IPSJ Transactions on Computer Vision and Image Media Vol. 1 No. 1 34–47 (June 2008)

Regular Paper

Synthesis of Dance Performance Based on Analyses of Human Motion and Music

Takaaki Shiratori[†1] and Katsushi Ikeuchi[†2]

Recent progress in robotics has a great potential, and we are considering to develop a dancing humanoid robot for entertainment robots. In this paper, we propose three fundamental methods of a dancing robot aimed at the sound feedback system in which a robot listens to music and automatically synthesizes dance motion based on the musical features. The first study analyzes the relationship between motion and musical rhythm and extracts important features for imitating human dance motion. The second study models modification of upper body motion based on the speed of played music so that a humanoid robot dances to a faster musical speed. The third study automatically synthesizes dance performance that reflects both rhythm and mood of input music.

1. Introduction

Since technology regarding humanoid robots is advancing rapidly, many research projects related to these robots have been conducted. To add to this research, we are considering to enhance human dance moiton for entertainment robots, and aiming at mimicing dance performance with a biped humanoid robot.

We developed an algorithm to enable a humanoid robot to represent dance performance [17]. This algorithm is based on a *Learning-from-Observation* (LFO) paradigm that has a robot directly acquire the knowledge of what to do and how to do by observing a human demonstration. This paradigm is necessary because the difference in body structure between a robot and a human performer makes it impossible to directly map human motion to a robot that needs to maintain its balance. We designed task models for leg motion of dance performance based on contact states, and we have used these to automatically modify human motion for use by a robot.

However, there still remains the problem that this algorithm is done offline and a robot cannot listen to music and respond to it when performing a dance. In this paper, we propose three fundamental techniques aimed at the achievement of *dancing-to-music* ability. This ability, which we call *sound feedback system*, indicates that people can synchronize their dance motion with various musical features such as rhythm, speed, and mood, even if they are novices. The ultimate goal is that a robot listens to currently played music and automatically synchronizes or composes dance motion synchronized with the music.

2. Overview of Proposed Methods

Considering characteristics of actual dance performances, there are various musical features affecting dance motion. We mainly focused on the following correspondences between music and motion:

Rhythm Rhythm is one of the most important features for dance performance. Even novices can recognize musical rhythm, and easily clap or wave their hands and legs in response to it.

Speed Dance motion should be synchronized with the speed of musical rhythm. Usually, when music gets faster, dance motion is modified to follow up the musical speed.

Mood Some dance performances are much affected by musical moods such as happiness and sadness. Even if we don't dance to music, we feel quiet and relaxed when listening to relaxing music such as a ballad, and we feel excited when listening to intense music such as hard rock music.

Based on these factors, we developed three methods to analyze and synthesize dance motion with musical features based on human perceptions.

The first study described in Section 3 is to analyze the relationship between motion and musical rhythm and to extract important stop features in order to mimic human dance motion. The goal of this study is to distinguish which features should be preserved for dance motion imitation. According to observation of human dance motion, motion rhythm is represented with a stop motion called a *keypose*, at which dancers clearly stop their movements, and the motion rhythm is synchronized with musical rhythm when performing a dance. The proposed

[†1] Department of Information and Communication Engineering, Graduate School of Information Science and Technology, The University of Tokyo

[†2] Interfaculty Initiative in Information Studies, The University of Tokyo

method aims to reveal this relationship.

The second study described in Section 4 is to model how to modify upper body motion based on the speed of played music. When we observed structured dance motion performed at a normal music playback speed and motion performed at faster music playback speed, we found that the detail of each motion is slightly different while the whole of the dance motion is similar in both cases. To prove this, we analyzed the motion differences in the frequency domain, and obtained two insights on the omission of motion details.

The third study described in Section 5 is to synthesize dance performance that is well matched to the mood of the input music. We mainly focus on *intensity* in dance performance as a mood feature. We designed an algorithm to synthesize new dance performance by assuming these relationship between motion and music rhythm mentioned in the first study, and the relationship between motion and music intensity. However, dance motion with high intensity is difficult to reproduce with a biped robot due to balance maintenance, and our target in this study is CG character animation. But we believe that this method will be applicable to a robot in the future.

3. Keypose Extraction for Imitating Dance Motion [26)]

Mapping human motion to a humanoid robot is a difficult problem, and understanding what features in human motion are important and using the features for reproduction with robots has been well studied. Some previous methods have actually extracted abstract models by recognizing what to do and how to do it, and generating motion for robots from the models [7),8),16),20),28)]. However, we found a problem that the traditional techniques tended to extract too many models from dance motion [19)], and it is nearly impossible to distinguish what is truly important for dance performance imitation. This section describes a novel method to analyze the relationship between important postures in dance motion and musical rhythms in order to understand the essential features of dance motion. We refer to these important postures as *keyposes*.

According to Flash, et al. [3)], every human motion consists of several motion primitives, which denote fundamental elements of human motion, and these primitives are segmented by detecting instances when hands and feet stop their movements. In whole body motion, there are many methods to segment human motion by detecting the local minima of end-effector speed and to classify the motion segment into several clusters by calculating co-occurrence [21)], by using Hidden Markov Models (HMM) [7)], or by applying a spatio-temporal isomap for dimensionality reduction [8)]. Kahol, et al. [9),10)] proposed a motion segmentation method using approximated physical parameters such as force, momentum and kinetic energy. We decided to follow this biomechanical concept basically, and thus we defined that keyposes in dance motion as stopping postures.

In addition, we focused on the rhythm of dance performance. The motion rhythm of most dance performance corresponds to music rhtyhm, and some prior work on character animation uses this property for animated motion synthesis [1),12),13)]. So our method consists of a motion analysis step that extracts stopping postures from motion and a music analysis step that extracts rhythm from music. Combining motion and musical information allows the motion's keyposes to be established.

3.1 Rhythm Tracking from Music Sequence

To estimate musical rhythm, we use the following known principles:

Principle 1: A sound is likely to be produced consistent with the timing of the rhythm.

Principle 2: The interval of the onset component is likely to be equal to that of the rhythm.

The onset component represents the spectral power increase from the previous temporal frame, and we use Goto, et al.'s method [5)] for the extraction of onset components. By applying an auto-correlation function to time series of onset components, we can estimate the rhythm of music.

3.2 Keypose Candidate Extraction from Motion Sequence

Our motion analysis method is based on the speed of a performer's hands, feet and center of mass (CM). In many forms of dance, including Japanese traditional dance, the movements of hands and feet have a strong relationship with the intended expression of the whole body. Therefore, the speed of the hands and feet is useful for extracting stop motions. However, this is not sufficient for keypose extraction because sometimes the dancer makes rhythm errors, or dances are varied by the preferences or the genders of performers, etc. So in addition to the

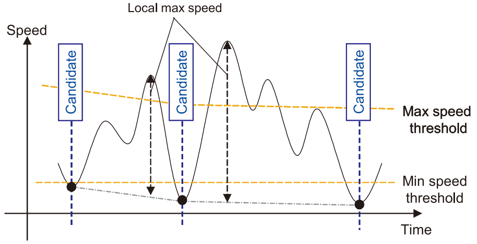


Fig. 1 Keypose candidate extraction for hand and CM motions.

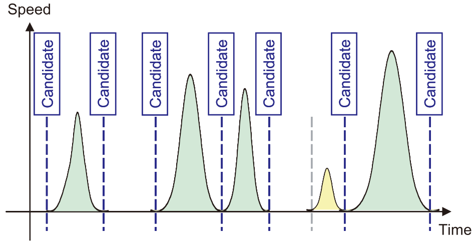


Fig. 2 Keypose candidate extraction for foot motions.

motion of the hands and feet, our algorithm uses the motion of the body's CM calculated from standard mass distribution. The motion of the CM represents the motion of the whole body; thus, the effects of missteps and individual differences are less. Through this step, we extract motion keypose candidates that satisfy the following criteria:

(1) Dancers clearly stop their movements.
(2) Dancers clearly move their body parts during neighboring keyposes.

The second criterion is necessary to avoid very small spikes of speed sequence produced by noise.

3.2.1 Hand and CM Motions

The speed of hand and CM motions approaches zero for stopping movements as shown in **Fig. 1**. To extract keypose candidates for hand and CM motions, we modify the criteria described above as:

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

3.2.2 Foot Motions

In foot motions, one leg often functions as a supporting leg while the other leg is functioning as a swing leg. Thus, the speed sequence for foot motions often consists of a series of bell-shaped curves, as shown in **Fig. 2**. To extract keypose candidates from foot motions, we first extract the rise and decay of the feet speed sequences. Then, the area between the rise and decay, which shows how far each foot moved while it was used as a swing leg, is calculated. If the area is larger than a trajectory length threshold, these rise and decay become candidates.

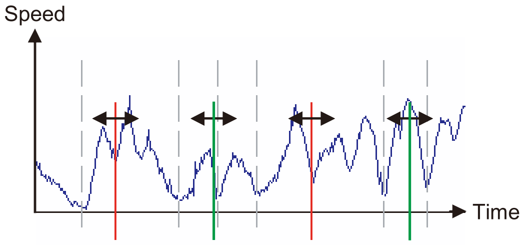


Fig. 3 Refinement of the keypose candidates using musical rhythm.

3.3 Keypose Candidate Refinement Using Musical Rhythm

The next step is to refine the keypose candidates by considering musical rhythm. For each speed sequence, our method tests whether there are candidates around musical rhythm inflection points as detected from the onset components. If there is a keypose candidate, it is possible that there is a stop point in the musical rhythm, and if so, this keypose candidate is retained for the final step.

Figure 3 illustrates the keypose candidate refinement process. Musical rhythms are represented as vertical solid lines. There are keypose candidates around second and fourth musical rhythms, represented in the figure by green vertical lines, and these candidates are preserved for the keypose extraction process.

3.4 Keypose Extraction

In the next phase, keypose candidates of a dance performance are subjected to two further criteria:

(1) Retained keypose candidates must include a match in time between more than two of the following: left hand, right hand, or feet.

(2) Retained keypose candidates must include a CM keypose match.

For example, the first criterion would be satisfied by keypose candidates of the left hand, the right hand, and one foot if these match in time. It would not be satisfied by keypose candidate time-matches in only one foot and one hand. In other words, the first criterion can extract poses at which dancers stop the movements of their hands and feet even when the stopping instance of each body part is slightly different. These poses are likely to be stop motions.

But this first criterion may extract poses that are not considered to be keyposes. For example, consider *walking* motion. In this motion, a performer's hands nearly stop their movements when his or her feet are on the ground. However, the body keeps moving in the forward direction, and this pose cannot usefully be considered a stop motion. Such translations are common in dance, so we define the second criterion to help eliminate false positives (keypose candidates labeled as valid poses, when in fact they are not); both criteria must be simultaneously satisfied to retain a keypose candidate.

3.5 Experiments

Our proposed method was evaluated using motion capture data and music data of two dance sequences: the *Aizu-bandaisan* dance and the *Jongara-bushi* dance. To evaluate the effectiveness of our proposed method, we compared the results of our keypose extraction method with those from Nakazawa, et al.'s method [19], which uses only motion capture data to extract keyposes. Additionally, we compared the results of our method with the important postures manually extracted by dancers. We refer to these dancers understandings as ground truth of the keyposes.

Aizu-bandaisan Dance

A subset of our method's analysis of a female dance master performing the Aizu-bandaisan dance is shown in **Fig. 4**. Our method correctly extracted all of these keyposes with no false positives and no mis-detected errors. The previous method which considers only motion capture data extracted 8 of the 9 true keyposes correctly, but generated 4 false-positives and mis-detected 5 errors.

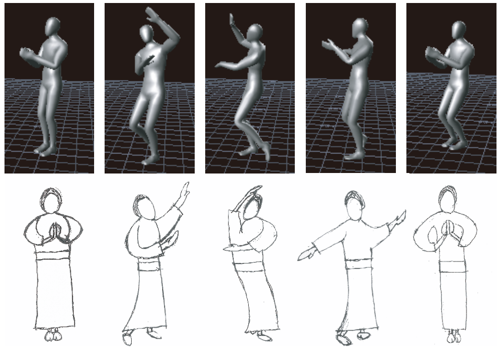


Fig. 4 Subset of extracted keyposes from the Aizu-bandaisan dance (top) and dancer's understanding (bottom).

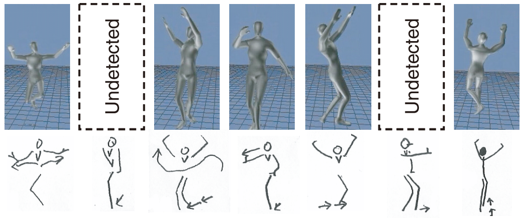


Fig. 5 Subset of extracted keyposes from the Jongara-bushi dance (top) and dancer's understanding.

Jongara-bushi Dance

A subset of results of our extraction method for a dancer performing the Jongara-bushi are shown in **Fig. 5**. This dance has 12 true keyposes. The previous method extracted 6 of these, and had no mis-detected errors. In contrast, our method extracted 9 correct keyposes, with no mis-detected errors. We believe that our method failed to detect 3 keyposes because of the high speed of this dance.

From these results, we concluded that our method was much more accurate than the previous method. We believe that the reason for our superior results is that our method considers not only motion capture data but musical information, while the previous method considers only motion capture data. By incorporating an analysis of a dance's musical rhythm, we reduced the number of false positives that previous methods have generated due to the high degree of freedom of any articulated figure.

4. Synthesis of Temporally-Scaled Dance Motion Based on Observation [25]

Temporal scaling of dance motion is very important for development of the sound feedback system, since there is a possibility that the speed of pre-recorded dance motion data is different from that of a live music performance. If this occurs, a robot cannot collaborate with music performers without this ability. In this section, we propose a novel method to temporally scale upper body motion involved in dance performance for synchronization with music.

Many researchers have attempted to efficiently edit and/or synthesize human motion from a single motion sequence through such procedures as editing motion capture data by signal processing techniques [2], retargeting motion to new characters [4], and modifying human motion to make it funny [29]. However, there are no methods to generate temporally scaled human motion except McCann, et al's method [15] that aimed at temporal scaling of jumping motion by considering physical laws. Their method does not work well for non-jumping motion.

To achieve this goal, we first observe how dance motion is modified based on played musical speed, and then we model the modification based on the acquired insights. When we observe structured dance motion performed by humans at normal music playback speeds versus motion performed using music that is 1.3 times faster, we find that the details of each motion sequence differ slightly, though the whole of the dance motion sequence is similar in both cases. An example of this type of motion modification, which is natural in humans, is shown in **Fig. 6**. This phenomenon is derived from the fact that dancers omit details of a dance, but retain its essential aspects, if this is necessary to follow faster music. If we therefore observe motion differences in dances performed at different speeds in the frequency domain, we can obtain useful insights regarding motion detail omission. Based on these insights, we propose a new modeling method and develop some applications useful for humanoid robot motion generation.

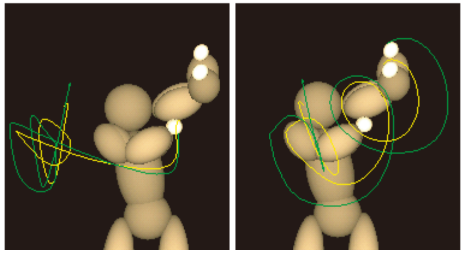


Fig. 6 Comparison of hand trajectory differences depending on music playback speed. The green and yellow curves represent the hand trajectories at a normal musical speed, and 1.3 times faster musical speed, respectively.

4.1 Observation of Human Dance Motion

This section denotes how we observed the relationship between human motion and musical rhythm speed.

4.1.1 Motion ObservationUsing Hierarchical B-spline

Hierarchical B-spline consists of a series of B-spline curves with different knot spacings; higher layers of a hierarchical B-spline are based on finer knot spacing that can preserve the higher frequency components of the original sequence. Each subject motion sequence has a different underlying musical rhythm, and a B-spline allows us to control frequency resolution by only setting its control points at desired temporal intervals. In our analysis, we considered musical rhythm for knot spacing, and we normalized temporal frames of motion sequences into the knot space with musical rhythm. Then we applied hierarchical B-spline decomposition to joint angles of dance motion via a least squares solution [14].

4.1.2 Observation Using Hierarchical B-spline

Using an optical motion capture system, we captured the *Aizu-bandaisan* dance, a classical Japanese folk dance, at three varying musical speeds for observation: the original speed, 1.2 times faster speed, and 1.5 times faster speed. Motion sequences at each speed were captured five times in order to investigate motion variance, so a total of 15 datasets were considered in this experiment. We set the knot spacing to the musical rhythm, and then applied a hierarchical B-spline decomposition technique. We used up to five layers in our motion decomposition and observed the mean and variance of each reconstructed motion.

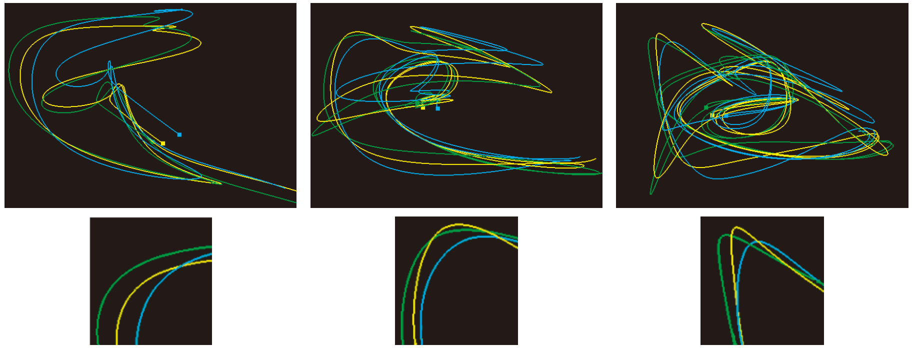

(a) (b) (c)

Fig. 7 Comparison of mean joint angle trajectories at the original musical speed (green), 1.2 times faster speed (yellow), and 1.5 times faster speed (light blue). Top: whole trajectories, and bottom: closer-look at top-left. (a) mean motion using a single-layer B-spline, (b) mean motion using a two-layer hierarchical B-spline, and (c) mean motion using a three-layer hierarchical B-spline.

Our choice of five layers was arbitrary, but it was empirically found to be enough to reconstruct high-frequency components of human motion.

Figure 7 shows mean joint angle trajectories of the left shoulder. With regard to motion reconstructed from a single-layer B-spline (Fig. 7 (a)), the motion at the 1.2 times faster musical speed is quite similar to the motion at the normal musical speed. The motion at the 1.5 times faster musical speed is also similar to the motion at the normal speed, but their details such as curvature differ slightly from each other. With regard to motion reconstructed from a two-layer hierarchical B-spline (Fig. 7 (b)), the shape of the joint angle trajectory at the normal musical speed differs slightly from that of the 1.2 times faster musical speed, especially in the trajectory's sharpest curves. On the other hand, the shape of the joint angle trajectory at the 1.5 times faster musical speed appears to be a smoothed version of the normal music playback speed trajectory. With regard to motion reconstructed from a three-layer hierarchical B-spline (Fig. 7 (c)), the differences among the joint angle trajectories become more noticeable. The shape of the joint angle trajectory at the 1.5 times faster musical speed is a much smoother version of the trajectory at the normal musical speed, whereas the shape at the 1.2 times faster musical speed is just a slightly smoother version of the trajectory at the normal musical speed. As for motion reconstructed from a four-layer hierarchical B-spline and a five-layer hierarchical B-spline, these phenomena appear more clearly.

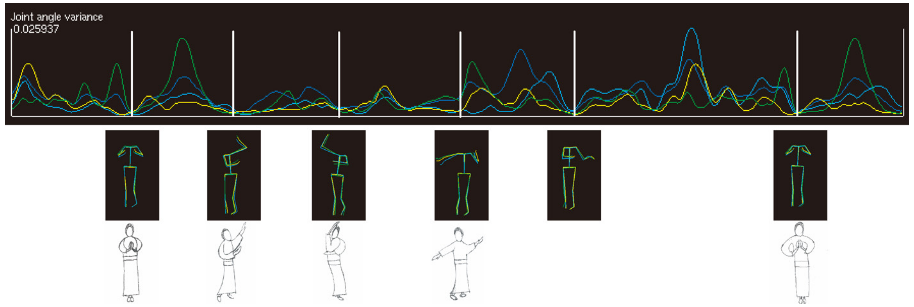


Fig. 8 Comparison of variance sequences at the original musical speed (green), 1.2 times faster speed (yellow), and 1.5 times faster speed (light blue). The blue line represents the variance calculated from all the motion sequences.

Figure 8 shows a comparison of variance sequences; the green, yellow, and light blue lines represent the variance sequences of the left shoulder joint angle at the normal musical speed, 1.2 times faster musical speed, and 1.5 times faster musical speed, respectively, and the blue line represents the variance sequence calculated from all the motion sequences. The joint angles for the variance calculation were reconstructed with a five-layer hierarchical B-spline, and were normalized by adjusting the knot of the estimated control points. From these variance sequences, it is confirmed that there are some valleys where each variance sequence is locally minimum. This means that the postures at these valleys (the middle row of Fig. 8) are preserved even if the musical speed gets faster and the high frequency components are attenuated. We found that most valleys represent important stop motions specified by the dance masters (the bottom row of Fig. 8), and that they are very close to the results of our keypose detection method described in Section 3.

From these observations, we obtained the following two insights:

Insight 1: High-frequency components of human motion will be attenuated when the music playback speed becomes faster.

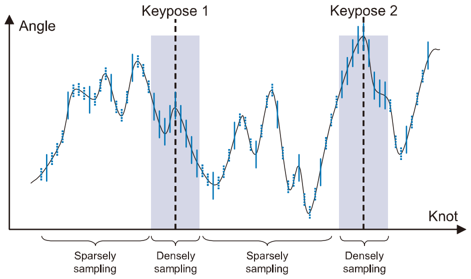


Fig. 9 Illustration of our sampling method to consider keypose information for hierarchical motion decomposition.

Insight 2: Keyposes will be preserved even if high frequency components are attenuated.

Based on these insights, we propose a method to model the temporal scaling of human dance motion.

4.2 Upper Body Motion Generation By Temporal Scaling

This section describes how to temporally scale human motion based on the insights.

4.2.1 Hierarchical Motion Decomposition Using Keypose Information

According to Insight 2, keypose information, including posture and velocity components, is preserved even if the musical speed is fast. Therefore, low frequency components of dance motion sequence must contain the keypose information. Remembering this insight, we improved the method of the traditional motion decomposition. To achieve this, our motion decomposition method needs to consider the posture and velocity information of the keyposes.

To consider posture information, we densely sample input motion sequence around keyposes, we sparsely sample it in other parts, and then we use these samples to form a linear system of equations. **Figure 9** provides an illustration of our data-sampling method for motion decomposition. All vertical lines in this illustration represent originally sampled data, and our method uses only the solid lines shown among them.

With regard to velocity information, the movements of a dancer's arms and hands stop around keyposes: the velocity of the hands and arms are approximately zero at keyposes. We exploit this useful property of keyposes as velocity information in our motion decomposition method. From all the keyposes, we form a linear system of equations to satisfy the velocity constraints. For each layer of hierarchical B-spline, we estimate the control points from posture and velocity constraints using pseudo inverse matrix and decompose the input motion sequence.

4.2.2 Motion Generation Based on Mechanical Constraints

The final step is to generate temporally-scaled motion for a humanoid robot. Simple temporal scaling can be done by adjusting the temporal frame of B-spline control points with the specified scaling ratio. However, the resulting motion may violate angular limitations such as joint angular velocity. To solve this, we consider Insight 1 and mechanical constraints that a humanoid robot has, and we modify upper body motion.

In this step, we first segment the motion sequence to correspond to music rhythm frames, and then we optimize weighting factors for each hierarchical B-spline layer in each motion segment so that a resulting motion sequence must satisfy certain mechanical constraints. The resulting joint angle θ_{opt} is represented as

$$\theta_{opt}(t) = \sum_{i=1}^{N} w_i f_i(2^{i-1}st), \qquad (1)$$

where s represents a temporal scaling factor (i.e., the resulting motion is s-times faster than the original motion), N represents the number of hierarchical B-spline layers, and f_i represents the i-th layer of the constructed hierarchical B-spline. $w_i \in [0,1]$ represents the weighting factor for the i-th layer to be detected via this optimization process.

According to Insight 1, the high frequency component is attenuated when the motion is beyond joint angle limitations. Therefore, this optimization process is done by attenuating the weighting factors from the finest layer. When the weighting factor reaches zero and the resulting motion does not satisfy mechanical constraints, the weighting factor for the next coarser layer is then gradually attenuated. Finally, when the resulting motion consists of n layers, the weighting

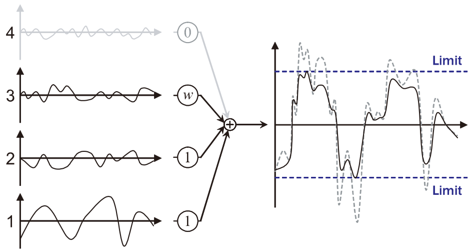


Fig. 10 Motion reconstruction considering mechanical constraints.

factors from the 1st to the $(n-1)$-th layers are 1, the factor for the n-th layer is with in a range of $(0, 1]$, and the factors from the $(n+1)$-th to the N-th layers are 0. This is illustrated in **Fig. 10**.

In this process, a discontinuity might develop between neighboring motion segments if there ends up being a difference in the weighting factors. So we simply apply motion blending around the discontinuities using spherical linear interpolation (SLERP) for joint angles. Through this interpolation process, there is a possibility of going beyond mechanical limitations. So we iteratively do the optimization and interpolation procedures until the resulting motion does not violate the mechanical constraints.

4.3 Experiments

In this section, we show the results of the experiments. We tested our algorithm by modifying the Aizu-bandaisan dance data through our algorithm. We applied the proposed method to the upper body motion, and applied Nakaoka, et al.'s method to generate leg motion[16]. Our experimental platform was HRP-2.

4.3.1 Result of Original-Speed Motion Generation

We first tested our algorithm by generating the dance motion for the HRP-2 at the normal speed. In this experiment, we compared our method with Pollard, et al.'s method that can adapt motion capture data for a humanoid robot using PD controller[22].

Figure 11 shows the experimental result with the actual HRP-2. It is confirmed that the robot can stably imitate the human dance motion.

Figure 12 shows the resulting joint angle trajectories of the left shoulder yaw. As for the joint angular velocity (Fig. 12 (b)), our method has two advantages.

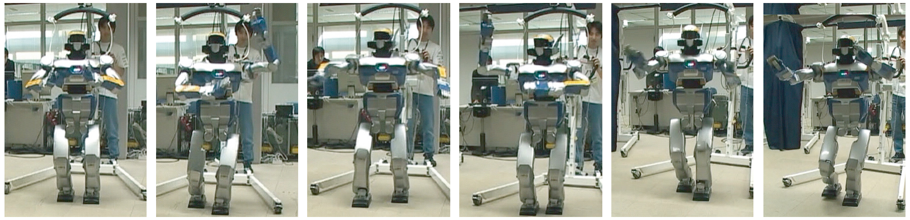
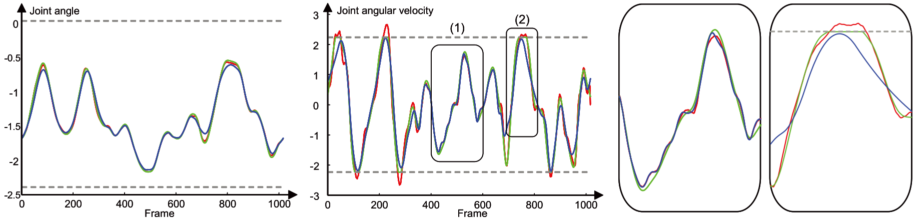


Fig. 11 Result of generating the Aizu-bandaisan dance motion at the original musical speed.



(a) (b) (b.1) (b.2)

Fig. 12 Comparison of left shoulder yaw angle trajectories of the original motion (red), generated by Pollard, et al.'s method (green), and generated by our method (blue). (a): joint angle trajectories, and (b): joint angular velocity. (b.1) and (b.2) represent the zoomed-in graph of part (1) and (2) in (b), respectively.

One is that our method can preserve more details than the trajectories resulting from Pollard, et al.'s method. The trajectories resulting from Pollard, et al.'s method often lack high frequency components, due to the PD control. This phenomenon is shown in Fig. 12 (b.1). The other is that the speed around motion frames when speed limitations are violated is fixed to a constant value in the motion generated by Pollard, et al.'s method. This phenomenon is shown in Fig. 12 (b.2). This can create two problems. One is that the humanoid robot cannot clearly reproduce a keypose if the posture and angular speed around the keypose violate kinematic constraints. The other is that the humanoid robot may fall because of the rapid changes in acceleration.

4.3.2 Simulation Result of 1.2 Times Faster Motion Generation

Next, we tested our algorithm by generating the dance motion whose speed was 1.2 times faster than the original speed in simulation. The upper body motion was

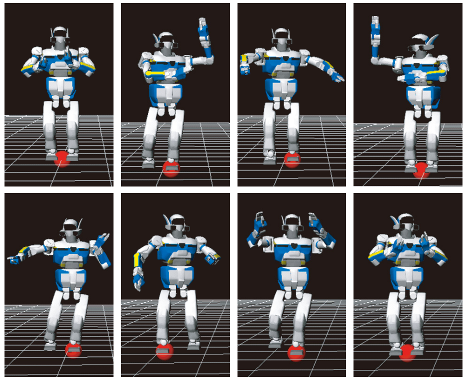


Fig. 13 Simulation result of generating 1.2 times faster dance motion. The red sphere represents the ZMP of the resulting motion.

generated by our proposed method, and as for leg motion, we first applied simple temporal scaling to the motion capture data and then applied Nakaoka, et al.'s method. **Figure 13** shows the simulation results, and the red sphere represents a Zero Moment Point. Our simulated motion satisfied the criterion for balance maintenance, and the humanoid robot successfully performed the dance.

5. Dancing-to-Music Character Animation Based on Aspects of Mood[27]

The previous section describes a method to enable a robot to dance to musical rhythm and speed. But some dance performances are improvisational based on music mood. The method described in this section focuses on this type of dances and its purpose is to synthesize dance motion well matched to music mood features. Though this method is only for CG characters because of intense motions such as jumping, this will be also applicable to a humanoid robot in the future.

Dancers can simultaneously compose a dance motion based on the musical sounds they are listening to. Although this ability may appear amazing, actually these performers do not create these motions, but rather combine appropriate motion segments from their knowledge database with music as their key to perform their unique movements. Considering this ability, we are led to believe that dance motion has strong connections with music in the two following aspects: 1) The rhythm of dance motions is matched to that of music, and 2) the intensity of dance motions is synchronized to that of music. The first assumption is derived from our work described in Section 3, while the second assumption is derived from the fact that people tend to feel quiet and relaxed when listening to relaxing music such as a ballad but may feel excited when listening to exciting music.

Kim, et al.[12] proposed a rhythmic motion synthesis method using the results of motion rhythm analysis. Alankus, et al.[1] and Lee, et al.[13] also proposed a method to synthesize dance motion by considering the rhythm of input music. The drawback of these methods is to consider only musical rhythm; because of this, it is very difficult to synthesize expressive dance motion.

Our approach consists of three steps: a motion analysis, a music analysis, and a motion synthesis based on the extracted features. In the motion analysis step, we analyze rhythm and intensity features of input dance motions, and assign the features to each motion in a database. In the music analysis step, first, we analyze a structure of input music sequence, and extract music segments based on the structure analysis results. Next, musical rhythm and intensity features are extracted, and are assigned to each music segment. Finally, our method automatically synthesizes new dance motion by interpolating between the motion segments.

5.1 Motion Feature Analysis

Our motion analysis method strongly relies on Laban's weight effort component. In this section, we describe our definition of the weight effort component and how to extract the motion features.

5.1.1 Weight Effort

According to Laban's theory, the emotion of human motion comes from motion features consisting of *effort* and *shape* components. The effort component is defined as the movements of body portions, and the shape component is defined as the shape of elements. More recently, Nakata, et al.[18] have tested the validity of Laban's theory by using their small robot and user studies. Although they could not find a significant relationship between the shape component and any emotions, they found that the *weight effort* component, one of the effort components, is closely related to the excitement of the motion. Thus, we define the weight effort component W as the linear sum of approximated instantaneous momentum magnitude calculated from the link and body directions:

$$W(f) = \sum_i \alpha_i \arccos\left(\frac{\mathbf{v}_i(f)}{|\mathbf{v}_i(f)|} \cdot \frac{\mathbf{v}_i(f+1)}{|\mathbf{v}_i(f+1)|}\right) + \sum_{j\in\{x,y,z\}} \arccos\left(\frac{\mathbf{r}_j(f)}{|\mathbf{r}_j(f)|} \cdot \frac{\mathbf{r}_j(f+1)}{|\mathbf{r}_j(f+1)|}\right), \quad (2)$$

where α_i is a regularization parameter for the i-th link, $\mathbf{v}_n$ is a unit vector representing the direction of the n-th body link in the body center coordinate system, and $\mathbf{r}_x$, $\mathbf{r}_y$ and $\mathbf{r}_z$ represent 3-dimensional orientation of body.

5.1.2 Motion Rhythm Feature

Considering the characteristics of the weight effort component, the local minimums of this component indicate stop motions, which are impressive instances for dance performance. We recognize these local minima as motion rhythm.

5.1.3 Motion Intensity Feature

It was validated that motion intensity is related to momentum and forward translation. We obtain instant motion intensity I from the momentum W and the speed of the forward direction:

$$I(f) = W(f) \cdot (1.0 + k \cdot \mathbf{r}_y(f) \cdot \dot{\mathbf{t}}(f)), \quad (3)$$

where k is a regularization parameter between the weight effort and the speed, and $\mathbf{r}_y \cdot \dot{\mathbf{t}}$ represents the speed of body direction change. Finally, we calculate the average of the instant motion intensity from the previous motion rhythm to the next one, and set it to the motion intensity.

5.2 Music Feature Analysis

When people listen or dance to music, they extract some musical features from an audio signal. The important features for dance performance are music structure, rhythm, and intensity. Regarding music structure, we focus on a repeating pattern of the melody line, we employ similarity measurements independently of timbre effects proposed by Lie, et al. [30], and we get music segments. We assign the extracted musical rhythm and intensity features described below to the music segments.

5.2.1 Music Rhythm Feature

To extract music rhythm, we employ the onset component-based rhythm estimation described in Section 3.1. After the music rhythm estimation process, the musical rhythm feature at time t is set to 1 when t is music rhythm time, and set to 0 otherwise.

5.2.2 Music Intensity Feature

To extract music intensity, we use the following:

Principle 3: The spectral power of a melody line is likely to increase during increasing intensity in the music.

Principle 4: A melody line is likely to be performed using a higher range than the C4 note.

Many surveys on auditory psychology [24] say that our ears tend to recognize only the sound whose spectral power is the strongest among the neighboring frequency sounds, which is often used in many audio signal compression algorithms such as MP3. Accordingly, for each music segment, we calculate a temporally average spectral power of each music note and extract their peaks to figure out which note sounds are mainly produced in the music segment. In order to extract the music intensity feature, we approximately calculate the *Sound Pressure Level* of the extracted musical note's power, which considers the humans' auditory properties and is related to both the amplitude and the frequency.

5.3 Motion Synthesis Considering Motion and Music Features

The final step of our approach is to synthesize new dance motions considering both the motion and music feature vectors. **Figure 14** gives an overview of our motion synthesis algorithm. From analysis steps, we have music segments with music rhythm and intensity features, and we have motion rhythm and intensity features for each motion sequence in the database. Since rhythm is one of the most important features in dance performance, we first evaluate the similarity of the rhythm components, and detect the candidate motion segments strongly corresponding to each music segment. Then, we apply connectivity analysis, which checks if synthesized transition motion between the neighboring motion segments looks natural, and we extract the possible sequences of motion segments. Finally, we analyze the similarity of their intensity components between the music segments and the selected motion segment sequences, and we synthesize new

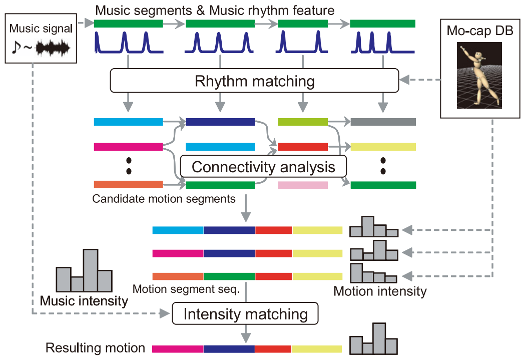


Fig. 14 Overview of our motion synthesis algorithm.

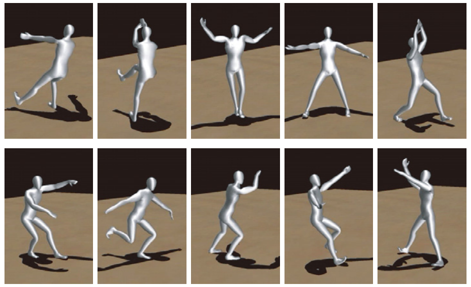


Fig. 15 The synthesis result for the dance music "*Kansho-odori.*"

dance motions by connecting the motion segments with each other.

5.3.1 Similarity Measurement of Rhythm Components

Through music analysis, we obtain music segments and then the music rhythm feature and music intensity feature for each segment. In addition, we obtain a motion rhythm feature and motion intensity feature through motion analysis. The aim of this procedure is to extract candidate motion segments by evaluating rhythm features. For each motion sequence in the database, we calculate the cross-correlation between motion and music rhythm features frame-by-frame, and we try to find partial motion sequences whose rhythm is well matched to music rhythm via a thresholding process. The partial motion sequences can be used as candidate motion segments for each music segment.

5.3.2 Connectivity Analysis of Motion Segments

Whether or not synthesized motion looks natural depends strongly on connectivity analysis. In this step, we consider both posture similarity and movement similarity. Posture similarity between the candidate motion segments is defined as the angular similarity of the body link direction, while movement similarity is calculated from the time derivative of the body link directions. We use these measurements between the end frame of one motion segment and the beginning frame of the neighboring motion segments. From the results of the connectivity evaluation, we obtain the candidate sequences of the candidate motion segments that satisfy the requirements for similarity with the rhythm features and naturalness of the synthesized motion.

5.3.3 Similarity Measurement of Intensity Components

Next, we evaluate the intensity components of the candidate sequences of the motion segments and input music. In order to find the semi-optimal solution, we consider the time series of the intensity features as a histogram, and calculate the Bhattacharyya coefficient [11)] to relatively evaluate the similarity between the motion and music intensity histograms. The motion segment sequence maximizing the coefficient becomes the final result which satisfies rhythm and intensity similarities.

Finally, the resulting motion sequence is acquired by connecting the motion segments via spline-based interpolation technique.

5.4 Experiments

We have experimented in our proposed method with our motion database consisting of several Japanese dance motion sequences.

Figure 15 shows the synthesized motion for music of the Japanese dance called *Kansho-odori*. **Figure 16** shows the features of the synthesized motion and the input music. We can easily confirm that most of the musical rhythm is matched to the motion rhythm, and the distributions of the intensity components are quite similar. Some artists working with the Japanese dance performance group *Warabi-za* told us that our technique was quite useful for a performance group such as theirs to compose new Japanese dances and this would be a new method for re-use of dance motion data.

Our proposed method is applicable not only to Japanese folk dance but also to

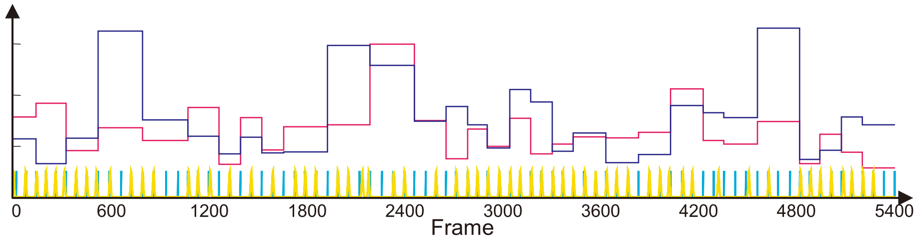


Fig. 16 The feature matching result for the music "*Kansho-odori*." Yellow and light blue lines represent motion and music rhythm components, and blue and red lines represent motion and music intensity components.

different styles of dance such as break dancing.

6. Concluding Remarks

The ultimate goal of our research is to develop a dancing entertainment robot with the achievement of the sound feedback system. For this purpose, this paper proposed methods to apply human perceptional models to human motion synthesis. Our research is strongly motivated by the fact that most previous work did not consider the perceptions of performers, but in fact human motion is highly affected by these perceptions. We have developed three fundamental methods for this purpose.

The first aspect of our proposed method, as described in Section 3, is to analyze the keyposes in dance motion. We exploit the relationship between motion rhythm and musical rhythm by detecting the stop motions in the dance motion data and by estimating the musical rhythm itself, in the form of its onset components. Dancers themselves have corroborated the results of our methods.

The second aspect of our proposed method, as described in Section 4, is to model how upper body motion can be modified depending on musical speed. Research in this arena is motivated by the observation that, as music speeds up, dancers omit details of dances to keep up with the musical rhythm. Using the insights obtained through this observation, we modeled our proposed algorithm for modification of upper body motion based on music speed.

The third aspect of our proposed method, as described in Section 5, is to synthesize dance motion using mood features. Our algorithm is designed such that motion rhythm is synchronized with musical rhythm, and that motion intensity is synchronized with musical intensity.

As future work, we will try to extend our proposed methods for sound feedback system of dancing CG characters and humanoid robots in real-time. In addition, we are thinking about devising an evaluation method of our work via subject tests. Perceptions of human motion depend upon many characteristics such as geometric models [6] and naturalness of motion after editing, and the evaluation is a very difficult problem [23]. We plan to start with subject tests to verify our assumption, and improve the proposed method.

Acknowledgments This work was supported in part by the Japan Science and Technology Corporation (JST) under the CREST project and Research Fellowships of Japan Society for the Promotion of Science (JSPS).

References

1) Alankus, G., Bayazit, A.A. and Bayazit, O.B.: Automated Motion Synthesis for Dancing Characters, *Computer Animation and Virtual Worlds*, Vol.16, No.3-4, pp.259–271 (2005).

2) Bruderlin, A. and Williams, L.: Motion Signal Processing, *Proc. ACM SIGGRAPH 95*, pp.97–104 (1995).

3) Flash, T. and Hogan, H.: The coordination of Arm Movements: An experimentally confirmed mathematical model, *Journal of Neuroscience*, Vol.5, pp.1688–170 (1985).

4) Gleicher, M.: Retargetting Motion to New Characters, *Proc. ACM SIGGRAPH 98*, pp.33–42 (1998).

5) Goto, M.: An Audio-based Real-time Beat Tracking System for Music With or Without Drum-sounds, *Journal of New Music Research*, Vol.30, No.2, pp.159–171 (2001).

6) Hodgins, J.K., O'Brien, J.F. and Tumblin, J.: Perception of Human Motion with Different Geometric Models, *IEEE Trans. on Visualization and Computer Graphics*, Vol.4, No.4, pp.307–316 (1998).

7) Inamura, T., Toshima, I., Tanie, H. and Nakamura, Y.: Embodied Symbol Emergence Based on Mimesis Theory, *Int'l Journal of Robotics Research*, Vol.23, No.4, pp.363–377 (2004).

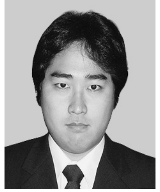
8) Jenkins, O.C. and Matarić, M.J.: Deriving Action and Behavior Primitives from Human Motion Data, *Proc. IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems*, pp.2551–2556 (2002).

9) Kahol, K., Tripathi, P. and Panchanathan, S.: Gesture Segmentation in Complex Motion Sequences, *Proc. IEEE Int'l Conf. on Image Processing*, Vol.2, pp.105–108

(2003).
10) Kahol, K., Tripathi, P. and Panchanathan, S.: Documenting Motion Sequences: Development of a Personalized Annotation System, *IEEE Multimedia Magazine*, Vol.13, No.1, pp.37–45 (2006).
11) Kailath, T.: The Divergence and Bhattacharyya Distance Measures in Signal Selection, *IEEE Trans. on Communication Technology*, Vol.COM-15, pp.52–60 (1967).
12) Kim, T., Park, S.I. and Shin, S.Y.: Rhythmic-motion Synthesis based on Motion-beat Analysis, *ACM Trans. on Graphics* (*Proc. ACM SIGGRAPH 2003*), Vol.22, No.3, pp.392–401 (2003).
13) Lee, H.-C. and Lee, I.-K.: Automatic Synchronization of Background Music and Motion in Computer Animation, *Computer Graphics Forum* (*Proc. Eurographics 2005*), Vol.24, No.3, pp.353–361 (2005).
14) Lee, S., Wolberg, G. and Shin, S.Y.: Scattered Data Interpolation with Multi-level B-Splines, *IEEE Trans. on Visualization and Computer Graphics*, Vol.3, No.3, pp.228–244 (1997).
15) McCann, J., Pollard, N.S. and Srinivasa, S.: Physics-Based Motion Retiming, *Proc. ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pp.205–214 (2006).
16) Nakaoka, S., Nakazawa, A., Kanahiro, F., Kaneko, K., Morisawa, M. and Ikeuchi, K.: Task model of Lower Body Motion for a Biped Humanoid Robot to Imitate Human Dances, *Proc. IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems*, pp.3157–3162 (2005).
17) Nakaoka, S., Nakazawa, A., Kanehiro, F., Kaneko, K., Morisawa, M., Hirukawa, H. and Ikeuchi, K.: Learning from Observation Paradigm: Leg Task Models for Enabling a Biped Humanoid Robot to Imitate Human Dances, *Int'l Journal of Robotics Research*, Vol.26, No.8, pp.829–844 (2007).
18) Nakata, T., Mori, T. and Sato, T.: Analysis of Impression of Robot Bodily Expression, *Journal of Robotics and Mechatronics*, Vol.14, No.1, pp.27–36 (2002).
19) Nakazawa, A., Nakaoka, S., Ikeuch, K. and Yokoi, K.: Imitating Human Dance Motions through Motion Structure Analysis, *Proc. IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems*, pp.2539–2544 (2002).
20) Ogawara, K., Takamatsu, J., Iba, S., Tanuki, T., Kimura, H. and Ikeuchi, K.: Acquiring Hand-Action Models in Task and Behavior Levels by a Learning Robot through Observing Human Demonstrations, *Proc. IEEE-RAS Int'l Conf. on Humanoid Robots* (2000).
21) Osaki, R., Shimada, M. and Uehara, K.: Extraction of Primitive Motions by Using Clustering and Segmentation of Motion-Captured Data, *Journal of Japanese Society for Artificial Intelligence*, Vol.15, No.5, pp.878–886 (2000). [in Japanese].
22) Pollard, N.S., Hodgins, J.K., Riley, M.J. and Atkenson, C.G.: Adapting Human Motion for the Control of a Humanoid Robot, *Proc. IEEE Int'l Conf. on Robotics and Automation*, pp.1390–1397 (2002).
23) Ren, L., Patrick, A., Efros, A.A., Hodgins, J.K. and Rehg, J.M.: A Data-Driven Approach to Quantifying Natural Human Motion, *ACM Trans. on Graphics* (*Proc. ACM SIGGRAPH 2005*), Vol.24, No.3, pp.1090–1097 (2005).
24) Roads, C.: *The Computer Music Tutorial*, The MIT Press (1996).
25) Shiratori, T., Kudoh, S., Nakaoka, S. and Ikeuchi, K.: Temporal Scaling of Upper Body Motion for Sound Feedback System of a Dancing Humanoid Robot, *Proc. IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems* (2007).
26) Shiratori, T., Nakazawa, A. and Ikeuchi, K.: Detecting Dance Motion Structure through Music Analysis, *Proc. IEEE Int'l Conf. on Automatic Face and Gesture Recognition*, pp.857–862 (2004).
27) Shiratori, T., Nakazawa, A. and Ikeuchi, K.: Dancing-to-Music Character Animation, *Computer Graphics Forum* (*Proc. Eurographics 2006*), Vol.25, No.3, pp.449–458 (2006).
28) Takamatsu, J., Tominaga, H., Ogawara, K., Kimura, H. and Ikeuchi, K.: Symbolic Representation of Trajectories for Skill Generation, *Proc. IEEE Int'l Conf. on Robotics and Automation*, pp.4077–4082 (2000).
29) Wang, J., Drucker, S.M., Agrawala, M. and Cohen, M.F.: The Cartoon Animation Filter, *ACM Trans. on Graphics* (*Proc. ACM SIGGRAPH 2006*), Vol.25, No.3, pp.1169–1173 (2006).
30) Wang, M., Lu, L. and Zhang, H.-J.: Repeating Pattern Discovery from Acoustic Musical Signals, *Proc. IEEE Int'l Conf. on Multimedia and Expo*, pp.2019–2022 (2004).

(Received September 10, 2007)
(Accepted March 9, 2008)

(Editor in Charge: *Nobutaka Shimada*)



Takaaki Shiratori received a B.E. in information and communication technology from The University of Tokyo, and M.I.S.T and Ph.D. both in information and communication technology from The University of Tokyo in 2004 and 2007 respectively. He is currently a Postdoctoral Fellow in School of Computer Science of Carnegie Mellon University. His research interests include human motion synthesis for computer animation and robotics. He is a member of the IEEE.

Katsushi Ikeuchi received a B.E. in mechanical engineering from Kyoto University, Kyoto, Japan, in 1973, and Ph.D. in information engineering from The University of Tokyo, Tokyo, Japan, in 1978. He is a professor at the Interfaculty Initiative in Information Studies, The University of Tokyo, Tokyo, Japan. After working at the AI Laboratory at the Massachusetts Institute of Technology for three years, the Electrotechnical Laboratory for five years, and the School of Computer Science at Carnegie Mellon University for 10 years, he joined The University of Tokyo in 1996. He was selected as a Distinguished Lecturer of the IEEE Signal Processing Society for the period of 2000–2001, and a Distinguished Lecturer of the IEEE Computer Society for the period of 2004–2006. He has received several awards, including the David Marr Prize in ICCV 1990, IEEE R&A K–S Fu Memorial Best Transaction Paper Award in 1998, and best paper awards in CVPR 1991, VSMM 2000, and VSMM 2004. In addition, in 1992, his paper, "Numerical Shape from Shading and Occluding Boundaries," was selected as one of the most influential papers to have appeared in the Artificial Intelligence Journal within the past 10 years. He is a fellow of the IEEE.