Learning to Synthesize Motion Styles

Lorenzo Torresani
Riya, Inc.
lorenzo@riya.com

Peggy Hackney
Integrated Movement Studies
PJHackney@aol.com

Christoph Bregler
New York University
chris.bregler@nyu.edu

Over the past decade research in computer animated human motion has progressed dramatically. Early automatic animation systems used either procedural, also known as rule-based, methods, or dynamic simulations. Recently, data-driven methods have become one of the prominent approaches for realistic animation. In particular, sample-based systems concatenating motion capture snippets are the most successful examples in this genre [6, 4]. Motion capture concatenation techniques can generate complex new motions but are restricted to producing exactly the subset of styles collected in the input database. A more desirable outcome is a method that has the ability to change the style of a motion sequence in a more general context. In recent years several machine learning animation systems [1, 5] have been proposed that attempt to generalize style differences and produce new styles. However, most of these methods learn simple parametric motion models that are unable to fully capture the subtleties and complexities of human movement. As a consequence, animations resulting from these systems are often plagued by low quality and scarce realism.

The technique introduced in this paper is a compromise between the pure concatenative approaches and the methods based on learned parametric models. The aim is to maintain the animated precision of motion capture data, while introducing the flexibility of style changes that can only be achieved by learned parametric models. Our system builds on the observation that stylistically novel, yet highly realistic animations can be generated via space-time interpolation of pairs of motion sequences. We propose to learn not a parametric function of the motion, but rather a parametric function of the interpolation or extrapolation weights applied to data snippets.

Motion style is not a very clearly defined term in computer animation literature. We base our definition of motion style on an established notation system, called Laban Movement Analysis or LMA [3]. A subset of this entire motion description system is the "LMA-Effort" dimension. The system is not attempting to describe the coarse aspects of a motion, i.e. whether someone is walking, or swinging his/her arm. Instead, it targets the subtle differences in motion style, i.e. is the movement "bound" or "free". We focus on three LMA-Effort factors: Flow, Weight and Time. Each of these factors varies in intensity in a continuous range between opposing poles. We represent these LMA-Effort parameters as points in a three-dimensional perceptual space.

In order to train our system to synthesize user-specified motion styles, we collected a vast corpus of human motion sequences with a state of the art marker-based motion capture system. We represent the motion as a time-varying vector of joint angles. In the training stage each motion sequence was manually segmented by an LMA human expert into fragments corresponding to fundamental actions or units of motions in the sequence. We denote $\mathbf{X}_i$ the joint angle data of the i-th fragment in the database. We apply a motion matching algorithm to identify fragment pairs $(\mathbf{X}_i, \mathbf{X}_j)$ containing similar actions. Our motion matching algorithm is based on dynamic-time warping. This allows us to compare kinematic contents while factoring out differences in timing or acceleration, more often associated to variations in style. We use these motion matches to augment the database with new synthetically-generated styles: given matching motion fragments $\mathbf{X}_i$, $\mathbf{X}_j$, and an interpolation parameter α, space-time interpolation smoothly blends the kinematics and dynamics of the two fragments to produce a new motion $\mathbf{X}_{i,j}^{\alpha}$ containing the original action, but having distinct style and timing. Both the synthesized animations $\mathbf{X}_{i,j}^{\alpha}$ as well as the "seed" motion capture data $\mathbf{X}_i$ were labeled with LMA-Effort values by an LMA expert. We denote $\mathbf{e}_i$ and $\mathbf{e}_{i,j}^{\alpha}$ the three-dimensional vectors encoding the LMA-Effort qualities of $\mathbf{X}_i$ and $\mathbf{X}_{i,j}^{\alpha}$, respectively. A non-linear regression model [2] was

Topics: visual processing and pattern recognition; estimation, prediction, and sequence modeling.
Preference: oral/poster.

fitted to the LMA labels and the parameters α of the space-time interpolation algorithm. This regression defines a function f predicting LMA-Effort factors $\mathbf{e}_{i,j}^{\alpha}$ from the style attributes and joint angle data of fragments i and j:

$$\mathbf{e}_{i,j}^{\alpha} \quad = \quad f(\mathbf{X}_i, \mathbf{X}_j, \mathbf{e}_i, \mathbf{e}_j, \alpha) \tag{1}$$

This function-fitting stage allows us to learn how the knobs of our animation system relate to the perceptual space of movement styles.

At testing stage we are given a motion sequence $\mathbf{Y}$, and a user-specified motion style $\bar{\mathbf{e}}$. The goal is to apply style $\bar{\mathbf{e}}$ to the input sequence $\mathbf{Y}$, without modifying the content of the motion. First, we use dynamic-time warping to segment the input sequence into snippets $\mathbf{Y}_i$ matching the content of fragment subsets $\{\mathbf{X}_i^{(k)}\}$ in the database. We then replace each snippet $\mathbf{Y}_i$ with the pairwise blend of examples in $\{\mathbf{X}_i^{(k)}\}$ that best approximates motion style $\bar{\mathbf{e}}$. This objective can be formulated as

$$\alpha^*, k^*, l^* \leftarrow \arg\min_{\alpha,k,l} ||\bar{\mathbf{e}} - f(\mathbf{X}_i^{(k)}, \mathbf{X}_i^{(l)}, \mathbf{e}_i^{(k)}, \mathbf{e}_i^{(l)}, \alpha)|| \tag{2}$$

The animation resulting from space-time interpolation of fragments $\mathbf{X}_i^{(k^*)}$ and $\mathbf{X}_i^{(l^*)}$ with parameter α^* will display motion style approximating the target $\bar{\mathbf{e}}$.

We demonstrate the system on a motion database consisting of 12 sequences performed by different professional dancers. Sample videos of the automatically generated motions can be viewed at `http://movement.nyu.edu/learning-motion-styles/`.

References

[1] M. Brand and A. Hertzmann. Style machines. In *Proceedings of ACM SIGGRAPH 2000*, Computer Graphics Proceedings, Annual Conference Series, pages 183–192, July 2000.

[2] H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola, and V. Vapnik. Support vector regression machines. In *Proc. NIPS 9*, 2003.

[3] P. Hackney. *Making Connections: Total Body Integration Through Bartenieff Fundamentals*. Routledge, 2000.

[4] L. Kovar, M. Gleicher, and F. Pighin. Motion graphs. *ACM Transactions on Graphics*, 21(3):473–482, July 2002.

[5] Y. Li, T. Wang, and H.-Y. Shum. Motion texture: A two-level statistical model for character motion synthesis. *ACM Transactions on Graphics*, 21(3):465–472, July 2002.

[6] K. Pullen and C. Bregler. Motion capture assisted animation: Texturing and synthesis. *ACM Transactions on Graphics*, 21(3):501–508, July 2002.