A STUDY OF SYNTHESIZING NEW HUMAN MOTIONS FROM SAMPLED MOTIONS USING TENSOR DECOMPOSITION

Rovshan Kalanov *Jieun Cho* *Jun Ohya*

Waseda University, Graduate School of Global Information and Telecommunication Studies
Bldg.27-7, 1-3-10 Nishi-Waseda, Shinjuku-ku, Tokyo, 169-0015, Japan

ABSTRACT

This paper applies an algorithm, based on Tensor Decomposition, to a new synthesis application: by using sampled motions of people of different ages under different emotional states, new motions for other people are synthesized. Human motion is the composite consequence of multiple elements, including the action performed and a motion signature that captures the distinctive pattern of movement of a particular individual. By performing decomposition, based on *N*-mode SVD (singular value decomposition), the algorithm analyzes motion data spanning multiple subjects performing different actions to extract these motion elements. The analysis yields a generative motion model that can synthesize new motions in the distinctive styles of these individuals. The effectiveness of applying the tensor decomposition approach to our purpose was confirmed by synthesizing novel walking motions for a person by using the extracted signature.

1. INTRODUCTION

In analogy with handwritten signatures people have characteristic motion signatures that individualize their movements. These signatures can be extracted from example motions and be used to synthesize new actions for those people in their distinctive styles.

Our study has a goal to synthesize human motion resulting from emotional state (say, person sad or happy walk). We extract a motion signature from a subset of actions for a new individual and synthesize the remainder of the actions using the extracted motion signature. The synthetic motions are then validated by classifying against a database of all the real motions.

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. Recent papers report the use of hidden Markov models [4], neural network learning models [5], three-mode principal component model [6] and tensor approach [1]. Synthesize problem is resolved by using techniques from numerical statistics. The mathematical basis of this approach is a technique known as n-mode analysis, which was first proposed by Tucker [3] and subsequently developed by others [7,8]. This multilinear analysis subsumes as special cases the simple, linear (1-factor) analysis associated with conventional SVD and principal components analysis (PCA), as well as the incrementally more general bilinear (2-factor) analysis that has recently been investigated in the context of computer vision [9]. For our purpose we use the algorithm proposed by M.A.O. Vasilescu [1] to extract human motion signatures individualizing their movements from example motions. M. Vasilescu also showed possibility of synthesizing a simple stair ascending-descending walking motion. In our case we applied the algorithm to motion data, previously stored in our database, of people of different ages walking under different emotional states to synthesize motions for new people.

According to this approach corpus of motion data spanning multiple people and actions is organized as higher-order array or tensor which defines multilinear operators over a set of vector spaces. Then *N-mode SVD* is applied to extract human motion signatures as well as other parameters inherent to human movement by decomposing this tensor. Using these parameters as a generative model to observe motion data for a new subject performing one of these actions we synthesize the remaining actions, which were never before seen, for this new individual.

2. THEORY

The starting point of our derivation of a N-mode SVD will be to consider an appropriate generalization of the link between the column (row) vectors and the left (right) singular vectors of a matrix. To be able to formalize this idea, we define "matrix unfoldings" of a given tensor, i.e., matrix representations of that tensor in which all the column (row, . . .) vectors are stacked one after the other. To avoid confusion, we will stick to one particular ordering of the column (row, . . .) vectors; for order three, these unfolding procedures are visualized in Figure 1.

0-7803-9332-5/05/$20.00 ©2005 IEEE

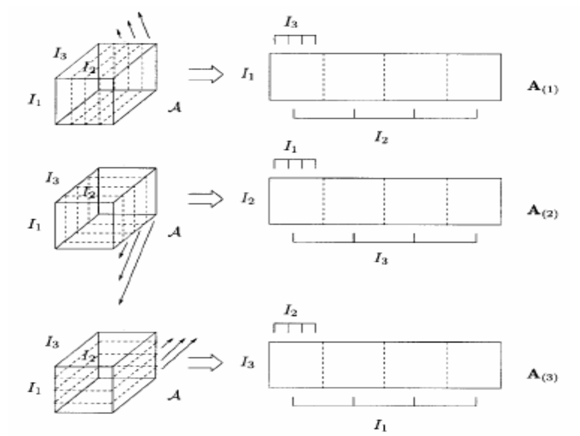


Figure 1. Unfolding of the $(I_1 \times I_2 \times I_3)$-tensor $\mathcal{A}$ to the $(I_1 \times I_2 I_3)$-matrix $\mathbf{A}_{(1)}$, the $(I_2 \times I_3 I_1)$-matrix $\mathbf{A}_{(2)}$ and the $(I_3 \times I_1 I_2)$-matrix $\mathbf{A}_{(3)}$ $(I_1=I_2=I_3=4)$.

SVD model for *N*th-order tensor is proposed as:

Every complex $(I1 \times I2)$-matrix $\mathbf{A}$[1] can be written as the product: $\mathbf{A} = \mathbf{U}_1 \mathbf{S} \mathbf{V}_2^T = \mathbf{S} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2$

in which : $\mathbf{U_1}$ is a unitary $(I_1 \times I_1)$ -matrix, $\mathbf{U_2}$ is a unitary $(I_2 \times I_2)$-matrix, $\mathbf{S}$ is an $(I_1 \times I_2)$-matrix with the properties of *pseudodiagonality* and *ordering* [2].

By extension, an order $N > 2$ tensor is N dimensional matrix comprising N spaces. "*N*-mode SVD" is a "generalization" of SVD that orthogonalizes these N spaces as the mode-*n* product of *N*-orthogonal spaces[2]:

$$\mathcal{A} = \mathcal{S} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \ldots \times_n \mathbf{U}_n \ldots \times_N \mathbf{U}_N \quad \textbf{(1)}$$

as shown in Fig. 2 for the case $N = 3$, where $\mathcal{S}$ is the *core tensor* and $\mathbf{U}$ is mode matrix. The core tensor governs the interaction between the *mode matrices* $\mathbf{U}_n$, for n = 1,…..,N. Mode matrix $\mathbf{U}_n$ contains the orthonormal vectors spanning the column space of the matrix $\mathbf{A}_{(n)}$ that results from the mode-*n* flattening of $\mathcal{A}$, as was depicted in Fig. 1.

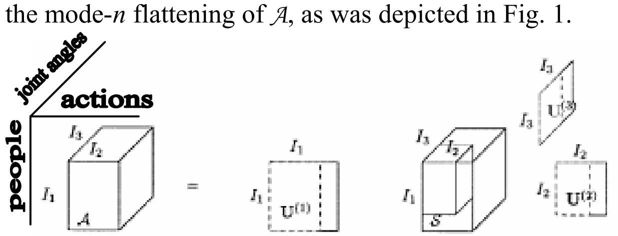


Figure 2. An *N*-mode SVD orthogonalizes the N vector spaces associated with an order-N tensor ($N = 3$).

---

[1] We denote scalars by lower case letters (*a*,*b*..), vectors by bold lower case letters (**a**,**b**..), matrices by bold upper-case letters (**A**,**B**..), high-order tensor by calligraphic upper-case letters ($\mathcal{A}$,$\mathcal{B}$,$\mathcal{C}$..).

[2] A matrix representation of the N-mode SVD can be obtained by:

$$\mathbf{A}_{(n)} = \mathbf{U}_n \mathbf{Z}_{(n)} (\mathbf{U}_{n-1} \otimes \ldots \otimes \mathbf{U}_1 \otimes \mathbf{U}_n \otimes ..\mathbf{U}_{n+2} \otimes \mathbf{U}_{n+1})^T$$

where $\otimes$ is the matrix Kronecker product.

N-mode SVD algorithm for decomposing $\mathcal{A}$ is:

1. For $n = 1,\ldots..,N$, compute matrix $\mathbf{U}_n$ in (1) by computing the SVD of the flattened matrix $\mathbf{A}_{(n)}$ and setting $\mathbf{U}_n$ to be the left matrix of the SVD.
2. Solve for the core tensor as follows

$$\mathcal{S} = \mathcal{A} \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \ldots \times_n \mathbf{U}_n^T \ldots \times_N \mathbf{U}_N^T \quad \textbf{(2)}$$

It can be computed in a matrix format, e.g.,

$$\mathbf{S}_n = \mathbf{U}_n^T \mathbf{A}_n (\mathbf{U}_{n-1} \otimes \mathbf{U}_{n-2} .. \otimes \mathbf{U}_n \otimes .. \otimes \mathbf{U}_{n+2} \otimes \mathbf{U}_{n+1})^T$$

Suppose given motion sequences of several people, we define a data set tensor $\mathcal{D}$ with size $(H \times E \times G)$, where H is the number of people, E is the number of action classes, and G is the number of joint angle time samples. We apply the *N*-mode SVD algorithm to decompose this tensor as follows: $\mathcal{D} = \mathcal{S} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3$

and denoting $\mathbf{U}_1$, $\mathbf{U}_2$, $\mathbf{U}_3$ as $\mathbf{P}$, $\mathbf{A}$, $\mathbf{J}$ respectively we get :

$$\mathcal{D} = \mathcal{S} \times_1 \mathbf{P} \times_2 \mathbf{A} \times_3 \mathbf{J} \quad \textbf{(3)}$$

The people matrix $\mathbf{P} = [\mathrm{p}_1 \ldots \mathrm{p}_n \ \ldots.\ \mathrm{P}_H]^T$ whose person specific row vectors $\mathbf{p}_n^{\mathrm{T}}$ span the space of person parameters, encodes the per-person invariances across actions. Thus $\mathbf{P}$ contains the human motion signatures. The action matrix $\mathbf{A} = [\mathrm{a}_1 \ \ldots.\ \mathrm{a}_m \ \ldots.\ \mathrm{A}_E]^T$, whose action specific row vectors $\mathbf{a}_n^T$ span the space of action parameters, encodes the invariances for each action across different people. The joint angle matrix $\mathbf{J}$ whose row vectors span the space of joint angles are the *eigenmotions* that are normally computed by PCA.

The tensor

$$\mathcal{B} = \mathcal{S} \times_2 \mathbf{A} \times_3 \mathbf{J} \quad \textbf{(4)}$$

contains a set of basis matrices for all the motions associated with particular actions. The tensor

$$\mathcal{C} = \mathcal{S} \times_1 \mathbf{P} \times_3 \mathbf{J} \quad \textbf{(5)}$$

contains a set of basis matrices for all the motions associated with particular people.

After extracting $\mathcal{S}$, $\mathbf{A}$ and $\mathbf{J}$ we have a *generative model* that can observe motion data $\mathcal{D}_{p,a}$ of a new person performing one of these actions (action a) and synthesize the remaining actions for this new person. The signature $\mathbf{p}$ for the new person is solved in the equation $\mathcal{D}_{p,a} = \mathcal{B}_a \times_1 \mathbf{p}^T$, where $\mathcal{B}_a = \mathcal{S} \times_2 \boldsymbol{a}_a^T \times_3 \mathbf{J}$– selected from $\mathcal{B}$ associated with the action of the new person and $\mathcal{D}_{p,a}$ is $1 \times 1 \times T$ tensor. Flattening this tensor in the people mode yields the matrix $\mathbf{D}_{p,a(\text{people})}$, actually a row vector which can be denoted as $\mathbf{d}_a^T$.

Therefore, in terms of flattened tensors, the above equation can be written as $\mathbf{d}_a^T = \mathbf{p}^T \mathbf{B}_{a\,(\text{people})}$

A complete set of motions for the new individual is synthesized as follows:

$$\mathcal{D}_p = \mathcal{B} \times_1 \mathbf{p}^T \quad \textbf{(6)}$$

where $\mathcal{B}$ is defined in (4) and the motion signature for the new individual is given by

$$\mathbf{p}^T = \mathbf{d}_a^T \times \mathbf{B}_{a\,(\text{people})}^{-1} \qquad \mathbf{(7)}$$

If several actions $\mathbf{d}_{ak}$ are observed, the motion signature is computed as follows:

$$\mathbf{p}^T = \mathbf{d}_{ak}^T \times \mathbf{B}_{ak\,(\text{people})}^{-1} \qquad \mathbf{(8)}$$

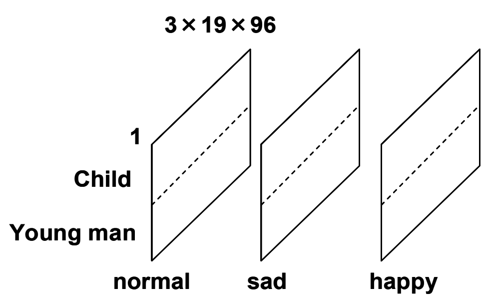
If we observe a known person (one who is already recorded in the motion database) performing a new type of action $\mathbf{d}_p$, we can compute the associated action parameters $\mathbf{a}^T = \mathbf{d_p}^T \mathbf{C}_{p(\text{action})}{}^{-1}$ and use them to synthesize that new action for all the people in the database as follows: $\mathcal{D}_a = \mathcal{C} \times_2 \mathbf{a}^T$, where $\mathcal{C}$ is given in (5). If several different people are observed performing the same new action $\mathbf{d}_{pk}$, the action parameters are computed as follows:

$$\mathbf{a}^T = \mathbf{d}_{pk}^T \times \mathbf{C}_{pk\,(\text{action})}^{-1}$$

Comparing this multilinear technique to conventional PCA, the latter represents each person as a set of E coefficient vectors, one for each action class, while multilinear analysis enables us to represent each person with single vector of coefficients relative to the bases comprising the tensor $\mathcal{B}$ in (4).

3. EXPERIMENT

The corpus of motion data (BVH files) was recorded by using VICON system from 2 different subjects, child and young man. The data was reduced to time-varying joint angles (19 markers, XYZ) for complete cycles (96 frames) of three types of motions: normal, sad and happy walks and was arranged as shown below.



All computations and manipulations are done by using Matlab 6.1.

Hence we have a 2×3×5472(19*3*96) tensor $\mathcal{D}$ or multidimensional array **D** in Matlab.

We flatten the tensor in 3 ways and get 3 matrices – $\mathbf{D}_1$, $\mathbf{D}_2$ and **D3** as:

D1 = reshape (permute (**D**, [1 3 2]), 2, 3*19*96*3);
D2 = reshape (permute (**D**, [3 2 1]), 3, 3*19*96*2);
D3 = reshape (permute (**D**, [2 1 3]), 5472, 6);

Applying SVD to these matrixes we get **P**, **A** and **J** mode matrices respectively.

[P, S, V] = svds(D1), **[A, S,V] = svds(D2)**,
[J, S, V] = svds(D3)

where **P** with size 2×2, **A** - 3×3 and **J** - 5472×6.

Then we compute core tensor $\mathcal{S}$ using matrix representation: $\mathbf{S} = \mathbf{P}^T \mathbf{D}_1 (\mathbf{A} \otimes \mathbf{J})$

In Matlab as: **S = P' * D1 * kron(A, J)** with size 2×18.

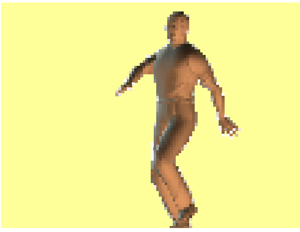
Then we calculate tensor $\mathcal{B}$ in (4), and Matlab representation is $\mathbf{B} = \mathbf{S} * \mathbf{kron}(\mathbf{U}_2, \mathbf{U}_3)$ with size 2×16416.

On the next step we synthesize motion for a new person (old man), for whom we have only normal walking motion, 2 remaining actions (sad and happy walking). We solve signature **p** for the new person in (7), where $\mathbf{d}_a$ is a 1×5472 (motion data for a new person) and $\mathbf{B}_{a(\text{people})}$ – is a 2×5472 matrix as: $\mathbf{p} = \mathbf{d_a} * \mathbf{pinv}(\mathbf{B_a})$ with size 1×2, where $\mathbf{pinv}(\mathbf{B_a})$ returns the Moore-Penrose pseudoinverse of $\mathbf{B_a}$. If $\mathbf{B_a}$ is square and not singular, then $\mathbf{pinv}(\mathbf{B_a})$ is an expensive way to compute $\mathbf{inv}(\mathbf{B_a})$. If $\mathbf{B_a}$ is not square, or is square and singular, then $\mathbf{inv}(\mathbf{B_a})$ does not exist. In these cases, $\mathbf{pinv}(\mathbf{B_a})$ has some of, but not all, the properties of $\mathbf{inv}(\mathbf{B_a})$.

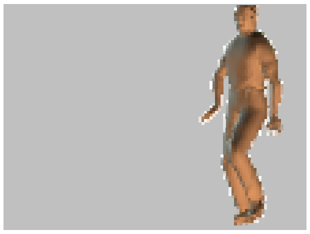
Then complete set of motions $\mathbf{D_p}$ for the new person is synthesized in (6) as $\mathbf{D}_\mathrm{p} = \mathbf{p} * \mathbf{B}$ with size 1×16416. Then it should be retensorized to form motion data for sad and happy walk as 1× 2 × 5472.

4. RESULTS

All matrix calculations and manipulations were done by using Matlab 6.1. The synthesized motion for sad and happy walk was modeled and rendered by Poser 5. The experiment showed positive result. Fig. 1, Fig. 2 and Fig. 3(a) show frames of motion data that was used to synthesize motion for a new individual (old man). These frames were taken by right-side camera. Each pose was taken when left leg was in the highest position from the ground to see differences between their walk. Motion signature of the old man was derived from his normal walk and general parameters of sad and happy walks were derived from child and young man walks. They were combined to synthesize sad and happy walk of old man Fig. 3 (a,b). Note that all sad walks differ subtly, they are slow with hands down. Happy walks differ more, and steps as it can be seen for happy walk of old man is some kind of combination of walking style of child and young man plus his own motion signature captured from his normal walk. The synthetic motions are then validated by classifying them against a database of all the real motions. We verified that the 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 real motion capture data of that individual.



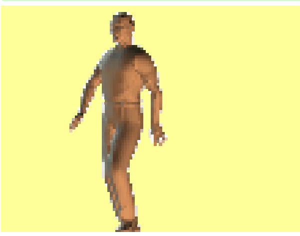
(a)



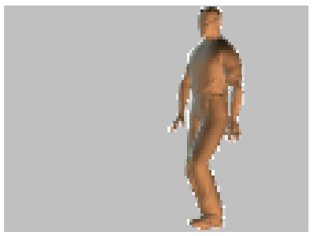
(b)

Fig. 1 (normal, sad and happy child walk)

(c)



(a)



(b)

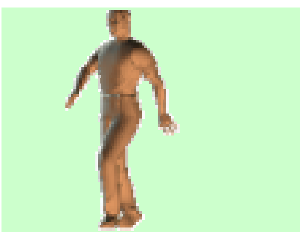
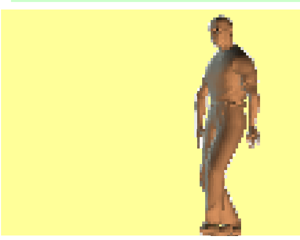
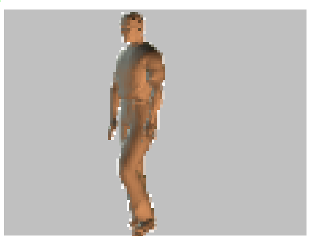


Fig. 2 (normal, sad and happy young man walk)

(c)



(a)



(b)

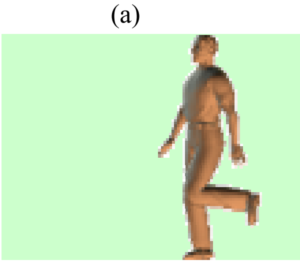


Fig. 3 (normal, sad and happy old man walk)

(c)

5. CONCLUSION

We have shown how motion data can be decomposed into primitives such as action parameters, and a motion signature. To achieve such decomposition we used an algorithm which is based on a numerical statistical analysis technique called n-mode analysis. It takes advantage of multilinear algebra in which motion data ensembles are represented as higher-dimensional tensors and an "N-mode SVD" algorithm is applied to decompose the tensor. This multilinear approach accommodates any number of factors by taking advantage of the mathematical machinery of tensors. It 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 for the synthesis of novel motions for animating articulated characters for motion recognition. Further study includes more different kinds of motions and motion recognition.

ACKNOWLEDGMENTS

We very appreciate valuable guidance M.A.O. Vasilescu gave us to overcome difficulties with matrix calculation.
Motion data was collected at the Waseda Unversity in Tokyo.

REFERENCES

[1] M. Alex O. Vasilescu, Human Motion Signatures: Analysis, Synthesis, Recognition. International Conference on Pattern Recognition (ICPR'02).

[2] L. de Lathauwer, B. de Moor, J. Vandewalle. A multilinear singular value decomposition. SIAM J. Matrix Anal. Appl., 21(4):1253–1278.

[3] L. R. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31:279–311, 1966.

[4] M. Brand and A. Hertzmann. Style machines. *Proc. ACM SIGGRAPH 2000*, New Orleans, LA, July 2000, 183–192.

[5] 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.

[6] James W. Davis, and Hui Gao. An expressive three-mode principal components model for gender recognition. Journal of Vision (2004) 4, 362-377.

[7] A. Kapteyn, H. Neudecker, and T. Wansbeek. An approachto 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] J.B. Tenenbaum and W.T. Freeman. Separating style and content. *Advances in Neural Information Processing Systems 10*, MIT Press, 1997, 662–668.

[10] M.A.O. Vasilescu. An algorithm for extracting human motion signatures. In *Proc. Tech. Sketches, IEEE Conf. Computer Vision and Pattern Recognition*, Kauai, HI, Dec., 2001.