K-087

Study of recognizing an unknown person action from motion capture data based on tensor decomposition

Kalanov Rovshan Jun Ohya Hiroyuki Ogata Hajime Takada
Waseda University Seikei University Yokohama National University

Abstract

This paper applies an algorithm, based on Tensor Decomposition, to human action recognition application: by using motion database of people producing different actions, to recognize the action depicted in motion data generated by an unknown person. 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 the core tensors that govern the interaction between the *mode matrices*. The analysis yields a generative motion model that can recognize distinctive motion style of unknown person. The effectiveness of applying the tensor decomposition approach to our purpose was confirmed by recognizing novel walking motions for a person by using motion database.

1. Introduction

At present human motion analysis are actively researched topics in computer vision area. The goal of this paper is to propose an algorithm to recognize a motion of unknown person different from the people

Several authors have introduced generative motion models for this purpose. Recent papers report the use of hidden Markov model [2], a two-mode PCA framework, a Fourier-based approach [1] and a tensor approach [4,5].

The latest approach, proposed by M.A. Vasilescu [4,5], is directly connected with our research and based on tensor decomposition and deals with human motion synthesis and recognition. In that work authors show how to recognize a unknown person from motion data of a known action and to recognize the action depicted in motion data generated by known person.

Unlike this work in our research we tried to apply tensor decomposition approach to recognize an unknown's person motion from motion database of other people. In this paper we have proposed the algorithm of human emotional motion recognition.

The mathematical basis of tensor approach is a numerical statistical technique known as n-mode analysis, which was first proposed by Tucker [3] and subsequently developed by others. It takes advantages of multilinear algebra in which motion data ensembles are represented as higher-order arrays or tensors and an N-mode SVD algorithm is applied to decompose the tensor.

2. Tensor and Decomposition

The starting point of our derivation of a N-mode SVD, in our case N=3, 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 Fig. 1.

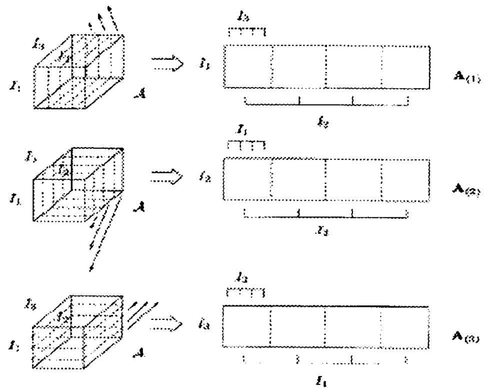

Figure 1. Unfolding of the $(I_1 \times I_2 \times I_3)$-tensor A to the $(I_1 \times I_2 I_3)$-matrix $\mathbf{A}_{(1)}$, the $(I_2 \times I_3 I_1)$-matrix $\mathbf{A}_{(2)}$ and the $(I_3 \times I_1 I_2)$-matrix $\mathbf{A}_{(3)}$ $(I_1{=}I_2{=}I_3{=}4)$.

SVD model for 3th-order tensor is proposed as:
Every complex (I1 × I2)-matrix A1 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, **U2** is a unitary $(I_2 \times I_2)$-matrix, **S** is an $(I_1 \times I_2)$-matrix with the properties of *pseudodiagonality* and *ordering* [2].

An order $N = 3$ tensor is 3 dimensional matrix comprising 3 spaces. "3-mode SVD" is a "generalization" of SVD that orthogonalizes these 3 spaces as the mode-3 product of 3-orthogonal spaces[2]:

[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 (*A,B,C*..).

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

$$\mathcal{A} = \mathcal{S} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times \mathbf{U}_3 \quad \textbf{(1)}$$

where S is the core tensor and U is mode matrix. The core tensor governs the interaction between the mode matrices Un, for n = 1,….,N. Mode matrix Un contains the orthonormal vectors spanning the column space of the matrix A(n) that results from the mode-n flattening of A, as was depicted in Fig. 1.

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

1. For n = 1,2,3 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

$$\mathcal{S} = \mathcal{A} \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \times_3 \mathbf{U}_3^T \qquad \textbf{(2)}$$

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

$$\mathbf{S} = \mathbf{U}_1^T \mathbf{A}_1 (\mathbf{U}_2 \otimes \mathbf{U}_2)^T$$

3. Recognition

Suppose D is $M \times N \times T$ tensor of motion data, where M (rows) is the number of people, N (columns) is number of actions and T (depth) is motion data size for each person (joint angles × number of frames), $\mathbf{D}_1$, $\mathbf{D}_2$ and $\mathbf{D}_3$ are flattened matrixes, Z is core tensor and $D_{p,a}$ is a $1 \times 1 \times T$ tensor of motion data for a new person. The algorithm of recognition as follows:

1. Compute SVD of the flattened matrix $\mathbf{D}_1$, $\mathbf{D2}_2$, $\mathbf{D}_3$
2. Solve for the core tensor Z
3. For each action data D_{ij}, where $i \in 1 \ldots M$ and $j \in N$, in tensor D is substituted by $D_{p,a}$ calculating core tensor Z_{new} each time, and therefore having MN tensors.
4. Tensor $Znew$ is calculated MN times (steps 2, 3) to find the least sum of element wise differences E_{ij} between tensor Z and new tensor $Znew$ as :

$$i^*, j^* = \arg\min_{i,j} \sum_{i=1}^{M} \sum_{j=1}^{N} \left\| \frac{Z - Znew}{Z} \right\|$$

4. Experiment results

For our experiments we used a magnetic sensor based motion capture system. In our experiment 19 makers were attached to a human body and corpus of motion data was collected from 10 persons walking under different emotional states – normal, sad and happy walk. Then motion data was converted to BVH format files for each person to get relative orientation of each limb. All matrix calculations and manipulations were done by using Matlab 6.1 and modeled and rendered by Poser 5.

To estimate effectiveness of our algorithm we performed a "leave-one-out" validation study and got 60 % of recognition. It is not so high due to some factors.

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.

The results show these factors could be variations in walk style, walk speed, body poise, noise in motion data and etc. For example a person's individual walk style differs from person to person for the same emotional state or does not differ for different emotional states. The more distinctive motions tend to give more accurate result.

For example person number 10 with normal walk in our experiment was recognized as person number 3 with normal walk. In Fig. 2 we can see his normal walk.

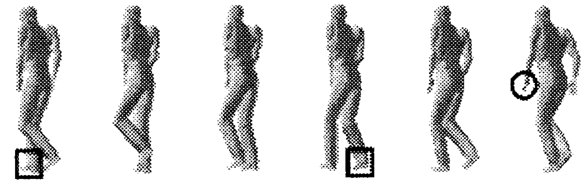

Figure 2. Normal walk of person number 10

The recognized human (num. 3) normal walk is depicted in Fig. 3, the similar parts are shown in boxes and circles.

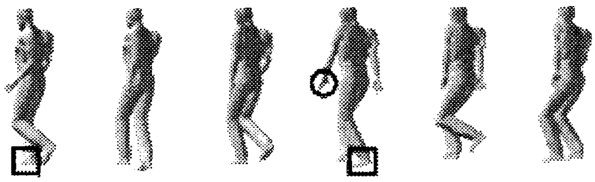

Figure 3. Normal walk of person number 3

5. Conclusion

We have showed how emotional motion - normal, sad and happy walk, can be recognized.

We have introduced the algorithm to human emotional motion recognition which is based on tensor decomposition approach and core tensor ability to govern interaction between matrices in multidimensional space. The remaining issue to be studied in the future is to improve motion recognition algorithm to archive more robust results.

References

1. M. Unuma, K. Anjyo, and R. Takeuchi. Fourier principles for emotion-based human figure animation. In Proc. SIGGRAPH, pages 91–96. ACM, 1995.
2. A.Wilson and A. Bobick. Parametric Hidden Markov Models for gesture recognition. IEEE Trans. Patt. Analy.
3. L. R. Tucker. Some mathematical notes on three-mode factor analysis. Psychometrika, 31:279–311, 1966.
4. M. Alex O. Vasilescu, Human Motion Signatures: Analysis, Synthesis, Recognition. International Conference on Pattern Recognition (ICPR'02).
5. M. Vasilescu. Human motion signatures for character animation. In SIGGRAPH Conference Abstracts and Applications, page 200. ACM, 2001.