EXTRACTION AND ALIGNMENT EVALUATION OF MOTION BEATS FOR STREET DANCE

Chieh Ho, Wei-Tze Tsai, Keng-Sheng Lin, and Homer H. Chen

National Taiwan University

ABSTRACT

The coordination between the dancer's movement and the accompaniment of music is an important element of dance performance. In this paper, we propose a system for coordination evaluation of street dance. Given a dance video clip as input, the system first extracts motion beats from the video and then measures how well the motion beats correlate with the music beats. The motion beats are obtained by analyzing the speed and the change of direction of the dancer's movement. Unlike most previous work, which mainly focuses on 2D motion trajectory analysis, our system provides a more efficient and accurate dance movement analysis by using the 3D joint data of the dancer acquired by Kinect. Another distinction of our system is that, for beat correlation, it considers not only the underlying steady beat of music but also the groove pattern that gives the propulsive rhythmic feel of the music. We believe the differentiation of music beats into these two categories leads to a finer dance coordination evaluation. The test video clips for performance evaluation are generated by professional dancers. The average F-score of our system is about 80%.

Index Terms— Motion beat, music beat, beat extraction, music and dance coordination, dance movement analysis.

1. INTRODUCTION

Dance is a performing art consisting of a successive group of rhythmical steps and/or bodily motions executed to the beats of the music. Dance performance evaluation is subjective and involves aesthetic elements. The harmony between the dance steps and the accompaniment of music is often a fundamental criterion. Imagine, if two songs with dissonant beats are played at the same time, it will definitely sound like a discord. Similarly, if a dancer's movements do not stay on count and with the beats of the music, the dance will be chaotic and unpleasant to watch.

The rhythm in music is characterized by a sequence of beats, so is the rhythm in dance movements. Motion beat is defined as the regular moment when the movement is changed significantly [6], [7]. An illustration of motion beat is shown in Fig. 1. Since the music and motion beats continually present in time, the concordance of the dancing piece can be achieved if these beats are aligned. Therefore,

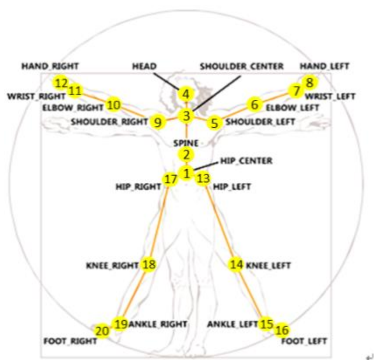
Fig. 1. Illustration of the video frame corresponding to a motion beat.

this paper focuses on the following two related issues: 1) beat extraction from dancer movements and 2) the measurement of alignment between motion beats and music beats.

Street dance is a dance style that is often accompanied with stylized rhythmic music. It includes the use of instruments such as guitar, bass, and drums to make strong rhythm. Unlike traditional ballad or ball room dance, which puts much emphasis on emotional flow, street dance emphasizes more on the synchronization of rhythm. Therefore, we choose to work on street dance first.

A key feature of this work is that we use 3D joint position data to extract motion beats. This is made possible by using a 3D sensor (Kinect) to acquire the joint data, which, in turn, are used to analyze dance movements. By utilizing the position of twenty joints recorded at a rate of 30 frames per second, we can then compute the velocity and directional changes to extract beats. Another key feature of our work is that, in addition to the underlying steady beat of music, it takes the groove pattern of music into consideration when analyzing the beat correlation, resulting in a finer evaluation of dance coordination. To recap, our system takes the position data of twenty joints as input, extracts motion beats of the dance steps, and finally evaluates the synchronization of dance movement and given music by analyzing the alignment of motion and music beats.

The remainder of the paper is organized as follows. Section 2 introduces related work. Section 3 details the approach of motion beat extraction. Section 4 describes the measurement of alignment between motion and music beats. Experiment results are shown in Section 5, and Section 6 concludes the paper.



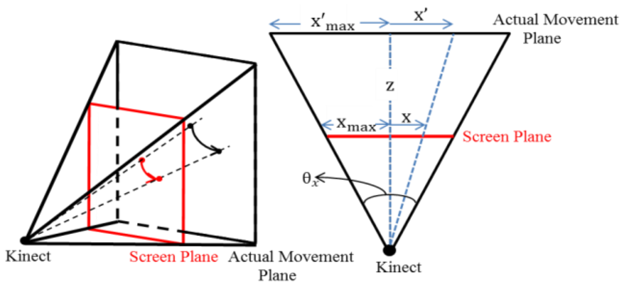


Fig. 3. Geometric relation between various coordinates.

2. RELATED WORK

2.1. Position detection using 3D data

3D data have long been used for object detection [1]–[3]. As an effective and affordable 3D sensing device, Kinect considerably facilitates interesting work in this area. For example, Tang et al. [4] use 3D data recorded by Kinect to detect hand gestures. They analyze RGB and depth data to detect skin area, locate hands, and recognize the specific poses. Xia et al. [5] present a novel human detection model using depth data acquired by Kinect. They use it to propose a model-based approach which uses a 2D head contour model and a 3D head surface model. These systems require the detection of object contour and shape from 3D data.

2.2. Motion beat extraction

Motion beat extraction is often an important issue for dance performance analysis. Chu et al. [8] detect motion beat from 2D video and then match it with music. The term "rhythm of motion" (ROM) adopted in their work to express the beat of dance movement is actually equivalent to the motion beat used in our work. But the way the motion beat is extracted is different. We obtain the position of joints from 3D data as opposed to 2D motion trajectory. Kim et al. [9] extract motion beat by finding the periodic pattern that resides in 2D motion data. They first detect directional changes of joints, and then determine the final beats based on the periodicity of the change pattern. It assumes that motion beats of a dance have the same periodic pattern throughout

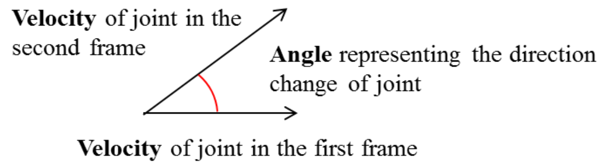


Fig. 4. Definition of the angle representing the change of direction for motion beat extraction.

the dance. However, the assumption is often not true for street dance because street dance steps are very diverse. Therefore, this method is only suitable for ballroom dance or marching bands performance that normally consists of regular steps.

3. MOTION BEAT EXTRACTION

This section describes the details of motion beat extraction, which contains data pre-processing, candidate beat selection, and final motion beat determination.

3.1. Pre-processing of data

As mentioned earlier, we use Kinect to record the x-, y-, and z-coordinates of twenty predefined joints of the dancer at a rate of thirty frames per second. Fig. 2 shows the twenty predefined joints of human body. However, the raw data of the x- and y-coordinates are different from the z-coordinate in scale. The first two are measured with respect to the screen size, but the latter is the actual depth from the dancer to the Kinect. To make the scales of the three coordinates consistent, the x- and y-coordinates are transformed with respect to the z-coordinate as follows:

$$x' = \frac{x x'_{\max}}{x_{\max}},\ x'_{\max} = z \tan\frac{\theta_x}{2} \qquad (1)$$
$$y' = \frac{y y'_{\max}}{y_{\max}},\ y'_{\max} = z \tan\frac{\theta_y}{2}$$

where θ_x and θ_y denote the angular field of view of Kinect in x and y directions, respectively, x_{max} and y_{max} are the recording ranges on the screen of Kinect in x and y directions, respectively, and x'_{max} and y'_{max} are the recording ranges of the actual movement. The relation expressed in (1) is derived by employing trianular similarity. More precisely, the ratio of x_{max} to x'_{max} equals the ratio of the distance between the screen plane and Kinect to the actual depth, so does the ratio of x to x'. A graphical illustration of the relation is shown in Fig. 3. In this way, we obtain the actual 3D position of the twenty joints.

We then calculate the velocity and the change of direction of the twenty joints for each frame. The velocity of each joint is obtained by taking the position difference of the joint between two consecutive frames and dividing it by the frame duration. The change of direction of each joint is obtained by taking the angular difference of the velocity of the joint between two consecutive frames, as shown in Fig. 4. As it will become clear, once the direction change of the

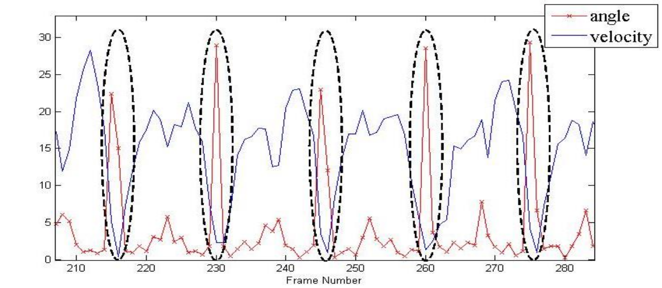


Fig. 5. The pattern of candidate motion beats.

joint is obtained, we no longer need to deal with the direction of velocity. Therefore, we only keep the magnitude of velocity for the remaining process.

3.2. Finding candidate motion beats

After pre-processing, we start to find candidate motion beats. It is known that distinctive direction changes periodically occur in dance movements [10]. After discussing with professional dancers, we learnthat the moment of rapid direction change is likely a candidate of motion beat. This is because dancers tend to have significant movements on motion beat frame (a frame when a beat occurs), resulting in direction changes. Each such change results in the co-occurrence of zero velocity and peak angle value. We select such frames as candidate beat frames, as indicated by the black ellipses in Fig. 5. In practice, however, zero velocity and peak angle value may not be aligned due to measurement error. To take measurement error into consideration, we set the tolerance of alignment between zero velocity and peak angle value to be 8 frames since human-being reaction time is about 0.25 second on average and the duration of a frame is 0.03 second.

3.3. Determining final motion beats

Our next step is to decide the final motion beats. Zero velocity and peak angle value alone are not enough to determine a motion beat. Due to some unintentional movements, not all frames with rapid direction changes necessarily represent real motion beats. A real motion beat corresponds to a significant dance movement and thus should have a significant velocity drop and rise before and after the beat frame. For example, consider a dancer making a big movement by swinging his arm from head to leg and back to head. The motion beat occurs at the time when his arm touches the leg. In this case, the velocity drops from a high peak to zero when the arm touches the leg then to another peak. On the other hand, if the dancer simply nods his head, the motion beat would have a smaller velocity drop. Therefore, not all frames with rapid direction changes correspond to actual motion beats, and our approach prunes the candidate beat frames based on the velocity drop at each candidate. We calculate the velocity drops of all candidate beat frames for the twenty joints. The velocity drop of non-candidate beat frame of a joint is assigned zero. Then, we sum up the values of the normalized velocity drops of the twenty joints in every candidate beat frame. The normalization is necessary since every joint has different range of movement.

Finally, we determine the final motion beats by comparing the sum of the velocity drops.

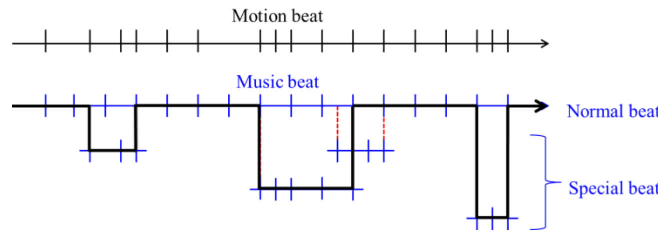


Fig.6. A motion beat sequence and its associated music beat sequences with a path of music beats aligned with motion beats.

4. ALIGNMENT EVALUATION OF BEATS

In this section, we discuss the alignment evaluation of motion and music beats. Although the procedure described in the previous section generates one sequence of motion beats, music beats come in the form of multiple sequences including normal beat sequence, which is the underlying steady beats of music, and special beat sequences, which represent groove patterns. Fig. 6 shows an example of a motion beat sequence and its associated music beat sequences along the time axis, where the vertical ticks represent beats.

We develop an approach to align the motion beats and music beats based on two common rules of dance. First, since it's not possible to dance to two kinds of beats simultaneously, and also it's unreasonable if none of the music beats is aligned, the motion beats should always be aligned with exactly one sequence of music beats. Second, the sequence of special beats should be aligned in a whole because special beats often represent a specific rhythm, and the rhythm is not separable. With the rules above, we turn the alignment of beats to a path finding problem. We need to find the best path on the map of music beats to be aligned with motion beats. An example is shown in Fig. 6.

To find the path of music beats that the dancer is most likely aligning with, we calculate scores for every path. The one with the highest score is considered the critical path. Due to the different characteristics of normal beats and special beats, we adopt different scoring rules. We use precision as score when aligned with normal beats, and use F-score when aligned with special beats. The two are different because when aligned with normal beats, we only care about how precise the alignment is. It doesn't matter if some of the normal beats have no motion beats to align with since normal beats don't have specific patterns. On the other hand, we care not only about the precision of special beats

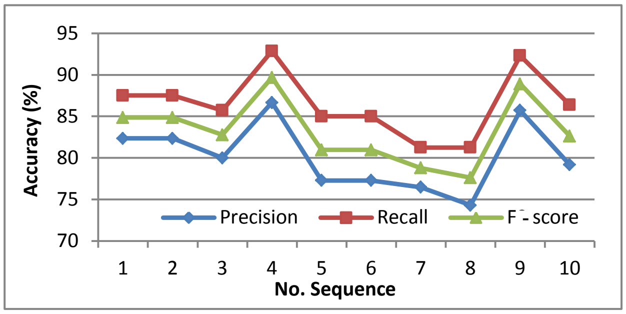


Fig. 7. Accuracy of motion beat extraction.

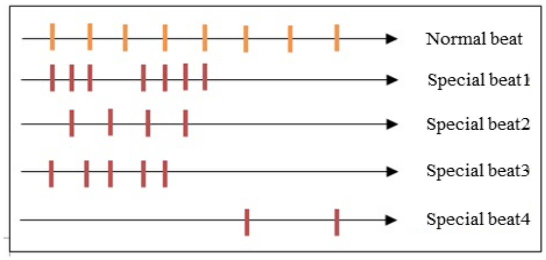


Fig. 8. Sequences of beats of the test song.

but also whether they are aligned completely. In this case, F-score is a good index describing the alignment. With the two rules, we separate a given path into several parts based on different kinds of beats, calculate the score of each part, and then use time duration as its weight to derive the weighted sum as the final score of each path, which ranges from 0 to 1.

5. EXPERIMENTS

We use several kinds of dance steps as our testing data, which are produced by professional dancers in the dance club of our university. All of them have over three years' experience of dancing. The dance steps are choreographed by themselves so that we believe our dataset is general enough for all kinds of street dance.

The experiment is divided into two parts. The first part tests the accuracy of motion beat extraction, and the second part tests the alignment evaluation. In the first part, the ground truth is recorded on the fly when the dancer is performing, and we generate the precision, recall, and F-score to evaluate our system. In the second part, the dancer performs sequences of steps with different beats. The goal of the experiment is to check whether our system can successfully determine which path of the music beats is aligned with the motion beats of the dancer.

5.1. Evaluation of motion beat extraction

In this part, the testing data contain ten sequences of dance steps, including Locking, Popping, Girl Style, and Freestyle. Each of the dance steps is about ten seconds long. Fig. 7 shows the performance of motion beat extraction of the dance sequences.

Table 1. Results of eight different paths (N stands for "Normal beat", S1 stands for "Special beat1", and so on)

No. sequence	Action path of motion beat	Average score of actual path
1	N	0.778
2	S1+N	0.4303
3	S2+N	0.7967
4	S3+N	0.8109
5	N+S4	0.7562
6	S1+S4	0.5223
7	S2+S4	0.800
8	S3+S4	0.9259

5.2. Performance of alignment evaluation

In the second part, we use the song "Die another day" with five different sequences of music beats including one sequence of normal beats and four sequences of special beats, as shown in Fig. 8. There are eight possible paths for these five sequences: N, S1+N, S2+N, S3+N, N+S4, S1+S4, S2+S4, S3+S4. (N stands for normal beat, and S1 stands for special beat1, and so on.) For each path, we record the same dance step ten times as the input. The performance is expressed by the average score that the testing input matches the correct path.

In Table I, we can see that the result is high enough to determine whether it is the right one or not. No matter how the sequences of beats are arranged, our system can find out the corresponding path by the path finding algorithm. However, the results of S1+N and S1+S4 sequence are not as good as the others because S1 contains more beats than others. It is thus harder for S1 to be aligned perfectly.

6. CONCLUSION

The alignment of dance steps with the accompanied music is an important factor of dance performance. In this paper, we have presented a system that is able to extract motion beats of street dance and evaluate the alignment of the motion beats with music beats. Unlike previous 2D approaches, the use of Kinect in the proposed approach allows 3D motion analysis of the dancer movement and leads to more accurate results. This two-stage approach first identifies candidate motion beats based on the direction change and speed of dancer movement and then removes the false candidates based on the velocity drop of motion beats. The alignment of motion and music beats is evaluated using a path finding technique. Our system achieves 80% accuracy on motion beat extraction, and 70% on beat alignment evaluation.

7. ACKNOWLEDGMENT

This work is supported by National Science Council under contract 100-2221-E-002 -198 -MY3.

8. REFERENCES

[1] H. Schneiderman and T. Kanade. “A statistical method for 3D object detection applied to faces and cars.” in *Proc. International Conference on Computer Vision*, vol. 1, pp. 746-751, 2000.

[2] S. Malassiotis, N. Aifanti, and M. Strintzis. “A gesture recognition system using 3D Data.” in *Proc. the First International Symposium on 3D Data Processing Visualization and Transmission*, pp. 190-193, 2002.

[3] K. Nickel and R. Stiefelhagen. “Pointing gesture recognition based on 3D-tracking of face, hands and head orientation”, in *Proc. International Conference on Multimodal Interfaces,* Vancouver, Canada, pp. 140-146, 2003.

[4] M. Tang, “Recognizing hand gestures with Microsoft’s Kinect,” [Online] http://www.stanford.edu/class/ee368/Project_11/Reports/Tang_Hand_Gesture_Recognition.pdf, 2011

[5] L. Xia, C.-C. Chen, and J. K. Aggarwal, "Human detection using depth information by Kinect", in *Proc. International Workshop on Human Activity Understanding from 3D Data*, Colorado Springs, CO, pp. 15-22, June 2011.

[6] L. Blom and L. T. Chaplin, *The Intimate Act of Choreography*, 1st ed. Pittsburgh: University of Pittsburgh Press, 1982.

[7] A. Hawkins, *Creating through dance*, Princeton Book Co ,U.S.; Revised edition, July 1988.

[8] W. -T. Chu, S.-Y. Tsai, “Rhythm of motion extraction and rhythm-based cross-media alignment for dance videos”, *IEEE Transaction. Multimedia,* vol. 14, no. 1, pp.129-141, January 2012.

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

[10] M. R. Jones and M. Boltz “Dynamic attending and responses to time.” Psychological Review, vol. 96, no. 3, pp. 459-491, 1989.

[11] J. Cao, “Kinect for Windows” 2012. [Online] http://msdn.microsoft.com/zh-tw/hh367958.aspx

210CA227