human poses as the alignment is an automatic process for synthetic images so that we can summarize a large number of poses in a database. Furthermore, instead of visualizing the mean image directly, we visualize the log density of the mean image to explore the subtleties of pose distribution in a database. Our work is

also related to motion-history images (MHIs) where a sequence of pose silhouettes in a video is summarized in a image and used as a motion template for activity recognition [7]. The main difference is that the image intensity of MHI encodes the recency of the motion rather than the silhouette distribution for the whole sequence.

Chapter 3

Natural Human Motion Classifier

In this chapter, we discuss our work on classifying human motion as natural or unnatural. Our basic assumption is that naturalness can be defined by a motion capture database that contains a variety of natural behaviors. Motions that similar to those in the database are regarded as natural. Similarity can be measured by statistical models that are learned from the database. In our work, we explore the performance of three classes of statistical machine learning techniques when trained on a large database of motion capture data and tested on sequences of unnatural and natural motion from a number of different sources. Because the validity of these results depends heavily on the training and testing datasets, we first describe those datasets and then explain the statistical techniques and show their performance.

3.1 Data

The training database consisted of 1289 trials (422,413 frames or about 4 hours) and included motions from 34 different subjects performing a variety of behaviors. Those behaviors included locomotion (42%: 5% jumping, 3% running, and 33% walking), physical activities (16%: basketball, boxing, dance, exercise, golf, martial arts), interacting with the

environment (7%: rough terrain, playground equipment), two subjects interacting (6%), and common scenarios (29%: cleaning, waiting, gestures).

The motion was captured with a Vicon motion capture system of 12 MX-40 cameras [67] at 120Hz and then downsampled to 30Hz. The subjects wore 41 markers, the three dimensional (3D) positions of which were located by the cameras. Using an automatically obtained skeleton for the user, the motion was further processed to the ASF/AMC format, which includes absolute root position and orientation, and the relative joint angles of 18 joints. These joints are the head, thorax, upper neck, lower neck, upper back, lower back, and left and right humerus, radius, wrist, femur, tibia, and metatarsal.

For the experiments reported here, we converted each frame of raw motion data to a high-dimensional feature vector of angles and velocities. For the root segment, we compute the angular velocity and the linear velocity (in the root coordinate system of each frame). For each joint, we compute the angular velocity. The velocities are computed as a central difference between the joint angle of the position on the previous frame and on the next frame. As a result, both joint angles and their velocities can be represented by unit quaternions (four components each). The complete set of joint angles and velocities, together with the root's linear velocity (three components) and angular velocity (quaternion, four components), form a 151-dimensional feature vector for each frame. The quaternions are transformed to be on one-half of the 4D sphere to handle the duplicate representation of quaternions. If the orientation of a joint crosses to the other half-sphere, we choose the alternative representation for that quaternion and divide the motion sequence at the boundary to create two continuous sequences. Fortunately this problem occurs relatively rarely in natural human motion because of human joint limits.

We generated a number of different test sets in an effort to span the space of natural and unnatural motions that might be generated by algorithms for producing human motion. Unlike our training data, the testing suite contains both positive (natural) and negative (unnatural) examples.

The negative testing sequences were obtained from a number of sources:

- Edited motions. Alias/Wavefront's Maya animation system was used to edit motion capture sequences to produce negative training examples. The editing was performed on either a joint or a limb using inverse kinematics.

- Keyframed motions. These motions were keyframed by an animator with significant Maya experience but limited keyframing experience.

- Noise. Noise has been used to generate human motion [47] and to improve the quality of captured motion by adding variation. We generate both positive and negative testing examples by varying the amount of noise and relying on a human observer to assess the naturalness of the motion.

- Motion transitions. These motions were computed using a commonly accepted metric for transitions (maintain contact constraints and keep the sum of the squared changes in joint angles below a threshold). Transitions above a high threshold and below a low threshold were then classified as good or bad by a human viewer. We included bad motion transitions in our negative test set.

- Insufficiently cleaned motion capture data. In the process of cleaning, motion capture data is transformed from the 3D marker locations to relative joint angles using a model of the subject's skeleton. For most marker sets, this process is accomplished through the use of inverse kinematics. If the markers have not been placed carefully or the kinematic chain is near a singularity, this process may result in unnatural motion (for example, knees that do not fully extend or swing out to the side if significantly bent).

The negative, or unnatural, testing set consisted of 170 trials (27774 frames or 15 minutes).

The positive testing set consisted primarily of motion capture data that was held out from the database. Additional positive testing data were created by adding noise to these motions and by generating motion transitions that were judged good by an expert human viewer. The natural motions consisted of 261 trials (92377 frames or 51 minutes).

3.2 Approach

The input data for our models, motion capture data, is a multivariate time series consisting of vectors of features (joint angles and velocities) sampled at discrete time instants. From this perspective, a model for natural motion must capture probabilistic dependencies between features across time. We construct this model in three steps. First, we select a statistical model to describe the variation in the data over time. We investigate three relatively standard techniques: mixtures of Gaussians (MoG), hidden Markov models (HMM) and switching linear dynamic systems (SLDS). Associated with each model is a set of model parameters and a likelihood function that measures the probability that an input motion sequence could be generated by the model. Second, we fit the model parameters using a corpus of natural human motion as training data. Third, given a novel input motion sequence, we compute a score which can be interpreted as a measure of naturalness.

By thresholding the naturalness score we obtain a classifier for natural motion. There are two types of classification errors: false positives (the classifier predicts natural when the motion is unnatural) and false negatives (the opposite case). By varying the threshold we can create a trade-off between these two types of errors. The *ROC curve* for a classifier summarizes its performance as a function of the threshold setting [66] (see Figures 3.2 and 3.3 for examples). Each threshold choice corresponds to an operating point on the ROC curve. By comparing the area under the ROC curve, we can measure the relative performance of a set of classifiers without the need to choose a particular threshold. In practice, the choice of operating point on the ROC curve will be dictated by the application requirements and will be assessed using a set of positive and negative examples that were not used for training.

We could construct a single statistical model of naturalness using the full 151-dimensional input feature vector for training. However, learning an accurate model for such a high-dimensional feature vector is difficult, even with a (relatively) large amount of training data. Therefore, we hierarchically decomposed the full body motion into its constituent parts and train an *ensemble* of statistical models, each responsible for modeling a particu-

lar part: joints, limbs, or the whole body. Given an input sequence, these smaller models would produce a set of likelihood scores and an ensemble rule would be used to combine these scores into a single naturalness measure. The ensemble approach has three potential advantages over creating a single model based on the complete feature vector:

- One potential problem in learning the parameters of statistical models is overfitting, which occurs when a model has excessive capacity relative to the amount of available training data. When overfitting occurs, the trained models do not generalize well during testing because they are excessively tuned to the training data set. The ensemble approach gives us flexibility in controlling the capacity of the individual models to prevent overfitting. In particular it allows us to control the degree of coupling between features in the model.

- In some motion sequences, the patterns of unnatural motion may be confined to a small set of joint angles. These cases can be difficult to detect with a single statistical model, because the small set of features with unnatural motion will be swamped by the majority of the features which are exhibiting natural motion. The ensemble approach avoids this problem because our method of combining the statistical models looks for an unnatural classification by *any* of the models, not an average classification of unnaturalness.

- The ensemble approach makes it possible to examine small groups of joints and identify the ones most strongly associated with the unnatural motion. This property should make it possible to provide guidance to the animator about what elements of the motion deserve the most attention.

We designed groups of features to capture dependencies between joints at different scales. Each group of features forms a feature vector that is associated with a single model in the ensemble. Specifically, given the input 151-dimensional feature vector, we define a set of 26 smaller feature vectors by combining joint angles and joint velocities into groups of features (figure 3.1). At the lowest level, we create an 8-D feature vector from each of

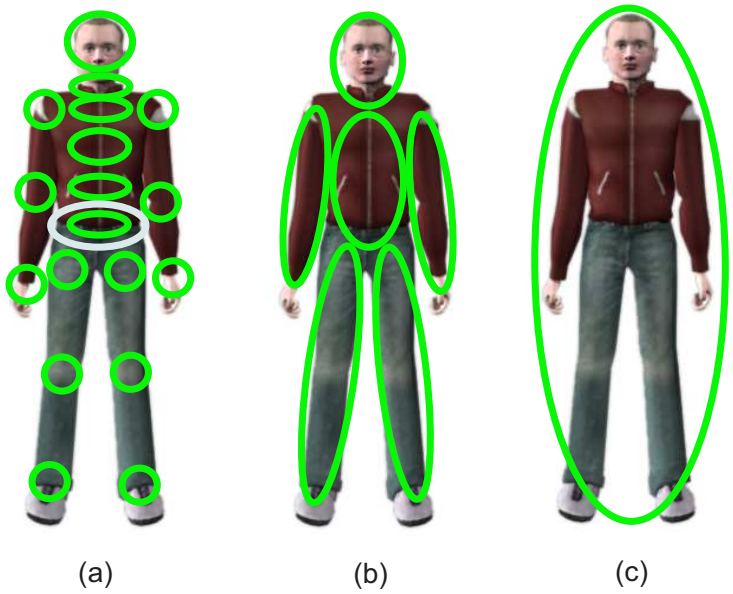


Figure 3.1: The three hierarchical groups of features. (a) At the lowest level each joint and its velocity form a feature group. Each feature group is illustrated as a green circle. The white circle represents the group of features from the root segment (linear velocity and angular velocity). (b) The next level consists of sets of joints grouped as limbs. (c) At the highest level, all the joints are combined into one feature group (without velocity information).

the 18 basic joints (angle and velocity). Another feature vector is created for the linear and the angular velocity of the root segment (seven features). To represent the aggregate motion of parts of the body, we assign a feature vector to each of the limbs: two arms (each three joints; 24 features), two legs (each three joints; 24 features), the head-neck group (head, upper neck, lower neck; 24 features) and the torso/root group (thorax, upper back, lower back, plus root; 31 features). Finally, at the top level, we define a feature vector representing the full body pose (rotation angles for all 18 joints but no velocities; 72 features). For the models created using HMM and SLDS, the feature vectors that comprise each of these feature groups are first processed with PCA (99% variance kept for the full-body model, 99.9% variance kept for the smaller models) to reduce the dimensionality.

Given an ensemble of models, we generate a naturalness measure for a motion sequence D of length T by first computing a score s_i for each model, where the model has parameters θ_i:

$$s_i = \frac{\log P(D \mid \theta_i)}{T} \tag{3.1}$$

The scores for each model will generally not be in the same range. Therefore we must normalize the scores before they can be combined. For each model, we compute the mean μ_i and standard deviation σ_i of the scores for the training data (after eliminating a small percentage of the high and low scores to reduce the effect of outliers). The final score for sequence D is then computed as follows:

$$s = \min_i \left(\frac{s_i - \mu_i}{\sigma_i} \right), i = 1, 2, \ldots, 26 \tag{3.2}$$

We choose the minimum (worst) normalized score from among the s_i because we assume that the entire motion should be labeled as unnatural if any of its constituent feature groups have a bad score.

We now describe the three statistical models used in our experiments, as well as a baseline method and a user study used for validating our results.

3.2.1 Mixture of Gaussians

We first experimented with a mixture of Gaussians (MoG) model because of its simplicity. The probability density of each feature vector was estimated using a mixture of 500 Gaussians, each with a spherical covariance. In this rudimentary representation, the dynamics of human motion are only encoded through the velocity components of the feature vector. As the result, this model is quite weak at modeling the dynamics of human movement.

3.2.2 Hidden Markov Models

Next, we experimented with a hidden Markov model (HMM) [49], because it explicitly encodes dynamics (change over time) and has been shown to work extremely well in other time-series domains such as speech recognition. In a HMM, the distribution of the body poses (and velocities) is represented with a mixture of Gaussians. In general, each hidden state in a HMM indexes a particular mixture density, and transitions between hidden states encode the dynamics of the data. Given positive training examples, the parameters of the HMM can be learned using the Expectation-Maximization (EM) algorithm. The parameters consist of the probabilities in a state transition matrix for the hidden state, an initial state distribution, and mixture density parameters. In the general case, this set of parameters includes mixture weights for each hidden state and the mean vectors and covariance matrices of the Gaussians.

For the full body HMM, we used a model with 180 hidden states. For the other feature groups comprising the ensemble of HMM, we used only 60 hidden states because the feature vectors were much smaller. Each hidden state in the HMM was modeled as a single Gaussian with a diagonal covariance matrix.

3.2.3 Switching Linear Dynamic Systems

A switching linear dynamic system (SLDS) model can be viewed as a generalization of a HMM in which each switching state is associated with a linear dynamic system (LDS) instead of a Gaussian distribution over the output space [46]. In a HMM, each switching state defines a "region" in the output space (e.g, poses and velocities), where the mean vector determines the location of the region and the covariance matrix determines its extent. In contrast, each LDS component in an SLDS model defines a family of trajectories with linear dynamics. We used a second-order auto-regressive (AR) model in our experiments. In this model, trajectories begin at an initial state that is described by a mixture of Gaussians. As the trajectory evolves, the state of the motion at time t is described by a linear combination of the state values at times $t-1$ and $t-2$ and the addition of Gaussian noise. By switching between these LDS components, the SLDS can model a system with nonlinear, non-Gaussian dynamics using a set of simple building blocks. Note that our application of SLDS does not require a separate measurement process, because we model the motion directly in the feature space.

Closely related to our SLDS model is the motion texture model [36]. The primary difference is that the motion texture approach confines each LDS element to a "texton" that is constrained to begin and end at specific keyframes, whereas we adopt the classical SLDS framework where transitions between LDS models can occur at each time step.

As in the HMM case, the SLDS model parameters are estimated using the EM algorithm. However, a key difference is that exact inference in hybrid dynamic models like SLDS is generally intractable [35]. We employed an approximate Viterbi inference algorithm which computes an approximation to the highest probability switching sequence [46].

Given a new motion sequence, we compute a score that corresponds to the log likelihood of the data under the SLDS model. This score is the sum of the log likelihoods for each frame of data. Per-frame scores depend on the cost of switching between models and the size of the one-step-ahead error between the model's prediction and the actual feature

vector.

For the full body SLDS, we used an SLDS model with 50 switching states. For the other groups of features comprising the ensemble model, we used five switching states each. We used diagonal covariance matrices for the noise process.

3.2.4 Naive Bayes (Baseline Method)

To establish a baseline for the other experiments, we also implemented a simple marginal histogram probability density estimator based on the Naive Bayes (NB) model. Assuming that the components of our 151-dimensional feature vector are independent (which is clearly wrong), we computed 1D marginal histograms for each feature over the entire training database. Each histogram had 300 buckets. Given this model, we estimated the score of a new testing sequence by summing over the log likelihoods of each of the 151 features for each frame and then normalizing the sum by the length of the motion sequence. Note that this method captures neither the dependencies between different features (even those comprising a single joint angle), nor the temporal dependencies between features at different frames (although velocities do provide some measure of dynamics). As expected, this method does not perform particularly well, but we included it as a baseline with which to compare the other, more complicated approaches.

3.2.5 User Study

To evaluate our results, we performed a user study approved by Institutional Review Board (IRB) of Carnegie Mellon University. Twenty-nine male subjects and twenty-five female subjects with different backgrounds and races were obtained by university-wide advertising.

We randomly selected and rendered 118 motion sequences from our testing set (approximately half from the positive testing set and half from the negative testing set). We showed the rendered videos to subjects in two segments with a 10 minute break between

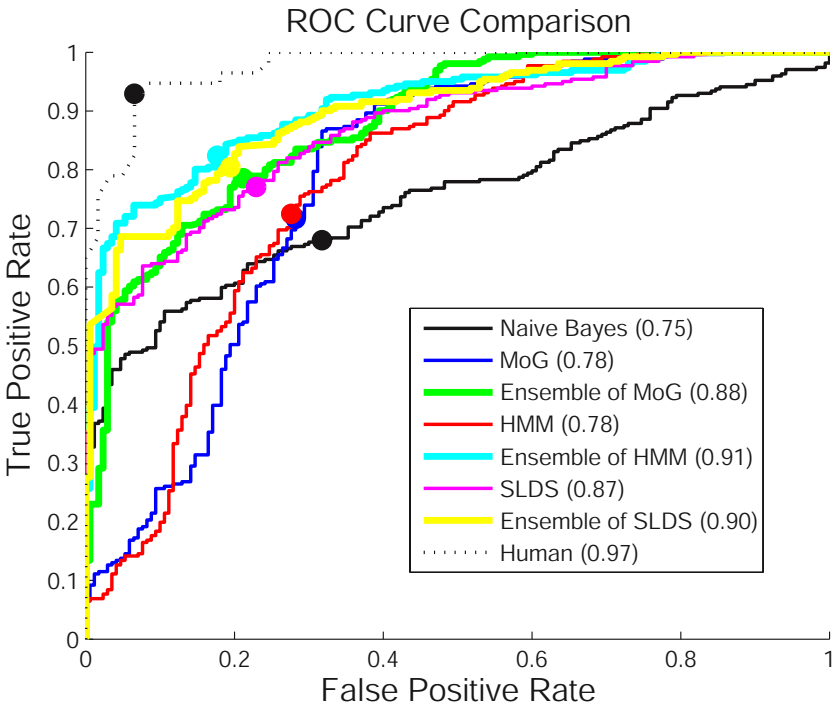


Figure 3.2: The ROC curves for each statistical model and for the human subjects in our user study. The circle on each curve represents the equal error rate. The area under the ROC curve is given in parentheses.

the segments. Each segment contained half the sequences in a random order and the ordering of the presentation of the two segments was randomized between subjects. After watching each motion, the subjects wrote their judgment about the naturalness of the motion (yes or no). The total length of the study (including the break) was about 30 minutes. For comparison with the statistical models, the results of the user study are summarized in Section 3.3.

Method	Positive Test Set (261)	Bad Motion Capture (37)	Edited (60)	Keyframed (11)	Noise (30)	Transition (32)	Area Under ROC	Number of Parameters
Naive Bayes	0.69	0.75	0.73	0.80	0.76	0.40	0.75	45,600
MoG	0.71	0.86	0.97	1.00	0.37	0.28	0.78	76,000
Ensemble MoG	0.74	0.89	0.80	1.00	0.80	0.40	0.88	201,000
HMM	0.72	0.78	1.00	1.00	0.53	0.22	0.78	21,087
Ensemble HMM	0.82	0.89	0.78	1.00	0.83	0.75	0.91	43,272
SLDS	0.76	0.78	0.75	1.00	0.43	1.00	0.87	333,150
Ensemble SLDS	0.82	0.76	0.82	1.00	0.67	0.97	0.90	159,340
Human Subjects	0.93	0.75	1.00	0.81	1.00	0.92	0.97	NA

Table 3.1: The percentage of each type of testing data that was classified correctly by each classification method (using the point on the ROC curve with equal error rate). The number of test sequences for each type of motion is given in parentheses.

3.3 Experiments

We trained the statistical models on the database of four hours of human motion and tested them on a set of 261 natural and 170 unnatural motions. Figure 3.2 shows the ROC curves for each method. The ROC curve for the user study was computed by varying the threshold for the number of subjects who must mark a motion as natural for it to be labeled as natural. The testing set for the human subjects was only 118 of the 431 testing motions to prevent fatigue.

Table 3.1 gives the area under the ROC curve for each method. For the single full-body models (151 features), SLDS had the best performance, followed by HMM and MoG. Each ensemble of 26 models performed better than the single model that used the same statistical technique. This improvement occurs largely because the smaller statistical models and our method of combining their scores makes the ensemble more sensitive to unnatural motion of a single joint than a single statistical model. The ensemble of HMM had the largest area under the ROC curve, although the performance of all three ensemble methods was similar. The human subjects performed significantly better than any of the methods, indicating that it may well be possible to develop better methods.

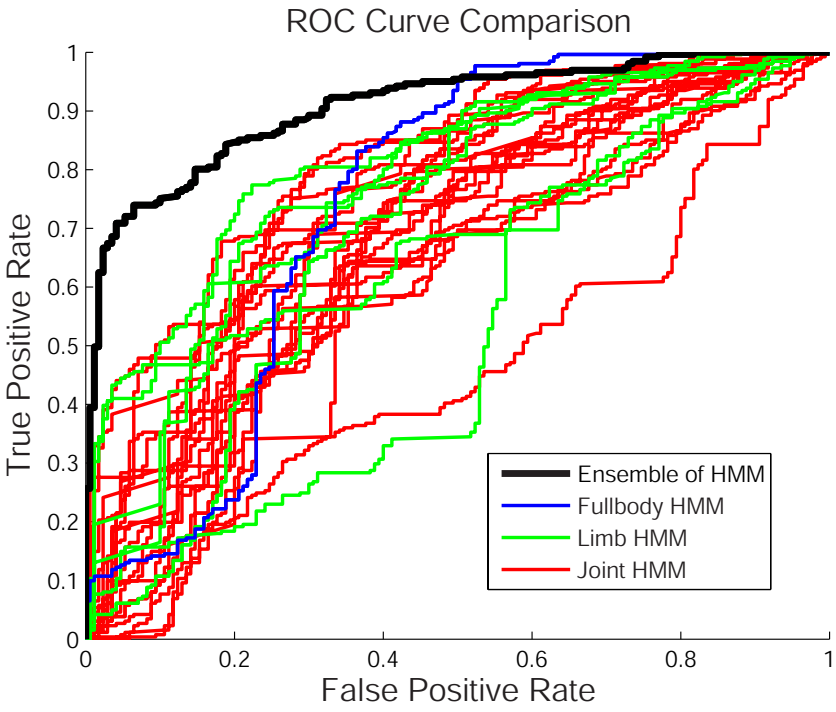


Figure 3.3: ROC curves for each of the 26 HMM and the combined ensemble HMM. The HMM for the individual joints are shown in red, for limbs in green, and for the full body in blue. The lowest curve corresponds to the right wrist which also causes the curve for the right arm to be low.

Table 3.1 also gives the percentage of the testing data that were classified correctly for each category of the test set and each model. The threshold setting for each classifier corresponds to the point of equal error rate on the ROC curve (see Figure 3.2). This point on the ROC curve is where the percentage of false positives equals the percentage of false negatives. Bad motion capture data was not easy for most of the classifiers to detect with only the ensemble of MoG and of HMM having success rates near 90%. The human subjects were also not particularly good at detecting those errors, perhaps because the errors were generally of short duration and the subjects did not have experience with the process of capturing or cleaning motion capture data. All of the methods were able

to correctly classify more than 70% of the edited motions as unnatural, and MoG, HMM, and the human subjects had a success rate of over 95% on those motions. The keyframed motions were small in number and were largely classified correctly as unnatural by all methods and the human subjects. The addition of sinusoidal noise was more difficult for most of the methods to detect with only ensemble MoG and ensemble HMM achieving scores near 80%. The human subjects, on the other hand, could easily discriminate these motions, scoring 100%. Motions with bad transitions were the most difficult type to identify for all of the methods, with the exception of SLDS and ensemble of SLDS. During a bad transition, the velocities change due to blending in a way that is locally smooth, but is inconsistent with the dynamics of the initial and final motion. We hypothesize that the good performance of the SLDS models can be attributed to their ability to correctly model longer-term temporal properties.

Table 3.1 also describes the number of parameters in each of the models. These parameters are the degrees of freedom that the model can exploit in fitting the data and provide a crude measure of the representational resources of the models. The ensemble of MoG and of SLDS have many more parameters than the ensemble of HMM but produce slightly inferior performance. This discrepancy is perhaps a sign that these more complex models may be overfitting the training data.

Figure 3.3 further explores the performance of the ensemble of HMM. Each type of model is shown in a different color: single joints, limbs, and full body. As expected, the ensemble model that is computed by combining the scores of the individual HMM has significantly better performance than any single HMM. The individual HMM are fairly tightly bunched indicating that each potentially has value in the computation of the overall score.

One advantage of the ensemble approach is that it can be used not only to detect unnatural motions but also to localize problem areas among the joints. This property is illustrated in Figure 3.3 where the color of a block indicates whether a particular HMM found each motion to be natural (black) or unnatural (red to yellow). In order to detect unnatural motion in an individual joint or limb, we compare the normalized score from

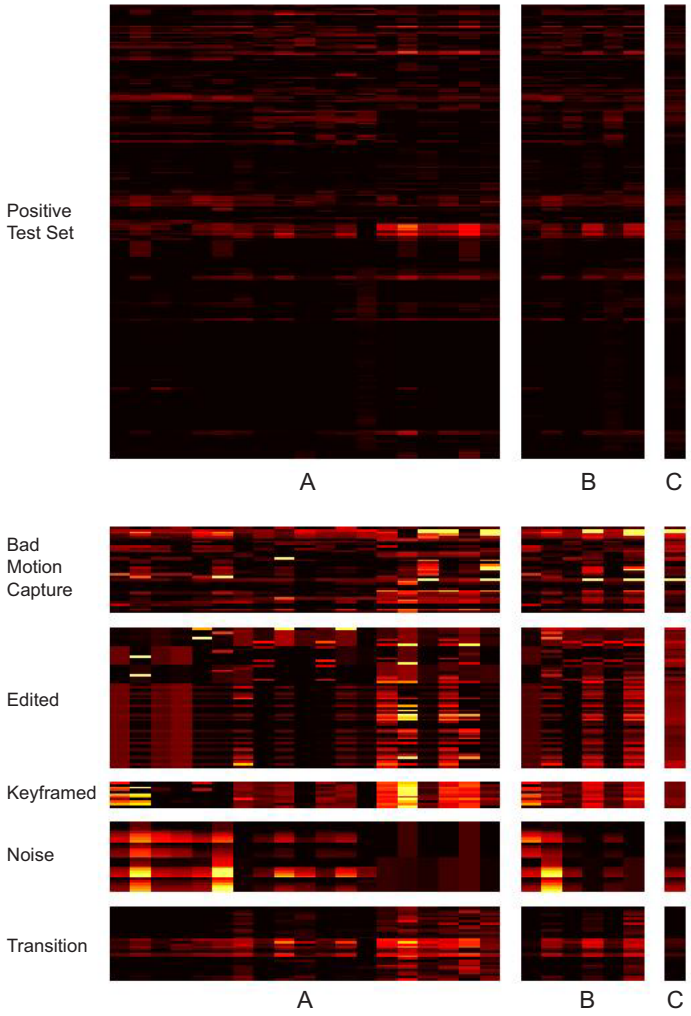


Figure 3.4: Response of the ensemble of HMM to the positive and the negative testing data. Each row shows the responses of all 26 models to a particular testing sequence. The intensity of the color (red to yellow) indicates a decreasing score (more unnatural). Each column corresponds to a single ensemble, grouped as follows: A-joints, B-limbs, and C-full-body (see Figure 3.1).

the corresponding smaller model with the threshold that gives the equal error rate for the

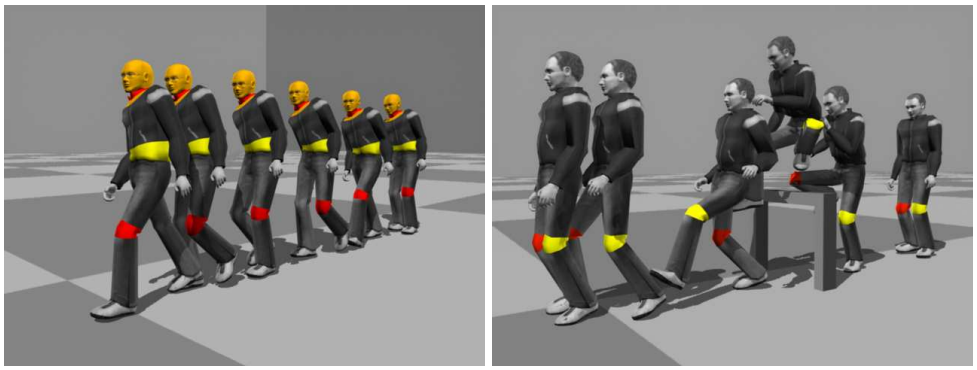


Figure 3.5: Two examples from our negative test set of motions. Both of them are unnatural motions. The sequence on the left is badly edited motion. The sequence on the right is incompletely cleaned. Joints that are marked in red-yellow (red is unnatural and yellow is most unnatural) were detected as having unnatural motion. Our scheme does not pinpoint the period of time when the unnatural motion happens.

ensemble classifier. Joints that are below threshold are flagged as unnatural and rendered with a color that is proportional to the score. Two unnatural motions are visualized in Figure 3.5 with the joints that were detected as unnatural shown in red-yellow. By localizing problem areas to particular joints or limbs, we found errors in our previously published database that had not been noticed when the data was cleaned and processed.

Our user study produced a true positive rate of 93% and a 7% false positive rate. The subjects were drawn from a variety of disciplines and had not spent any significant time studying human motion data so it is perhaps not surprising that their classification did not agree completely with that of the authors when they assembled the testing database. Informal interviews with the subjects indicated that they were sometimes confused by the absence of objects that the character should have been interacting with (a box that was stepped onto, for example). If the semantics of the motion was not clear, they were likely to label it as unnatural. The subjects also missed some errors in the motion, most commonly those of short duration.

The training time for each of these statistical methods was significant, ranging from a few hours for the simpler methods to several days for the ensemble methods. The testing time is not long however, we were able to test the entire set of motions in 20 minutes.

3.4 Discussion

The quality of our measures is likely dependant on the quality of the motion database of positive examples used to train them. Motions that are quite distant from those in the training set will be easily judged unnatural even if they are in fact natural. In our experiments, we have seen that unusual motions that have little in common with other motions in the database are sometimes labeled unnatural. For example, we have only a few examples of falling in the motion database and examples of that behavior were judged as unnatural by our measures. On the other hand, we have also seen evidence that the measures do generalize. For example, our testing set included walking while picking up a coffee mug from a table. This motion was judged natural by most of the methods although based on a visual inspection, the closest motions in the training data set were a two arm reach while standing, walking, and sweeping with a broom.

Negative examples often bear the imprint of the algorithm used to create them. For example, carelessly edited motions might evidence unbalanced postures or foot sliding if inverse kinematics was not used to maintain the foot constraints. Similarly, motions that include bad transitions often have significant discontinuities in velocity as the algorithm attempts to blend between two distant poses with differing velocities. We have attempted to address this concern by testing on a wide variety of common errors: motions that were aggressively edited in a commercial animation package, motions that were keyframed by an inexperienced animator, badly cleaned motion capture data, bad (and good) transitions, and motions with synthetic noise added. A larger variety of negative training examples would allow a more rigorous assessment of competing techniques.

Despite our attempt to span the space of motion errors with our negative testing set,

other common errors may not be reliably detected. For example, our methods will likely not detect very short errors because the score on a motion is computed as an aggregate over an entire sequence of motion. The magnitude of the error caused by a single glitch in the motion will be reduced by the high percentage of good, natural motion in the sequence. We could potentially reduce this problem by only computing naturalness scores for short sequences. For a longer sequence, we could first divide it into short sub-sequences, then evaluate the naturalness score for each sub-sequence individually, and finally take the minimum as the score for the original sequence. This approach could also help improve joint error identification in time. Specific sub-sequences of joint motion can also be flagged as unnatural if their scores are below the common threshold that gives equal error rate for the ensemble classifier.

Our measures are also not very effective at detecting otherwise natural motion that has been slowed down by a factor of two. Such a slow-down is sometimes difficult for human observers to detect as well, particularly for behaviors that do not include a flight phase to provide decreased gravity as a reference. We believe that our methods do not perform well on these motions because the poses and lower velocities seen in these motions are "natural" in the sense that they would be seen in such natural behaviors as slow walks. Furthermore, the HMM have self-loops that allow slower motions to pass without significant penalty. Even though SLDS is powerful, its performance is not better than that of the HMM on these examples. The reason could be that the approximate inference of this complex model can easily get stuck in local minima.

Our approach to measuring the naturalness of a motion via ensembles of smaller models was quite successful. However, it is likely that the methods could be improved, given that human observers perform significantly better on our test set. In the approach reported here, we used our knowledge about the synergies of human motion to pick appropriate feature groups. We expect that feature selection from among a larger set of features might produce better results.

Chapter 4

Exploring the Statistics of Natural Human Motion

In this chapter, we discuss our work on exploring the statistics of natural human motion. We introduce the motion data and their representations in our statistical analysis in the next section. In Section 4.2, we present statistics on the range of joint motion. We look at different aspects of dimensionality reduction in Section 4.3 and develop a statistical analysis method to identify joint angles that have synergies in human motion in Section 4.4. In Section 4.5, we develop summary images to intuitively represent motion data sets in a single image. Finally, we discuss the insights gleaned from these statistics in the last section.

4.1 Motion Data

Our analysis is based on a large and representative motion capture database. The database contains more than four hours of motions captured from 34 different subjects performing a variety of behaviors such as locomotion (42%), common scenarios such as daily activities (29%), physical activities (16%), interacting with the environment (7%) and two subjects

interacting (6%). For comparison, we also perform some of our analysis using behavior-specific data sets: walking, running, forward jumping and swing dancing.

All the motions were captured with a Vicon motion capture system with 12 MX-40 cameras [67]. The motion was captured at 120Hz and then downsampled to 30Hz. The subjects wore 41 markers, the 3D positions of which were located by the cameras. The skeleton of the user is obtained automatically from a subject calibration process (“T” pose, motorcycle pose, and joint range of motion capture) and stored in the ASF format. Each motion is cleaned (gaps are filled, marker correspondence problems are fixed) and then the motion is converted from 3D marker locations (C3D format) to AMC, the joint angle format.

This pipeline is a standard one for processing motion capture data. However, this marker set and processing has several flaws which may affect the statistics that we wish to compute. First, 41 markers are not complete in that there are not enough markers to form an independent coordinate system for each body part. Therefore the processing of the data includes assumptions about the placement of markers (at the joint axis or along a line between two joints, for example) and the motion is constrained to a simplified skeleton when it is converted to the AMC format. For example, the knee is modeled as a simple 1-DOF joint but the axis of rotation of the human knee depends on the joint angle [39]. Motion outside of the DOFs in the ASF skeleton we use might occur due to muscle or skin movement relative to the skeleton, off-axis joint motion, or noise in the capture of the 3D location of the markers. A data set based on a more complete marker set would be better for our statistical analysis, however, such a data set is not available.

We first discuss the reference skeleton which is based on the ASF file format. We then introduce Euler angles, the joint orientation representation used in the AMC file format. We finally compare Euler angles to other joint orientation representations.

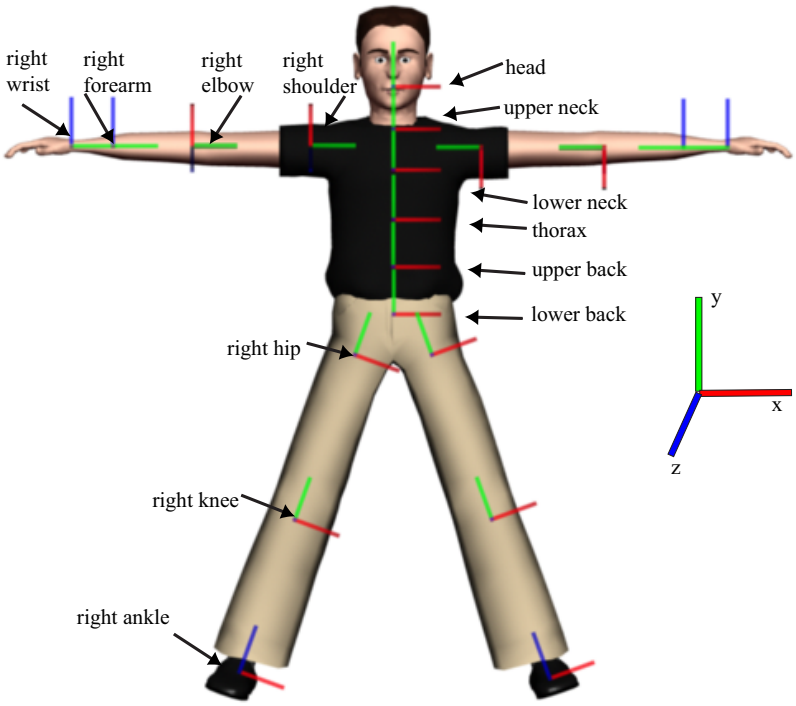


Figure 4.1: The reference human skeleton. A local coordinate system is established at the end of the inboard bone for each joint. The movement of the outboard bone is represented as an orientation with respect to this local coordinate system creating a hierarchical structure.

4.1.1 Reference Skeleton

The human skeleton for our motion data is based on VICON's ASF file format (Figure 4.1). However, in this chapter, we do not use the joint names in the ASF/AMC file format but instead use more common and hopefully more intuitive names. The skeleton used in the Carnegie Mellon database has 18 joints, which can be classified into four groups. The torso joint group includes the lower back, upper back and thorax. The head joint group includes the lower neck, upper neck and head. The arm joint group includes the shoulder,

elbow, wrist, and hand for the left and right arms. The leg joints group includes the hip, knee, and foot for the left and right legs. Based on the ASF file format, all the joints are organized in a hierarchical tree structure with the root node located at the lower back. Each body part has two joints (aside from the end effectors). Following the convention used in character animation, we call a joint the inboard joint if it is closer to the root and the outboard joint otherwise. Each joint connects two body parts except for the lower neck and lower back which connects three bones. Similarly, we call a body part inboard if it is closer to the root in the hierarchy and outboard if it is further from the root. A local coordinate system is anchored at the end of the inboard bone of each joint (Figure 4.1).

Because our motion data were captured from 34 different subjects, we have 34 different skeletons which have the same DOF but different limb lengths. We simplify the analysis by using only one reference skeleton. The limb length variations may influence the results of statistical analysis, however, the impact should be minor as our analysis uses only relative joint angles, a representation which is not particularly sensitive to the skeleton size variations found in adult humans. For example, the inter-subject variance has been shown to be extremely small for the range of motion at the shoulder joint [71].

We follow the AMC convention and represent the joint angles using the XYZ fixed angle representation or equivalently the Euler angle representation ZYX. The 42 DOFs are described in Table 4.1. The body configuration at each frame is represented as a 42-dimensional pose vector in our statistical analysis. The Euler angles used in the ASF/AMC format are only one possible representation for joint angles. Other common representations are unit quaternions and exponential maps and we use these representations in some of our experiments to understand the effect of representation on the computed statistics.

Euler angles are often preferred because they are intuitive. They are widely used in robotics where the physical design of the hardware usually prevents problems with gimbal lock. Although singularities will always exist in any 3D parametrization of orientation, we can carefully design the reference skeleton so that the range of joint angles seen in natural human motion avoids the singularities as much as possible. The strategy was used in the ASF/AMC file format and it is quite successful for all joints except for the shoulder.

Joint Group Name	Joint Name	Number of DOFs	Joint Angle ID
Torso	Lower Back	3 (X, Y, Z)	1-3
	Upper Back	3 (X, Y, Z)	4-6
	Thorax	3 (X, Y, Z)	7-9
Head	Lower Neck	3 (X, Y, Z)	10-12
	Upper Neck	3 (X, Y, Z)	13-15
	Head	3 (X, Y, Z)	16-18
Arm	Right Shoulder	3 (X, Y, Z)	19-21
	Right Elbow	1 (X)	22
	Right Forearm	1 (Y)	23
	Right Wrist	1 (X)	24
	Left Shoulder	3 (X, Y, Z)	25-27
	Left Elbow	1 (X)	28
	Left Forearm	1 (Y)	29
	Left Wrist	1 (X)	30
Leg	Right Hip	3 (X, Y, Z)	31-33
	Right Knee	1 (X)	34
	Right Ankle	2 (X, Z)	35-36
	Left Hip	3 (X, Y ,Z)	37-39
	Left Knee	1 (X)	40
	Left Ankle	2 (X, Z)	41-42

Table 4.1: The DOFs of each joint in each joint group. The joint angle IDs represent the ordering for the joint angles in the pose vector used in the statistical analysis.

Unit quaternions are located on a hypersphere in R^4 called S^3 and are free of singularities. The representation is particularly suited for animation applications that are based on motion interpolation as the interpolation metric is well defined in S^3. One redundancy exists in this representation because the *antipodal symmetry* of the unit quaternion allows any unit quaternion q and its negative $-q$ to represent the same orientation. We preprocess the data to remove discontinuities caused by this redundancy by choosing a single hemisphere in S^3 and then mapping all joint angles to this hemisphere. We first compute the approximate mean orientation from the data using a fast approach [45] and then use the

approximate mean as a reference quaternion to map all joint orientations to the single hemisphere centered at this reference point in S^3.

Exponential and logarithmic mappings take vectors in R^3 into unit quaternions in S^3 and vice versa. More specifically, the logarithmic mapping takes a unit quaternion q into a point, $log(q)$, in the tangent space (R^3) at the identity point, I, thus providing a 3D parameterization of orientation. The local linearity in 3D makes this representation suitable for animation applications that are based on additive motion editing operations. A log map can take each orientation in SO(3) to an infinite number of points in R^3. We follow the technique of Yahia and Gagalowicz [75] and limit the magnitude of the log map to $|log(q)| \leq \pi$ so that we can obtain a one-to-one mapping. Any 3D vector in the tangent space describes the axis and magnitude of a 3-DOF orientation.

The distance metric used for each orientation representation plays a key role in the statistical analysis. The Euclidean distance metric, which is commonly used with Euler angles, is problematic because it does not reflect the correct distance—the shortest rotation between any two orientations (the geodesic distance in S^3). This property may cause problems with common operations in statistical analysis such as interpolation or averaging between distant orientations. The spherical linear interpolation or slerp, introduced to the graphics community by Shoemake [57] can correctly interpolate between two quaternions. For exponential and logarithmic mappings, the Euclidean distance metric in the tangent space only approximates the geodesic distance metric in S^3. If the orientations are far from the reference point of the tangent space (the identity point) or close to the singularities (at orientations with rotation magnitude of $2n\pi$ (for $n = 1, 2, 3, ...$)), this approximation will not be accurate. On the other hand, this problem is reduced if the distribution of orientations is compact as they are for most human joints. In this case, we can set the reference point (p) of the tangent space as the approximate mean orientation of the data. Then we can map a unit quaternion q into the tangent space at a location p by rotating the hypersphere to align p with the identity and then taking the logarithm ($log(p^{-1}q)$).

In this chapter, we compute the basic joint statistics based on Euler angle representations because that is the most common representation in animation algorithms. We also

compare Euler angles to other representations for some of the statistics as we compute. For example, we test different joint angle representations to understand their effects on the compression performance of PCA.

4.2 Joint Statistics

Based on the full motion capture database, we compute statistical quantities for joint angles, angular velocities and angular accelerations. With the XYZ fixed angle representation, we can compute the limits and the 1D distributions of the joint angles, the angular velocities, and the angular accelerations (Figure 4.2). The angular velocities and the angular accelerations are computed from discrete joint angles using a central difference approximation. Because differentiating will amplify noise, we apply a low-pass filter with a five frame window to the original motion capture data at 120Hz before computing angular velocities and angular accelerations. In Figure 4.2, we visualize the distributions (1D histograms) of these statistical quantities using a color map whose value is proportional to the log density of the distribution. We report the limits for the joint angles, the angular velocities and the angular accelerations in Table 4.2. Because the absolute limits may suffer from outliers in the data, we report the minimum value in the 1st percentile of the data and the maximum in the 99th percentile. We also fit Gaussian distributions to the data and report these parameters in Table 4.3.

We compare some of those statistics with these computed from a behavior-specific data set. In Figure 4.3, we visualize the distributions (1D histograms) of joint angles for typical locomotions such as walking and running and complex behaviors such as swing dancing. We compare these behavior-specific joint angle distributions (Figure 4.3) with those computed from the full database (Figure 4.2) and observe that the range of joint movements in each behavior is smaller than that of general human motion. We also find that the leg joints exhibit similar joint angle distributions in these three behaviors, which may reflect the fact that balancing is essential in all three behaviors. On the other hand, the arm joints

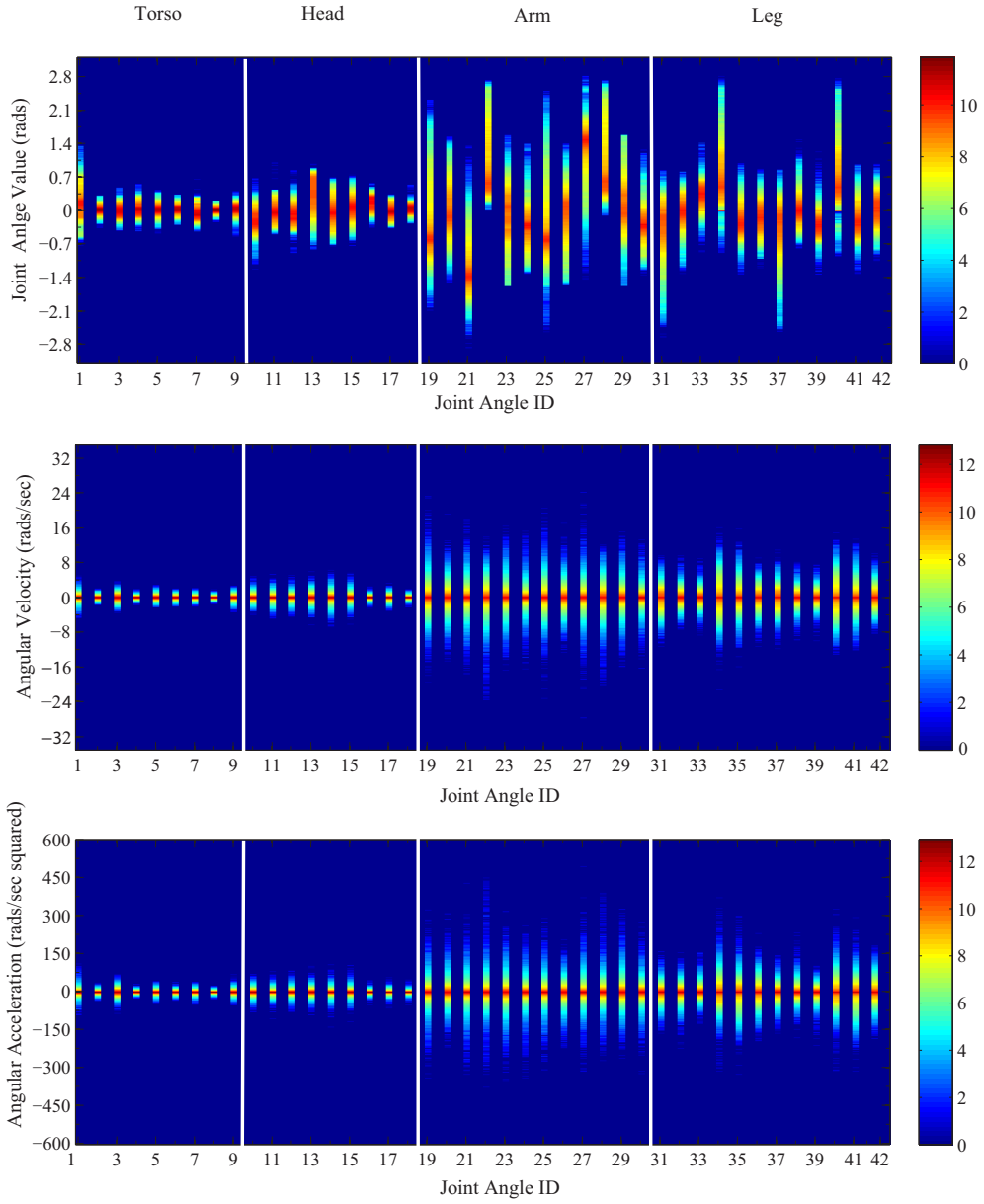


Figure 4.2: The limits and distributions of joint angles, angular velocities and angular accelerations. The color value of the bar on the right is proportional to the log of the number of entries in each bin. The bin sizes are 0.015 rads, 0.165 rads/sec and 3.00 rads/sec.

Joint Angle ID	01	02	03	04	05	06	07	08	09
Angle Min	-0.74	-0.35	-0.41	-0.43	-0.48	-0.29	-0.43	-0.23	-0.53
Angle Min (1%)	-0.40	-0.11	-0.19	-0.16	-0.15	-0.13	-0.24	-0.07	-0.14
Angle Max	1.42	0.31	0.48	0.56	0.43	0.33	0.66	0.21	0.40
Angle Max (99%)	0.73	0.14	0.18	0.25	0.19	0.16	0.22	0.09	0.17
Velocity Min	-6.77	-2.74	-4.33	-1.64	-2.68	-2.43	-2.00	-1.76	-3.42
Velocity Min (1%)	-0.84	-0.46	-0.97	-0.36	-0.63	-0.51	-0.44	-0.31	-0.64
Velocity Max	5.03	2.05	4.28	1.87	2.97	1.91	2.28	1.74	2.79
Velocity Max (99%)	0.92	0.47	0.98	0.36	0.64	0.50	0.40	0.32	0.63
Acceleration Min	-104.95	-37.57	-93.27	-29.32	-51.59	-35.40	-52.58	-42.86	-61.78
Acceleration Min (1%)	-12.06	-6.30	-15.90	-5.27	-8.09	-6.83	-7.59	-4.09	-9.50
Acceleration Max	110.23	37.28	83.93	33.93	49.12	38.37	41.62	30.37	78.85
Acceleration Max (99%)	12.57	6.18	16.00	5.35	7.93	6.91	7.31	4.02	9.49

Joint Angle ID	10	11	12	13	14	15	16	17	18
Angle Min	-1.31	-0.48	-0.88	-0.81	-0.74	-0.73	-0.47	-0.36	-0.26
Angle Min (1%)	-0.53	-0.34	-0.38	-0.31	-0.47	-0.34	-0.07	-0.24	-0.10
Angle Max	0.67	1.01	0.87	0.97	0.69	0.74	0.57	0.34	0.53
Angle Max (99%)	0.17	0.31	0.24	0.79	0.46	0.43	0.36	0.21	0.16
Velocity Min	-4.66	-6.05	-4.78	-6.81	-8.28	-6.07	-2.87	-3.54	-2.18
Velocity Min (1%)	-0.82	-0.94	-0.81	-1.13	-1.28	-0.94	-0.40	-0.64	-0.34
Velocity Max	5.12	5.26	5.14	6.42	7.02	6.07	3.18	3.32	2.03
Velocity Max (99%)	0.86	0.88	0.78	1.13	1.18	0.97	0.42	0.59	0.34
Acceleration Min	-119.88	-105.90	-112.25	-165.48	-141.57	-121.23	-79.56	-65.30	-52.19
Acceleration Min (1%)	-12.66	-9.35	-11.37	-14.00	-11.58	-12.64	-5.82	-6.20	-5.09
Acceleration Max	128.99	116.31	99.81	152.50	154.78	117.09	74.22	72.33	49.00
Acceleration Max (99%)	12.80	9.21	11.30	14.45	11.43	12.68	5.81	6.16	5.08

Joint Angle ID	19	20	21	22	23	24	25	26	27	28	29	30
Angle Min	-2.20	-1.52	-3.04	-0.03	-1.57	-1.36	-2.59	-1.55	-1.53	-0.11	-1.57	-1.37
Angle Min (1%)	-1.24	-0.85	-1.90	0.25	-1.29	-1.02	-1.23	-1.18	0.00	0.22	-0.72	-0.93
Angle Max	2.31	1.54	1.55	2.72	1.57	1.48	2.81	1.50	2.92	2.71	1.57	1.51
Angle Max (99%)	1.41	1.04	-0.02	2.44	0.79	0.42	1.41	0.75	1.94	2.36	1.33	0.37
Velocity Min	-25.06	-15.86	-25.11	-25.76	-23.11	-18.18	-28.83	-13.89	-30.14	-23.79	-20.07	-22.31
Velocity Min (1%)	-4.12	-3.49	-3.55	-4.50	-3.49	-2.25	-3.89	-3.15	-3.47	-4.02	-3.16	-2.15
Velocity Max	26.76	20.48	21.72	14.78	21.16	18.88	29.74	16.98	27.18	12.68	30.49	20.24
Velocity Max (99%)	4.34	3.55	3.58	4.04	3.57	2.36	4.17	3.23	3.37	3.68	3.08	2.17
Acceleration Min	-497.87	-331.09	-519.01	-375.91	-456.49	-399.66	-499.27	-303.09	-509.22	-292.39	-460.72	-487.60
Acceleration Min (1%)	-41.74	-40.02	-43.02	-45.18	-52.75	-37.62	-39.32	-41.53	-38.55	-38.59	-45.09	-36.91
Acceleration Max	443.08	350.93	506.97	476.01	396.09	349.84	471.26	342.64	495.82	508.10	462.77	325.94
Acceleration Max (99%)	41.38	42.40	41.40	47.99	48.58	36.65	39.32	37.25	41.27	41.67	50.28	36.08

Joint Angle ID	31	32	33	34	35	36	37	38	39	40	41	42
Angle Min	-2.64	-1.35	-1.04	-0.88	-1.37	-1.20	-2.62	-0.80	-1.33	-0.95	-1.33	-1.06
Angle Min (1%)	-1.60	-0.59	-0.10	-0.00	-0.65	-0.68	-1.56	-0.43	-0.63	-0.00	-0.63	-0.50
Angle Max	0.83	0.85	1.41	2.81	0.99	0.86	0.86	1.51	1.23	2.76	0.99	1.07
Angle Max (99%)	0.33	0.43	0.66	2.06	0.37	0.36	0.34	0.57	0.12	2.05	0.41	0.52
Velocity Min	-11.67	-8.44	-8.66	-21.23	-16.01	-12.78	-12.25	-9.75	-11.75	-17.87	-13.82	-9.72
Velocity Min (1%)	-3.64	-2.12	-1.99	-5.26	-3.52	-3.05	-3.54	-2.16	-1.91	-5.15	-3.67	-2.92
Velocity Max	10.36	9.92	9.45	17.75	17.81	11.37	10.92	8.35	10.34	16.94	15.21	10.50
Velocity Max (99%)	2.93	2.24	1.95	5.79	4.18	2.91	2.86	2.09	2.01	5.76	4.21	3.13
Acceleration Min	-203.95	-223.43	-190.63	-366.32	-363.97	-275.82	-181.93	-201.89	-249.48	-365.65	-340.27	-270.18
Acceleration Min (1%)	-37.16	-35.02	-24.31	-62.29	-78.29	-53.48	-36.76	-30.21	-29.13	-60.77	-81.30	-51.57
Acceleration Max	210.82	177.42	191.48	417.58	358.17	256.86	190.52	240.48	123.04	370.87	346.57	208.71
Acceleration Max (99%)	33.08	30.82	29.56	70.61	54.39	51.92	32.30	34.34	24.27	71.27	55.14	53.97

Table 4.2: The limits of the joint angles, the angular velocities and the angular accelerations for the 42 joint angles in the torso, head, arm and leg joint group. The limits shown are Minimum, Maximum, 1-percentile-Minimum and 99-percentile-Maximum.

Joint Angle ID	01	02	03	04	05	06	07	08	09
Joint angle $\mu(rads)$	0.14	0.01	-0.01	0.02	0.01	0.02	-0.06	0.01	0.03
Joint angle $\sigma(rads)$	0.20	0.04	0.08	0.08	0.06	0.05	0.08	0.03	0.06
Angular Vel. $\mu(rads/sec)$	0.00	0.00	0.00	0.00	0.00	-0.00	-0.00	0.00	-0.00
Angular Vel. $\sigma(rads/sec)$	0.30	0.16	0.34	0.12	0.22	0.17	0.15	0.11	0.22
Angular Acc. $\mu(rads/sec^2)$	-0.00	-0.00	-0.00	-0.00	-0.00	0.00	0.00	-0.00	0.00
Angular Acc. $\sigma(rads/sec^2)$	4.34	2.19	5.36	1.83	2.83	2.33	2.60	1.45	3.30

Joint Angle ID	10	11	12	13	14	15	16	17	18
Joint angle $\mu(rads)$	-0.21	-0.03	-0.09	0.35	-0.04	0.05	0.17	-0.02	0.03
Joint angle $\sigma(rads)$	0.14	0.12	0.12	0.26	0.18	0.14	0.10	0.08	0.05
Angular Vel. $\mu(rads/sec)$	0.00	-0.00	0.00	-0.00	-0.00	0.00	-0.00	-0.00	0.00
Angular Vel. $\sigma(rads/sec)$	0.28	0.29	0.26	0.37	0.40	0.32	0.14	0.19	0.11
Angular Acc. $\mu(rads/sec^2)$	0.00	-0.00	-0.00	-0.00	-0.00	0.00	-0.00	-0.00	0.00
Angular Acc. $\sigma(rads/sec^2)$	4.37	3.25	4.00	4.86	4.14	4.45	2.06	2.14	1.81

Joint Angle ID	19	20	21	22	23	24	25	26	27	28	29	30
Joint angle $\mu(rads)$	-0.37	-0.01	-1.26	0.91	-0.10	-0.33	-0.41	-0.07	1.35	0.90	0.12	-0.30
Joint angle $\sigma(rads)$	0.54	0.37	0.36	0.52	0.40	0.25	0.51	0.36	0.36	0.53	0.39	0.21
Angular Vel. $\mu(rads/sec)$	0.00	0.00	0.00	0.00	0.00	-0.00	0.00	-0.00	-0.00	0.01	0.00	-0.00
Angular Vel. $\sigma(rads/sec)$	1.33	1.12	1.13	1.35	1.12	0.77	1.26	1.04	1.11	1.23	1.03	0.75
Angular Acc. $\mu(rads/sec^2)$	0.00	-0.00	0.00	-0.01	0.01	0.00	-0.00	-0.00	0.00	-0.01	-0.00	0.00
Angular Acc. $\sigma(rads/sec^2)$	14.36	14.07	14.77	16.64	17.10	13.16	13.61	13.42	14.01	14.21	16.31	13.07

Joint Angle ID	31	32	33	34	35	36	37	38	39	40	41	42
Joint angle $\mu(rads)$	-0.35	-0.03	0.33	0.67	-0.23	-0.15	-0.35	0.01	-0.30	0.66	-0.22	0.03
Joint angle $\sigma(rads)$	0.37	0.19	0.16	0.42	0.20	0.20	0.37	0.18	0.15	0.41	0.20	0.21
Angular Vel. $\mu(rads/sec)$	-0.00	-0.00	-0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.00
Angular Vel. $\sigma(rads/sec)$	1.16	0.73	0.67	1.79	1.22	0.97	1.14	0.71	0.66	1.77	1.25	1.00
Angular Acc. $\mu(rads/sec^2)$	-0.00	0.00	0.00	-0.00	-0.01	-0.01	-0.00	0.00	0.00	0.01	0.00	-0.00
Angular Acc. $\sigma(rads/sec^2)$	11.47	10.83	8.94	21.90	20.48	16.55	11.29	10.67	8.80	21.60	20.83	16.56

Table 4.3: The parameters (μ and σ) of Gaussian distributions fit to the joint angles, angular velocities and angular accelerations of the 42 joints in the torso, head, arm and leg joint group.

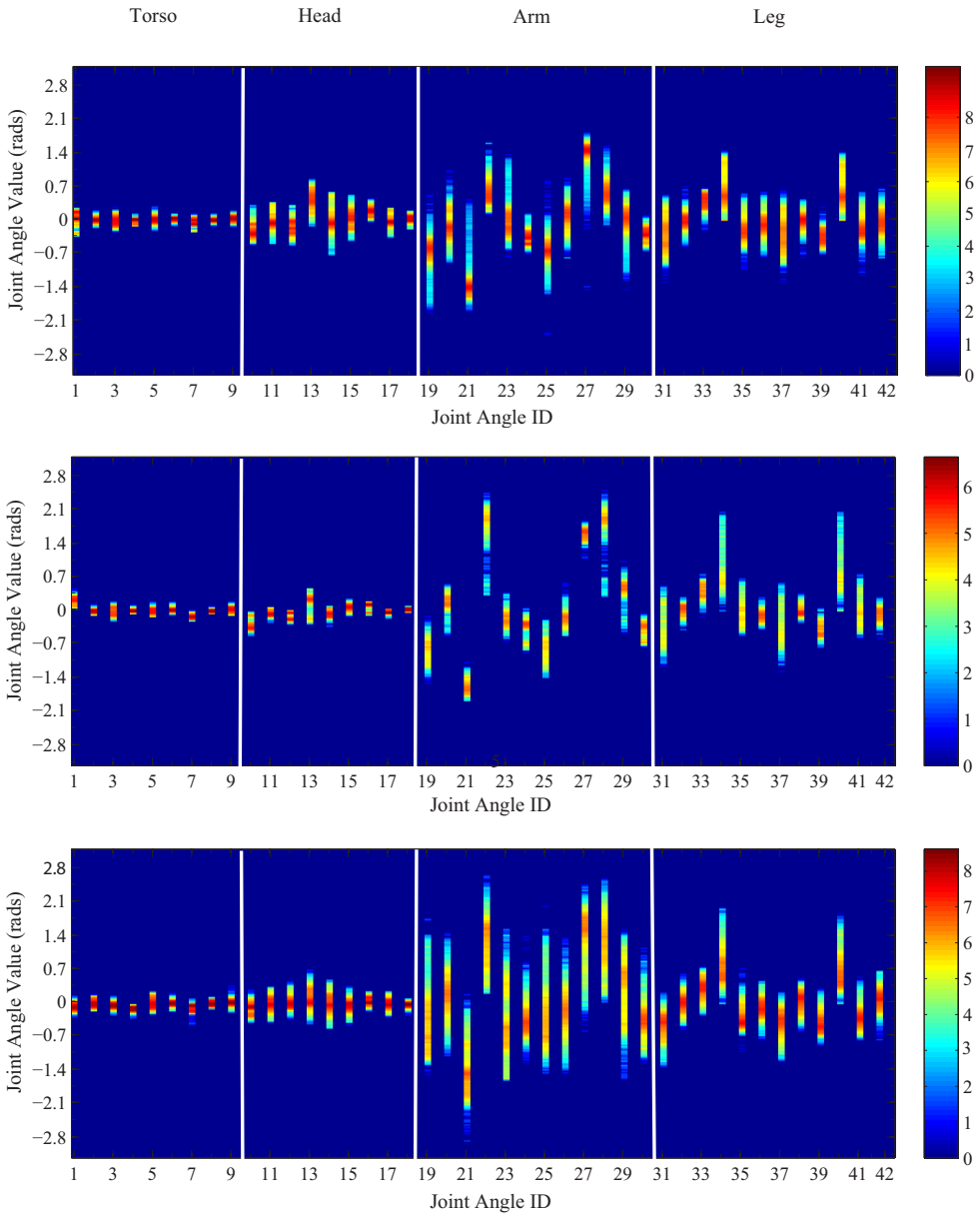


Figure 4.3: The limits and distributions of the joint angles for three behavior-specific data sets: walking, running and swing dancing. The color value of the bar on the right is proportional to the log of the number of entries in each bin. The bin size is 0.03 rads.

show significantly different distributions. For example, swing dancing demonstrates the largest range of arm motions, which illustrates the rich set of arm movements in this behavior. When we compare the joint angle distributions of the arms in running with those in walking, we find most arm joints show a larger range of motion in walking than in running except that the running motion demonstrates a larger range of motion in the elbow joints.

The key insight provided by these data is that the distributions of joint angles, angular velocities and angular acceleration are better modeled by a Gaussian than by the uniform distribution that is usually assumed. The Gaussian density function or its log form can be used to generate a "soft" joint limit and replace the widely used "hard" limit. A "soft" limit provides additional information about how likely or natural a joint configuration is, which would be useful in quantifying natural human motion and in IK-based applications.

We also visualize the range of joint motion to better understand the workspace of each limb. Visualizing the range of joint motion for a 1-DOF joint (hinge joint) or 2-DOF joint (saddle joint) is straightforward but a 3-DOF joint (ball-and-socket joint) is more challenging. We employ a twist-and-swing parametrization [17, 5]. In this parametrization, a 3-DOF joint performs an axial motion (or twist) of the body part (1 DOF) and a spherical motion (or swing) that determines the direction of the bone (2 DOFs). As the Y axis in the local coordinate system of the reference skeleton is aligned along the principal axis of the bone, we can first convert the XYZ fixed angle to a YXZ fixed angle and then take the first component as the twist angle. After we factor out the twist component for 3-DOF joints, we can compute its distribution (1D histogram). We use a 2D distribution (histogram) of X and Z angles to visualize the range of the swing. We apply a similar method to the 1-DOF or 2-DOF joints ignoring the missing DOFs.

We visualize the range of motion for joints in the torso, head, arms, and legs (Figures 4.4–4.7). The range of joint motion is consistent with the common assumptions about joint limits. For example, each joint in the torso or the head group has a small range of motion, whereas the shoulder and hip joints have a much larger range of motion.

The range of motion for the twist components is generally small which is consistent

with previous assumptions [17]: The range of twist for the head and hip joints has a symmetric distribution but the range of twist for the shoulder joints has an asymmetric distribution, implying that people tend to work in front of their body rather than behind.

These results for the range of joint motion may provide useful insight for the development of better algorithms for motion interpolation [53, 31]. The performance of motion interpolation is affected by the distance between the point of interest and the nearby pose examples. Given a fixed number of pose examples, the joints at the hip and shoulder are more likely to have a sparse set of orientation examples than the head or torso joints because the shoulder and hip joints have a larger range of motion. If the orientation examples are dense, we can safely choose any orientation representation for interpolation. On the other hand, if the orientation examples are sparse, the performance will be sensitive to errors introduced by the distance metric [17]. As a result, we should carefully choose the orientation representation for the hip and shoulder joints. Euler angles could be a bad choice but unit quaternions, equipped with the geodesic distance metric, are likely a better choice [17]. Exponential maps are also better than Euler angles if the orientations are not too far away from the reference point.

4.3 Dimensionality Reduction

Dimensionality reduction techniques have been widely used in character animation to reduce a high dimensional motion space to a lower dimensional subspace. For example, this compact subspace can constrain the search space for optimization and potentially lead to a more natural solution [54].

PCA is the most common dimensionality reduction method in the animation literature. However, the mean pose and the eigenposes of natural human motion have not been reported in the literature, perhaps due to the lack of a large and representative motion capture database. We apply PCA analysis to the full database and compute the mean pose and eigenposes based on the XYZ fixed angle representation. Figure 4.8 illustrates the mean

Joint Name	Shoulder	Elbow	Forearm	Wrist	Hip	Knee	Ankle
DOFs	X, Y, Z	X	Y	X	X, Y, Z	X	X, Z
Joint Angle Difference (rads)	0.037, 0.066, 0.101	0.012	0.021	0.037	0.005, 0.037, 0.030	0.012	0.017, 0.113

Table 4.4: The absolute joint angle difference between the left and right joints in the mean pose. Because the ASF skeleton has asymmetric local coordinate systems for the limbs on opposite sides of the sagittal plane, we negate these joint angles before taking the absolute difference.

pose of the entire database from two views. Table 4.4 quantifies the symmetry of joints in the mean pose. The mean pose shows left/right symmetry especially for the legs. Perhaps this is because each leg shares the workload of balancing in most human motions. The two arms show slightly more asymmetry, which may reflect the fact that most people are right-handed and use that hand more.

We explored the effect of joint angle representation on the mean pose and found it was robust to the representation. The mean poses computed based on Euler angles, quaternions and exponential maps are visually similar. We also found that the mean pose was robust to the behavior distribution. For example, the mean pose for our database with locomotion excluded is similar to the one for the entire database.

Figure 4.9 visualizes the magnitudes of the 42 PCA bases for the entire database. These PCA bases are sorted by the amount of variance in the data accounted for by each basis. The first set of bases captures the coordination between the motion of the arm joints and that of the leg joints. The second set of bases focuses mostly on the leg joints. The last set of bases captures primarily the motion in the torso and the head joints. We were surprised by the block layout of the matrix. We speculate that this is because of the strong coordination between the flex/extend joints of the legs and the arms. At a lower magnitude, we also see strong correlations between the back and neck DOF. We further explore these synergies in the next section.

Figure 4.10 illustrates the first six eigenposes, which capture 64% of the variance of the entire motion data set. The pose variations captured by the first three eigenposes

are symmetric whereas the 4th–6th eigenposes represent asymmetric pose variations. To capture 95% of the variance, we need 22 dimensions are required, and 28 dimensions are needed to capture 99% of the variance.

We also studied the effect of joint angle representation on the compression performance of PCA. We first generated a full motion database for each joint angle representation. We then applied PCA to each database. We computed the average pose reconstruction error as an indication of compression performance. When more eigenposes are included in the pose reconstruction, the average pose reconstruction error decreases. For each frame in the database, we converted the reconstructed pose based on a joint angle representation to a virtual marker representation [2] so that the pose reconstruction error could be computed as a distance between the reconstructed and true maker positions and was therefore independent of the joint angle representation. In the virtual marker representation, the local coordinate system of each body part is converted into a set of three virtual markers placed at fixed positions [2]. Transforming from other representations into this virtual marker representation is a lossless procedure. For the quaternion representation, we had to normalize the reconstructed quaternions because the linear reconstruction does not guarantee that the reconstructed pose vector consists of unit quaternions.

One standard preprocessing step for PCA analysis is to normalize the data by mean and standard deviation so that each component of the data vector has the same range. We generated a full motion database based on normalized Euler angles to evaluate the effect of normalization on the compression performance of PCA.

We compare virtual markers, Euler angles, normalized Euler angles, exponential map, and unit quaternions on three different data sets: the full database, one particular behavior (swing dancing), and one motion sequence (swing dancing). We can draw several conclusions based on these graphs (Figure 4.11). Virtual markers achieve the best compression performance when fewer than 35 eigenposes (99.9% data variance accounted for) are used for pose reconstruction. The performance improvement is more significant for the single behavior dataset of swing dancing. The reason why the virtual markers achieve better performance might be because other representations demonstrate more nonlinearity

in representing human motion. Normalized euler angles have the worst performance in most cases. The curves in Figure 4.11 demonstrate that increasing the number of eigenposes does not always improve the performance. This unintuitive behavior is because we compress in the particular representation being tested but measure the error in the virtual marker space for consistency across all representations. Finally, Euler angles, exponential maps and quaternions have no significant difference in performance. When more than 35 eigenposes are used for pose reconstruction, both Euler angles and exponential maps achieve better performance than virtual markers. Quaternions have a slightly worse performance than Euler angles and exponential maps and the difference is greater at higher dimensions (> 29).

Compression is only one possible metric and might not be the most important one for computer animation. For example, the smoothness of the space is also important for optimization algorithms and that also likely changes based on the representation.

4.4 Human Motion Synergies

Synergies represent coordination between the movements of a group of a few joints. In this section, we present a new statistical approach to identify groups of joints that exhibit synergies. We compute a 42x42 similarity matrix $m(i, j)$ that encodes the synergies between any two joint angles by using mutual information to measure the statistical dependency between each pair of joint angles i and j. We then apply a segmentation algorithm to divide the joints into groups that demonstrate synergies. We apply our algorithm to the entire database and to behavior-specific data sets of walking and forward jumping in which synergies can be easily observed.

The similarity matrix requires the computation of the mutual information between each pair of joint angles. The mutual information of two random variables, X and Y, is a quantity that measures the mutual dependence of the two variables [44]. Intuitively, mutual information measures the information about X that is shared by Y. If X and Y are indepen-

dent, then X contains no information about Y and vice versa, so their mutual information is zero. If X and Y are identical then all information conveyed by X is shared with Y and the mutual information is the same as the information conveyed by X (or Y) alone, namely the entropy of X.

The mutual information of two random variables X and Y is defined as

$$I(X;Y) = \int_Y \int_X p(x,y) \log \frac{p(x,y)}{f(x)\,g(y)}\, dx\, dy, \tag{4.1}$$

where p is the joint probability density function (PDF) of X and Y, and f and g are the marginal PDFs of X and Y respectively. $I(X;Y) = 0$ if and only if X and Y are independent random variables. Mutual information is also nonnegative ($I(X;Y) > 0$) and symmetric ($I(X;Y) = I(Y;X)$).

An alternative method for computing the similarity matrix would be to compute the product-moment correlation coefficient between the joint angles. The product-moment correlation coefficient measures how well a linear equation describes the relation between two variables X and Y and can be obtained by dividing the covariance of these two variables by the product of their standard deviations [10]. However, the product-moment correlation coefficient only handles linear dependency whereas mutual information handles nonlinear dependency and other more general relationships.

We apply a segmentation algorithm to the similarity matrix and use normalized cut to divide the joint angles into groups that demonstrate synergies [56]. Normalized cut uses a criterion that measures both the total dissimilarity between the different groups as well as the total similarity within the groups. It formulates the segmentation problem as a generalized eigenvalue problem that allows an efficient optimization of this criterion.

Figure 4.12 illustrates the similarity matrix computed from the entire data set and the grouping result (six groups) indicated by color in the color bar at bottom. As there is no automatic rule for determining how many groups to use, we choose this number manually. We find there are synergies between the left and right arm for the full database. Synergies also exist between the left leg and right leg. The grouping result also shows a strong dependance between the X angles in the head and torso joints and all the joint angles in

the arm but not other components for the head and torso joints. This synergy might occur because the torso is used for balance as the arms are moved fore and aft. The X angles of the ankles are also grouped with the legs. Figure 4.12 also shows that joints that are directly linked are likely to be grouped together. For example, the Y and Z angles of the head joints (head, upper neck, lower neck) are grouped together.

Figure 4.13 illustrates the similarity matrix and four groups for a data set of forward walking. The X angles of hip joints are grouped together with all the joint angles in the two arms, which characterizes the typical fore/cft oscillation we can see in forward walking. Only X angles from the hip joints are included in the group because in the walking rotation around the X axis dominates the motion in these joints and the rotations around the other two axes are very small and less cyclical.

Figure 4.14 illustrates the similarity matrix and four groups for a data set of forward jumping. The similarity matrix shows a strong dependency between most joint angles because the motion shows a strong coordination. The motions of the two arms are grouped together as they show a strong interdependency in forward jumping.

This method can only identify synergies at the level of joint components because it relies on mutual information computed between two joint angles (two DOFs) rather than treating all the DOFs of a joint as a single entity. Although generalizations of mutual information to more than two random variables have been proposed, a widely agreed on definition has not yet emerged [62].

The proposed method can also be used to identify synergies in the motion of markers. Given a group of markers that move in a coordinated fashion, we can pick one and use it to predict the movements of the others. Motion estimation from a reduced marker set is required for performance animation synergies and other motion tracking applications [37]. It is also useful in cleaning motion capture data, where the missing trajectory of one marker can be inferred from the trajectory of the correlated markers.

4.5 Motion Summarization

We would like to be able to provide intuition about the motion contained in a particular sequence or database. We do this via summary images, which average a set of images representing the motion. To generate the summary image, we render each pose excluding global translation and yaw orientation in the sequence or database using a human-like figure from a fixed viewpoint. We then compute the average of the resulting images and take the logarithm of each pixel intensity. Finally we generate a color image by applying a linear mapping from the logarithmic intensity of each pixel to a value in a color map. The resulting image is expressive because it portrays the distribution of human poses in that data set. The summary image is view dependent, so we show it from a front and a side view.

We use summary images to illustrate the full motion database and several behavior-specific data sets (Figure 4.15-4.17). The summary images demonstrate the typical characteristics of these behaviors. For example, the pitching in forward jumping and the arm motion for swing dancing can be clearly observed in their summary images in Figure 4.16. We can also capture a sense of various styles of human walking (Figure 4.18).

As the size of available libraries of motion data increase, it will become harder to quickly explore a database. Navigation could be made easier by creating a hierarchy of summary images allowing the user to focus on motion sequences which contain a lot of upper body motion, for example.

4.6 Discussion

In this chapter, we explored the statistics of natural human motion and their applications in algorithms for generating human motion. We reported a set of statistics about basic joint motion, dimensionality reduction, human motion synergies and motion summarization computed on a large and representative motion capture database. We expect these statistics

will benefit such applications in character animation as motion editing, motion retargeting and motion capture. Because our goal is to provide statistics that other researchers can use, we have placed the raw numbers needed to create the images and graphs in this paper online at http://graphics.cs.cmu.edu/projects/statistics/. The motion capture data is online at http://mocap.cs.cmu.edu.

Although the human skeleton used in our motion data is widely used in character animation, it contains fewer degrees of freedom than would be ideal. For example, the skeleton only assigns one DOF to the joints at the knee and three DOFs to the joints at the shoulder. Some natural human motions such as a shoulder shrug cannot be well modeled by this skeleton. The number of joints might not be enough to accurately represent the subtleties of human motion. For example, the anatomical structure of human back has 33 vertebrae while our model has only 5. These approximations had an effect on the computed statistics. For example, the back DOF likely had a wider range of motion than they would have had if the motion had been apportioned across all of the vertebrae in a complete model of the human back.

The motion capture data used in our statistical analysis are representative as they contain a variety of behaviors from 34 different subjects. However, most of these subjects are young and of a healthy weight and therefore are not representative of the population overall. This limitation prevents us from comparing the statistics between different body types or age groups.

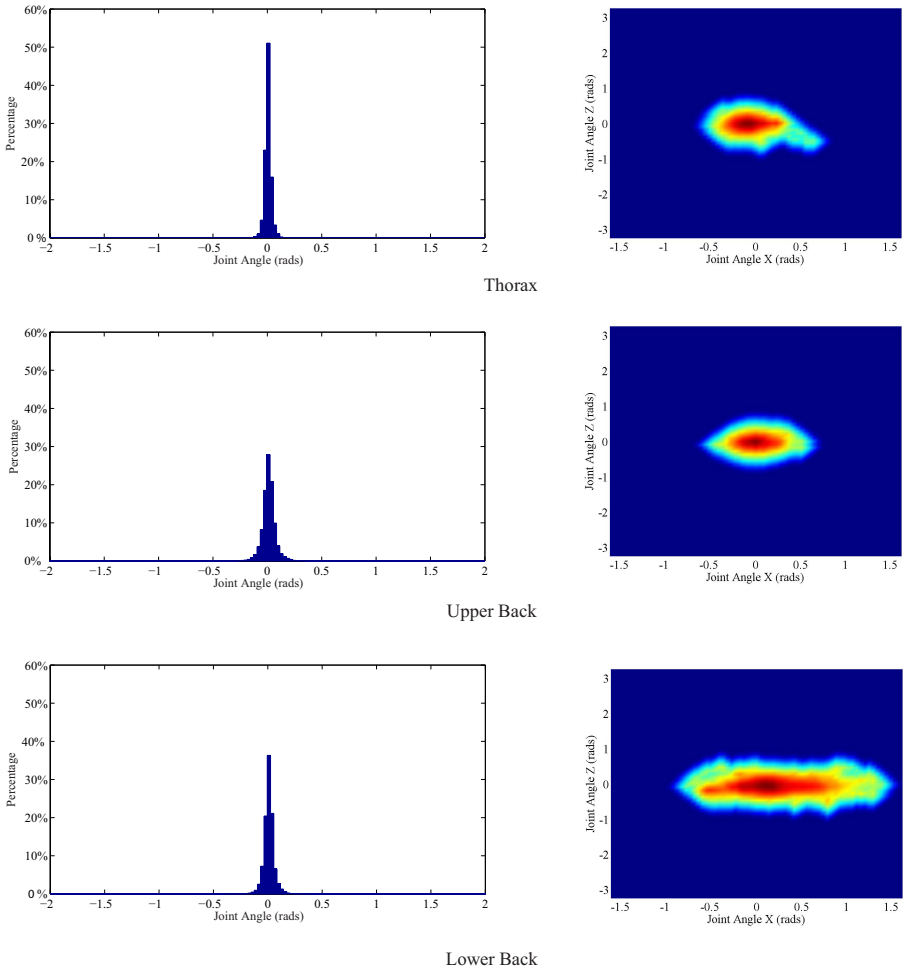


Figure 4.4: The range of motion for each joint in the torso.

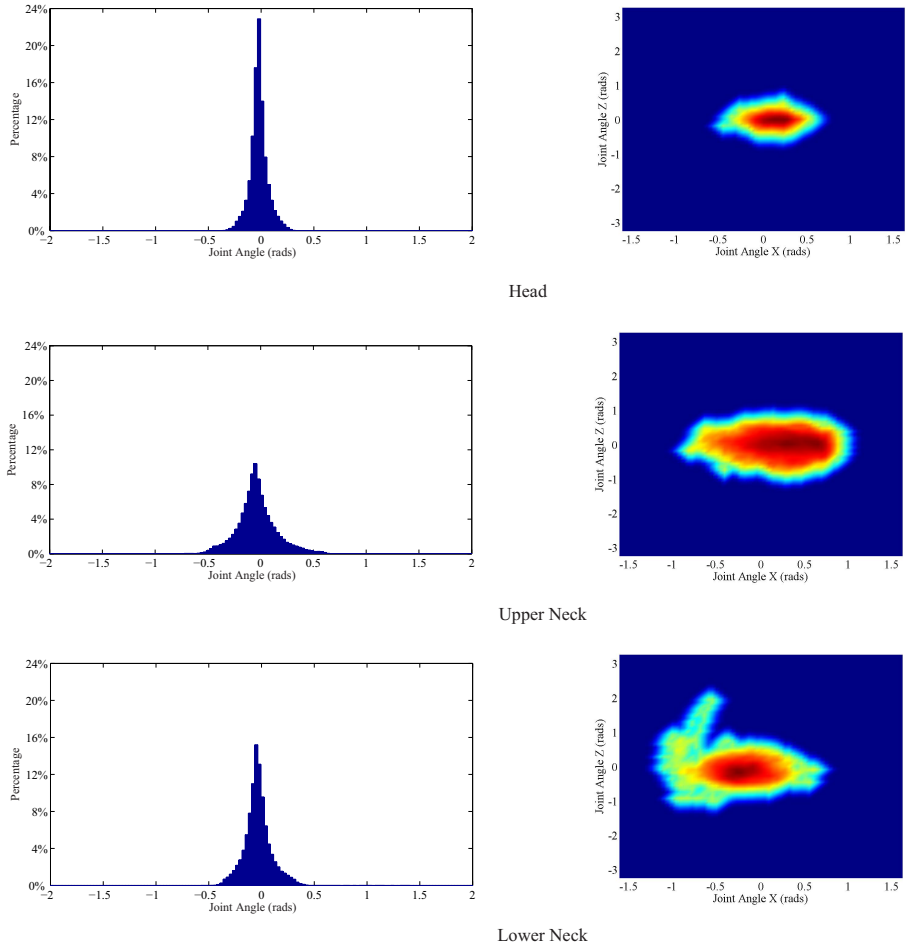


Figure 4.5: The range of motion for each joint in the head joint group.

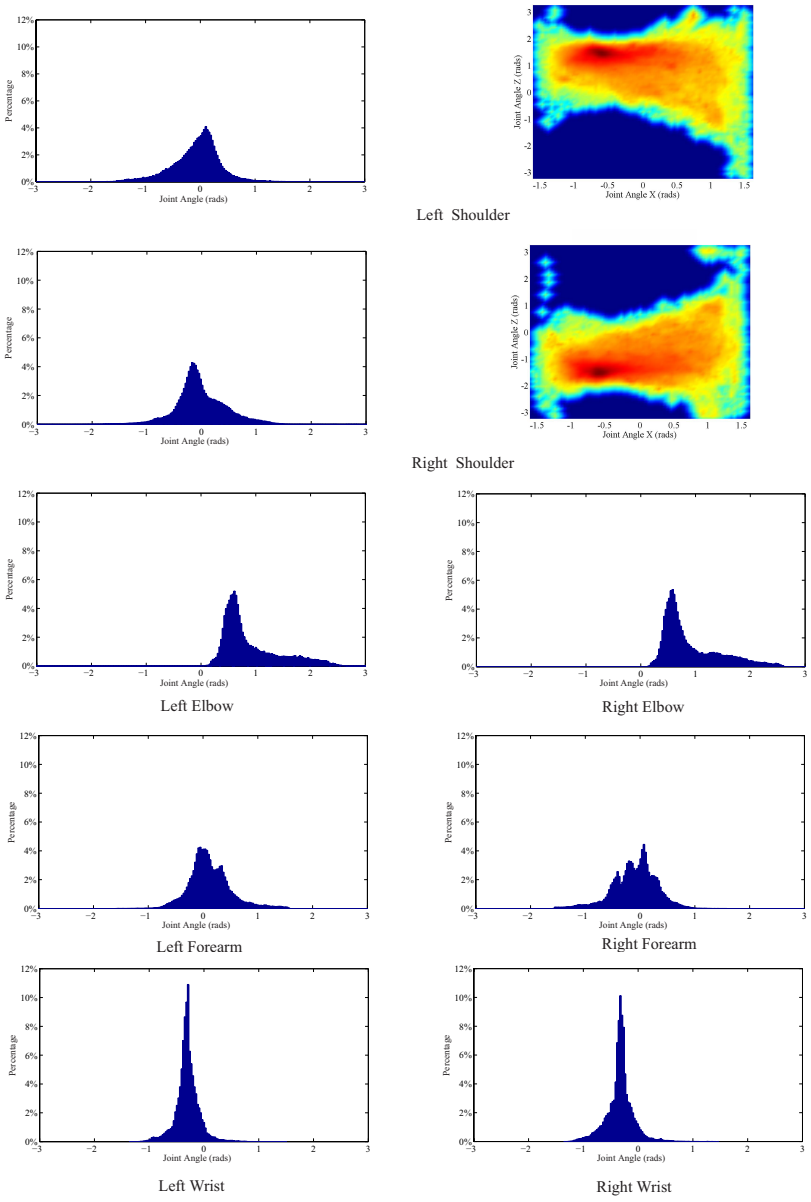


Figure 4.6: The range of motion for each joint in the two arms. We only illustrate the range of swing motion for elbows and wrists because twist (Y) is not defined for these joints in our motion database. We only show the twist distribution for the two forearms joints because only twist (Y) is defined for these 1-DOF joints.

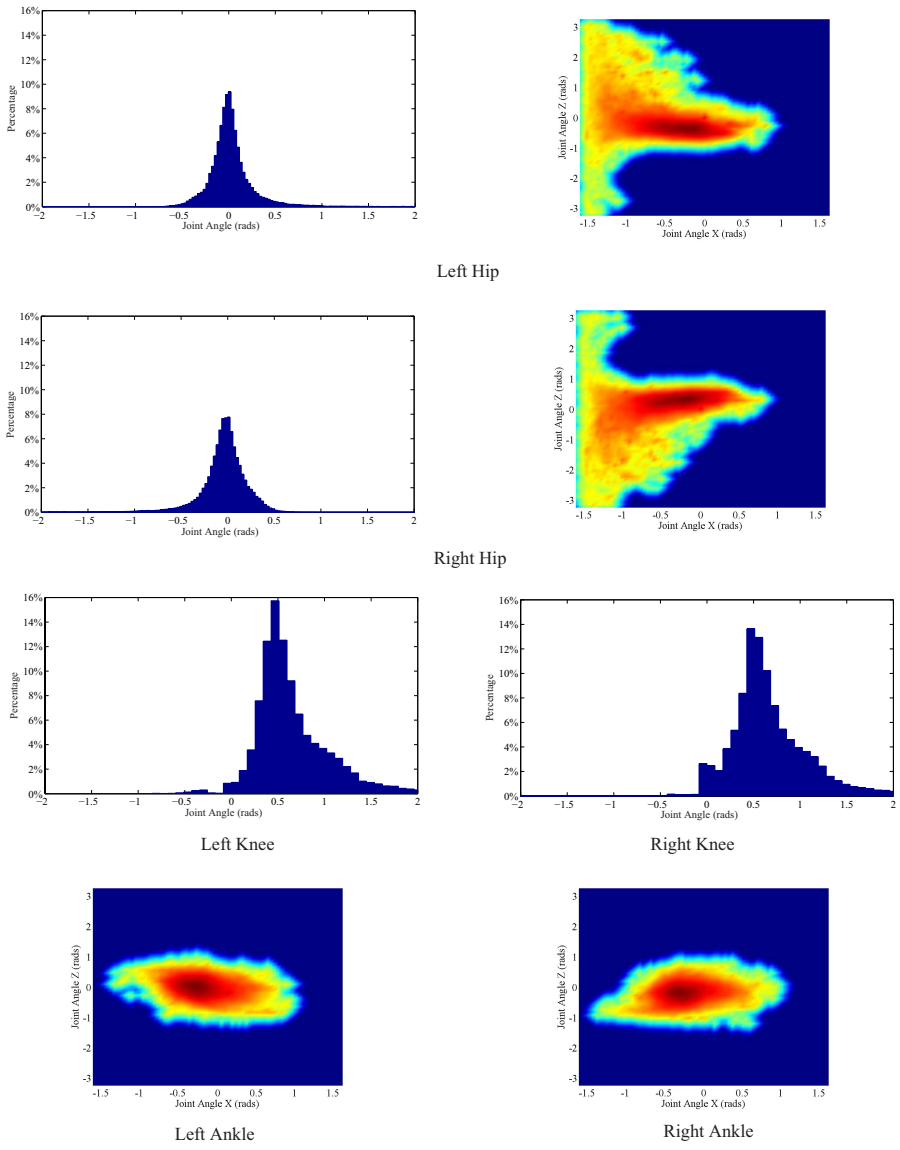


Figure 4.7: The range of motion for each joint in the legs. We only illustrate the range of swing motion for the knees (1-DOF joints) and the ankles (2-DOF joints) because the twist (Y) is not defined for these joints in our dataset.

Figure 4.8: The mean pose of the entire data set. The computation is based on the XYZ fixed angle representation.

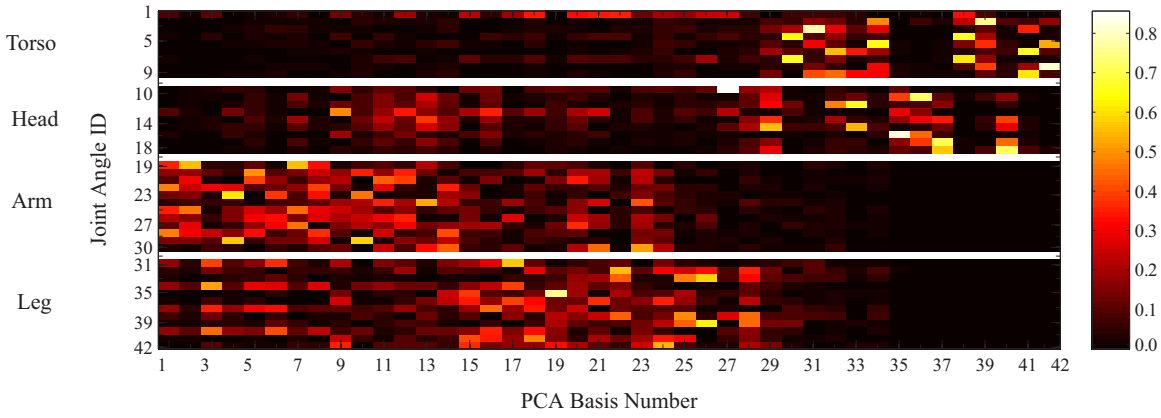
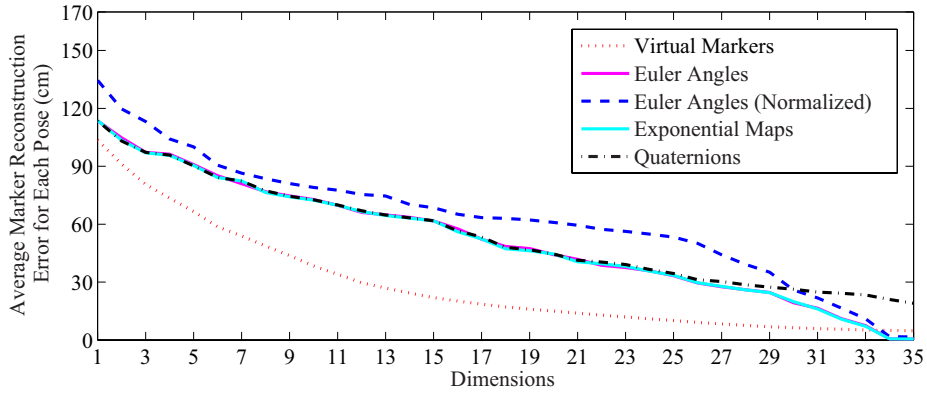
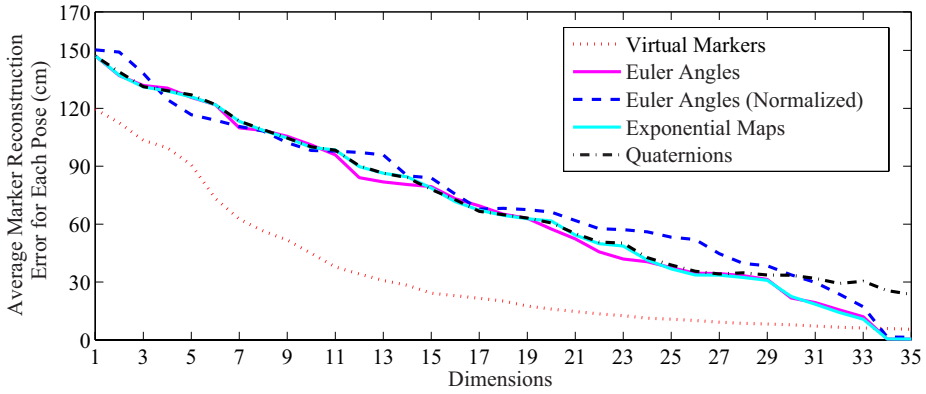


Figure 4.9: The 42 PCA bases for the entire motion data set. Each basis consists of 42 coefficients that correspond to the 42 joint angles in a pose. We take the absolute value of each coefficient in each basis for this visualization so that the intensity is proportional to the magnitude.

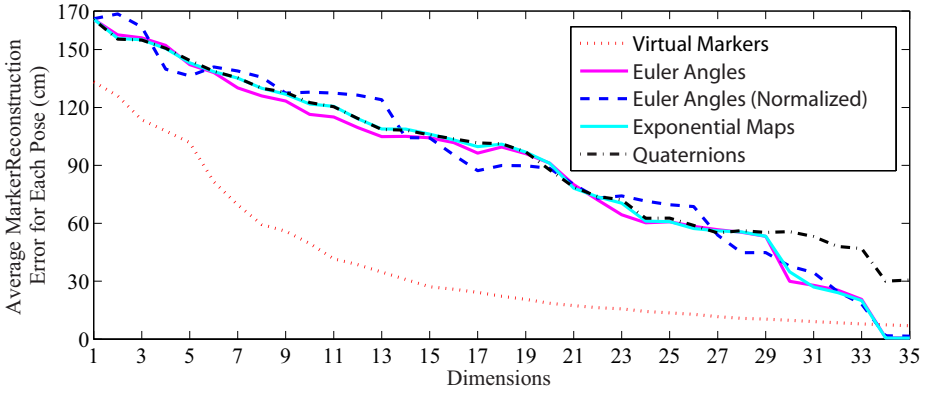
Figure 4.10: The first six eigenposes of the entire motion database using the XYZ fixed angles. In each row, we show the mean pose (middle column) and the pose as the values of the first six bases are increased.



Full motion capture database



Behavior-specific data set (swing dancing)



A single motion sequence (swing dancing)

Figure 4.11: PCA compression comparison based for three data sets.

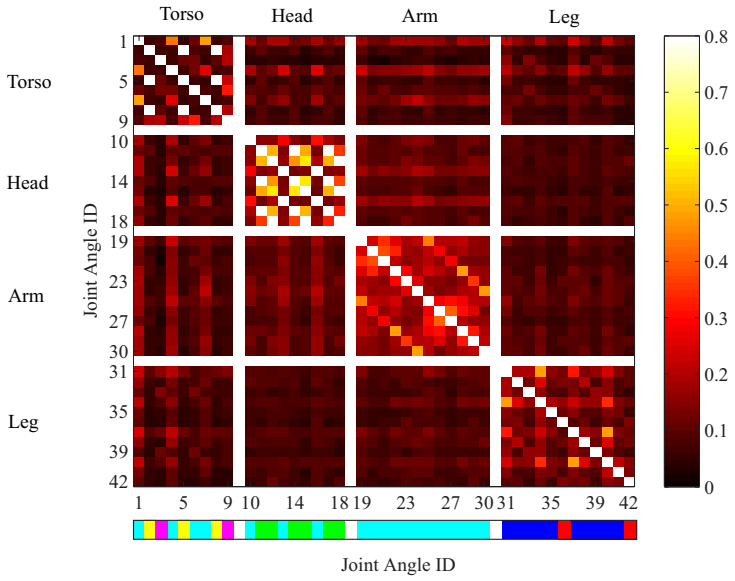


Figure 4.12: The similarity matrix for the entire database. Each element of the matrix is the pairwise mutual information between two joint angles. The joint angles in each group share the same color in the bar underneath the matrix.

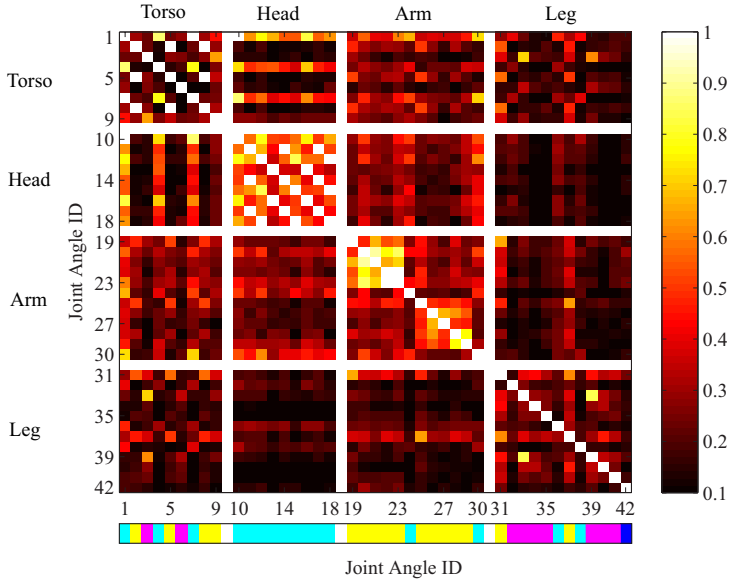


Figure 4.13: The similarity matrix for a walking data set.

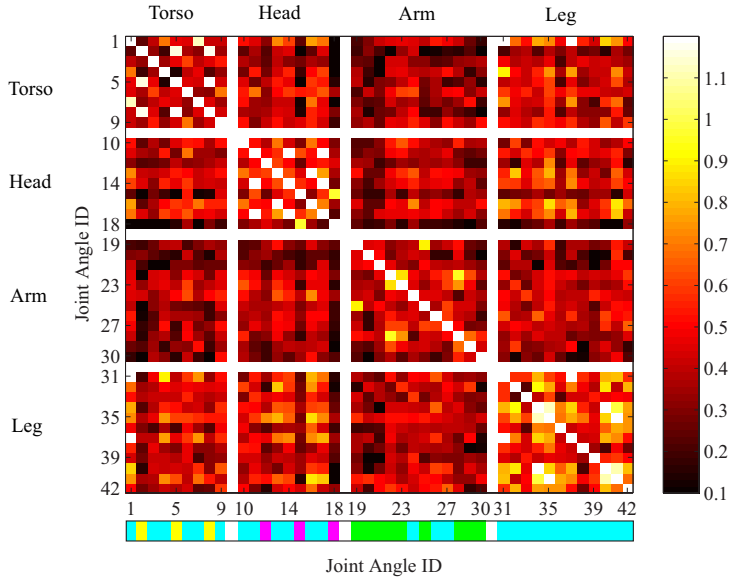
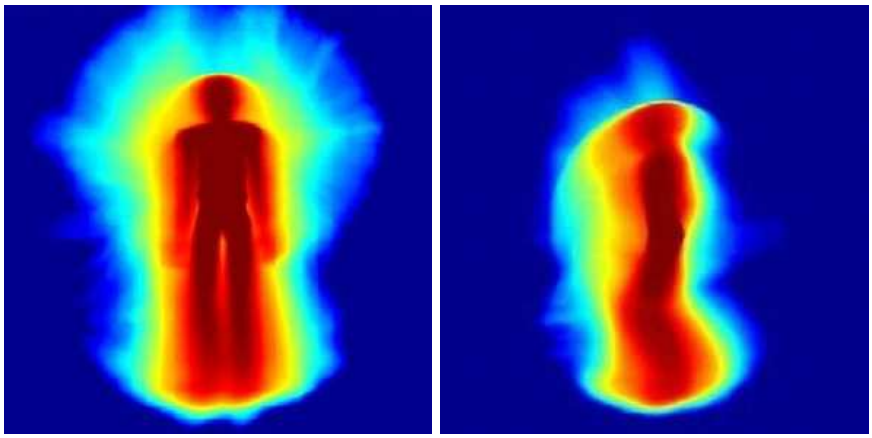


Figure 4.14: The similarity matrix for a forward jumping data set.



Front view Side view

Figure 4.15: Summary images for the full database.

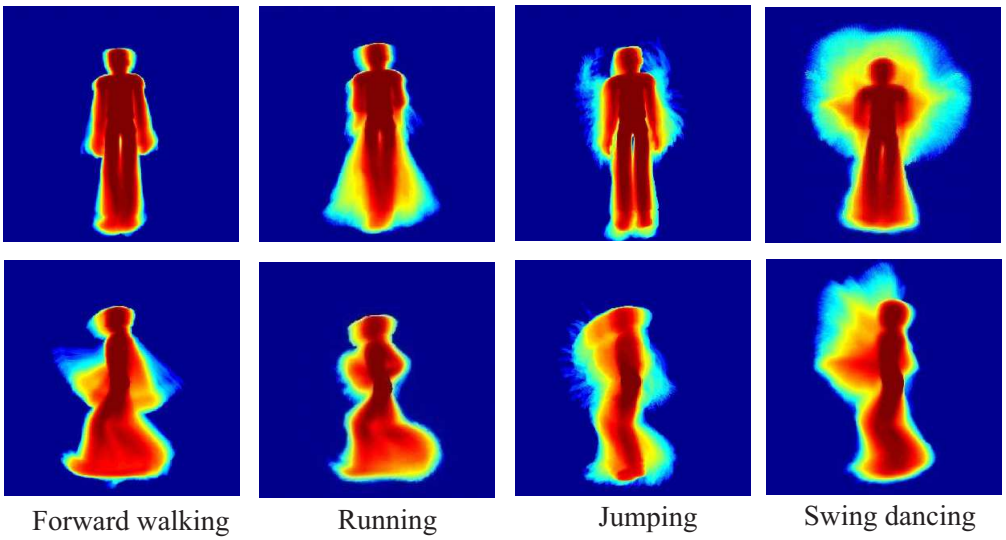


Figure 4.16: Summary images for different behaviors. The images in the first row are rendered from the front and those in the second row are generated from the side.

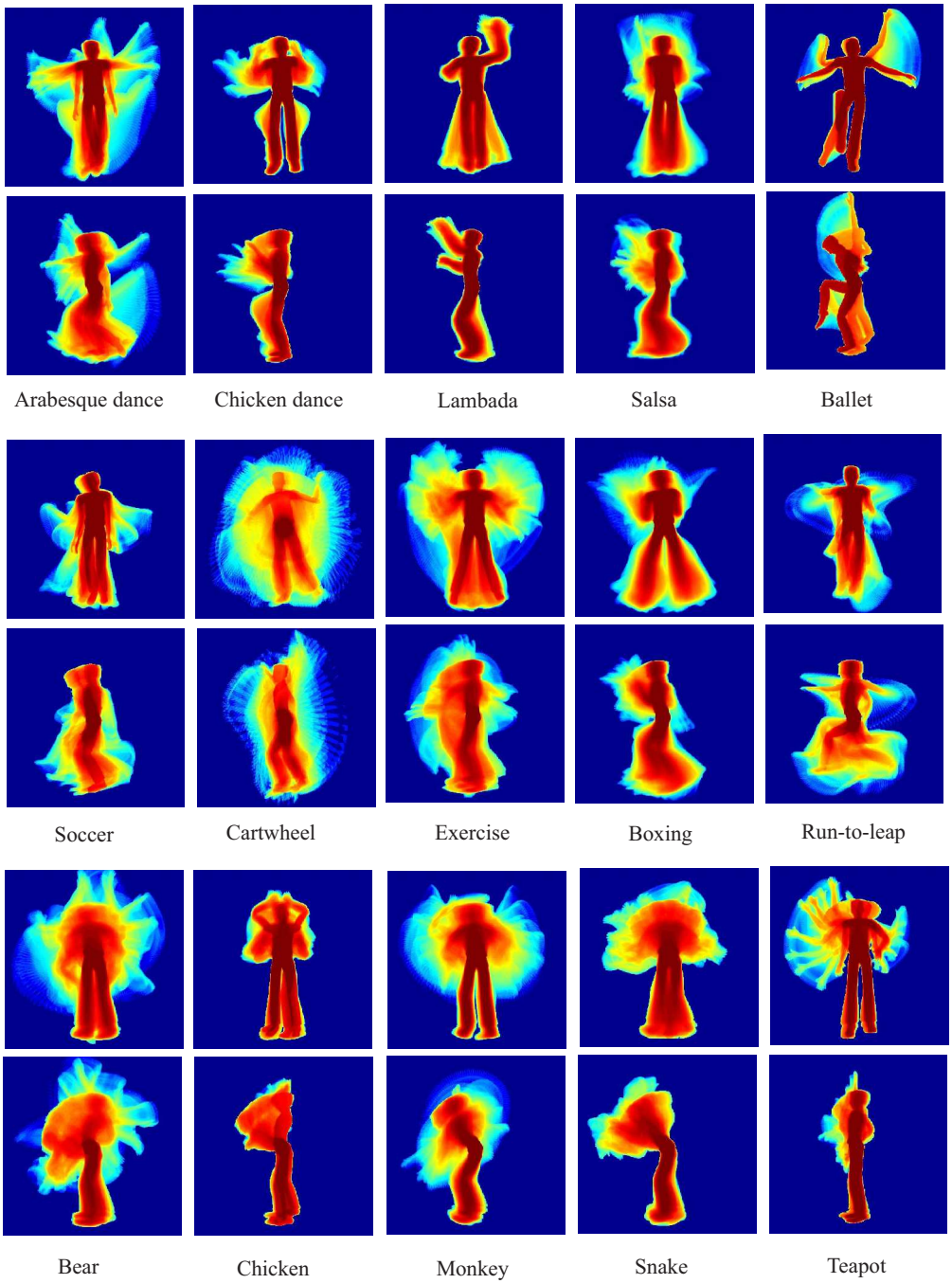


Figure 4.17: The summary images for individual sequences of motion: dancing, physical activities, and pantomime. We show a front and a side view for each sequence.

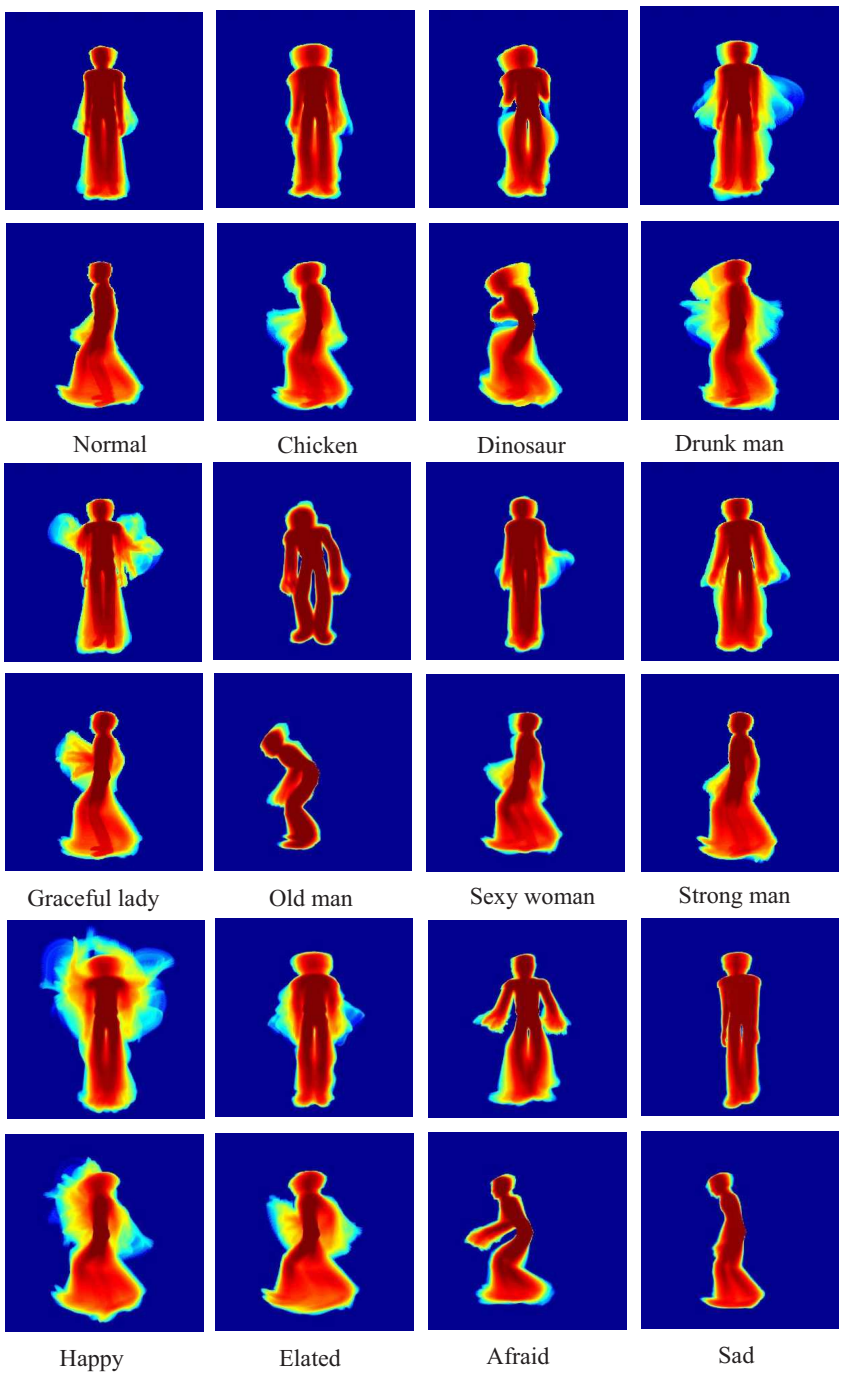


Figure 4.18: Summary images for stylized walking motions.

Chapter 5

Conclusion

The thesis includes three major contributions. First, we demonstrated that an implementable definition for the naturalness of human motion. Naturalness could be defined based on the statistical properties of a large and representative motion capture database. We implemented several techniques to quantify natural human motion independent of any specific motion source. We presented ROC curves to demonstrate the performance of these techniques on a broad set of test sequences and compared the results to human performance in a user study. Our ensemble method achieved the best performance among these automatic methods. It first hierarchically decomposed human motion into its constituent parts (individual joints, limbs, and full body), then built a statistical model of each one using existing machine learning techniques, and finally combined these models into an ensemble model for classification of the motion as natural or unnatural. Compared with approaches that only employed a single statistical model, the ensemble method improved the classification performance. It also located the bad section of the motion automatically, thus facilitating the process of interactive motion editing.

Second, we explored the statistics of natural human motion via a comprehensive statistical analysis. Even though the aggregate statistics about properties of human motion are needed to guide animation algorithms for human figures toward natural looking so-

lutions, these statistics had not been comprehensively studied or reported. In this thesis, we computed these statistics using a large and representative motion capture database and provided insights into their use in character animation. We evaluated the effects of different joint angle representations on commonly used statistical algorithms such as linear dimensionality reduction of human motion. We also presented a statistical definition for synergies in human motion, which had not been rigorously defined in character animation or biomechanics. Based on this definition, we presented a new algorithm for synergy identification in natural human motion. Finally, we proposed a method for summarizing and visualizing a motion capture dataset, which should be useful for providing thumbnail images to aid in browsing and navigating a large motion capture database. All the resulting statistics and insights should be useful in designing and implementing a large variety of algorithms for character animation.

Finally, we contributed a substantial database of human motion and a testing set that would enable others to apply their algorithms to the problem of quantifying natural human motion: http://graphics.cs.cmu.edu/projects/natural/.

The thesis takes one step toward quantifying natural human motion in character animation by performing a series of statistical analyses on a large and representative motion capture database. In this thesis, basic joint angle statistics (distributions of joint angles and velocities) were regarded as the most important factor in characterizing the naturalness of human motion. However, many other factors also play a role in defining natural human motion. For example, statistical models of dynamical quantities (joint forces and joint torques) could be employed for natural vs. unnatural motion classification or for guiding the search of an optimizer. They can also be computed and reported for the synthesis of natural human motion. Other factors may include local features related to a specific artifact in unnatural human motions. For example, foot sliding is a visually perceptible artifact. Such a specific artifact is not likely to be easily detected by a general statistical classifier but would require the detection of intended foot contacts [27]. These local features could be included in a naturalness measure and used to improve algorithms for animation generation and evaluation.

The long term goal of quantifying natural human motion is to build an automatic naturalness measure that will achieve a level of performance that is comparable to that of humans. One approach to this problem is to better model what is perceptually important about human motion. The animation community has begun to assemble data about the perception of human motion. For example, the rendering style or appearance of the character likely affects the perception of the naturalness of a human motion sequence [24]. People are sensitive to physical errors in the character motion such as gravity [50] and changes in limb length [22]. We would like to combine the results of these perceptual studies with the statistical analysis of motion capture data to construct better classifiers and collect additional statistics that are relevant for the generation of natural human motion.

Our motion capture database plays an important role in our statistical analysis of natural human motion. We choose the largest motion capture database (publicly available) to perform our statistical analysis, which should make our statistical analysis more representative and convincing. However, there are still limitations. For example, though the database includes a variety of behaviors, any natural behavior would still be classified as unnatural if the behavior is significantly different from what we have in our database. Because most of our motion capture subjects are young adults, the statistics we computed from all these different behaviors mostly represent the motion characteristics of young adults and would differ from statistics computed for heavier or old populations.

The results of our statistical analysis for quantifying natural human motion cannot be used in all applications in character animation because “naturalness” may not always have a single definition. For example, a “natural” motion for a cartoon character could violate the laws of physics yet remain compelling and appealing. However, as long as we can build a motion database containing the motion of a particular character, we should be able to employ the same set of statistical analysis methods to quantify the “naturalness” of the character’s motion and provide statistics that aid in the generation of motion that is natural or appropriate for that character.

Bibliography

[1] Karim Abdel-Malek, Jingzhou Yang, Richard Brand, and Emad Tanbour. Towards understanding the workspace of human limbs. *Ergonomics*, 47(13):1386–1405, 2004. 2.2

[2] Okan Arikan. Compression of motion capture databases. *ACM Transactions on Graphics*, 25(3):890–897, 2006. 4.3

[3] Okan Arikan and David A. Forsyth. Interactive motion generation from examples. *ACM Transactions on Graphics*, 21(3):483–490, 2002. 1.1, 2.1

[4] Jackie Assa, Yaron Caspi, and Daniel Cohen-Or. Action synopsis: Pose selection and illustration. *ACM Transactions on Graphics*, 24(3):667–676, 2005. (document), 1.1, 1.2, 2.2

[5] Paolo Baerlocher and Ronan Boulic. Parametrization and range of motion of the ball-and-socket joint. In *DEFORM'00/AVATARS'00: Proceedings of the IFIP TC5/WG5.10 DEFORM'2000 Workshop and AVATARS'2000 Workshop on Deformable Avatars*, pages 180–190, 2001. 4.2

[6] Jernej Barbič, Alla Safonova, Jia-Yu Pan, Christos Faloutsos, Jessica K. Hodgins, and Nancy S. Pollard. Segmenting motion capture data into distinct behaviors. In *Proceedings of the 2004 Conference on Graphics Interface*, pages 185–194, 2004. 1.2, 2.2

[7] Aaron F. Bobick and James W. Davis. The recognition of human movement using temporal templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(3):257–267, 2001. 2.2

[8] Oren Boiman and Michal Irani. Detecting irregularities in images and in video. In *ICCV '05: Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05)*, volume 1, pages 462–469, 2005. 2.1

[9] Matthew Brand and Aaron Hertzmann. Style machines. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, pages 183–192, 2000. 1.1, 1.2, 2.2

[10] Jacob Cohen. *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates, 1988. 4.4

[11] Ronald A. Cole. Survey of the state of the art in human language technology. http://cslu.cse.ogi.edu/HLTsurvey, 1996. 2.1

[12] Hany Farid and Siwei Lyu. Higher-order wavelet statistics and their application to digital forensics. In *IEEE Workshop on Statistical Analysis in Computer Vision (in conjunction with CVPR 2003)*, 2003. 2.1

[13] Kevin Forbes and Eugene Fiume. An efficient search algorithm for motion data using weighted PCA. In *Proceedings of the 2005 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 67–76, 2005. 1.2, 2.2

[14] Michael Gleicher. Motion editing with spacetime constraints. In *I3D '97: Proceedings of the 1997 Symposium on Interactive 3D Graphics*, pages 139–147, 1997. 1.2, 2.2

[15] Michael Gleicher. Retargetting motion to new characters. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*, pages 33–42, 1998. 1.2, 2.2

[16] Michael Gleicher. Comparing constraint-based motion editing methods. *Graphical Models*, 63(2):107–134, 2001. 1.1, 2.1

[17] F. Sebastin Grassia. Practical parameterization of rotations using the exponential map. *Joural of Graphics Tools*, 3(3):29–48, 1998. 4.2, 4.2

[18] Keith Grochow, Steven L. Martin, Aaron Hertzmann, and Zoran Popović. Style-based inverse kinematics. *ACM Transactions on Graphics*, 23(3):522–531, 2004. 1.2, 2.2

[19] Vijaykumar Gullapalli, Jack J. Gelfand, Stephen H. Lane, and Wade W. Wilson. Synergy-based learning of hybrid position/force control for redundant manipulators. In *IEEE Robotics and Automation Conference*, pages 3526–3531, 1996. 2.2

[20] R. Hamid, A. Johnson, S. Batta, A. Bobick, C. Isbell, and G. Coleman. Detection and explanation of anomalous activities. In *IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, CA, 2005. 2.1

[21] K. Hara, T. Omori, and R. Ueno. Detection of unusual human behavior in intelligent house. In *Neural Networks for Signal Processing XII-Proceedings of the 2002 IEEE Signal Processing Society Workshop*, pages 697–706, 2002. 2.1

[22] Jason Harrison, Ronald A. Rensink, and Michiel van de Panne. Obscuring length changes during animated motion. *ACM Transactions on Graphics*, 23(3):569–573, 2004. 1.1, 5

[23] Lorna Herda, Raquel Urtasun, and Pascal Fua. Hierarchical implicit surface joint limits for human body tracking. *Computer Vision and Image Understanding*, 99(2):189–209, 2005. 2.2

[24] Jessica K. Hodgins, James F. O'Brien, and Jack Tumblin. Perception of human motion with different geometric models. *IEEE Transactions on Visualization and Computer Graphics*, 4(4):307–316, 1998. 5

[25] Jessica K. Hodgins and Nancy S. Pollard. Adapting simulated behaviors for new characters. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, pages 153–162, 1997. 1.2

[26] Jessica K. Hodgins, Wayne L. Wooten, David C. Brogan, and James F. O'Brien. Animating human athletics. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*, pages 71–78, 1995. 1.2, 2.2

[27] Leslie Ikemoto, Okan Arikan, and David Forsyth. Knowing when to put your foot down. In *SI3D '06: Proceedings of the 2006 symposium on Interactive 3D graphics and games*, pages 49–53, 2006. 5

[28] Leslie Ikemoto, Okan Arikan, and David Forsyth. Quick motion transitions with cached multi-way blends. Technical Report EECS-2006-14, Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, February 2006. 2.1

[29] Leslie Ikemoto and David A. Forsyth. Enriching a motion collection by transplanting limbs. In *Proceedings of the 2004 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 99–108, 2004. 1.1, 1.1, 2.1

[30] Jason Salavon. 100 special moments. http://salavon.com/SpecialMoments/SpecialMoments.shtml, 2004. 1.2, 2.2

[31] Lucas Kovar and Michael Gleicher. Automated extraction and parameterization of motions in large data sets. *ACM Transactions on Graphics*, 23(3):559–568, 2004. 1.1, 4.2

[32] Lucas Kovar, Michael Gleicher, and Frederic Pighin. Motion graphs. *ACM Transactions on Graphics*, 21(3):473–482, 2002. 1.1, 2.1

[33] Jehee Lee, Jinxiang Chai, Paul Reitsma, Jessica Hodgins, and Nancy Pollard. Interactive control of avatars animated with human motion data. *ACM Transactions on Graphics*, 21(3):491–500, 2002. 1.1, 2.1

[34] Philip Lee, Susanna Wei, Jianmin Zhao, and Norman I. Badler. Strength guided motion. In *Proceedings of the 17th Annual Conference on Computer Graphics and Interactive Techniques*, pages 253–262, 1990. 1.2, 2.2

[35] Uri Lerner. *Hybrid Bayesian Networks for Reasoning about Complex Systems*. PhD thesis, Stanford University, 2002. 3.2.3

[36] Yan Li, Tianshu Wang, and Heung-Yeung Shum. Motion texture: a two-level statistical model for character motion synthesis. *ACM Transactions on Graphics*, 21(3):465–472, 2002. 1.1, 1.2, 2.2, 3.2.3

[37] Guodong Liu, Jingdan Zhang, Wei Wang, and Leonard McMillan. Human motion estimation from a reduced marker set. In *SI3D '06: Proceedings of the 2006 Symposium on Interactive 3D Graphics and Games*, pages 35–42, 2006. 4.4

[38] Karen Liu, Aaron Hertzmann, and Zoran Popović. Learning physics-based motion style with nonlinear inverse optimization. *ACM Transactions on Graphics*, 24(3):1071–1081, 2005. 1.2

[39] Anderson Maciel, Luciana Nedel, and Carla Freitas. Anatomy-based joint models for virtual humans skeletons. In *Proceedings of Computer Animation 2002*, 2002. 2.2, 4.1

[40] Walter Maurel and Daniel Thalmann. Human shoulder modeling including scapulothoracic constraint and joint sinus cones. *Computers & Graphics*, 24(2):203–218, 2000. 2.2

[41] Victor Ng-Thow-Hing and Wei Shao. Modular components for detailed kinematic modelling of joints. In *Proceedings of XIX Congress of the International Society of Biomechanics*, 2003. 2.2

[42] Cynthia Norkin and Joyce White. *Measurement of joint motion: a guide to goniometry*. F.A. Davis Company, 1995. 1.2, 2.2

[43] C. O'Sullivan, J. Dingliana, T. Giang, and M. K. Kaiser. Evaluating the visual fidelity of physically based animations. *ACM Transactions on Graphics*, 22(3):527–536, 2003. 1.1

[44] Athanasios Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 1984. 1.2, 4.4

[45] Sang Il Park, Hyun Joon Shin, and Sung Yong Shin. On-line locomotion generation based on motion blending. In *Proceedings of the 2002 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 105–111, 2002. 4.1.1

[46] V. Pavlović, James M. Rehg, and J. MacCormick. Learning switching linear models of human motion. In *Proceedings of Advances in Neural Information Processing Systems (NIPS 2000)*, pages 981–987, 2000. 3.2.3

[47] Ken Perlin. Real time responsive animation with personality. *IEEE Transactions on Visualization and Computer Graphics*, 1(1):5–15, 1995. 2.1, 3.1

[48] F. Pollick, J. G. Hale, and P. McAleer. Visual perception of humanoid movement. In *Proceedings Third International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems 101*, pages 107–114, 2003. 1.1

[49] L. R. Rabiner and B-H. Juang. *Fundamentals of Speech Recognition*. Prentice Hall, 1993. 3.2.2

[50] Paul S. A. Reitsma and Nancy S. Pollard. Perceptual metrics for character animation: Sensitivity to errors in ballistic motion. *ACM Transactions on Graphics*, 22(3):537–542, 2003. 1.1, 5

[51] Liu Ren, Alton Patrick, Alexei A. Efros, Jessica K. Hodgins, and James M. Rehg. A data-driven approach to quantifying natural human motion. *ACM Transactions on Graphics*, 24(3):1090–1097, 2005. 2.2

[52] Charles Rose, Michael F. Cohen, and Bobby Bodenheimer. Verbs and adverbs: Multidimensional motion interpolation. *IEEE Computer Graphics and Applications*, 18(5):32–40, 1998. 1.1

[53] Charles Rose, Michael F. Cohen, and Bobby Bodenheimer. Verbs and adverbs: Multidimensional motion interpolation. *IEEE Computer Graphics and Applications*, 18(5):32–41, 1998. 4.2

[54] Alla Safonova, Jessica K. Hodgins, and Nancy S. Pollard. Synthesizing physically realistic human motion in low-dimensional, behavior-specific spaces. *ACM Transactions on Graphics*, 23(3):514–521, 2004. 1.2, 2.2, 4.3

[55] Yasuhiko Sakamoto, Shigeru Kuriyama, and Toyohisa Kaneko. Motion map: Image-based retrieval and segmentation of motion data. In *Proceedings of the 2004 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 259–266, 2004. 2.2

[56] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000. 4.4

[57] Ken Shoemake. Animating rotation with quaternion curves. In *Proceedings of the 12th Annual Conference on Computer Graphics and Interactive Techniques*, pages 245–254, 1985. 4.1.1

[58] Adnan Sulejmanpašić and Jovan Popović. Adaptation of performed ballistic motion. *ACM Transactions on Graphics*, 24(1):165–179, 2005. 1.2, 2.1

[59] Harold C. Sun and Dimitris N. Metaxas. Automating gait generation. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, pages 261–270, 2001. 2.2

[60] Seyoon Tak and Hyeong-Seok Ko. A physically-based motion retargeting filter. *ACM Transactions on Graphics*, 24(1):98–117, 2005. 1.2, 2.2

[61] Emanuel Todorov and Zoubin Ghahramani. Analysis of the synergies underlying complex hand manipulation. In *26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2004. 2.2

[62] Kari Torkkola. Feature extraction by non parametric mutual information maximization. *The Joural of Machine Learning Research*, 3:1415–1438, 2003. 4.4

[63] Antonio Torralba and Aude Oliva. Statistics of natural image categories. *Network: Computation in Neural Systems*, 14:391–412, 2003. 2.2

[64] N. K. Troje. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision*, 2:371–387, 2002. 2.1

[65] Raquel Urtasun, Pascal Glardon, Ronan Boulic, Daniel Thalmann, and Pascal Fua. Style-based motion synthesis. *Computer Graphics Forum*, 23(4):799–812, 2004. 2.2

[66] H. L. Van Trees. *Detection, Estimation, and Modulation Theory*, volume 1. John Wiley, 1968. 3.2

[67] Vicon Motion Systems. http://www.vicon.com/, 2005. 3.1, 4.1

[68] Vicon Motion Systems. Personal communication with Vicon customer support team, 2005. 1.2

[69] Jing Wang and Bobby Bodenheimer. An evaluation of a cost metric for selecting transitions between motion segments. In *Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 232–238, 2003. 1.1, 2.1

[70] Jing Wang and Bobby Bodenheimer. Computing the duration of motion transitions: an empirical approach. In *Proceedings of the 2004 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 335–344, 2004. 2.1

[71] X. Wang, M. Maurin, F. Mazet, N. De Castro Maia, K. Voinot, J. P. Verriest, and M. Fayet. Three-dimensional modelling of the motion range of axial rotation of the upper arm. *Journal of Biomechanics*, 31(10):899–908, 1998. 4.1.1

[72] Douglas J. Wiley and James K. Hahn. Interpolation synthesis of articulated figure motion. *IEEE Computer Graphics and Applications*, 17(6):39–45, 1997. 1.1

[73] Jane Wilhelms and Allen Van Gelder. Fast and easy reach-cone joint limits. *Journal of Graphics Tools*, 6(2):27–41, 2001. 2.2

[74] David Winter. *Biomechanics and Motor Control of Human Movement*. Wiley-Interscience, 1990. 1.2, 2.2

[75] Hussein Yahia and Andre Gagalowicz. Interactive animation of object orientations. In *Proceedings of the 2nd International Conference. Pixim 89*, pages 265–275, 1989. 4.1.1

[76] Jianmin Zhao and Norman I. Badler. Inverse kinematics positioning using nonlinear programming for highly articulated figures. *ACM Transactions on Graphics*, 13(4):313–336, 1994. 1.2, 2.2

[77] Hua Zhong, Jianbo Shi, and Mirko Visontai. Detecting unusual activity in video. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 819–826, 2004. 2.1