

Imitating Human Dance Motions through Motion Structure Analysis

Atsushi Nakazawa Shinichiro Nakaoka[†] Katsushi Ikeuchi[†] Kazuhito Yokoi^{††}

Japan Science and Technology Corporation

[†]Institute of Industrial Science, The University of Tokyo

^{††}Intelligent Systems Institute, National Institute of Advanced Industrial Science and Technology

Abstract

This paper presents the method for importing human dance motion into humanoid robots through visual observation. The human motion data is acquired from a motion capture system consisting of 8 cameras and 8 PC clusters. Then the whole motion sequence is divided into some motion elements and clustered into some groups according to the correlation of end-effectors' trajectories. We call these segments as 'motion primitives'. New dance motions are generated by concatenating these motion primitives. We are also trying to make a humanoid dance these original or generated motions using inverse-kinematics and dynamic balancing technique.

Keywords: human motion, humanoid robot, motion primitive, motion capture data

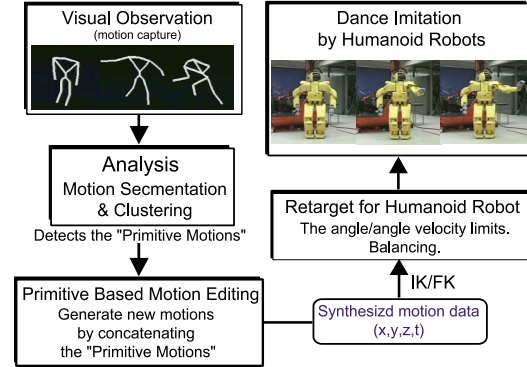


Figure 1: Overview of our project.

1 Introduction

Importing human motions into a robot through visual observation is one of the final problem in the humanoid robot studies [1]. This technology enables robots to imitate human motions easily, and be useful for programming the skill of the robot which works around our life-space. This also interests us from AI aspects because we must have acquired our motion skills in similar way.

Many studies have been done for this issue. The Jenkins' approach starts with analyzing the silhouette of a motion of movements [2]. The human hands' movements are divided into some basic motions (lines, circles, etc) and its parameters. Then whole human movements are described with these basic motions. Finally original motion is re-generated from the sequence of the basic motions. Inamura et.al proposes the idea of the 'minesis', that are the sets of basic movement of the human joint angle [5]. The human joint angle movements are divided into the list of minesis and its parameters. Our group have proposed the idea of 'learning from observation', such as APO (Assembly Plan from Observation) and Attention Point Analysis [4][3]. A robot observes and imitates a human performing an assembly task by analyzing the trajectory of human hand movements and the contact states

of the objects. Basically, all these concepts indicate that the human motions consist of some variations of simple motions. It is natural to think that the human motion consists of the limited number of the basic motions, not made from scratch. We also employ this idea for this study. We call these basic motions as "motion primitives".

We tried to apply this idea for importing human dance motions into humanoid robots. Our project overview is shown in Fig.1. The dance motions are one of the good example of whole body motion, and its characteristic is having the scenario. This means they must have a structure of the motion primitives. Our first try is to detect both ones, the motion primitives and the structure. To generate the robot movement from the motion primitives, we developed enhanced methods: the concatenation of the motion primitives and the modification technique for a humanoid robot.

2 Acquisition of the human motions

The human dance motion is acquired by the motion capture system that consists of 8 cameras (SONY DXC-9000) and PC clusters (Pentium III-800MHz Dual). The cameras are arranged to surround a person and PCs can

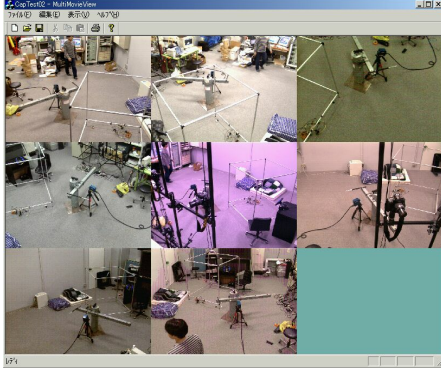


Figure 2: Acquire images from eight cameras.

acquire image frames in the size of 720x480 pixels at near 30Hz. All internal clock of PCs are rectified by NTP protocol in advance, and acquisition time of each image frames are also recorded in millisecond's accuracy. We also acquire blur-free images of moving objects by using the frame-shutter functionality with which this camera is equipped (Fig.2). All cameras' calibration parameters are acquired, then the geometrical relations between cameras and the internal parameter of each cameras can be determined.

The human motion acquisition is carried out by attaching lighting markers on the desirable positions of human body. During the actor performing the dance motion, all PCs only acquires the multi-viewpoint images and after that, depth measurement is done by matching markers between the images [6].

3 Analysis: detecting 'motion primitives'

Our aim of motion analysis is to detect similar motion elements ('motion primitives') and describe the whole dance motion with the sequence of them. According to the analysis result, we can recognize the structure of the dance motion, such as the same motion segments, iterative motion sequence or other kinds of regularities. Furthermore, it becomes possible to detect the mutual relations of different dancing by comparing the primitive motion of them.

To detect the motion primitives, we paid our attention to the local velocity minimum frames of the end effectors (hands and feet). Because it represents the human motion segment: the start points and the end points of the motion primitives. Many researchers notice this value for motion segment points [7] [8].

To evaluate the similarities of the motion segments, we used the DP distance of the target points' trajectories in 3D space. According to this value, the motion segments

are clustered into some groups. Consequently, same motions of the target points has same labels. They are registered as the 'minimum motion primitives'.

In the structural motions such as the dances, much longer motion sequences (the regularities of the minimum motion primitives) can be seen. Our algorithm can detect these primitive motion patterns and detect final motion primitives by equalizing the same motion sequences.

3.1 The motion analysis algorithm

We use 15 measurement points for analysis: hands(L,R), elbows(L,R), shoulders(L,R), head, hip, body center, waists(L,R), thighs(L,R) and feet(L,R). The analysis pipeline consists of following steps (Fig.3).

(1) Define the body center coordinate system.

We define the body center coordinate system which set the X-axis as the direction of the waist and Z-axis as the perpendicular direction. In order to detect the symmetry of the right and left arms/feet movements, we used the symmetry coordinate system for right/left half of the body portions.

(2) Coordinates conversion of target points.

The target points (both hands and feet) is changed into a body center coordinate system.

(3) Preliminary segmentation.

Calculate the velocities of the target points and detect the local minimum. Gaussian filter is applied in advance to prevent segmentation errors.

(4) Evaluate the correlation between the segments.

Evaluate the correlation between the target points' trajectories according to the DP matching distance, which is calculated by following equations.

Assume that segment m , n are described as $V_m = \{vm_1, vm_2, \dots, vm_{im} | vm_i \in R^3\}$, $V_n = \{vn_1, vn_2, \dots, vn_{in} | vn_i \in R^3\}$, then the distance between these segments $D(M, N)$ can be calculated with following :

$$D(m, n) = S(V_m, V_n)$$

$$S(k, l) = d_{k,l} + \min(S_{k,l-1}, S_{k-1,l-1}, S_{k-1,l})$$

$$d_{i,j} = |vm_i - vn_j|$$

(5) Cluster and label the segments.

The detected segments are clustered with nearest neighbor algorithm. So the segments in which the target points passes the similar locus have the same labels. The symmetrical motions are also detectable because we use the symmetrical coordinate system for right and left side of the body. Using these preliminary analysis procedures, whole motion sequence is segmented and clustered into the segments in which a target point draws same trajectory. We call these segments as the 'minimum motion segments'. Figure 4 shows the preliminal analysis result of the Japanese folk dance 'Soran-Bushi'. The minimum motion segments represent very simple motions such as "Swing down the left arm", "Steps forward the right leg".

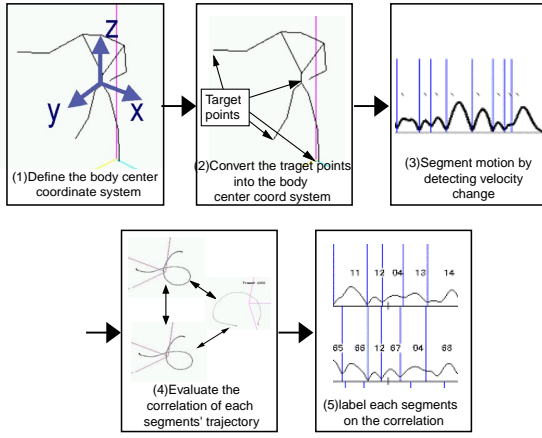


Figure 3: The preliminary motion analysis algorithm.

In additions, much long motion units exist for dance motions. To find these ones, following steps are applied for extracting the frequently appearing sequence of the minimum motion segments.

(6) Find the frequently appearing minimum motion segment sequences.

From the labeled segment sequence within the same portions, frequently appearing sequences are detected by using the apriori algorithm [9]. These results are registered as 'higher level motion segments'. This processing is performed to the segment sequence of all possible length in a part. Consequently, we can acquire multi-length and multi-hierarchical motion segments.

(7) Find the motion primitives among the different target points.

In order to find the motion primitive of the whole body portion, the correlations between different target points are evaluated. For any level motion primitives, coincidence probabilities between the primitives of different target points are calculated. If this value is higher than the threshold, they have a relation and be defined as

$$co - occurrence(X_A \cap Y_B) = \frac{f(X_A \cap Y_B)^2}{f_X(X_A)f_Y(Y_B)} > thresh$$

where $f(p)$: the frequency of the label p .

(8) Equalization.

Finally, the motion segment sequences that are labeled to the same motion are equalized for its 3D trajectories. In this process, the DP matching result is used to find the temporal matching points. The equalized results are preserved as the final primitive motion of this motion sequence.

Figure 5 shows the final analysis result of the "Soran

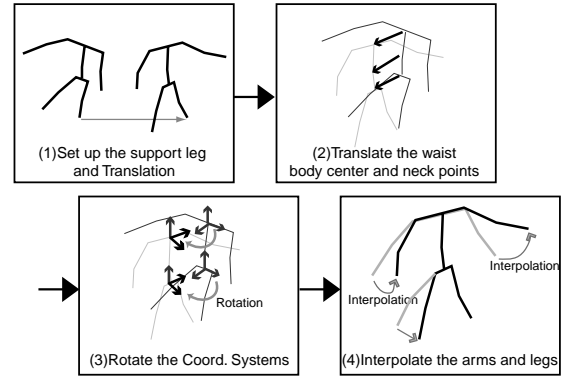


Figure 6: The motion generation algorithm.

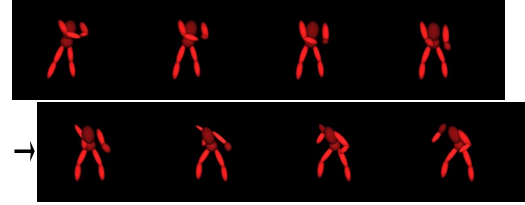


Figure 7: Generated Movements of the two motion primitives in which two different folk dances "Soran Bushi" and "Harukoma".

Bushi". We can notice following features:

- (a) The whole dance motion consists of the iterative motion primitives and unique motion segments.
- (b) Iterative motion primitives are connected by a unique primitive motion.
- (c) The longer unique primitive motion sequences exist in a part of whole dance motion.

According to these results, we understand the structure of this dance. This consists of some variations of iterative motion primitives and unique motion sequences.

4 Generate new motions from motion primitives

As described in the last section, dance motions consist of the iterative motion primitives and unique motions that connect the iterative motions. We also noticed that this structure exists not only in this example but almost in most of Japanese folk dances. According to this fact, new dance motion can be generated by concatenating the motion primitives. For this issue, the study for the human motion planning is a great help. Horgan and Flash proposed the idea of the minimum jerk model for planning the arm movements. In this theory, human arm passes to satisfy the equations: $C_j = \frac{1}{2} \int_0^{t_f} \{ (\frac{d^3x}{dt^3})^2 + (\frac{d^3y}{dt^3})^2 \} dt$.

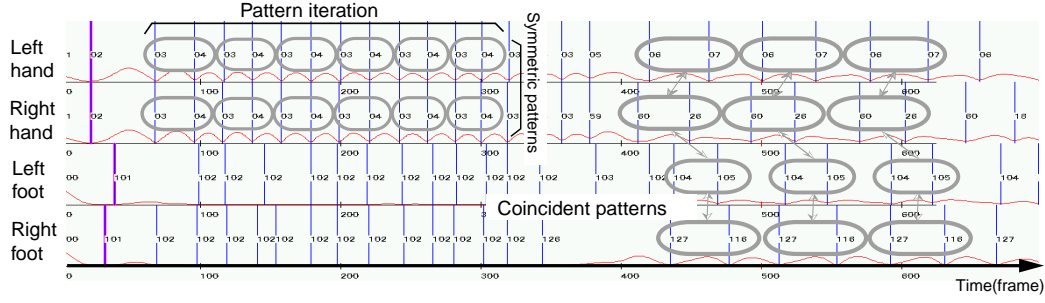


Figure 4: The preliminary detection result of the minimum segments and the correlation between the minimum segments.

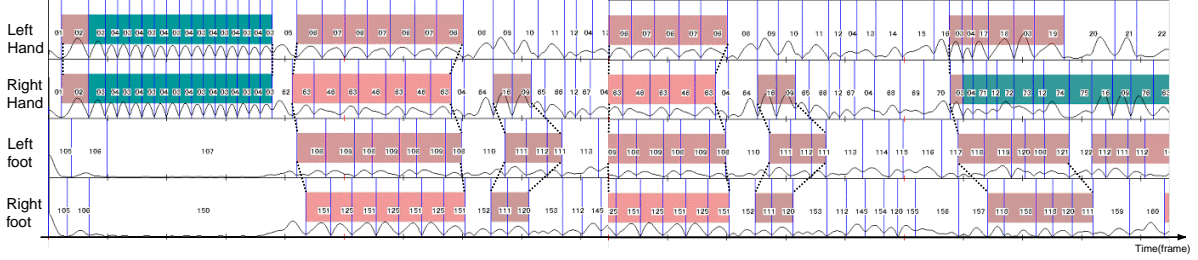


Figure 5: The analysis result of the "Soran Bushi". Colored portions indicates that they has high relations.

Where t_f is the interval time of the motion primitive. Similar theories are also proposed by Uno and Kawato [8], they proposes the minimum torque change model.

These theories indicate that we can generate new motions from the border conditions (the posture parameters at the start and the end). Our motion generation algorithm is shown in Fig.6. We employed the minimum joint angle jerk model because its simple and useful enough to apply to our purpose.

4.1 The motion generation algorithm

Assume that two motion primitives are selected to be concatenated. The transition of these motions is generated with following steps (Fig. 6).

(1) Set up a support leg during the transition. The latter primitive is translated so that this leg comes to a same position.

(2) Calculate the positions of the unsupported foot, waist, body and neck during transitions. We employed 2nd order polynomials to keep the continuity of the position and velocity.

$$v(t) = a(t - \frac{T}{2})^2 + b$$

where $v(0) = v_0, v(T) = v_1$

$$x(t) = x(0) + \int v(t)dt$$

(3) Linear interpolation is applied for the waist and neck coordinate system with following steps.

(a) Assume that the coordinate systems of the start and

end time as ${}^0R_1, {}^0R_2$. Then the rotation matrix between these coordinates is calculated ${}^1R_2 = {}^0R_1^{-1} {}^0R_2$.

(b) Convert 1R_2 into the quaternion description $Q_{12} = \{x, y, z, w\}$

(c) Interpolating rotation matrix is calculated as $Q_{12}(t) = \{x, y, z, w(t)\}$, where $w(t) = w * t/T$.

(d) Convert $Q_{12}(t)$ into the rotation matrix $R_{12}(t)$.

(e) Interpolated coordinate system is acquired as ${}^0R(t) = {}^0R_1 R_{12}(t)$

(4) For the movements of the arms and the feet, the minimum joint angle jerk model is employed for interpolation. We assume each portion has 4-DOF(2-DOF on a shoulder, thigh, 2-DOF on a elbow, knee), the joint angle parameter vectors during transition as

$\theta(t) = (\theta_1(t), \theta_2(t), \theta_3(t), \theta_4(t))$. To generate the transition motions is to determine these functions. We define each ones as the 5 order polynomials $(\theta_n(t) = a_n t^5 + b_n t^4 + c_n t^3 + d_n t^2 + e_n t + f_n)$ and set the border conditions $\theta_4(0)$ and $\theta_1(T)$. Finally, each parameters are determined to minimize the joint angles' jerk during transitions :

$$\int_{t=0}^{t=T} \sum_n \frac{\partial^3 \theta_n(t)}{\partial t^3} dt \rightarrow \min.$$

The duration T is determined by the equalization of the former and latter segments' time. Figure 7 shows the interpolation of the motion primitives from Japanese Folk Dances 'Soran Bushi' and 'Harukoma'.

5 Presenting dance motions by a humanoid robot

In this section, we show a method to import the dance motions into a humanoid robot.

A similar study has been done by Pollard et.al [10]. They used robot arms that has the same DOF of humans' ones. In our study, we employed 28-DOF whole body humanoid robot HRP-1S [12] and try to imitate whole body motions. As the first trial, we have assumed that the feet are fixed and imitate upper body motions. Although only imitating hands movements, we noticed the whole body balancing control is necessary to keep standing this robot. For this issue, we propose a new method which enables robot to imitate human dance motions as similar as possible while keeping its body standing, through the dance motion structure analysis we presented the last sections.

For these trials, we used OpenHRP simulator [11] and HRP-1S virtual model. These ones enable to test whether our data works on the real humanoid robot.

5.1 Acquiring joint angles; limiting them and their velocities

The joint angles of dance motions can be solved by using original motion capture data, simple inverse kinematics algorithm and humanoid robots connection models. But these angle values cannot be imported directly because of these restrictions : the singularity and the limits of joint angle/joint angle velocity.

Pollard et.al proposed a method to solve these problems. On their methods, joint angle values are deformed so that they become within the limits, by applying this filter like the PD control model:

$$\dot{\theta}_i = \theta_i - \theta_{i-1}, \quad (1)$$

$$\ddot{\theta}_{F,i+1} = 2\sqrt{K_s}(\dot{\theta}_i - \dot{\theta}_{F,i} + K_s(\theta_i - \theta_{F,i})), \quad (2)$$

$$\dot{\theta}_{F,i+1} = \max(\dot{\theta}_L, \min(\dot{\theta}_U, \dot{\theta}_{F,i} + \ddot{\theta}_{F,i+1})) \quad (3)$$

$$\theta_{F,i+1} = \theta_{F,i} + \dot{\theta}_{F,i+1} \quad (4)$$

$$\dot{\theta}_i = \theta_i - \theta_{i+1} \quad (5)$$

$$\ddot{\theta}_{B,i-1} = 2\sqrt{K_s}(\dot{\theta}_i - \dot{\theta}_{B,i} + K_s(\theta_i - \theta_{B,i})) \quad (6)$$

$$\dot{\theta}_{B,i-1} = \max(\dot{\theta}_L, \min(\dot{\theta}_U, \dot{\theta}_{B,i} + \ddot{\theta}_{B,i-1})) \quad (7)$$

$$\theta_{B,i-1} = \theta_{B,i} + \dot{\theta}_{B,i-1} \quad (8)$$

$$\theta_V = 0.5(\theta_{F,i} + \theta_{B,i}) \quad (9)$$

where θ_i is the original joint angle, $\dot{\theta}_L$ and $\dot{\theta}_U$ are the lower and upper velocity limits. Equations(1)~(4) are solved from the start frame to the end, and equations (5)~(8) are solved backward. The final joint angle $\theta_{V,i}$ is the average of the forward and backward passes (9).

Using this method, the joint angle velocities are kept within the given limits. The problem of singularity is also resolved at the same time because a part of the sequence near the singular point can be regarded as the area where the velocity is very fast. As a result, a gimbal locked part is decomposed to properly interpolated angles.

5.2 Keeping balance

Although we are assuming that the robot's feet are fixed while dancing, the balancing problem still exists. When the robot swings its arm in a wide arc, (Fig.8-a), it cannot keep balance and falls down (Fig.8-c). To keep the robot standing during the whole dance sequence, its ZMP (Zero Moment Point), which indicates a balanced force point existed between the robot and ground, must be within a support area enclosed by its soles [13]. The movements of the arms are main factors to move ZMP outside the support area because the portions are far from feet contact points ('the fulcrum' of the body). As a result, the robot must fall down like Fig.8-c.

In order to keep ZMP within the support area, the motion must be modified to compensate for the ZMP trajectory.

On the Pollard's filter, the stiffness parameter K_s controls the motion dynamics. When K_s is reduced, the joint angle accelerations are limited and the whole motion becomes loose and compact. Then the ZMP position is kept so that the robot can remain standing. But too small K_s results in a much different motion from the original ones (Fig.8-b). So finding most suitable and optimal K_s is very important for good motion imitation while keeping balance.

The first idea of our method is maximum K_s during the whole motion sequence as follows:

- (1) Detect frames where ZMP is outside the support area.
- (2) Reduce K_s value on the primitive segments which includes ZMP deviation.

- (3) Iterate the above process until ZMP is inside over the whole motion sequence.

In this process, K_s values are optimized and we achieve a proper motion between similarity and balancing.

We have simulated the motions generated by this method and they have realized balance keeping. Then we have experimented with the real robot (Fig.8-d). It has completed the whole motion standing by itself.

5.3 Clarify the dance motions' poses

The problem of the last algorithm is making the motion 'ambiguous'. This is because such kinds of filter loses the 'Stop Motions' of the dance motions. As described in section 3, the local minimum frames of the target points' velocity represent the borders of primitive motions, and the dancer takes some particular and important body poses on these frames.

Based on this idea, we have proposed a method for clar-

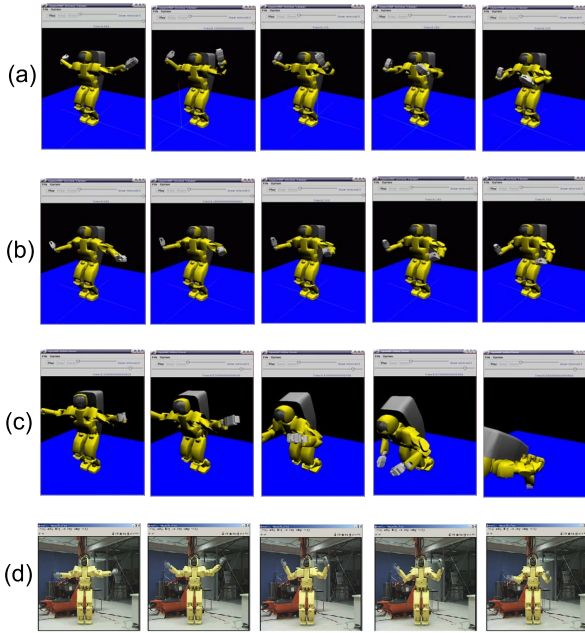


Figure 8: Motion sequences: (a) is the original motion. (b) is the motion generated with small K_s . (c) is same as (a), this falls down on dynamics simulation. (d) is the real humanoid (HRP-1S). Our method enables it to keep standing.

ify the dance motions. The following steps are applied to each joint sequence:

- (1) When the joint angle velocity is not nearby zero at the primitive boundary, force the velocity and the acceleration to be zero.
- (2) In a joint angle graph, find the nearest extreme points both backward and forward from the boundary point. If the distance between them is shorter than a threshold, a value of the boundary point is replaced with the extreme point, and the next extreme point is taken.
- (3) Boundary point is smoothly connected with the extreme points by a proper polynomial equation with constraints that the velocity and the acceleration are zero at the end points. This connection override former values.

Figure 9 shows the results. We have confirmed this method is useful for imitating much human-like dance motions.

6 Conclusion

In this paper, we propose the dance motion imitation method for humanoid robots through visual observation. The human motion is acquired by the motion capture systems that consists of 8 cameras and 8 PCs. Using motion analysis method, we can recognize the structure of the human dance motion and the motion primitives. We also develop the motion concatenation method and show that the original motion is recovered by concatenating the mo-

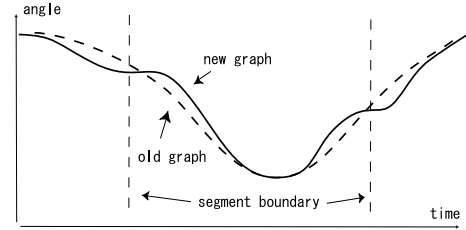


Figure 9: Example of primitive distinguished graph; In the new graph, a gradient is flattened and the velocity is zero at the segment boundary.

tion primitives. For importing human motions into the robot, two problems are there : the limit of the angles and their velocities and the balancing. For later problem, we present a new method that controls the hand's movement by using ZMP trajectories and motion analysis results. Final simulation results proved our method can keep both the robot's balancing and the shape of original motion.

References

- [1] Y.Kuniyoshi et.al, "Learning by watching", IEEE Trans. of Robotics and Automation, No.10, Vol.6, 1994
- [2] O. C. Jenkins, M. J. Metaric, S. Weber: "Primitive based movement classification for humanoid imitation", Humanoids 2000, 2000
- [3] Koichi Ogawara et.al: "Acquiring hand-action models in task and behavior levels by a learning robot through observing human demonstrations", Humanoids 2000, 2000
- [4] Jun Takamatsu et.al: "Symbolic Representation of Trajectories for Skill Generation", ICRA2000, Vol.4 pp.4077-4082, 2000
- [5] T. Inamura et.al: "Imitation and Primitive Symbol Acquisition of Humanoids by the Integrated Minesis Loop", ICRA2001, pp.4208-4213, 2001
- [6] R.F.Rashid: Towards a system for the interpretation of moving light display, IEEE PAMI vol 2, no 6, pp574-581, 1980
- [7] T.Flash, H.Hogan: The coordination of Arm Movements, J. Neuroscience, pp.1688-1703, 1985
- [8] Y.Uno, M.Kawato, R.Suzuki: Formation and Control of Optimal Trajectory in Human Multi-Joint Arm Movement-Minimum Torque Change Model, Biological Cybernetics 61, pp.89-101, 1989
- [9] R. Agrawal et.al: Fast Discovery of Association Rules, Advances in Knowledge Discovery and Data Mining, MIT Press, 1996
- [10] Nancy S. Pollard et.al, Adapting Human Motion for the Control of a Humanoid Robot, ICRA2002, 2002
- [11] Kanehiro,F. et al, Virtual humanoid robot platform to develop controllers of real humanoid robots without porting, Proc. IEEE/RSJ IROS 2001, 2001.
- [12] K. Yokoi, et al: A Honda Humanoid Robot Controlled by AIST Software, Humanoids 2001, pp.259-264.
- [13] Seyoon Tak, Oh-young Song, and Hyeong-Seok Ko: Moiton Balance Filtering, EUROGRAPHICS2000