Structure of a Motion Capture Database under Different Similarity Metrics

Donovan Parks (donovan.parks@gmail.com)

October 25, 2007

1 Introduction

Large motion capture databases are becoming commonplace. These databases are widely used to achieve realistic human motion animation in both the film and video game industries. Recently, significant research effort has been devoted to developing methods for efficiently searching these databases in order to allow animators to quickly find similar motions. Identifying similar motion sequences is a challenging problem since *logically* similar motion sequences are often not *numerically* similar. For example, motion sequences of individuals walking at different speeds are logically similar although they differ significantly if numerically compared on a frame-by-frame basis. Thus, the essential problem is to define a similarity metric that can “bridge the semantic gap between logical similarity as perceived by humans and computable numerical similarity”[5]. A robust similarity metric must also properly normalize the motion sequence data and take into account motion sequences of different lengths.

The goal of this project is to develop an InfoVis environment for analyzing the structure that is imposed on a human motion database by a given similarity metric. It is believed that such an environment will allow for the rapid understanding of the strengths and weaknesses of a similarity metric along with suggesting ways in which the metric could be improved. To illustrate the use of this visualization environment, the similarity metric proposed by Li and Prabhakaran [3] will be analyzed on a subset of the CMU Motion Capture (MoCap) database[1].

2 CMU MoCap Database

This project will consider walking, running, and jumping motion sequences from the CMU MoCap database. These three classes of locomotion are being considered as they are well represented in the CMU database and they form an

---

[1] CMU MoCap database: mocap.cs.cmu.edu

interesting subset of the entire database. In particular, walking quickly is logically more similar to running than walking slowly and jumping can be considered fairly distinct from both walking and running. To what extent this logical (dis)similarity is captured by a similarity metric is a key question that the proposed environment will facilitate in answering. There are also interesting subsets within each of these locomotion classes. For example, there are 17 jumping sequences that are labeled either “forward jumping”, “jumping”, or “high jump”. The database contains 44 running sequences (labeled as “run/jog”, “run”, “run” with various turns) and over 100 walking sequences (labeled as “slow walk”, “walk”, “walk” with various turns).

We now introduce some terminology that is useful for discussing motion capture data. Human motion can be described by considering a simplified model of the human skeleton. A skeleton consists of *body segments* (bones) that are connected by *joints*. Motion capture can be thought of as a process which records a temporal sequence of 3D joint positions. The position of all joints at a given time is known as a *pose*. A pose can be described as a matrix $P \in \mathbb{R}^{3\times|J|}$, where $|J|$ is the number of joints in the skeletal model and column i specifies the 3D coordinates of joint i. A *motion sequence* can then formally be described as a time-dependent sequence of poses. This can be represented by a matrix $S \in \mathbb{R}^{3\times|J|\times T}$, where T is the number of poses (frames) in the motion sequence. For convenience, it is assumed poses are sampled at a fixed sampling rate.

The motion sequences in the CMU MoCap database are of varying lengths (i.e., T varies between motion sequences) and consider a skeletal model consisting of 32 joints (i.e., $|J| = 32$). The placement of markers used to record this skeletal model can be found on the CMU MoCap website[2]. This model is relatively detailed. For example, the skeletal model considers the position of the thumb joint. This level of detail is unnecessary when comparing locomotion sequences, so we will use a simplified model consisting of only the 18 most significant joints. This should result in a skeletal model similar to that considered by Li and Prabhakaran [3] who indicated they use a skeletal model with 18 joints, although do not specify which joints were used.

3 Previous Experience

I do not have any previous experience using MoCap data. My interest in this project stems largely from the possibility of pursuing computer vision based pose estimation and gesture recognition for my PhD research.

4 Proposed Solution

Numerous similarity metrics have been defined for comparing human motion sequences [1, 2, 3, 4, 5]. Which of these should be preferred? What are their respective strengths and weaknesses? How can a given metric be improved?

---

[2] CMU MoCap Skeletal Model: http://mocap.cs.cmu.edu/markerPlacementGuide.pdf

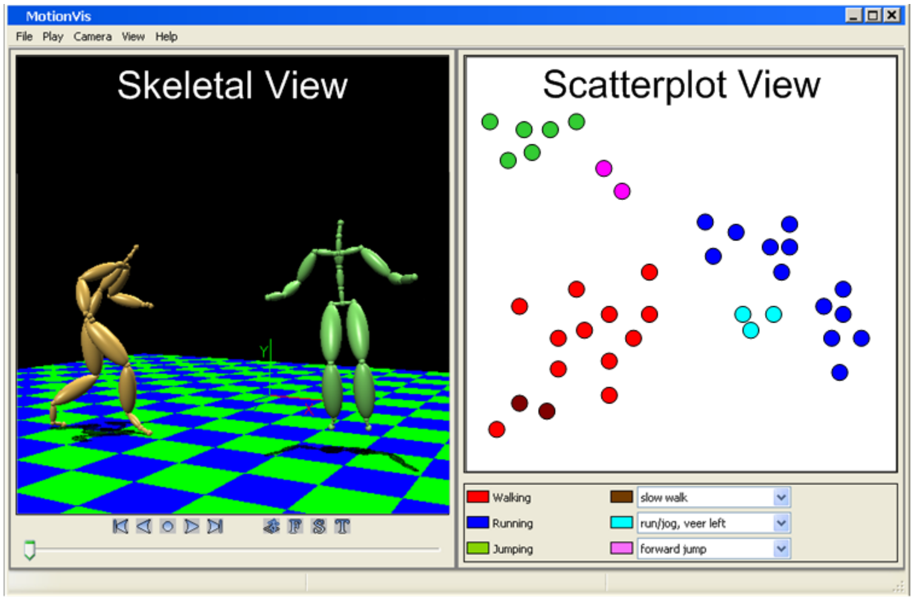

Figure 1: Mockup of proposed GUI. In the skeletal view motion sequences are animated with a simplified human model. The scatterplot view indicates the structure of the motion database under a given similarity metric. Tight coupling between these two views allows for the rapid exploration of the motion sequences in the scatterplot.

Is the structure imposed on a motion database by a given similarity metric useful for classifying novel motion sequences or indexing into the database? The purpose of this project is to design an InfoVis environment that can aid in answering questions of this nature. Given a similarity metric a similarity matrix can be constructed for the motion capture database. Multi-dimensional scaling (MDS) can then be applied to the similarity matrix in order to allow its approximate structure to be visualized using a simple 2D scatterplot. By tightly coupling this scatterplot view with a 3D view capable of animating selected motion sequences, critical information about a similarity metric can be quickly obtained. This 3D skeletal view can be seen as providing details-on-demand.

4.1 Interface Mockups

Figure 1 gives a mockup of the proposed interface. It consists of two main views: the skeletal view and the scatterplot view. In the skeletal view motion sequences are animated with a simplified human model. Below this view, standard movie controls are provided to allow for easy navigation through an animation sequence. The user can move the camera within this skeletal view in order to view a motion sequence from any position. Buttons below the skeletal view allow rapid switching between viewpoints (i.e., front, top, side).

The scatterplot view indicates the structure of the motion database under a given similarity metric. Locomotion classes and sub-classes are colour coded to allow similar motions to be easily identified. Choice boxes below the scatterplot view allow the user to select which sub-classes they are interested in viewing. There are far too many sub-classes to be viewed at once, so this selection scheme has been employed.

Selecting points in the scatterplot view will cause the motion sequence associated with the point to be animated in the skeletal view. A colouring scheme will be used to help the user associate points in the scatterplot view with skeletons in the skeletal view. This colour scheme is not illustrated in Figure 1 since it has not been finalized. The general idea will be to indicate selected points by drawing a heavy border around them. This border colour will be unique for each selected point. The center portion of a skeleton will be coloured to correspond with the selection colour whereas the peripheral portions of the skeleton (i.e. arms and legs) will reflect the locomotion class or sub-class.

4.2 Scenario of Use

Dr. Williams has just finished implementing a similarity metric for human motion sequences. She is very interesting in understanding the strengths and weaknesses of this similarity metric. Using the proposed solution, she can rapidly investigate different aspects of this similarity metric and determine if it will meet her needs. The most significant question is whether the similarity metric successfully distinguishes between different motion classes by causing them to form distinct clusters. When Dr. Williams loads her motion database she is immediately presented with a scatterplot indicating the structure her similarity metric has imposed on the database (see Figure 2). Use of colour to distinguish between different motion classes makes it easy to identify 3 distinct clusters corresponding to the walking, running, and jumping locomotion sequences she is considering.

Looking more carefully at the scatterplot, Dr. Williams notices a few interesting anomalies. She is very concerned that there is a running motion sequence near the center of the cluster corresponding to jumping sequences (blue point in upper, left of Figure 2). Clicking on this anomaly causes the motion sequence associated with the point to be animated in the skeletal view. Dr. Williams is relieved to find that this is a simple labeling error. The motion sequence is actually of someone jumping. Hovering her mouse over the anomalous point causes the name of the motion file associated with it to be displayed (Figure 3). She can now go into her motion database to re-label this motion sequence.

She is also interested by the group of 3 walking sequences that are relatively isolated from the main walking cluster (Figure 4). She selects these 3 points in order to view their motion in the skeletal view. After watching the animation she understands that these 3 motion sequences are all of people walking briskly. This explains why they have formed their own cluster and why they are relatively close to the cluster of running sequences.

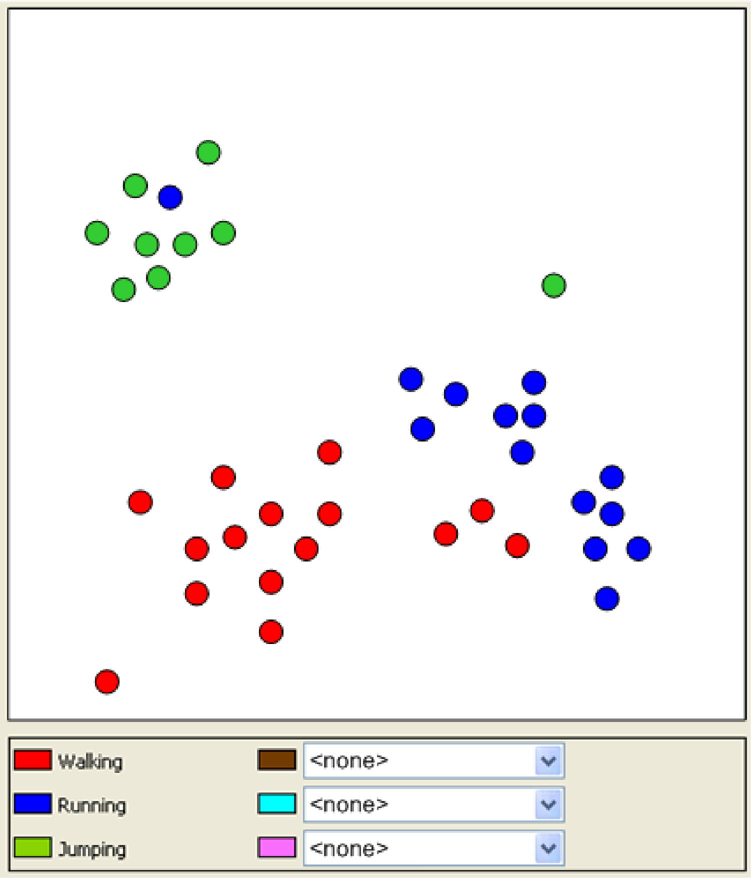

Figure 2: The scatterplot view uses colour to quickly allow related motion sequences to be identified.

Figure 3: Tooltips allow the motion file associated with a point to be quickly identified.

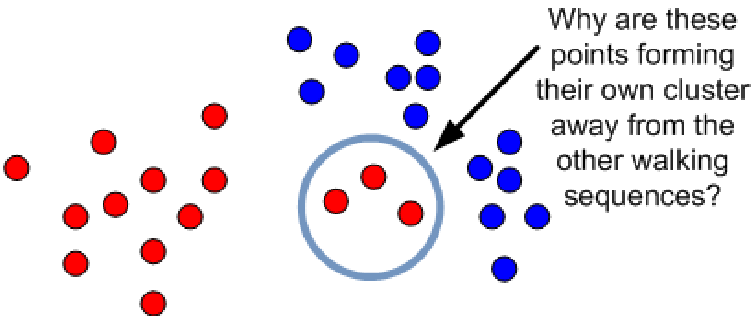

Figure 4: Selecting the 3 circled points allows their motion to be viewed in the skeletal view. This will help to determine why these points have formed their own cluster away from the other walking sequences.

The last anomaly that draws Dr. Williams' attention is the relatively isolated jumping sequence (green point in middle, right of Figure 2). She hopes this point is simply a mis-labeled running sequence since it is closer to the "running" cluster than to the "jumping" cluster. Unfortunately, after clicking on this point to view its motion in the skeletal view she finds that it is of someone jumping. By comparing this sequence to other jumping sequences, Dr. Williams begins to understand that it is unique in that it is the only one where a person is jumping to the side. This helps explain why it is not in the "jumping" cluster, but does little to explain why it would be considered similar to a running sequence. After examining the running sequences closest to this anomalous point, Dr. Williams is left to concluded that this problem is likely a result of her similarity metric. She will need to carefully consider her metric and why such an error could occur. Although the proposed environment can not provide her with a full understanding of this problem, it has drawn her attention to it and given her a strong qualitative understanding of the situation.

4.3 Implementation Details

A number of open source MoCap players have been identified. All of the identified players are written in C/C++. Implementing a MoCap player is non-trivial so the proposed solution will be implemented in C/C++ in order to build upon this existing codebase. Specifically, I plan to make use of the MotView player[3].

The next most challenging aspects of the proposed solution are implementing the dimensionality reduction algorithm and similarity metric. C/C++ implementations of SVD (used by Li and Prabhakaran's similarity metric) and MDS (most likely candidate for performing dimensionality reduction, although SVD could also be used) are readily available. This has further prompted the use of C/C++.

---

[3] MotView: www.cs.wisc.edu/graphics/Courses/cs-838-2000/Students/gardner/motView/

In order to take advantage of my existing skill set, the project will be implemented using Visual C++ under Windows XP and make extensive use of OpenGL and wxWidgets for implementing the GUI.

5 Previous Work

This work is unique in that its goal is to provide an interactive environment for analyzing the structure of a human motion database under a given similarity metric. However, the individual elements of the proposed environment are in mainstream use. Several commercial programs (e.g. Maya, Poser) and several open source programs (e.g., MotView, Blender, CMU MoCap Player) exist for visualizing MoCap data in the form of an animated 3D skeleton. As previously mentioned, this project will make use of MotView as an initial codebase.

Many methods have been proposed for visualizing similarity matrices (most notably, graph layout algorithms [6] and MDS [8]). This project makes use of MDS in order to allow the structure of a similarity matrix to viewed as a simple 2D scatterplot. A critical aspect of the proposed environment, is the tight coupling between the 3D skeletal view and the scatterplot view. The use of coordinated views for analyzing data is well studied in InfoVis (see [7] for a recent review). We refrain from reviewing the literature on dimensionality reduction and coordinated views since the focus of this project is not on how to perform InfoVis, but rather on how established InfoVis practices can aid in the exploration and qualitative understanding of the structure of a similarity matrix defined over a large human motion database.

6 Milestones

The following milestones will be used to track the progress of this project:

- October 31: Finish all required reprocessing of CMU database.
- November 6: Familiarize myself with MotView code and implement 3D skeletal view.
- November 10: Implement similarity metric proposed by Li and Prabhakaran. Test on simple dataset with clear ground truth.
- November 14: Implement scatterplot visualization (including required dimensionality reduction algorithm). Test on simple dataset with clear ground truth.
- November 20: Apply Li and Prabhakaran's similarity metric to locomotion sequences in the CMU database at varying window sizes along with applying dimensionality reduction to the resulting similarity matrices. It is expected that calculating the similarity metric along with the dimensionality reduction will be sufficiently computationally expensive to require it to be done as offline preprocessing.

- November 25: Implement coordination between the scatterplot view and the 3D skeletal view.
- November 27: Use implemented solution to perform an initial investigation of Li and Prabhakaran similarity metric.
- December 1: Refinements to solution based on the investigation performed above.
- December 5: Use solution to fully investigate the structure of the CMU MoCap database under the Li and Prabhakaran similarity metric.
- December 12: Complete project write-up.

7 Possible Extensions

Time permitting the following extensions will be considered:

- Extending the scatterplot view to allow for 2D *and* 3D visualization of the similarity matrix.
- Examining an additional similarity metric.
- Providing a direct view of the similarity matrix where the similarity measure is colour coded and the ordering of the columns and rows can be sorted to aid in visualizing clusters and/or outliers.
- Consideration of additional motion sequences from the CMU MoCap database.

8 Conclusion

A wide range of similarity metrics have been developed for comparing human motion sequences. The goal of this project is to design and develop an interactive environment that will allow for the rapid understanding of the structure imposed on a human motion database by a given similarity metric. This goal will be achieved by taking advantage of established InfoVis practices (e.g., multi-dimensional scaling, colour coding of nominal data, coordinated views, details-on-demand).

References

[1] Eamonn Keogh, Themistoklis Papanas, Victor B. Zordan, Dimitrios Gunopulos, and Marc Cardle. Indexing large human-motion databases. In *Proceedings of the 30th VLDB Conference*, pages 780–791, 2004.

[2] Lucas Kovar and Michael Gleicher. Automated extraction and parameterization of motions in large data sets. *ACM Trans. Graph.*, 23(3):559–568, 2004.

[3] Chuanjun Li and B. Prabhakaran. Indexing of motion capture data for efficient and fast similarity search, 2006.

[4] Guodong Liu, Jingdan Zhang, Wei Wang, and Leonard McMillan. A system for analyzing and indexing human-motion databases. In *SIGMOD '05: Proceedings of the 2005 ACM SIGMOD international conference on Management of data*, pages 924–926, New York, NY, USA, 2005. ACM Press.

[5] Meinard Müller, Tido Röder, and Michael Clausen. Efficient content-based retrieval of motion capture data. *ACM Trans. Graph.*, 24(3):677–685, 2005.

[6] Chris Roussin Rich DeJordy, Stephen P. Borgatti and Daniel S. Halgin. Visualizing proximity data, 2007.

[7] Jonathan C. Roberts. State of the art: coordinated and multiple views in exploratory visualization. *Proc. Conference on Coordinated and Multiple Views in Exploratory Visualization*, 2007.

[8] W. S. Torgerson. *Theory and methods of scaling*. New York: John Wiley, 1958.