### An Algorithm for Extracting Human Motion Signatures

M. Alex O. Vasilescu Department of Computer Science, University of Toronto Toronto, ON M5S 3G4, Canada

### Abstract

Human motion is the composite consequence of multiple elements-the action performed, an expressive cadence, and a motion signature that captures the distinctive pattern of movement of a particular individual. We develop a new algorithm that is capable of extracting these motion elements and recombining them in novel ways. The algorithm is based on the numerical statistical analysis of motion data spanning multiple subjects performing different types of motions. In particular, we demonstrate that, after our algorithm analyzes a corpus of walking, stair ascending, and stair descending data collected over a group of subjects, it can then observe a sample of walking motion for a new subject and recognize never before seen ascending and descending motions for this new individual. Our approach also yields a generative motion model that can synthesize unseen motions in the distinctive style of this individual. We validate our algorithm using a standard pattern classifier.

## 1. Introduction and Background

In analogy with handwritten signatures, do people have characteristic motion signatures that individualize their movements? If so, can these signatures be extracted from example motions? Can extracted signatures be used to recognize, say, a particular individual's walk subsequent to observing examples of other movements produced by this individual?

Our ability to perceive motion signatures seems wellgrounded from an evolutionary perspective, since it is clearly conducive to survival to be able to recognize the motions of predator or prey, or of friend or foe, especially at a distance and in the absence of other details. In the 1960s, the psychologist Gunnar Johansson performed a series of famous experiments in which he attached reflectors to people and recorded videos of them performing different activities, such as walking, running, dancing, and so on [5]. Observers of these videos, in which only the reflective dots are visible, were asked to classify the activity performed and to note certain characteristics of the motion, such as a limp or an energetic/tired walk. Observers can usually perform this task with ease and they could even recognize specific individuals in this way. This may corroborate the hypothesis that the signature of a motion is a tangible quantity that can be separated from the actual motion type.

Our research in progress has three goals. The first is to develop a new algorithm that can analyze everyday human motions to separate out distinctive motion signatures. The second is to recognize specific individuals performing motions not included in our database, using extracted motion signatures associated with these individuals. The third goal is to synthesize novel motions that are in accord with extracted motion signatures. The mathematical basis of our approach is a statistical numerical technique known as *n*-mode analysis, a multi-linear singular value decomposition (SVD) applicable to higher-order data arrays.

Our algorithm exploits corpora of motion data which are now reasonably easy to acquire through a variety of modern motion capture technologies developed for use in the entertainment industry [3]. Motion synthesis through the analysis of motion capture data is currently attracting a great deal of attention within the computer graphics community as a means of animating graphical characters. Several authors have introduced generative motion models for this purpose. Several recent papers report the use of hidden Markov models [1]. Howe *et al.* [4] analyzes motion from video using a mixture-of-Gaussians model. Grzeszczuk et al. develop neural network learning models to emulate physically simulated motions [2].

We address the motion analysis/synthesis problem using techniques from numerical statistics. The n-mode analysis algorithm that we adapt to our purposes was described for scalar observations by Kapteyen *et al.* [7, 8]. Marimont and Wandell [9] extended it to 2-mode vector observations in the context of characterizing color surface and illuminant spectra. Freeman and Tenenbaum [10] used this extension, their so-called bilinear model, in three different perceptual domains, including the translation of faces to novel illuminations.

# 2. Motion Data Acquisition

Human limb motion was recorded using a VICON system that employs four video cameras. The cameras detect infra-



Figure 1: The motion capture facility

red light reflected from markers placed on the limbs of a human subject. The system computes the 3D position of the markers,  $\vec{m}_i$ , relative to a fixed lab coordinate frame,  $F_L$ . The video cameras are positioned on one side of a 12 meter long walkway such that each marker can be observed by at least two cameras during motion.

To extract the three angles spanned by a human joint, we must define a plane for each limb whose motion can be measured relative to the sagittal, frontal and transverse planes through the body. A total of 18 markers were used, 9 to measure the motion of each leg. In general, we must define a plane for each limb, using three markers. We placed a marker at each end of the limb (at the joints) and we attached a third marker to a 15 cm long stick strapped half way down the limb extending away from the body, such that the plane formed with the other markers are parallel to the frontal plane. For the foot plane, markers were placed on the ankle joint (at the base of the fifth metatarsal), laterally on the heel and on the lateral malleolus (right before the toes). The pelvis was defined by markers placed on the iliac crest, anterior superior spine and greater trochanter.

Each subject was asked to perform three types of motions—walking, ascending and descending stairs. Each motion was repeated 10 times. A motion cycle was segmented from each motion sequence. To suppress noise, the collected motion data were low-pass filtered by a fourth-order Butterworth filter at a cut off frequency of 6 Hz and missing data were interpolated with a cubic spline.

After the marker positions were measured, filtered, and interpolated, the time-varying rotations of each of the limbs were calculated with respect to the lab frame,  $F_L$ . These transformations are used to compute intermediate transformations between the limb segments from which joint angles can be computed. Pelvic rotation is the rotation with respect to lab frame,  $F_L$ . The rotation of the thigh is the rotation with respect to the pelvis through which the hip angles are obtained. The knee angles are defined as the rotations of the shank with respect to the thigh. The ankle angles are defined as the rotations of the foot with respect to the shank.

Therefore, we first calculate the frame coordinate transformation for each limb with respect to the lab, next we calculate the relative orientation of each limb in the kinematic chain, and finally we solve for inverse kinematic equations to compute the joint angles.

# 3. Motion Analysis

Given motion sequences of several people walking on level ground, as well as ascending and descending stairs, we define a data set D which takes the form of a  $nt \times m$  matrix, where n is the number of people, t is the number of joint angle time samples, and m is the number of motion classes.<sup>1</sup> The first column of D stacks the mean walk of every person, the second column stacks the mean ascending motion and the third stacks the mean stair descent, as follows:

$$D = \begin{bmatrix} D_{1} \\ \vdots \\ D_{i} \\ \vdots \\ D_{n} \end{bmatrix}$$
(1)
$$D_{i} = \begin{bmatrix} \rightarrow & \rightarrow & \rightarrow \\ walk_{i} & ascend_{i} & descend_{i} \end{bmatrix},$$
(2)

The columns of matrix  $D_i$  are the average walk, ascend and descend of stairs of the  $i^{th}$  person. Each motion is defined as the angles by every joint over time.

Motivated by the *multi-mode component analysis* or *n*mode component analysis from the numerical statistics literature [7, 8], we decompose the complete data set into the "product" of a core matrix Z, a people parameter matrix P, and an action parameter matrix A, as follows:

$$D = (Z^{VT} P^T)^{VT} A^T$$
 (3)

$$= SA^T, (4)$$

where the VT-operator is a matrix transpose T followed by a "vec" operator that creates a vector by stacking the columns of the matrix. The signature matrix

$$S = (Z^{VT} P^T)^{VT} = [S_1 \dots S_i \dots S_n]^T$$
(5)

is composed of person-specific signatures  $S_i$ . The people matrix

$$P = [\vec{p}_1 \dots \vec{p}_i \dots \vec{p}_n]^T, \tag{6}$$

whose row vectors  $\vec{p}_i$  are person specific, encodes the invariances across actions for each person. The action matrix

<sup>&</sup>lt;sup>1</sup>Note that the natural structure of our acquired motion data is a 3dimensional  $n \times m \times t$  array — i.e., a rank-3 tensor — as opposed to the "flattened" data matrix form *D*. De Lathauwer *et al.* present a direct, tensor reformulation of n-mode analysis that they call higher-order singular value decomposition (HOSVD). Their formalism is suitable in our application. It is straightforward to specify our algorithm in their formalism and we will do so in a subsequent paper.

whose row vectors  $\vec{a}_c$ , contain the coefficients for the different action classes c, encodes the invariances across people for each action. The core matrix

$$Z = [Z_1 \dots Z_i \dots Z_n]^T \tag{8}$$

represents the basis motions which are independent of people and of actions. It governs how P and A interact to produce the observed motions.

We solve for Z, P and A by applying the 2-mode vector analysis algorithm, which minimizes

$$E = \|D - (Z^{VT}P^{T})^{VT}A^{T}\| + \lambda_{1}(\|P^{T}P - I\| + \lambda_{2}(\|A^{T}A - I\|),$$
(9)

where I is the identity matrix. The algorithm computes P as follows:

$$D = (Z^{VT} P^T)^{VT} A^T, (10)$$

$$(DA)^{VT} = Z^{VT}P^T \tag{11}$$

$$USV^{I} = Z^{VI}P^{I} \text{ Compute SVD of LHS}$$
(12)

$$V = P$$
 Set P to first r-columns of  $V(13)$ 

and A as follows:

$$D^{VT} = (ZA^T)^{VT} P^T (14)$$

$$(D^{VT}P)^{VT} = ZA^T (15)$$

$$USV^T = ZA^T$$
 Compute SVD of LHS (16)

V = A Set A to first r-columns of V.(17)

Note that we compute the singular value decompositions (SVD) of the left hand sides (LHS) of (11) and (15). The matrix Z is computed as follows:

$$Z = (D^{VT}P)^{VT}A, (18)$$

where P and A are orthogonal matrices.

#### 4. Generative Motion Model

We define the new signature model for a new person

$$D_{new} \left\{ \left[ \begin{array}{c} ? \\ \end{array} \right| \\ \left| \begin{array}{c} \end{array} \right] = \underbrace{\left[ \begin{array}{c} ? \\ \end{array} \right]}_{S_{new}} \\ A^T$$
(19)

for whom we have examples of only some of the motion classes, as the linear optimal combination of known signatures:

$$S_{new} = \underbrace{\begin{bmatrix} W_1 & \cdots & W_i & \cdots & W_n \end{bmatrix}}_{W} \underbrace{\begin{bmatrix} S_1 \\ \vdots \\ S_i \\ \vdots \\ S_n \end{bmatrix}}_{S}$$
(20)

where W is a weight matrix.

We solve for the weight matrix W using iterative gradient descent of the error function

$$E = \|D_{new} - WSA_{inc}^T\|,\tag{21}$$

where  $A_{inc}^T$  has only the columns corresponding to the motion examples available in  $D_{new}$ .

The gradient update of the signature matrix,  $S_{new}$  for the new person is given as follows:

$$Q = SA_{inc}^T \tag{22}$$

$$W(t+1) = W(t) + \gamma (D_{new} - WQ)Q^T$$
(23)

$$S_{new}(t+1) = W(t+1)S$$
 (24)

For example, to synthesize new data for walking, we multiply the newly extracted motion signature  $S_{new}$  from equation (24) and the action parameters for walking,  $\vec{a}_{walk}$ , as follows:

$$wa\overline{lk}_{new} = S_{new}\vec{a}_{walk} \tag{25}$$

### 5. Results

A corpus of motion data was collected from 6 subjects. Three motions were collected for each person: walk, ascend-stairs, descend stairs. Given a sufficient quantity of motion data, our human motion signature extraction algorithm can consistently produce walks and stair ascend/descend motions in the styles of individuals.

In a "leave-one-out" validation study, we verified that our algorithm was able to compute motion signatures sufficiently well to synthesize all three types of motions in the distinctive style of each individual compared against ground-truth motion capture data of that individual. If the motion signature  $S_{new}$  captures the distinctive pattern of movement, the synthesized walk would best match the actual walk of the new person. Using a nearest neighbor classifier, the synthesized walk was indeed recognized against a complete database that includes the actual walk data for the new person.

Fig. 2(a) shows, in frontal view, the synthesis of three different styles of walking motion given only examples of descending stairs in those corresponding styles. Note that the walking styles differ subtly: The woman on the left walks in a pigeon-toed style, the clown struts, and the skeleton on the right walks with knocked knees. Fig. 2(b) shows a side view of the motions; the figures animated using synthesized motions are in the foreground. Fig. 3 shows a stair ascending motion synthesized for one of the individuals. Our algorithm extracted the motion signature from a sample walk from this individual. We then used the extracted motion signature to synthesize the stair-ascending motion for this individual. The motion signature was combined with



Figure 2: Synthesizing 3 styles of walking motions from example motions of ascending stairs in those corresponding styles. (a) Comparing synthesized walking motion data against ground truth (the synthesized data is depicted by the characters without hair), our method captures stylistic differences in motion such as pigeon-toed walking, knocked-knees or strutting. (a) The synthesized motions are depicted by the characters in the foreground and, for comparison, the captured walking motions are depicted by the characters in the background.



Figure 3: A synthesized stair-ascending motion.

general stair ascending parameters which were extracted a priori from our database.

In [11] we presented an animation short that was created using motion data synthesized by our algorithm. The graphical characters shown are modeled and rendered by the *MetaCreations Poser* system.

### 6. Conclusion

We have introduced the notion of decomposing motion data into primitives such as action parameters, temporal (cadence) parameters, and most importantly a motion signature. To achieve such a decomposition, we have proposed an algorithm which is based on a numerical statistical analysis technique called multi-mode analysis. Our algorithm robustly extracts signature parameters from a corpus of motion data spanning multiple subjects performing different types of motions. We have shown that the extracted signatures are useful in the recognition and synthesis of novel motions for animating articulated characters.

#### Acknowledgments

Professor Demetri Terzopoulos provided invaluable guidance during the writing of this paper. Motion data were collected at the Gait Laboratory of the Bloorview MacMillan Medical Centre in Toronto. The data acquisition work was done with the permission of Professor Stephen Naumann, Director of the Rehabilitation Engineering Department and with the helpful assistance of Mr. Alan Morris.

# References

- M. Brand and A. Hertzmann. Style machines. Proc. ACM SIGGRAPH 2000, New Orleans, LA, July 2000, 183–192.
- [2] R. Grzeszczuk, D. Terzopoulos, and G. Hinton. NeuroAnimator: Fast neural network emulation and control of physicsbased models. *Proc. ACM SIGGRAPH* 98, July 1998, 9–20.
- [3] M. Gleischer (ed.) Making motion capture useful. ACM SIG-GRAPH 2001, Course 51, Los Angeles, CA, August, 2001.
- [4] N.B. Howe, M. Leventon and W.T. Freeman. Beyesian reconstruction of 3D human motion from single-camera video. *Advances in Neural Information Processing Systems 12*, MIT Press, 1999, 820–826.
- [5] G. Johansson. Visual motion perception. Scientific American, June 1974, 76–88
- [6] L. de Lathauwer, B. de Moor, J. Vandewalle. A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.*, 21(4):1253–1278. On the best rank-1 and rank-(R<sub>1</sub>, R<sub>2</sub>,...R<sub>N</sub>) approximation of higher-order tensors. *SIAM J. Matrix Anal. Appl.*, 21(4):1324–1342.
- [7] A. Kapteyn, H. Neudecker, and T. Wansbeek. An approach to n-mode component analysis. *Psychometrika*, 51(2):269– 275, June 1986.
- [8] J.R. Magnus and H. Neudecker. Matrix Differential Calculus with Applications in Statistics and Econometrics. Wiley, New York, 1999.
- [9] D.H. Marimont and B.A. Wandell. Linear models of surface and illuminance spectra. J. Optical Society of America, A., 9(11):1905–1913, 1992.
- [10] J.B. Tenenbaum and W.T. Freeman. Separating style and content. Advances in Neural Information Processing Systems 10, MIT Press, 1997, 662–668.
- [11] M.A.O. Vasilescu, Human motion signatures for character animation. ACM SIGGRAPH 2001 Conf. Abstracts and Applications, August, 2001, pg. 200.