5-1

Analysis and Synthesis of Human Dance Motions[1]

Atsushi Nakazawa, Shinichiro Nakaoka, Takaaki Shiratori and Katsushi Ikeuchi
Cybermedia Center, Osaka University
Institute of Industrial Science, University of Tokyo
nakazawa@ime.cmc.osaka-u.ac.jp
{nakaoka,siratori,ki}@cvl.iis.u-tokyo.ac.jp

Abstract—**This paper presents the method for synthesizing stylistic human motions through visual observation. The human motion data is acquired from a motioncapture system. The whole motion sequence is divided into some motion elements and clusterd into some groups according to the correlation of end-effectors' trajectories. We call these segments as 'motion primitives'. Concatenating these motion primitives, we can generates new dance motions. We also think of a motion primitive consists of a basic motion and a motion style. The basic motion is common to all dancers, and the style represents their characteristics. We extracted these two components through the further analysis steps. The experiment results shows the validity of our approach.**
Keywords: Human Motion, Humanoid Robot, Motion Primitive, Motion Capture

Keywords— **human motion, dance motions, motion style, motion synthesis, humanoid robot**

I. INTRODUCTION

Synthesizing realistic human motion is very necessary for humanoid robot's motion planning and computer animation. Compared to the industrial purposes, such applications highly need human-like reality for generated motions. Our group aims to develop the total techniques to imitate stylistic human motions such as the dance motions by humanoid robots, for the purpose of the digital human motion archive [1]. The overview of our project is shown in figure 1. The human motion is acquired by the motion capture systems, then the motion analysis methods are applied for the purpose of motion searching and editing. Finally, original or synthesized motions are displayed by using humanoid robots. Users can recognize the motion effectively through the realistic display.

In this paper, we describe the motion synthesis method that can generate realistic human motion based on the analysis of the captured human motions. On the motion analysis step, the original motion sequence is divided into the motion segments, and the fundamental motions are extracted through the motion analysis. According to the results, characteristic poses and body movements are retrieved. These poses and movements are unique to the motions' scenario, such as the different kinds of dances, personality or other factors. On the synthesis step, the user designs the motion by directing the end effector's positions of the key frames. The system generates the key poses and the transition motion between them, finally new motion sequence can be generated.

The study for human motion synthesis has been done in the robotics and computer animation fields; they are categorized in two types. The most basic researches are based on finding the pose or transition by optimizing the particular evaluation values, such as jerk of joint angles [3], integration of the joint torque [2]. They are mainly considering the arm movement and proposed models are evaluated by the simulation result and the actual human motion. For the whole body motion synthesis, Tak et. al proposed the method by keeping whole body balancing [4]. Their method uses the ZMP (Zero Moment Point) as the evaluation value, which indicates if they can keep the dynamic balance.

These methods are based on the finding optimal solution of pre-defined evaluation values, but they have other problems that generated motion always draws a single path. In reality, a human body motion has characteristics according to scenarios. For example, a person performing dance motions, the motion will not always follows the optimal trajectory.

The solution for this problem is using some example motions of actual humans'. The new motion can be synthesized by using them. Ijspeert et. al proposes the method that the robot acquire its control command by self learning method. The robot system learn its control by comparing its movement and the actual human motion [5]. Bland et.al has propose the idea of the "Style Machine" [6]. In this method, the same class motions with different style (such as walking slowly, rapidly, etc.) has described by Hidden Malkov Model(HMM), and the other stylistic motions can be generated by the analysis result of this HMM. Lee et,al proposed the model that the motions consists of the hierarchy [7]. The human body motion is described by the direction of body links, and they are also expressed by B-spline. Because this B-spline is described by hierarchical knot vectors, new motions can be synthesized by manipulating these vectors.

All these methods can not design the particular desired human poses and motions, but combine or deform slightly existing motion segments. Our method aims to solve this issue that enables the user to design the motion sequence much directly. The user can direct the end effectors' positions of the key frames key positions, the resulting motion are synthesized as its naturally being in the desired sce-

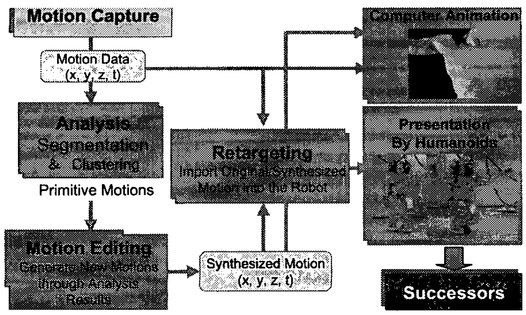

Fig. 1. Overview of our project.

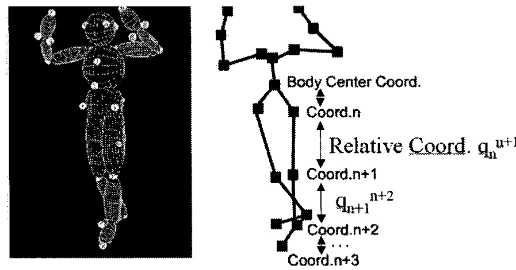

Fig. 2. The human body model.

nario. Even if the user directed the same positions on the key frames, the generated motions are different if the different database has used for the synthesis.

II. The Basic Idea

We think that the human motion sequence is consisting of some kinds of the motion elements, namely primitive motions. The primitive motions are defined on every body portions (arms and legs), and the same primitive motions appear more than a time on a motion sequence. Therefore, the whole motion sequence is also the sequence of the primitive motions:

$$\begin{aligned} HumanMotionSequence &= PrimitiveMotionA \\ &+PrimitiveMotionB + ... \end{aligned}$$

Moreover, a primitive motion consists of a motion base and a motion style:

$$PrimitiveMotion = MotionBase \oplus MotionStyle$$

The motion base is calculated from the border conditions of the motion segments (start and end poses). The motion style is the displacement between the real human motion and the motion base. Accordingly, the original motions are recovered from the key poses and the motion styles. On the analysis stage, these key poses and motion styles are stored into the motion database.

On the synthesis step, users design the human motion by directing end-effectors positions of the human figure. The two border conditions - the first and last key poses of human figure - are obtained from the key pose database, and then the motion base is calculated. Finally, the system retrieves the most suitable motion style and applies it to the motion base, which is calculated by the key poses. Because the motion styles are unique to the kinds of the original human motions, the generated motions are different between the used motion styles. And they also keep the features of the original human motions.

III. Motion Structure Analysis

We used 19 measurement points for whole body representation: fingers, hands, elbows, shoulders, head, hip, body center, waists, thighs, feet, and feet fingers. On each body joints, we set the coordinates and whole body are represented with 15 coordinates. The root coordinate is set in the body center, and other coordinates are represented with relative coordinate to the parent node (figure 2). Because the length of each body links are fixed, all body shape is represented by one translation vector on root coordinate X, and 15 quaternion vectors representing the 3x3 rotation matrixes $q_0 - q_{14}$. On the motion analysis, the input data is segmented and primitive motions are detected and the key poses and motion styles are retrieved and stored as the motion database.

A. Segmentation

The aim of the motion segmentation is to find the start and stop frame of the end effectors' movements. On the top of the analysis, we define the body center coordinate system at each motion frames. This coordinate has the waist direction as the X-axis and the perpendicular direction as the Z-axis. The four end effector's positions are mapped onto these coordinate systems, and their velocities are calculated.

The segmentation is done by detecting the local minimum of the velocity. To prevent the over segmentation, the gaussian filter has applied beforehand and these terms are checked for each segments.

1. The velocity of the segmented frames is less than the threshold.

2. The amplitude of the velocity inside of the segment is larger than the threshold.

The poses on the segmented frames are registered into the database as the key poses.

B. Detecting Primitive Motions

For detective primitive motion of the motion sequences, the correlations between the motion segments are evaluated by comparing the end effectors' trajectory. We use the DP matching distance for this purpose. Assume that the end effectors positions in the motion segment m, n are described as $V_m = \{vm_1, vm_2, ..., vm_{im} | vm_i \in R^3\}$, $V_n = \{vn_1, vn_2, ..., vn_{in} | vn_i \in R^3\}$, then the distance between these segments $D(m, n)$ can be calculated with following :

$$D(m,n) = S(V_m, V_n)$$
$$S(k,l) = d_{k,l} + min(S_{k,l-1}, S_{k-1,l-1}, S_{k-1,l})$$
$$d_{i,j} = |vm_i - vn_j|$$

After the correlations between all combinations of the segments are calculated, they are clustered by using the nearest neighbor algorithm. As the result, the segments in which the end effectors pass the similar trajectory are in the same cluster. Finally, the segments in the same clusters are equalized and registered as the primitive motions. Moreover, we can easily recognize the motion sequence is the repetition of the primitive motions, and also having some kinds of structures (see figure 4).

C. Extracting the motion base and the motion style

The motion bases and the motion style are extracted from all primitive motions. The motion base can be generated by interpolating the body links at two segment frames (the start and end frames). The human body link n on the motion primitive m is described as $Q = \{q_n(0), ..., q_n(T)\}$. For this motion segment, we define the motion base $Q^B = \{qb_n(0), ..., qb_n(T)\}$ as the SLERP (Spherical Liner Interpolation) of the segment from $q_n(0)$ to $q_n(T)$, and the motion style qs_n is the difference between the motion base and original motion (figure 3). We can formulate as following:

$$qb_n(t) = \frac{\sin(\delta\theta(1 - \frac{T-t}{T}))}{\sin(\delta\theta)} q_n(0) + \frac{\sin(\delta\theta(\frac{T-t}{T}))}{\sin(\delta\theta)} q_n(T).$$

$$qs_n(t) = \frac{q_n(t)}{qb_n(t)}.$$

The motion base is calculated from the border conditions, the resulting motion seems to be artificial and mechanical. If the same border conditions (key poses) are given, the same motion can be generated. On the other hand, the motion styles are different between the indivisuals, representing the twist or speed of the body portion's movement. We think of this information is unique feature of dancers. By applying the motion styles onto the motion base, the result becomes similar to the human motion, because these parameters describe the distortion and fluctuation of the human motion. We also thinks the motion styles are characteristic between the scenario of the original motions. This means the motion style is a kind of the human skill.

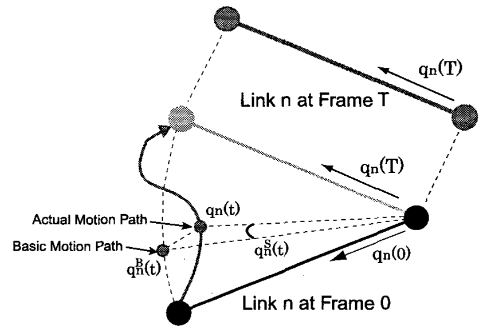

Fig. 3. The Motion Base and the Motion Style.

IV. Motion Synthesis

With the acquired poses and motion style parameter database, new stylized motions can be synthesized. The user indicates the end effectors' positions and the durations of the motion segments. According to these values, the system can synthesize new motion by using the database.

A. Synthesize Key Poses

At the beginning, the first and the end poses of the motion segment is determined. The user indicates the end effectors' positions of each frame in body center coordinate. The system searches the pose that has the nearest end effector's position from the database. If some candidates are found, the user can design the desired poses by blending them. The result is used for the initial value for synthesizing poses. To 'fit' the initial pose to the user's desired one, Jacobian based inverse kinematics solver is employed.

The initial joint angles $\theta_0 \in R^N$ are given by the found pose, and the desired joint angles $\theta_{designed} \in R^N$ can be acquired by following algorithm (N is the number of DOF of the portions):

$\theta = \theta_0$
x = ForwardKinematics(θ)
while(| x - $x_{designed}$ | > thresh)
    $J^W = W^{-1}J^T(JW^{-1}J^T)^{-1}$
    $\dot{\theta} = J^W (x_{designed} - x)$
    $\theta = \theta + \alpha \cdot \dot{\theta}$
    x = ForwardKinematics(θ)
end
$\theta_{designed} = \theta$

Where $x_{designed}$ is the user designed position, J is the Jacobian matrix of the target body portion and α is the constant (0.01 - 0.1). Because this algorithm minimize the equation $\dot{\theta}^T W \theta$, the matrix W is the diagonal matrix which indicate the weight of each joints. For the arm movement, we used smaller value for body-to-shoulder joints than other ones because of the observation of the motion. Using this algorithm, each portions' start and end poses of the motion segments can be obtained. Because we uses the pose database of original motion sequence as the initial poses, the resulting ones keep the characteristics of the original motion sequence.

B. Synthesize the Transition Motion

We developed two-step approach for synthesizing transition motion. First, the motion base is calculated from the border conditions. After that, the appropriate motion style is retrieved from the motion database and applied. The resulting motion has human-like reality due to the motion style.

The motion base is synthesized by using the algorithm described in the last section and the border conditions obtained by the key pose synthesis. Assume that the user designed the motion segment that the end effector is moving from $x(0)$ to $x(T_{org})$ in the body center coordinate. And the stored transition motion in the motion database is the path from $y(0)$ to $y(T_{ref})$, the distance between them are obtained by :

$$d = \sum_{i=1}^{N} ||\{(1-\frac{i}{N})x(0) + \frac{i}{N}x(T_{sty})\} - \{(1-\frac{i}{N})y(0) + \frac{i}{N}y(T_{ref})\}||$$

According to the nearest motion primitive, the motion style is retrieved and applied the synthesized motion base. The final synthesized motion of the body link n is acquired by following:

$$\begin{aligned} q_n(t) &= qb(t) * qs(t) \\ qb(t) &= Slerp(q^A(0), q^A(T_{org}) \\ qs(t) &= q^B * Styleof(q^B) \end{aligned}$$

This algorithm is based on the idea that the body movement that draws the similar end effector's path would have the similar motion styles. In this case, we do not consider the actual transition path of the end effectors, because we would like to synthesize the motion's characteristics only from the border conditions.

V. EXPERIMENTS

A. Experiments of Motion Analysis

We have captured three kinds of Japanese fork dances at 60Hz frequency as the motion database. The length of these ones are 1500(harukoma), 1517(nishi-monai-ondo) and 4500 frames(soran-bushi), totally 7517 frames. Figure 4 and table I show the detection results of the motion segments and primitive motions of these motions. We can recognize the number of the primitive motions is about 10% to 50% of the number of the original motion segments. Especially in Harukoma dance motion, all hands motions are described about 1/10 numbers of primitives.

Name	Segment	Primitive	Ratio
Soran-Bushi(left hand)	140	58	41.4%
Soran-Bushi(right hand)	150	61	40.7%
Soran-Bushi(left foot)	153	26	17.0%
Soran-Bushi(right foot)	153	30	19.6%
Harukoma(left hand)	86	9	10.4%
Harukoma(right hand)	89	8	9.0 %
Harukoma(left foot)	57	8	14.0%
Harukoma(right foot)	50	8	16.0%
Nishi-Monai-Ondo(left hand)	46	21	45.7%
Nishi-Monai-Ondo(right hand)	45	22	48.9%
Nishi-Monai-Ondo(left foot)	39	7	17.9%
Nishi-Monai-Ondo(right foot)	41	7	17.1%

TABLE I
THE ANALYSIS RESULT OF THE THREE DANCE MOTIONS (ON THE LEFT HAND MOVEMENT).

B. Experiments of Motion Synthesis

Figure 5 shows the key pose frames segmented by the left hand's velocity. As we can see, each poses has the characteristic body shape if the end effector is located at near positions. Figure 6 show the synthesis result of the stylized motion. These results are base on the synthesis result of the poses in the last experiment. We can recognize the difference between the motion base and stylized motions in the arm movement. Because the motion base is the linear interpolation of the border conditions, the resulting motions are always short-cut paths between them. And the transition angle speeds are constant, they seems to be a kind of mechanical motion. On the other hand, the stylized motions not only draw the shortest path between the start and end frames. For example, the stylized motion according to the Aizu Bandaisan dance database, the left hand motion is far longer way compared to the motion base, and velocity of the right hand motion are very different. We can see the similar characteristics in other stylized dance motions and proposed stylized parameter can actually synthesize the feature of the characteristic poses and transition features.

Figure 7 to 9 show the synthesis result of the stylized motion by using different motion-style database. We can recognize the difference between the motion base and stylized motions in the arm movement. Because the motion base is the linear interpolation of the border conditions, the resulting motions are always short-cut paths between them. And the transition angle speeds are constant, they seems to be a kind of mechanical motion. On the other hand, the

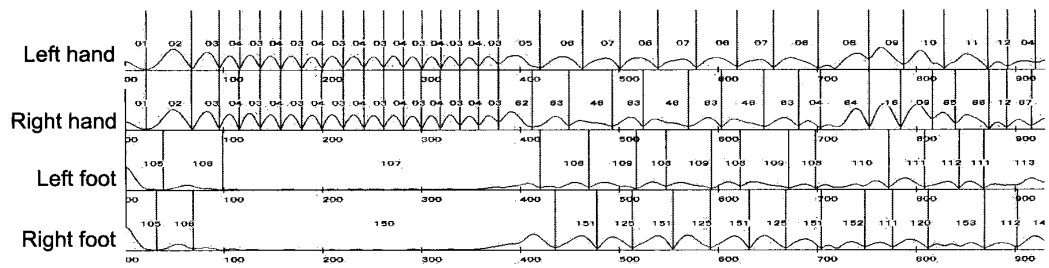

Fig. 4. The result of the segmentation and detection of primitive motions on Soran-Bushi motion sequence. The number of each segments indicate the ID of the primitive motions.

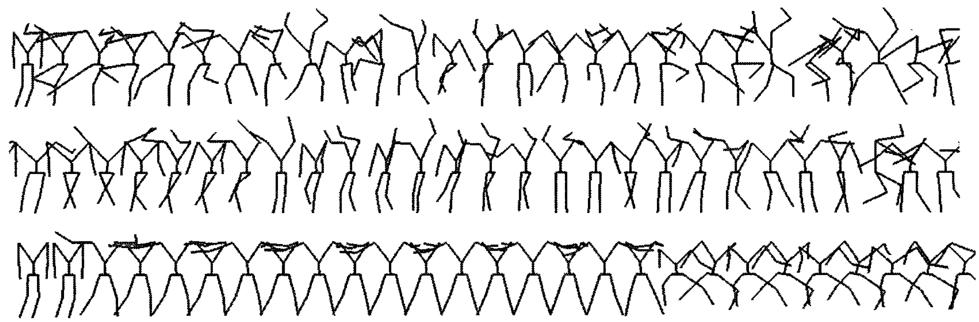

Fig. 5. The key poses of the three kinds of dances (segmented according to the left hand's movement).

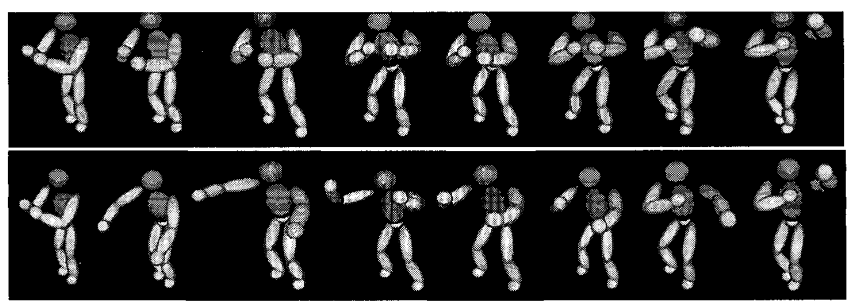

Fig. 6. The synthesized results with Aizu-Bandaisan motion database. Upper: Motion Base, Lower: Motion Base + Motion Style.

stylized motions not only draw the shortest path between the start and end frames. For example, the stylized motion according to the Harukoma dance database, the left hand motion is far longer way compared to the motion base, and velocity of the right hand motion are very different. We can see the similar characteristics in other stylized dance motions and proposed stylized parameter can actually synthesize the feature of the characteristic poses and transition features.

VI. Conclusion

In this paper, we proposed the motion synthesis method that can generate realistic human motion according to the motion capture databases. This method is based on the idea that the human motion is consisting of the characteristic poses, motion base and the motion style. For synthesizing poses, original human motions are segmented to detect the stop motion frames. After the motion base has generated mathematically from the border conditions (synthesized start and end poses), the motion style is applied.

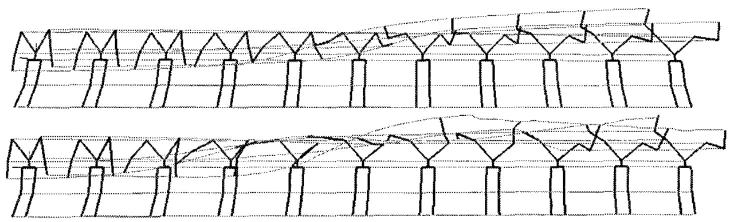

Fig. 7. The synthesized results with Harukoma motion database. Upper: Motion Base, Lower: Motion Base + Motion Style.

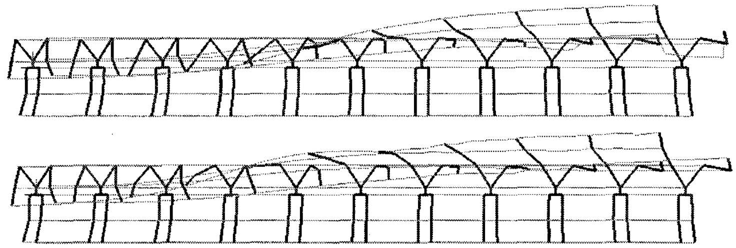

Fig. 8. The synthesized results with Nishi-Monai motion database. Upper: Motion Base, Lower: Motion Base + Motion Style.

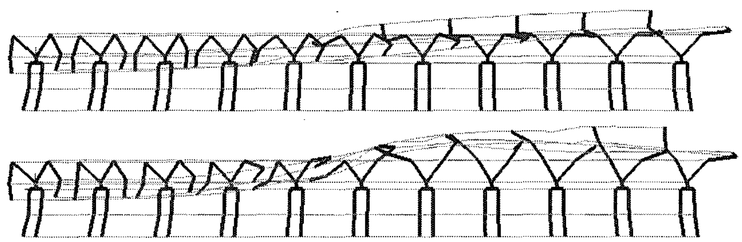

Fig. 9. The synthesized results with Soran-Bushi motion database. Upper: Motion Base, Lower: Motion Base + Motion Style.

The motion style can be obtained from the motion database according to the desired transition path. We can confirm this motion style has large effect for increasing the human-like reality through the experiment results. Moreover, the synthesized results are different according to the motion database, even if the user directs the same information for the synthesis. This feature can be useful for computer animation and imitating human skill by the robots. Now we can generate the motion which the user designed their end effectors positions under the constraint that all of the body portions have the stop frame at the same time. But in reality, this constraint cannot be always appropriate. For this issue, we are now trying to detect the relation between the stop frame of each portion, and use for the designing.

Acknowledgement

This research was, in part, supported by Japan Science and Technology Corperation under CREST Ikeuchi Project.

REFERENCES

[1] Nakazawa A., Nakaoka S., Kudoh S., Ikeuchi K. Yokoi K.: Imitating Human Dance Motions through Motion Structure Analysis, Proc. of International Conference on Intelligent Robots and Systems (IROS2002), 2002.

[2] Y.Uno, M.Kawato, R.Suzuki: Formation and Control of Optimal Trajectory in Human Multi-Joint Arm Movement-Minimun Torque Change Model, Biological Cybernetics 61, pp.89-101, 1989.

[3] T.Flash, H.Hogan: The coordination of Arm Movements, Journal of Neuroscience, pp.1688-1703, 1985.

[4] Tak S., Song. O., Ko H.: Motion Balance Filtering, Proc. of Eurographics 2000, Vol.19, No.3, pp.437-446, 2000.

[5] Ijspeert A.J., Nakanishi J., Schaal S.: Movement imitation with nonlinear dynamical systems in humanoid robots, International Conference on Robotics and Automation (ICRA2002), pp 1398-1403, 2002.

[6] Brand, M.E.; Hertzmann, A.: Style Machines, Proc. of ACM SIGGRAPH2000, pp. 183-192, 2000.

[7] Lee J. and Shin S.Y. : A Hierarchical Approach to Interactive Motion Editing for Human-like Figures, Proc. of ACM SIGGRAPH'99, pp.39-48, 1999.