Detecting Dance Motion Structure through Music Analysis

Takaaki Shiratori† Atsushi Nakazawa‡ Katsushi Ikeuchi†

† Institute of Industrial Science
The University of Tokyo
4-6-1 Komaba, Meguro-ku,
Tokyo 153-8505, Japan
{siratori, ki}@cvl.iis.u-tokyo.ac.jp

‡ Cybermedia Center
Osaka University
1-32 Machikaneyama, Toyonaka,
Osaka 560-0043, Japan
nakazawa@ime.cmc.osaka-u.ac.jp

Abstract

In these days, many important intangible cultural properties of the world are being lost because of the lack of successive performers. Digital archiving technology is one of the effective solutions for this issue, and we have started our digital archiving project of cultural properties including these intangible ones. For these human motion archives, the method of automatic motion structure analysis is vital for a variety of purposes. We believe that the dance motion consists of "primitive motions" and that motion analysis is necessary to detect these components. Particularly for dance motions, we think these primitives must be synchronized to the musical rhythm. In this paper, we introduce musical information for motion structure analysis. This method automatically detects the musical rhythm and segments the original motion, and classifies them as to the primitive motions. The experimental results confirm that our motion analysis yielded the primitive motions in accordance to the musical rhythm.

1 Introduction

Intangible cultural heritages such as folk dancing have traditionally been taught from one person to another. Unfortunately, due to the decreasing number of performers, the number of intangible heritages is also decreasing. In order to conserve the intangible heritages, we have developed a digital archiving method for human motions.

A motion capture system is one of the most effective methods for digitizing human motions and computer graphics applications. But the system provides only low–level information. More variable motions can be generated by "motion graph"[1][7][8], using methods of signal processing[2] or using spacetime constraints[4][9].

We have developed a method to convert dance motions to symbolic descriptions like music scores[10]. The symbolization made it easier to reconstruct or to edit dance motions. However, the captured raw data includes individual differences or gender differences, and raw motion data is not always appropriate for symbolization of dance motion. On this symbolization step, we introduced an assumption of "Primitive Motion." According to Flash and Hogan[3], every motion is represented as a sequence of primitive motions. Particularly for dance motion, the stop motions are key poses for extracting primitive motions. Tae-hoon et al.[6] proposed an analysis method that estimates rhythm of motion to extract primitive motions. They assumed that the intervals of segmented motions are the same. We, however, think that most key poses in dance motions appear at the timing of some rhythmic pattern based on musical rhythm, because generally people dance to the music. So considering musical rhythm makes it possible to refine the segmentation results(Figure 1).

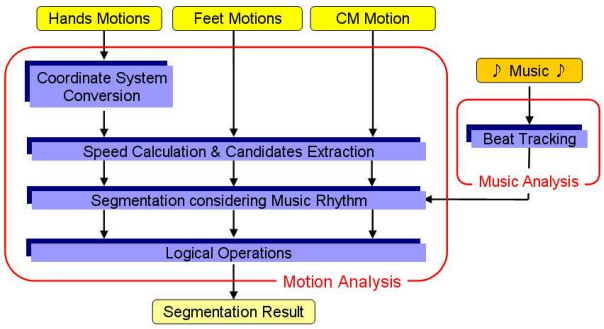

Figure 1: Algorithm Overview

In this paper, we propose a method to detect key poses and to segment dance motions through beat tracking from music data for extracting primitive motions. We detect the musical rhythm by the beat tracking method based on [5] in Section 2. The segmentation method for extracting the primitive motions is described in Section 3. In Section 4, we describe how our method was verified and compare it with another method. The paper is concluded in Section 5.

2 Beat Tracking Method

We have developed the beat tracking system based on [5], which was used for detecting rhythm in western music.

2.1 Musical Elements for Beat Tracking

Experiments in [5] showed that there are three musical elements for beat tracking of popular music; "onset times", "chord changes" and "percussion patterns". However, it is not appropriate to use chord changes and percussion patterns for beat tracking of Japanese dance music, because the frequency characteristics of chord and percussion sounds in western music are different from those in Japanese dance music. So, we conclude that the rhythm in Japanese dance music can be effectively extracted by onset times.

2.2 Onset Component Extraction

When a sound is produced, the spectral power corresponding to the frequency of that sound will increase. Each onset component is extracted as shown in Figure 2. The magnitude of increasing spectral power $d(t, f)$ at time t for frequency f is defined as:

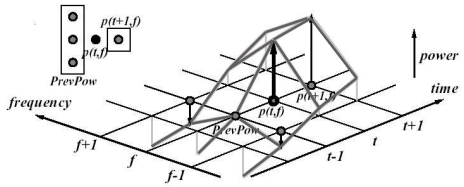

Figure 2: Onset Component Extraction: Calculating how much power increases from "PrevPow"

$$d(t,f) = \begin{cases} \max(p(t,f), p(t+1,f)) - \mathrm{PrevPow} \\ \quad (\min(p(t,f), p(t+1,f)) \geq \mathrm{PrevPow}), \\ 0 \quad (\text{otherwise}) \end{cases} \tag{1}$$

where

$$\mathrm{PrevPow} = \max(p(t-1,f), p(t-1, f \pm 1)), \tag{2}$$

and $p(t, f)$ is the spectral power at time t and frequency f.

PrevPow considers the spectral power between $f-1$, f and $f+1$ at time $t-1$. Therefore, an error onset time is not picked up when the musical frequency is deviated by noise, players' skill, etc.

Next, at the frequency bands a user selected, beat tracking system calculates onset component $D(t)$ defined as follows:

$$D(t) = \sum_{\mathrm{ChosenFreqBands}} d(t,f), \tag{3}$$

Frequency range is divided into 7 bands(0-125Hz, 125-250Hz, 250-500Hz, 500-1000Hz, 1k-2kHz, 2k-4kHz and more than 4kHz). Each band corresponds to one octave as sensed by humans.

2.3 Estimation of Beat Start and Average Beat Interval

In this section, we describe how we estimate beat start and average beat interval.

First, the auto-correlation function($R_{DD}(\tau)$) of $D(t)$ is calculated. Auto-correlation function indicates the periodicity of $D(t)$ and is defined as follows:

$$R_{DD}(\tau) = \frac{1}{T}\sum_{t=1}^{T} D(t)D(t+\tau), \tag{4}$$

where T is the number of the data.

In $R_{DD}(\tau)$, if a peak appears at $\tau_{\max}$, $D(t)$ is very similar to $D(t+\tau_{\max})$ and $\tau_{\max}$ is an average beat interval(τ_{rhythm}).

Then, cross-correlation function($R_{DP}(\tau)$) of $D(t)$ and $P(t)$, whose pulse interval is τ_{rhythm}, is calculated:

$$R_{DP}(\tau) = \frac{1}{T}\sum_{t=1}^{T} D(t)P(t+\tau). \tag{5}$$

The time when a peak of R_{DP} appears is beat start(t_{st}).

2.4 Beat Tracking

Even if music sounds as if its rhythm has not changed, rhythm sometimes changes slightly because of players' sense, etc. and errors caused by these slight rhythm changes make beat tracking impossible. This problem is solved by the following steps:

1. For timing of musical rhythm t_i, the system finds a local maximum of onset component around $t_i + t_{\mathrm{rhythm}}(t_0 = t_{\mathrm{st}})$. Then, the next timing of musical rhythm t_{i+1} is time of the local maximum of onset component.

2. For t_{i+1}, 1st process is done and repeated until music ends.

3 Motion Analysis Considering Musical Rhythm

Keyframes are important for recognizing whole motion sequences and are used for motion editing. According to Flash and Hogan[3], the end effectors' movements consist

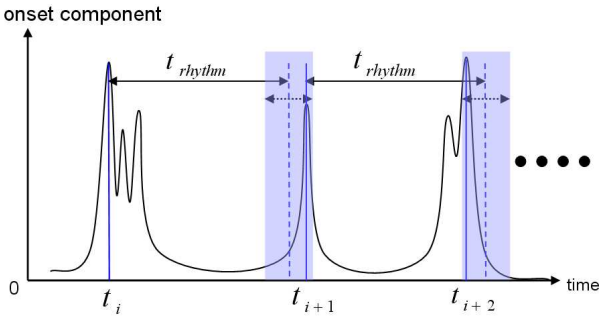

Figure 3: Beat Tracking

of the motion segments and they are divided by their velocity change. The motion segments are called "primitive motions." Using this knowledge, we developed an algorithm for segmenting the original motion. Particularly for dance motions, the keyframe is when a performer stops moving and this stop point often appears at musical rhythm, because the intervals of the primitive motions are based on the musical rhythm and correspond to the rhythmic pattern underlying the motion. So considering musical rhythm refines the segmentation results efficiently.

The overall analysis method for human dance motions is presented in this section.

3.1 Motion Elements for Analysis

Our analysis method is based on the speed of a performer's hands, feet and center of mass(CM). In most dances including Japanese traditional dance, movement of hands and feet is related with the whole body expression. Therefore, the speed of the hands and feet is useful for extracting the stop motions. However, it is not sufficient for primitive extraction, because sometimes the dancer loses their sense of rhythm, or dances are varied by the preference or gender of performers, etc.

In addition to the motion of the hands and feet, our algorithm uses the motion of the body's CM. The motion of CM represents the motion of the whole body; thus, the effects of misstep and individual difference are less. Primitives can then be effectively extracted.

3.2 Segmentation of Dance Motions considering Musical Rhythm

There are three steps to segment the motion sequence.

1. Speed calculation
2. Segmentation candidates extraction
3. Refinement of the segmentation with musical rhythm

3.2.1 Speed calculation

The captured motion data is recorded in global coordinate system. To calculate speed of the hands, we define the local coordinate system as follows. The origin of the local coordinate system is the middle of the human waist. The three axes of the local coordinate are shown in Figure 4. x axis is parallel to the waist, y axis is perpendicular to the waist and z axis is vertical. This coordinate system enables the system to understand the relative motion of hands to the entire body.

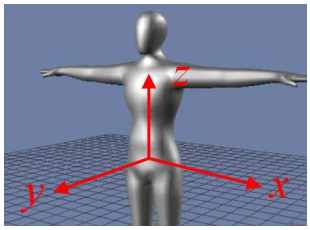

Figure 4: Body Center Coordinate System

On the other hand, speed of feet and the CM is calculated in the global coordinate system. In the global coordinate system, the speeds of feet and CM is nearly 0 when these parts stop. So it is very easy to analyze the motion of these parts.

The effect of noise is reduced by smoothing motion sequence with gaussian filter before extracting candidates.

3.2.2 Segmentation candidates extraction

After calculating speed, the system extracts the candidates for segmentation. The candidates of hands and CM are defined as the local–minimum point which satisfies the following two criteria(Figure 5):

1. Each candidate is a local minimum in speed sequence and the local minimum is less than the minimum–speed threshold.
2. The local maximum between two successive candidates is larger than the maximum–speed threshold.

To extract the candidates of feet, the system extracts rise and decay of the feet speed sequence. Then, the area between between rise and decay, which means how far each foot moves, is calculated. If the area is larger than length–threshold, the rise and decay become candidates(Figure 6).

3.2.3 Refinement of the segmentation with musical rhythm

Finally, system refines the segmentation candidates. At each speed sequence, our method tests whether there are

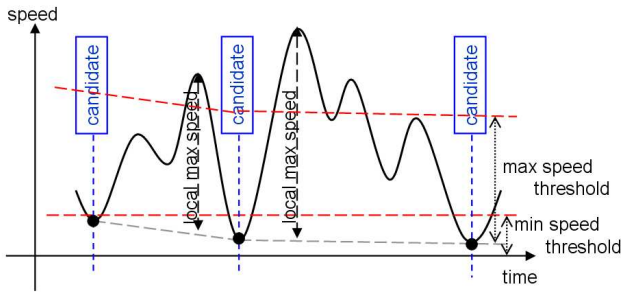

Figure 5: Segmentation Candidates Extraction - Hands and CM

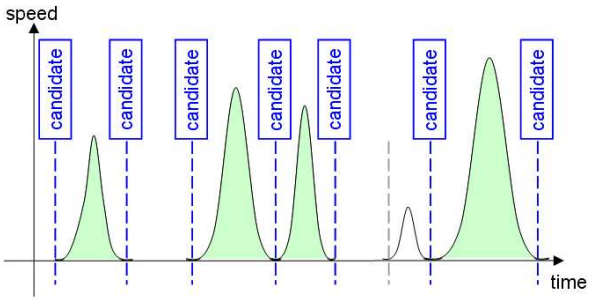

Figure 6: Segmentation Candidates Extraction - Feet

candidates around musical rhythm(t_{beat}) detected by onset components. If there is a candidate, it is possible that there is a stop point around t_{beat} and the motion sequence is segmented at t_{beat}. Figure 7 illustrates the key pose extraction process. In the figure, there are no candidates around 1st and 3rd musical rhythm. So they are not stop motions and each motion sequence is not segmented for extracting primitive motions. On the other hand, because there are candidates around 2nd and 4th musical rhythm, each motion sequence is segmented at these points.

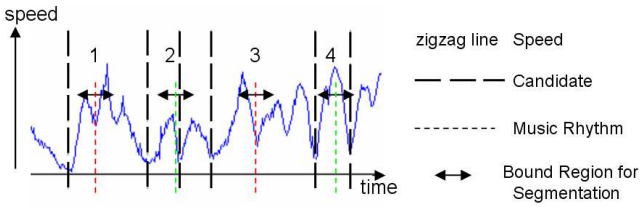

Figure 7: Refinement of the segmentation with musical rhythm

Then, logical operation is used to confirm whether t_{beat} is the stop point of the entire body. The operation checks the harmony of hands, feet and the CM and is defined as:

CM Result AND
(Result that more than 2 of L. Hand Result, R. Hand Result and Feet Result match)

This criterion means that the entire body of the dancer is stopped at timing of musical rhythm it enables to extract stop motions and to segment dance motion.

4 Results

Our proposed method was evaluated using three dance sequence: *Aizu-bandaisan* performanced by a male dancer and a female dancer, and *Jongara-bushi*. These motion data were captured by Vicon Motion Systems, an optical motion capturing system that recorded the position of 33 markers on a person. The sampling rate of *Aizu-bandaisan* dance and *Jongara-bushi* were 120 fps and 200fps, respectively.

The music was converted into the *wav* format by USB Audio device. The data size per 1 sample was 16 bits and the sampling rate was set at 32000Hz.

4.1 Results of Beat Tracking

To extract the onset components, frequency spectrum was calculated by FFT. The number of samples for FFT was 1024 samples. The window function for FFT was gaussian function which is suitable for audio signal processing and was shifted by 256 samples.

The estimated average beat interval of *Aizu-bandaisan* and *Jongara-bushi* dance music were 0.704 seconds = 85M.M.(Mälzel's Metronome : the number of beat in 1 minute) and 0.576 seconds = 104M.M. respectively. In Figure 8, the upper left window shows the spectrogram and vertical lines were the estimated rhythm. The lower left window shows the $D(t)$ of each frequency band described. The rhythm appeared at the onset times, which are represented by the deep gray in the spectrogram.

4.2 Results of Motion Segmentation considering Musical Rhythm

In all dances, we compared the result from our key pose extraction method with the result from the other method. To evaluate our proposed method, we compare the results of our method with the key poses extracted by dancers.

4.2.1 Aizu-bandaisan - A Female Dancer

The result of our method is shown in Figure 9. The segmentation based on the speed of hand and feet, without musical rhythm, oversegmented the motion. This method extracted 8 correct key poses(accuracy:73%) and there were 3 undetected errors and 5 mis-detected errors. On the other hand,

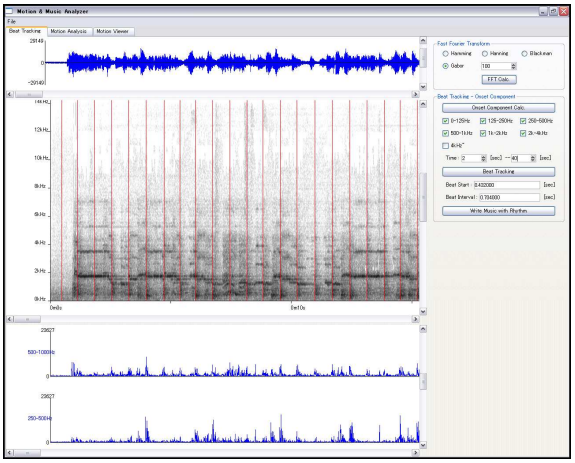

Figure 8: Result - Beat Tracking of Aizu-bandaisan

our method extracted 9 correct key poses(accuracy:82%) and there were only 2 undetected errors. It is caused by the reason that the undetected points are not stop motions but key poses for dancers though our method aims to extract stop motions.

4.2.2 Aizu-bandaisan - A Male Dancer

In this dance, a male dancer mis-steped many times. During the mis-steps, he retuned his motion to the music, and could not completely stop his motion.

The segmentation based on the hands, feet and music without CM could extract only 3 correct key poses(accuracy:27%) and there were 8 undetected error. On the other hand, our proposed method extracted 7 correct key poses successfully(accuracy:64%), and only 4 were undetected(Figure 10). These results imply that the errors caused by individual difference can be reduced by considering the CM motions. This shows that using CM motions is very useful in segmenting motions.

4.2.3 Jongara-bushi

The results of our method is shown in Figure 11. The segmentation based on the hands, feet and CM motions failed to extract 6 stop points(accuracy:50%). On the other hand, our proposed method extracted 9 correct key poses(accuracy:75%) and failed to detect 3 key poses in the result because of the quickness of this dance.

5 Conclusion

In this paper, we have described a new analysis method for the motions of human dance. Our method segments motion sequence by considering musical rhythm to detect primitive motions of dance motion. The experiment on the three dance sequences indicate that our method efficiently segmented the dance sequence into primitives(short dances). The proposed method considering musical rhythm makes it possible to divide the dance sequence into the appropriate segments.

Our proposed method also considers the speed of center of mass to understand the movement of the body; it can deal with the case that there are some variations derived from misstep or gender difference etc.

We have also attempted to realize a dancing humanoid robot. The processes for a humanoid robot to imitate dance performance are not realtime and it seems that humanoid robots do not listen to the dance music. So we will develop our proposed method to apply an intelligence that people dance to the music to a humanoid robot as future work.

Acknowledgments

This work is supported in part by the Japan Science and Technology Corporation (JST) under the Ikeuchi CREST project.

References

[1] O. Arikan and D. A. Forsyth. Interactive motion generation from examples. *In Proc. of ACM SIGGRAPH 2002*, pages 483–490, 2002.

[2] A. Bruderlin and L. Williams. Motion signal processing. *In Proc. of ACM SIGGRAPH 1995*, pages 97–104, 1995.

[3] T. Flash and H. Hogan. The coordination of arm movements. *J. Neuroscience*, pages 1688–1703, 1985.

[4] M. Gleicher. Retargetting motion to new character. *In Proc. of ACM SIGGRAPH 1998*, 1998.

[5] M. Goto. An audio-based real-time beat tracking system for music with or without drum-sounds. *Journal of New Music Research*, 30(2):159–171, June 2001.

[6] T. Kim, S. I. Park, and S. Y. Shin. Rhythmic-motion synthesis based on motion-beat analysis. *In Proc. of ACM SIGGRAPH 2003*, 2003.

[7] L. Kovar and M. Gleicher. Motion graph. *In Proc. of ACM SIGGRAPH 2002*, pages 473–482, 2002.

[8] J. Lee, J. Chai, P. S. A. Reitsma, J. K. Hodgins, and N. S. Pollard. Interactive control of avatars animated with human motion data. *In Proc. of ACM SIGGRAPH 2002*, 2002.

[9] J. Lee and S. Y. Shin. A hierarchical approach to interactive motion editing for human-like figures. *In Proc. of ACM SIGGRAPH 1999*, pages 39–48, 1999.

[10] A. Nakazawa, S. Nakaoka, K. Ikeuchi, and K. Yokoi. Imitating human dance motions through motion structure analysis. *In Proc. of International Conference on Intelligent Robots and Systems*, pages 2539–2544, October 2002.

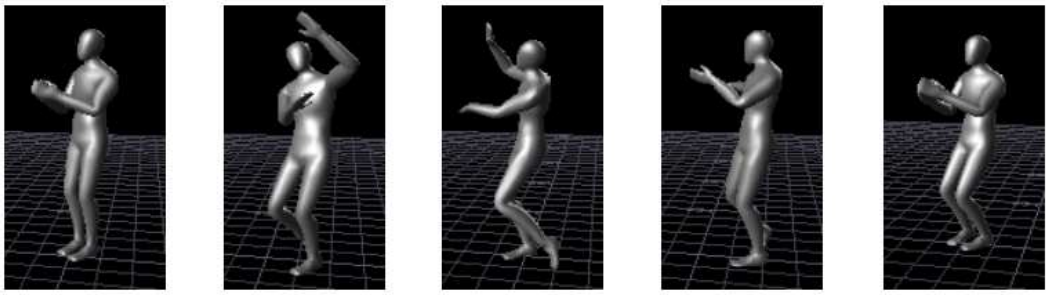

Figure 9: Segmentation result - Aizu-bandaisan dance(female)

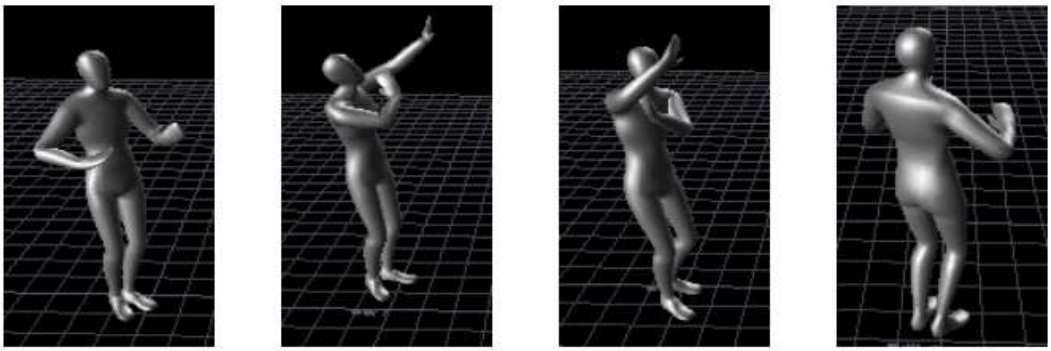

Figure 10: Segmentation result - Aizu-bandaisan dance(male)

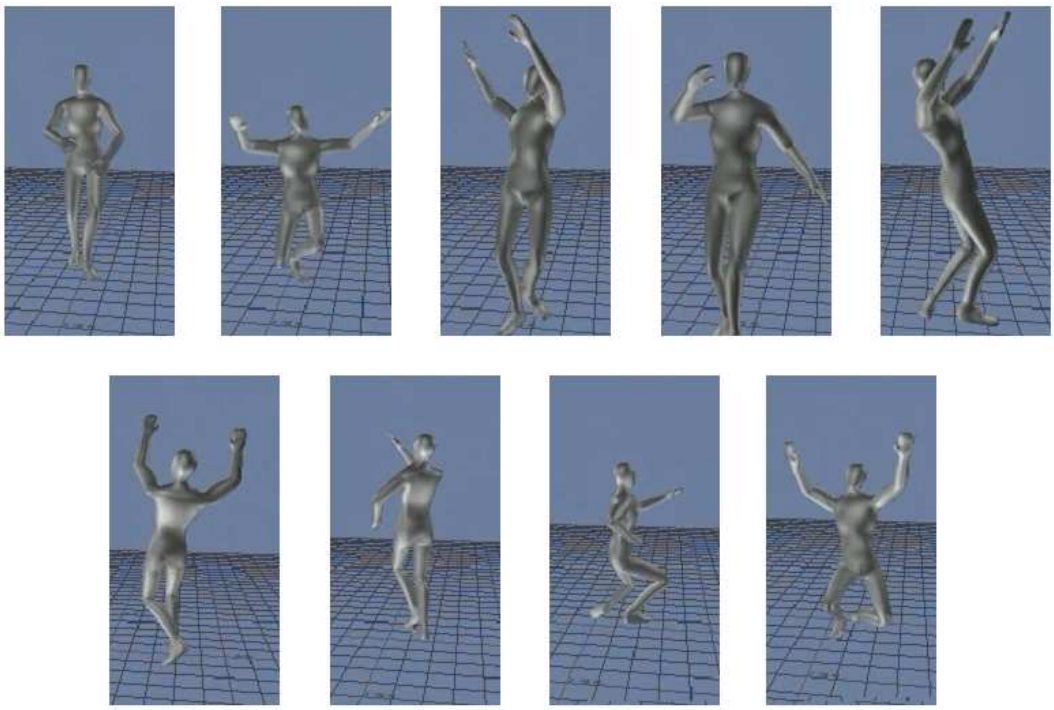

Figure 11: Segmentation result - Jongara-bushi