2008

Mixture Model Segmentation for Gait Recognition

Matthew Field
University of Wollongong, field@uow.edu.au

David A. Stirling
University of Wollongong, stirling@uow.edu.au

Fazel Naghdy
University of Wollongong, fazel@uow.edu.au

Zengxi Pan
University of Wollongong, zengxi@uow.edu.au

http://ro.uow.edu.au/engpapers/603

Publication Details
Mixture Model Segmentation for Gait Recognition

Matthew Field, David Stirling, Fazel Naghdy, Zengxi Pan
University of Wollongong, Australia
{mf91, stirling, fazel, zengxi}@uow.edu.au

Abstract
Modeling of human motion through a discrete sequence of motion primitives, retaining elements of skilful or unique motion of an individual is addressed. Using wireless inertial motion sensors, a skeletal model of the fluid human gait was gathered. The posture of the human model is described by sets of Euler angles for each sample. An intrinsic classification algorithm known as Minimum Message Length encoding (MML) is deployed to segment the stream of data and subsequently formulate certain Gaussian Mixture Models (GMM) that contain a plausible range of motion primitives. The removal of certain less seemingly important modes has been shown to significantly affect the fluidity of a gait cycle. The approach is described and the outcomes so far are provided.

1. Introduction

Human motion is a ubiquitous element of life that has attracted widespread research, from the analysis of sporting techniques to designing humanoid robots. Gesture and posture recognition has been an increasingly popular research field due to enhanced image processing, camera resolution and new motion capture sensors. Motor skill encompasses the ability and quality in achieving a particular motion task. As such, motor skill can be divided into two parts, gross motor skill and fine motor skill. Gross motor skill refers to major translations of a body that define the motion and usually follow a pattern. Fine motor skill involves precise movements that entail accurate positioning in a delicate task.

In this work, the focus is on quantifying human motor skill by capturing motions with an array of inertial sensors and articulating the motions as a set of patterns. This will ultimately be directed towards a motion model and control scheme for a robot, which can be programmed through demonstration from motion capture information.

1.1. Human Motion Capture Data

There are a range of motion sensing technologies available for data capture with different research areas concentrating on particular systems depending on suitability. Optical motion capture systems employ either infrared emitting diodes (IRED) or reflective markers to determine dynamic positions of fixed points. They are popular in biomechanics due to the fixed point accuracy and reliability of results [1]. However, the markers are prone to occlusion and the subject is confined to a small camera range.

Another popular approach is optical capture with a digital camera, which is very common in gesture recognition and gross motion tracking [2]. This method is very practical and increases the freedom of movement but is also affected by lighting conditions and current analyses have limited capacity for monitoring complex behavior.

Inertial sensors are becoming more practical due to the miniaturization to micro electromechanical systems (MEMS) allowing for easier mounting on the body. Solid state accelerometers have had some use in biomechanics experiments for estimating peak forces in jumping [3], analyzing gait patterns [4] and walking stability in the aged [5]. MEMS sensors are more flexible and are suitable in arrays for more complex motion capture [6] providing accurate orientation and accelerometer data.

1.2. Perception of Motion

The perception of human motion and its structured representation requires a method that essentially compresses multidimensional time series data into a symbolic form. In order to describe the complex data stream it is divided into a set of motion primitives which can be inferred from inherent patterns in the motion data. Several approaches have been explored to achieve this symbolic representation.

Kahol and colleagues [7] used an algorithm called Hierarchical Activity Segmentation to implement a dynamic hierarchical structure that describes the human body segments. The structure was characterized by demonstrator's profiles from a naive Bayesian classifier. The approach successfully recognized new gestures. Hidden Markov Models (HMM) have also been extensively employed in gesture recognition with Green [8] using particle filtering to fine tune classification with
blurred images. Saboune [2] employed a modified particle filtering algorithm named Interval Particle Filtering to analyze gait disorders from markerless motion capture. Inamura [9] also proposed a layered architecture encouraging symbolic representations of motion primitives with HMM.

In related work Sun [10] applied a Minimum Message Length (MML) encoding algorithm to inertial data gathered from different hand movements. This formulated a Gaussian Mixture Model for each segment of the data, which defined each mode or motion primitive in a motion sequence. By training the model with pre-defined motions and refining the cluster sequences, further motions of the same style could be recognized with high accuracy..

1.3. Motion Transformation

The programming of humanoid robots with human motion capture data (HMCD) is a particularly challenging extension from motion recognition. Despite the difficulties due to the high degrees of freedom (DOF) and the discrepancies in structure there has been promising results. Lopez [11] showed through simulation that HMCD provided a much smoother robot gait than the mathematical-foot trajectory calculation. Naksuk [12] was able to successfully transfer a walking motion to a stable simulation model by compensating the angular momentum at the centre of mass (CoM).

In the next section, the equipment and data preparation will be described. In Section 3 the basic concept behind the approach will be introduced. In Section 4, the different methods and ideas employed in the data analysis will be outlined. Section 5 will illustrate the results from segmentation and skill reduction. Finally, section 6 will discuss the implications of the results and future directions.

2. Experimental Rig

The challenging goal of this work is to model human behavior from an array of inertial sensors, resulting in an extensive motion database. The expression of locomotive behavior in terms of motion primitives provides a symbolic data compression approach to recognize regular sequences in multiple data streams.

The sensors used in this work were a set of sixteen inertial measurement units, the MTx, which are embedded in a spandex body suit named Moven from Xsens Technologies [13] as illustrated in Figure 1. Each sensor itself contains nine MEMS sensors to provide accurate 3D orientation to 2° RMS [14] and kinematics data from 3D acceleration and 3D rate of turn. With this array of orientation sensors mounted on fixed points on the body, an accurate kinematic model of the human body can be resolved. Although this equipment can provide accurate 3D skeletal models in real-time, unlike optical motion capture, it does require post-processing in the event of unknown contact points.

The system works via a Bluetooth wireless link allowing complete freedom of movement. The motion capture data is supplied in quaternion form for orientation and Cartesian coordinates for position for each body segment. Other available kinematic data includes acceleration, velocity and angular acceleration and velocity all sampled with the position and orientation at 100Hz. For our data analysis the orientation was converted from quaternion form to Euler angles of Roll, Pitch and Yaw because this form provides more immediate visual meaning and a reduction in dimensionality.

3. Approach

The underlying premise of this study is that human behavior can be expressed explicitly as sets of motion primitives, which in different configurations display a skill. This notion of skill identifies the expertise or technique of a motion that is not readily modeled with classical mechanics. Mathematical models are commonly employed in analysis of postural stability and gait although these models often either localize the analysis or sacrifice complexity for a more generalized model. However, if a locomotive behavior can be expressed in the segmented motion primitive form with sufficient accuracy it can be assumed that all human motion skills, from gross body movements to fine idiosyncratic behavior can be captured.

The approach in this study is to segment the orientation Euler angles of each major body part to
describe an individual pose or motion primitive. Conceptually the segmented poses should vary in time throughout a particular motion and form a proportion of the complete behavior. It is assumed that more abundant poses translate to the major movements of the limbs whereas the less abundant poses tend to be slight variations in these movements which translate to a motion style or technique. Furthermore, the removal of variant numbers of the least abundant poses results in a reduction in skill and fluidity of movement and further reduction results in a disjointed unnatural movement.

In order to judge the quality of a gait cycle, some integrated modeling technique is required to deal with the complex nature of the human form. The problem of analyzing dynamic human behavior is the multidimensional and high degrees of freedom (DOF) of the human body. This complexity increases the mathematical computation exponentially with every additional DOF. However, in gait and posture research, there are a few integrated physical parameters which have been identified as crucial to maintaining balance and ensure a stable walking motion. In this study, the centre of mass will be used to indicate the fluidity of gait for different skill levels.

4. Methodology

This section outlines the background theory relevant to the results of data analysis. It also introduces the software used and outlines the practical steps taken in analyzing the large corpus of data.

4.1. Machine Learning

An unsupervised classification method is required for partitioning unlabelled, complex data so that the model fits the observations. Without a pre-defined model an appropriate number of motion modes cannot be determined. Through Gaussian Mixture Modeling and inductive Bayesian probability based segmentation, this problem can be overcome. The approach deployed in this work is based on Minimum Message Length encoding (MML).

Minimum Message Length encoding is, as the name implies, based on evaluating a models ability to compress a message expressing the data [15]. Quantitatively, MML embodies 'Occam's razor' where the most complex model is automatically discarded despite the 'goodness of fit'. The algorithm first postulates a model of the data and evaluates it based on the estimated code length to express the model and the accuracy of the model. From information theory the minimum code length is:

\[ \text{Length}(E) = -\log P(E) \]  

where \( E \) is the evidence or data and from Bayes' theorem

\[ P(E | H) \cdot P(H) = P(H \cap E) \]  

where \( H \) is a probable hypothesis. Maximizing the probability that the evidence supports the hypothesis is equivalent to minimizing the total message length.

\[ \arg \max (P(H \cap E)) \equiv \arg \min (-\log(P(E | H)) - \log(P(H))) \]  

Figure 2 illustrates the message length output of clustering algorithm.

Figure 2: Message length output of clustering algorithm.

The resultant modes are a trade off between accuracy and message compression. Figure 2 illustrates the message length compression of the model spectrum.

Although greater compression theoretically provides better prediction, one of our aims in model selection is to abstract skills in a manner that produces results that can be understood by a human. The dimensional complexity of this data leads to overly complicated models with diminishing returns in compression. A further trade-off was introduced as a selection criterion. The exponent of the message difference or gradient was weighted by the model complexity (number of clusters, \( N \)). In this way, minimal improvements in message length are outweighed by increasing model complexity. Figure 3 shows the result of this selection.
4.2. Centre of mass

The centre of mass was estimated from data in biomechanics literature [16] and projected onto the feet locations. The purpose of the ground support calculation is to determine whether the body is balanced based on the condition that the CoM must lie over the support base. Assuming that a walking gait is described, the ground support can only come from within the convex hull of the feet positions. Bearing this in mind, the feet are tracked by an ellipse with the midpoints of the feet as the foci as illustrated in Figure 2. With this method the elliptical ground support can be transformed to a unit circle allowing a constant parameter space to monitor the center of mass. It, however, does not represent true ground support, rather (in a gait) the support for one foot and the intentional support for the other.

The semi-major and minor axes can be found from standard ellipse equations,

\[
a = \sqrt{\frac{\text{mid}_x^2 - \text{mid}_y^2}{2} + \frac{x_{fl} + x_{fw}}{2}}
\]

\[
b = a \sqrt{1 - \frac{\text{mid}_x^2 - \text{mid}_y^2}{4a^2}}
\]

which gives

\[
x = \frac{x_G - \text{mid}_x}{a}
\]

\[
y = \frac{y_G - \text{mid}_y}{b}
\]

where \(x_{fl}\) and \(x_{fw}\) stand for foot length and width, \(x_G\) and \(y_G\) is the global centre of mass position and \(x\) and \(y\) are these positions transformed to a unit circle space.

5. Validation

In the initial experiments with the motion capture, a volunteer was fitted with the equipment and recorded a collection of gait cycles for analysis. The objective was to gather data that could be segmented into posture sets and evaluated at different segment resolutions. With different numbers of resulting modes or clusters a technique for measuring the gait performance was investigated. The volunteer walked in a consistent forward trajectory in three recordings totaling 30 seconds and 45 strides.

5.1. Motion Capture

Using the orientation output the CoM was calculated with respect to the ground support as mentioned earlier. Figure 5 shows the part of the CoM path on the unit circle for each axis. The recordings indicated accurate translations consistent with the length of the designated walking area. No post-processing was required since the motion involved no unexpected external contacts, such as stairs or sloped surfaces.

Figure 3: (a) Exponential of message difference, (b) Trade off between model complexity and message improvement

Figure 4: Derivation of ground contact ellipse.

Figure 5: Center of mass trajectory from complete motion capture.
5.2. Segmentation

The orientation data, including 69 Euler angles, during this sequence was processed by the MML algorithm resulting in a significant number of modes or clusters. Each mode therefore represents the static posture at every time instant in the sequence. Due to the complexity of the human gait, the modes were limited using the information gain trade-off in Section 4.1. The algorithm found 24 unique poses to articulate the subject’s gait. Training is done off-line and produced the complete model spectrum, shown in Figure 2, in approximately 20 minutes. Most of this time, however, is spent searching among complex models (due to increased candidature of modes for splitting). Figure 6 illustrates a small portion of the resulting sequence of modes. Since each mode can be described by CoM position the sequence was recalculated at this reduced resolution only using the mean values of each mode. Figure 7 shows the mode abundance graph and the X and Y axis center of mass trajectories for this model.

5.3. Skill Reduction

It is proposed that a collection of the least abundant clusters form the fine skills of the subject whereas the...
majority of the abundant clusters account for the gross motor skill. To remove one of the motion primitives the instances classified as that mode were replaced with closest temporal mode in an attempt to retain a more natural mode transition.

Figure 8 illustrates the reduction in skill level with the abundance graph nullifying the rejected motion primitives. Under the condition shown in this graph with 16 modes left remaining, the CoM trajectory was recalculated. The CoM plot shows that the mode reduction had eliminated a significant section of the gait cycle corresponding to a leg extension along the X-axis.

6. Conclusions

This work illustrated that the human behavior can be segmented into a sequence of motion primitives which directly correspond to a set of static poses. This provides a starting point for developing a structured hierarchy of motion primitives for synthesizing new behaviors. However, in order to generate dynamic sequences other influences such as inertia must also be accommodated within the motion primitives. This research is building towards a heuristic control scheme for a humanoid robot based on human demonstration, which also requires a detailed study of the structural transformation between human and humanoid.

The ideas provided in this paper are of relevance to the treatment of diseases such as Parkinson’s disease and the development of balance prostheses or aids for elderly patients. People with Parkinson’s disease may be able to control the gross motor skills but often gradually lose the influence of fine motor skills for precise movement. One could also use these techniques to diagnose locomotor disorders, analyzing the effects on behavior after rehabilitative treatment or for developing intelligent prostheses and potentially for Functional Electrical Stimulation (FES) feedback [17].

Future work in this area will involve investigating the dynamic motion primitives by incorporating acceleration data or combined force vectors, and investigating a performance measure which responds to the reduction in fine motor skills.

7. References


