H-039

Motion Beat Induction by Short-Term Principal Component Analysis

Jianfeng Xu† Koichi Takagi† Akio Yoneyama†

1． Introduction

This paper presents a novel tool called Short-Term Principal Component Analysis (ST-PCA) to analyze human motion, which records realistic movements in a high dimensional time series and is widely used in animation, game, and movie [1]. Our ST-PCA is successfully applied to beat induction, which is an important perception of human motion especially in dances and is required by many applications such as synchronization with music [2,3]. Following Kim et al. [2], *motion beats* are defined as the regular moments when the movement is changed significantly in direction or magnitude. Different from the previous approaches [2,3] that analyze motion signals in each channel, we estimate the motion beats by our ST-PCA that utilizes motion signals as a whole. Our experiments demonstrate that such an ST-PCA can estimate much more accurate motion beats in a wide range of motion categories including complicated dances than the state-of-the-art alternatives.

2． Beat Induction Approach

We observe that the major movement variance reveals motion beats well, where PCA is potentially suitable as the most popular dimension reduction method [4]. By PCA, the raw data are transformed into an uncorrelated space with a descending variability, making it possible to discard most of the dimensions and extract the major variance from the data set. However, PCA does not consider temporal coherence among the data samples in human motion. Moreover, the conventional global PCA does not work well in a complicated motion [5]. The proposed ST-PCA, on the other hand, is demonstrated to be effective even in a complicated motion such as dances because our method utilizes effectively the three characteristics in human motion: hierarchical structure, spatial correlation, and temporal coherence. Our basic idea is similar to piece-wise linear approximation to a non-linear problem. PCA is a linear model while a complicated motion is highly non-linear [5]. However, in the short term, motion data are almost linear due to the strong temporal coherence [6], which forms the base of our ST-PCA algorithm.

The proposed approach works in three stages. (1) **Conversion from angles to positions:** The raw data (i.e., Euler angles) are non-linear in the sense of the least square terms, which is employed in PCA. Moreover, Euler angles have difficulty in reflecting spatial correlation in human motion. Therefore, using the hierarchical structure, we transform Euler angles to joint positions by forward kinematics. Because the root translation is the major variance but contains little beat information, the related joint positions to the root joint are employed. We have observed that the related joint positions have more regular signals in PC space than Euler angles.

†KDDI R&D Laboratories Inc.

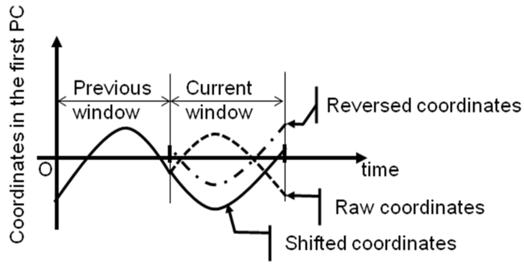

Fig. 1 The coordinates of the first PC in the current window are connected smoothly with those in the previous window by reversal and shifting operations.

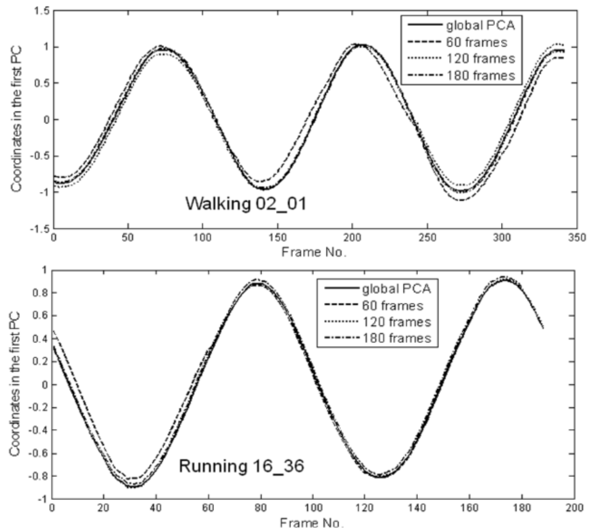

Fig. 2 The coordinates of the first PC in Walking 02_01 (top) and Running 16_36 (bottom), calculated by different window lengths (see the legend).

(2) **Dimension reduction by ST-PCA**: Perform PCA in a sliding window and connect properly the coordinates in the first PC. Firstly, we divide the entire motion into small windows with a *fixed* length (e.g. 0.5 seconds or 60 frames in our implement) and perform PCA in each window as Eq. (1), where we only select the coordinates in the first PC because they reveal the major movement variance.

$$\mathbf{D} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^T \quad (1)$$

where the matrix $\mathbf{D}$ denotes the related joint positions in a window after removing the average value, columns of $\mathbf{U}$ and $\mathbf{V}$ are orthogonal unit vectors, and $\boldsymbol{\Lambda}$ is a diagonal square matrix with eigenvalues. Columns of $\mathbf{V}$ form the bases of transformation to principal component space. Thus, the coordinates $\mathbf{y_1}$ in the first PC can be obtained by Eq. (2).

$$\mathbf{y_1} = (\mathbf{D}\mathbf{v_1})^T \quad (2)$$

where $\mathbf{v_1}$ denotes the first column in $\mathbf{V}$ or the first basis.

Secondly, a connection operation is required as discontinuity happens due to the independent operation of ST-PCA in each window, see Fig. 1. We reverse the basis $\mathbf{v_1}$ and its coordinates $\mathbf{y_1}$ of the first PC in the current window if the basis has a negative

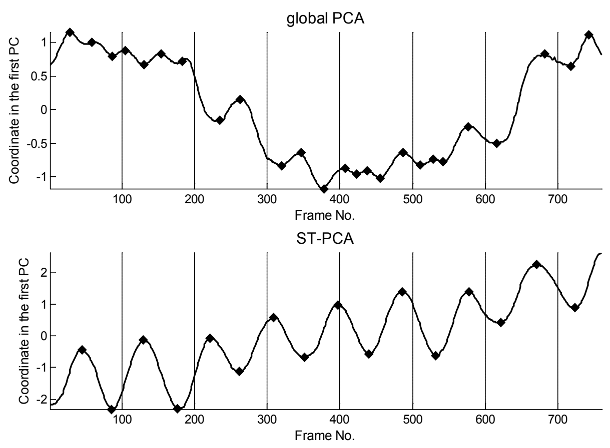

Fig. 3 The coordinates of the first PC in a running motion on a circle path (38_03), calculated by global PCA (top) and ST-PCA (bottom). ST-PCA extracts local movement variance successfully while global PCA fails.

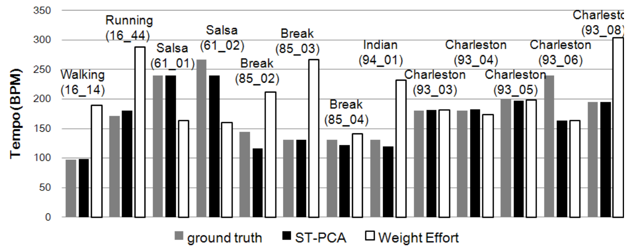

Fig. 4 Estimated tempi in walking, running, and some dancing motions including Salsa, Break, Indian, and Charleston.

inner product with the corresponding basis $\mathbf{v}_1^{\mathbf{pre}}$ in the previous window as Eq. (3). Also, we shift the coordinates in the current window to connect smoothly those in the previous window.

$$\mathbf{v}_1 \cdot \mathbf{v}_1^{\mathbf{pre}} < 0 \quad (3)$$

(3) **Beat Induction**: Extract motion beats and refine them. From the connected coordinates of the first PC, we regard the peaks and nadirs as our coarse beats, see the diamonds in Fig. 3 bottom. Assume the beat intervals are invariable in a motion, which is reasonable if the motion is not so long (e.g. less than 30 seconds) [2]. Then, we regularize the beat intervals by their maximal frequency as Kim et al. [2]. Finally, the estimated tempo is calculated from the beat interval *TB* and the frame rate *Fs* (120 frames per second in CMU database [7]) as Eq. (4).

$$tempo = 60Fs/TB \quad (4)$$

3. Simulation Results

We test our approach on the CMU Motion Capture database [7]. In a simple motion such as Walking 02_01 [1] or Running 16_36, ST-PCA achieves almost the same results as the global PCA, see Fig. 2, no matter how long the window is. Therefore, it is not necessary to know the boundary of a single motion in the case of a complicated motion if the window length is shorter than the single motion, allowing a fixed-length window effective.

On the other hand, our ST-PCA extracts successfully the local major movement variance while the global PCA fails in a complicated motion. Figure 3 shows the results in a running motion on a circle path (38_03), which demonstrates that ST-PCA achieves a much more regular curve that reveals accurately the motion cycles in the motion 38_03.

For the computational complexity, ST-PCA is the same as the global PCA except the connection operation because it is linear to the total frame number. Moreover, ST-PCA is obviously more suitable for parallel computing and consumes much less memory than the global PCA.

Using the ground truth derived from human subjects, we compare the estimated tempi with Shiratori et al. [3], where *Weight Effort* features in Laban's theory are employed in each channel. In our experiments of 13 motions, see Fig. 4, the average of relative errors from the ground truth is much smaller in our method (6.67% vs. 43.3%).

However, our method has limitations in those motions where the major movement variance does not reveal beat information. For example, in Charleston 93_06, the body rotation becomes the major movement variance while the leg movements reveal the motion beats, where our ST-PCA cannot extract correct motion beats. Although such cases are rarely found, we observe that the coordinates in the second PC can solve the problem.

4. Conclusions

In this paper, our main contribution is to propose a novel ST-PCA method to extract motion beats, considering the hierarchical structure, spatial correlation, and temporal coherence in human motion. Our algorithm improves the robustness and accuracy with less parameter by utilizing motion signals as a whole. Moreover, it is also possible to apply ST-PCA in other applications such as key pose extraction and motion compression.

As our future work, we would like to develop a music synchronization system based on our beat induction algorithm, which can automatically generate a new motion from a motion database that is synchronized with a piece of input music.

[1] Motion ID is compatible to CMU database [7].

References

[1] C. Bregler, “Motion capture technology for entertainment,” *IEEE Signal Processing Magazine*, vol.24, no.6, pp.160-158, Nov. 2007.

[2] T. H. Kim, S. I. Park, and S. Y. Shin, “Rhythmic-motion synthesis based on motion-beat analysis,” *ACM Transactions on Graphics*, Vol. 22, No. 3, pp. 392-401, July 2003.

[3] T. Shiratori, A. Nakazawa, and K. Ikeuchi, “Dancing-to-music character animation,” *Computer Graphics Forum*, Vol. 25, No. 3, pp. 449-458, 2006.

[4] R. Urtasun, P. Glardon, R. Boulic, D. Thalmann, and P. Fua, “Style-based motion synthesis,” *Computer Graphics Forum*, vol.23, no.4, pp.799-812, 2004.

[5] H. J. Shin and J. Lee, “Motion synthesis and editing in low dimensional spaces,” *Computer Animation and Virtual Worlds*, vol.17, no.3-4, pp.219-227, 2006.

[6] J. Xu, T. Yamasaki, and K. Aizawa, “Summarization of 3D video by rate-distortion trade-off,” *IEICE Trans. Inf. & Syst.*, vol.E90-D, no.9, pp.1430-1438, 2007.

[7] CMU Motion Capture Database, http://mocap.cs.cmu.edu/.