Motion Recognition and Generation for Humanoid based on Visual-Somatic Field Mapping

[1]Masaki Ogino, [1]Shigeo Matsuyama, [1]Jun'ichiro Ooga, and [1,2]Minoru Asada
[1]Dept. of Adaptive Machine Systems, [2]HANDAI Frontier Research Center,
Graduate School of Engineering, Osaka University
2-1, Yamada-Oka, Suita, Osaka, Japan
{ogino, shigeo, ooga}@er.ams.eng.osaka-u.ac.jp, asada@ams.eng.osaka-u.ac.jp

Abstract

This paper presents a method of imitation learning based on visuo-somatic mapping from observing the demonstrator's posture to reminding the self posture via mapping from the self motion observation to the self posture for both motion understanding and generation. First, various kinds of posture data of the observer are mapped onto posture space by self organizing mapping (hereafter, SOM), and the trajectories in the posture space are mapped onto a motion segment space by SOM again for data reduction. Second, optical flows caused by the demonstrator's motions or the self motions are mapped onto a flow segment space where parameterized flow data are connected with the corresponding motion segments in the motion segment space. The connection with the self motion is straightforward, and is easily acquired by Hebbian Learning. Then, the connection with the demonstrator's motion is automatic based on the learned connection. Finally, the visuo-somatic mapping is completed when the posture space (the observer: self) and image space (the demonstrator: other) are connected, which means observing the demonstrator's posture associcates the self posture. Experimental results with human motion data are shown and the discussion is given with future issues.

1 Introduction

Humanoid robot is expected to have behaviors like human as it is supposed from its appearance. The motion programming for such a robot with multiple joints is a hard task, therefore, imitation is one of the plausible solutions for humanoid motion programming [9]. This attempt has already achieved success to some extent in real robots. Nakazawa et al. [5] have realized a dancing humanoid robot that can imitate human dance performances. They segment human dancing motion into typical motion primitives with parameters. Ijspeert et al. [2] have focused on dynamical aspects of imitation and proposes the methods to describe the observed motion using the basic non-linear dynamics primitives.

On the other hand, imitation is also supposed to a fundamental framework for motion recognition in biological system. Billard and Mataric [1] emphasized importance of motion primitives, and constructed the motion control system based on the CPG modules and the learning modules. Inamura et al. [7] proposed a system that describes the self and the demonstrator's motions in the same mimesis loop, in which motions are recognized and generated in the hidden Markov models.

However, almost existing approaches to imitation in robotics assume that the angles of others' links are available. The somatosensory signals or motion commands of others are not accessible and it is necessary to have a mechanism that converts visual information observing others to self motion. Recently, Kuniyoshi et al. [4] proposed a learning system for early imitation. They suppose the optical flow information is the key to induce the self motion corresponding to the observed motion. However, they didn't mention how the early imitation can be extended to the higher level of learning.

This paper presents a method of imitation learning based on visuo-somatic mapping from observing the demonstrator's posture to reminding the self posture via mapping from the self motion observation for both motion understanding and generation. First, various kinds of posture data of the observer are mapped onto a *posture space* by self organizing mapping [3] (hereafter, SOM), and the trajectories in the posture space are mapped onto a *motion segment space* by SOM again for data reduction. Second, optical flows caused by the demonstrator's motions or the self motions are mapped onto a *flow segment space* where parameterized flow data are connected with the corresponding motion seg-

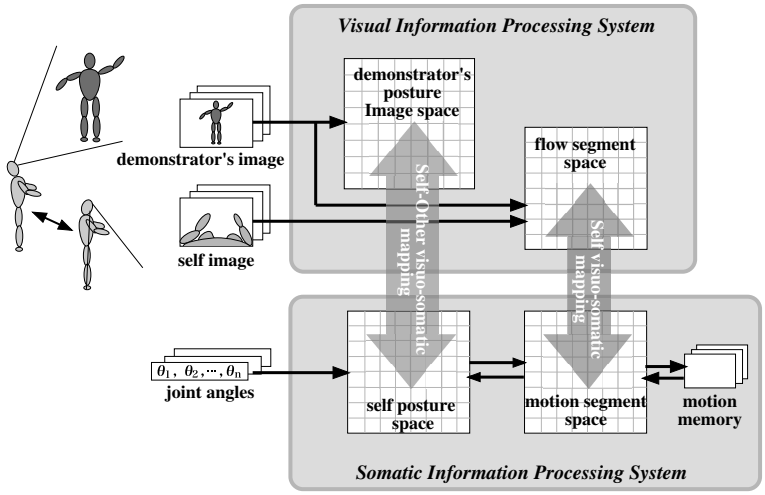


Figure 1. System overview

ments in the motion segment space. The connection with the self motion is straightforward, and is easily acquired by Hebbian Learning. Then, the connection with the demonstrator's motion is automatic based on the learned connection. Finally, the visuo-somatic mapping is completed when the posture space (the observer: self) and image space (the demonstrator: other) are connected, which means observing the demonstrator's posture associates the self posture like a mirror system [10]. Experimental results with human motion data are shown and the discussion is given with future issues.

2 A System Overview

2.1 Basic assumptions

Here, we assume the followings to realize the visual imitation based on the visuo-somatic mapping:

1. No *a priori* knowledge on the link structure, that is, connections between joints.
2. No *a priori* knowledge on the body part (joint) correspondence between the demonstrator and the observer.
3. Both the demonstrator's and the self motions can be observed in terms of a temporal sequence of joint vectors.
4. The joint angles of the self posture can be observed, but no relationship between the self posture and the flow segment space is given.
5. Currently, we focus on the mirror image imitation. This means the right (left) side of the demonstration corresponding to the left (right) side side of the observation.

2.2 Imitation system

Fig. 1 shows the proposed system, consisting of two sub-processing systems, *Visual Information processing system* and *Somatic Information processing system*. In these processing sub-systems, row sensory data are mapped onto the corresponding two dimensional Self-Organizing Maps (SOMs) [3]. The images observing the demonstrator's motion are first mapped onto *image space* which includes the posture image of the demonstrator, and then *flow segment space* in which the changes in posture are represented. The visual feature space is also utilized to represent the self motion, too. On the other hand, the self somatic sensory data are mapped onto *posture space* and their changes are mapped onto *motion segment space*. After generation of these maps independently, the flow segment space and the motion segment space are connected based on Hebbian learning.

The connection between the visual feature space and the motion segment space is easily carried out by using Hebbian learning based on the simultaneous activations of segments in both spaces during the self motions. Once this connection is acquired, the connection between the flow segment space for the demonstrator's motion and the motion segment space is automatic based on the learned connection between the visual feature space and the motion segment space. Through these connections, the mapping from the image space of the demonstrator's posture to the self posture space is enabled, that is, visuo-somatic mapping can be obtained.

In the followings, the details of each sub processing system are explained in section 3, and the mapping among them is shown in section 4.

2.3 Sensory data

We prepare the sensory data by using human motions acquired by a motion capturing system. The captured data, which are three dimensional data sets in the global coordinate system, are converted to the two dimensional data on a virtual camera images captured by the observer (self). The angles between links in a human model are also calculated to be used as the self posture data. Figs. **??**(a) and **??**(b) show the attached place of labels as joints to be captured and a sample of captured data of the demonstrator's motion, respectively. A joint angle vector is mapped onto the self posture space and the segmented trajectories on the map are mapped on motion segment space. The spherical image projection from the camera position at the observer head is assumed to capture the whole self body image. Fig. 3 shows examples of the self body image (a) and the demonstrator's one (b) on the spheres, and their development onto a plane (c).

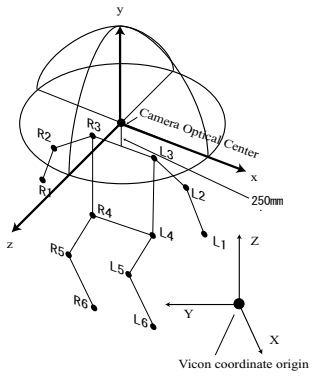
Twelve kinds of motions are captured from the human motion performances. They are combinations of motion, side, and part such as "raise," "wave," and "rotate" as motions, "left," "right," or "both" as sides, and "hand," and "knee" as parts. Also a "walking" motion is added as a whole body motion.

3 Construction of SOMs for behavior recognition and generation

3.1 Posture space and motion segment space

We construct a *posture space* SOM from the *somatic sensation information*, which is the sequence of the vectors consisting of sixteen angles calculated from the captured motion data. The size of *posture space* SOM is 15×15, and it is constructed by 240 [frames] (8 [sec]) per each motion. Fig. 4 (a) shows the resultant SOM.

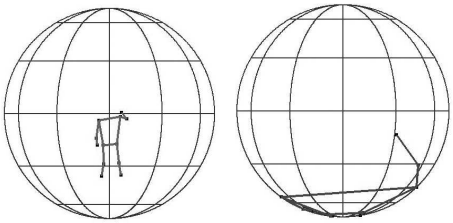
Since the posture data are input sequentially, we can visualize how posture data are connected each other in the posture space. Fig. 4 (b) shows such data indicating that the trajectries of motions are roughly segmented and construct the clusters corresponding to performed actions. These trajectories are divided into small segments, each of which includes 10 [frames] of trajectories on the posture space SOM, and are clustered in the upper layer SOM, motion segment space, (Fig. 4 (c)).



(a) Position of the camera optical center

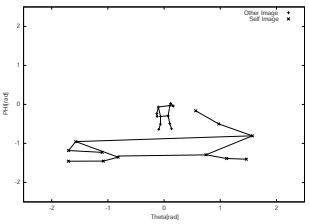
(b) An example of captured motions

Figure 2. Capturing data



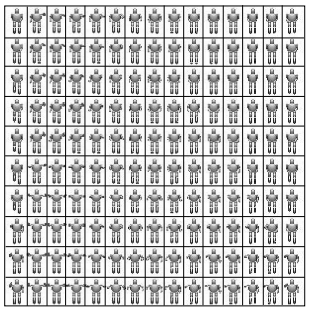
(a) Image of a demonstrator

(b) Image of self body

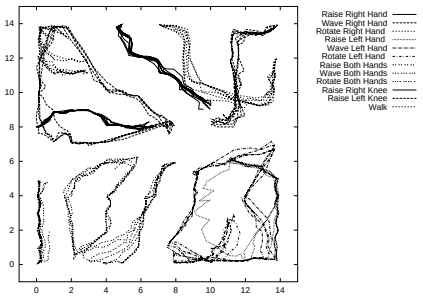


(c) The development of the spherical projection

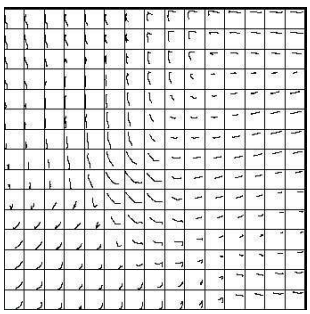
Figure 3. Image data to be input to the system



(a) Posture space



(b) Trajectories of motions on posture space

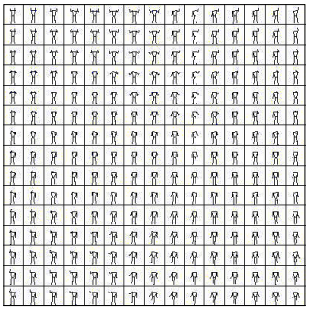


(c) Motion Segment Space

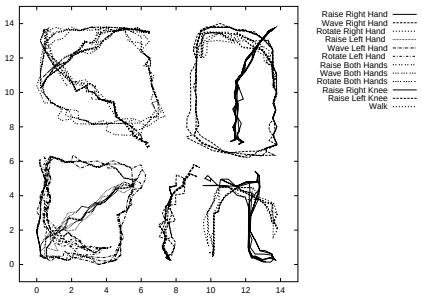
Figure 4. The soms in self somatic sensation information processing system

3.2 Demonstrator's posture image space

A demonstrator's posture image space (hereafter, image space in short) consists of the representative image position vectors obtained by self organizing mapping of image positions of joints of the human model in Fig. **??**. Fig. 5 (a) shows the image space where various postures are clustered into 15×15 representative postures. Similar to the posture space based on the somatic information, we can visualize how posture image data are connected each other in this space. Fig. 5 (b) shows such data which indicate that the trajectories of motions are roughly segmented and construct the clusters corresponding to performed actions.



(a) Demonstrator's posture image space



(b) Trajectories of motions on demonstrators posture image space

Figure 5. A demonstrator's posture image space

3.3 The flow segment space

The problem here is how to associate the observed flow caused by the demonstrator's motion with the self motion. If the flows by the demonstrator's motions are similar to the flows by the self motions, the desired association seems

easy to find because the connection between the observed self motion and the self motion segment can be easily found based on the simultaneous activations during the self motion. However, it would not be so due to the viewpoint difference. Then, as the common features of the flow segments, we chose the direction and the relative position of the flow segments. Fig. 6 indicates the directions of flow segments by the demonstrator (other) and the observer (self), where we can see that the directions are very similar to each other although there are slight differences in the directional changes between them. The relative positions are quantized into four regions (top left, top right, bottom left, and bottom right) by setting the centroid of posture image vectors as the origin. These four regions are called attention areas. Fig. 7 shows that the positions of joints are similar to each other between the demonstrator's and the observer's in spite of large shape difference.

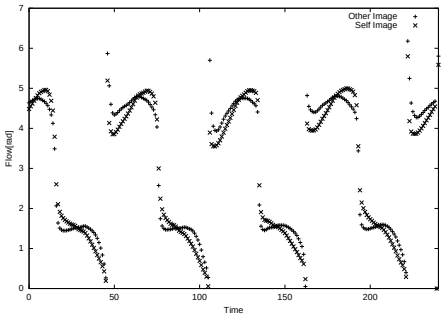
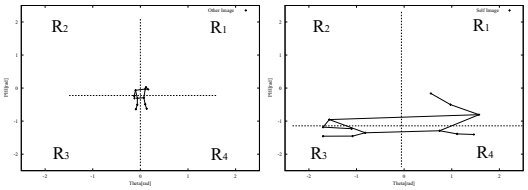


Figure 6. The flow directions from different viewpoints



(a) Attention area of Other Image

(b) Attention area of Self Image

Figure 7. Attention Area

By using the quantized directions and the normalized magnitudes of the flows, and the attention area, the flow segment space is constructed. A data structure for the flow segment space is shown in Fig. 8 (a).

Although Fig. 6 shows the directions of flow vectors in the same parts of self and a demonstrator are almost the same in spite of the camera position, the correspondence between self and a demonstrator's body are unknown. We construct the perceptive field of motion, flow segment space, based on the flow directions and the relative magnitude of flow vectors. The flow segment space has the same number of layers as observed labels (joints), which consists of F_p unit. Each unit has the representative direction correspondent to quantized direction ranging from 0 to 2π (see Fig. 8).

The directions of the flows are segmented when the sign of horizontal or vertical element of flow vector is inverted. In each segment, the directions are averaged. Suppose the time when n-th flow inversion happens is T_n, then the averaged flow direction is given by

$$\S_i^{\phi}(t) = \frac{1}{T_{n+1}-T_n}\int_{T_n}^{T_{n+1}} \phi_i^F(s)ds \quad (T_n < t < T_{n+1}), \qquad (1)$$

where $\phi_i^F(t)$ indicates the flow direction of body segment i at time t. The averaged direction data are sorted by their length of the flow vectors. And the i-th data is assigned to the i-th layer in flow segment space. In each layer, the unit that has the nearest direction to the input data is activated.

3.3.1 Attention area

Although the positions of flow vectors in the robot's view are quite different between the self and the demonstrator, the relative positions among them (upper right, upper left, lower right and lower left) are roughly maintainted well as shown in Fig. 7. Using this feature, the *attentional area* describes what part of the self image includes first flow vectors.

Let N_f the number of observed points of the self and the demonstrator's body and the regions around the center of observed points R_1, R_2, R_3 and R_4 as shown in Fig. 7, then the total flow speed included in each region R_i is given by

$$F_j(t) = \int_{i=1}^{N_f} p_i(t)||v_i(t)||, \qquad (2)$$

$$p_i(t) = \begin{cases} 1 & if \quad u_i(n) \in R_j \\ 0 & else \end{cases} \quad (j = 1, \cdots 4), \qquad (3)$$

where $v_i(t)$ is the observed flow vector, and $u_i(t)$ is the position vector of observed point i at time t. Note that i does not correspond to the labeled point of the body. We define the relative total strength of flow among regions as

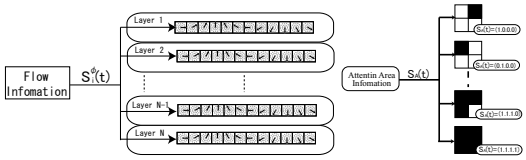
$$A_j(t) = \frac{F_j(t)}{\sum_{n=1}^{4} F_n(t)} \quad (j = 1, \cdots 4). \qquad (4)$$

The input vector to attention area, $S_A(t)$, consists of binarized $A_j(t)$,

$$S_A(t) = (A_1^S(t), A_2^S(t), A_3^S(t), A_4^S(t)) \qquad (5)$$

$$A_j^S(t) = \begin{cases} 1, & if \quad A_j(t) \geq 0.20 \\ 0, & else \end{cases} \qquad (6)$$

Attention area space consists of all the combinations of activated areas, $2^4 = 16$, as shown Fig. 8.



(a) Flow direction (b) Attention Area

Figure 8. The flow segment space

4 Mapping between visual and somatic field

4.1 Self visual-somatic sensation mapping

The simultaneous activations of the units in the flow segment space and the self posture space during self motion make it possible to find correspondence between the units in those spaces (see Fig. 9). The connection coefficients

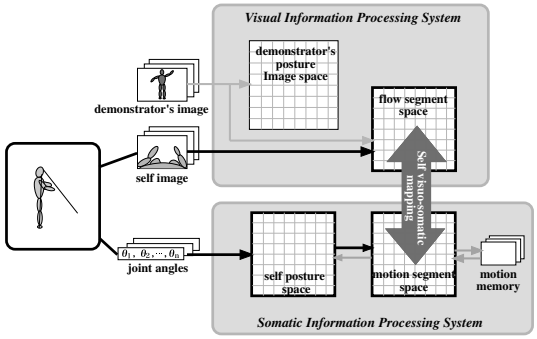


Figure 9. Self visual-somatic sensation mapping

between the units in each space are learned based on Hebbian learning. All the connection coefficients are initialized to 0s, and during the self motion the coefficient, w^{AB}, which is the connection coefficient betwen the i-th unit in space A and the j-th unit in space B, is updated during self motion, as follows,

$$w_{ij}^{AB}(t+1) = w_{ij}^{AB}(t) + \varepsilon(y_i^A(t)y_j^B(t) - y_i^A(t)^2 w_{ij}^{AB}(t)). \quad (7)$$

At the same time, the time sequences of the activated units in motion segment space during various motions are memorized as the motion modules in *motion memory*.

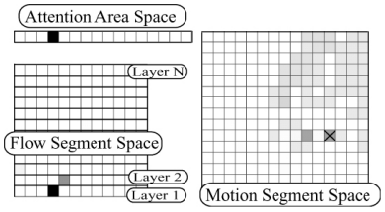


Figure 10. Arrangement of each space

4.2 Recognition of other person's motion

After acquisition of self visual-somatic mapping, the input of image data observing a demonstrator's motion activates the units in the motion segment space through the flow segment space via connections between them (Fig. 11). Let the activation level of the i-th unit in flow direction $y_i^F(t)$ and that of the j-th unit in attention area $y_j^A(t)$, the activation level of the k-th unit in motion segment space, $y_k^M(t)$ is given by

$$y_k^M(t) = \sum_{i=1}^{N_F} w_{ik}^{FM} y_i^F + \sum_{j=1}^{N_A} w_{jk}^{AM} y_j^A \quad (8)$$

The quantization in the flow segment space is coarse and the mapping between the flow segment space and the motion segment space is not one-to-one mapping. The motion of a demonstrator activates multiple units in the motion segment space at a time, which makes it difficult to identify the corresponding motion module. So, we compare the temporal sequences of activated units of observed motion with those of memorized motion modules in the motion segment space. To do that, we define the evaluation function, E_m, which indicates the similarity of the time sequence of acvtiated units of an observed motion to that of the m-th memorized motion as follows,

$$E_m = \max_{s_t} \int_0^T \sum_{i=1}^{N_M} y_i(t) m_i(t_s + t) dt. \quad (9)$$

Thus, the observed motion is recognized as the same as the motion module that maximizes E_m.

4.3 Mapping between a self posture image space and a demonstrator's posture image

Recalling the self motion from the observation of a demonstrator's motion makes it possible to correlate the demonstrator's posture image space in visual information processing system with the self posture space in somatic sensation information processing system (Fig. 12). When observing a demonstrator's motion, the unit in the image

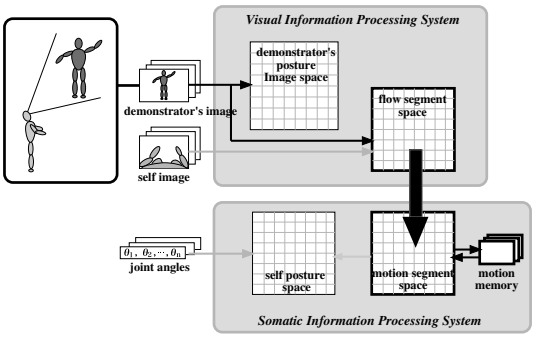


Figure 11. recalling the motion

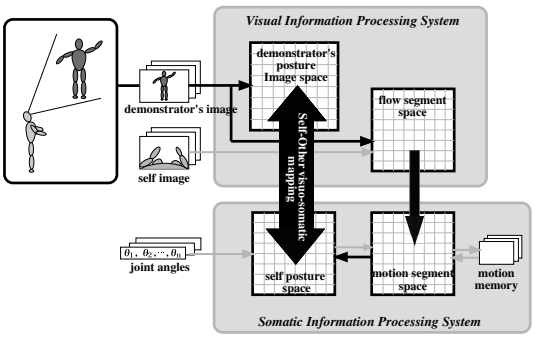


Figure 12. Self-Other Visual-Somatic sensation mapping

space and the unit in the posture space activate simultaneously. So we can use Hebbian learning again between these two maps.

Fig. 13 shows the recalled posture (the rightmost figure) from the observed image (the leftmost figure). The two maps in the middle of the figures describe the activated units in image space(left) and that posture space (right) after hebbian learning.

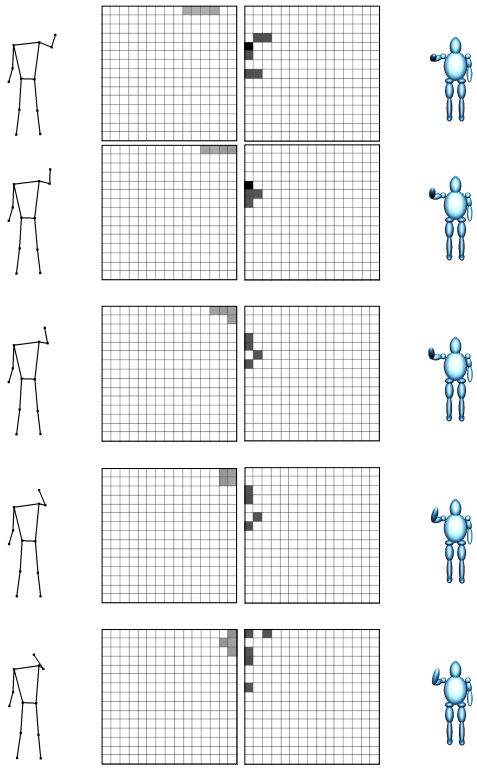


Figure 13. Recalling somatic sensation by Self-Other visual-somatic sensation mapping

5 Conclusions

In this paper, we proposed a learning system for imitation based on visuo-somatic mapping. This system excludes the pre-designed model of a demonstrator as much as possible. The demonstrator's model is made through demonstrator's images in the demonstrator's posture image space. The model of self is not pre-designed, either. It is constructed by self-organizing the self motion information in self posture space and motion segment space. The primitive visual features are related to the representative vectors in motion segment space during self motion. This connection induces the self motion when observing demonstrator's motions and

further mapping between demonstrator's posture image and self posture space is made. After constructing the visuo-somatic mapping, this system can directly activate the self posture corresponding the observed demonstrator's image.

Although initial aims to construct the visuo-somatic mapping through learning are accomplished in this system, it has many problems for practical use as an imitation system. First this system assumes that an observer always stands face to face with a demonstrator, and this sytem does not have concept about the translation or rotation to the ground of the demonstrator. An observer can recognize only jestures of the demonstrator. Second, the resultant visuo-somatic mapping is not so accurate as to make new motion modules only from observation of demonstrator's motion, because the sequence of the activated postures is not smooth.

For the first problem, we are now extending our model so that it can describe the transition and rotation of a demonstrator relative to the ground. The second problem can be solved by using velocity information acquired by another pathway. Acquiring new motions which are not experienced through observation is the next challenge for us.

Acknowledgments

This study was performed through the Advanced and Innovational Research program in Life Sciences from the Ministry of Education, Culture, Sports, Science and Technology of the Japanese Government.

References

[1] A. Billard, M. Mataric: Learning human arm movements by imitation: Evaluation of a biologically inspired connectionist architecture, *Robotics and Autonomous Systems*, 37, pp.145–160, 2001.

[2] A. Ijspeert, J. Nakanishi, S. Schaal, Movement Imitation with Nonlinear Dynamical Systems in Humanoid Robots, In *Proc. of IEEE International Conference on Robotics and Automation*, pp. 1398-1403, 2002.

[3] T. Kohonen: The Self Organization and Associative Memory, Springer-Verlag, 1989.

[4] Y. Kuniyoshi, Y. Yorozu, M. Inaba, H. Inoue: From visuo-motor self learning to early imitation -a neural architecture for humanoid learning-, In *Proc. of IEEE International Conference on Robotics and Automation*, pp. 3132-3139, 2003.

[5] A. Nakazawa, S. Nakaoka, K. Yokoi, T. Harada, and K. Ikeuchi: Imitating human dance motion through motion structure analysis, In *Proc. of 2002 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 2539–2544, 2002.

[6] E. Oja, J. Karhunen: On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix, *J. Math. Anal. and Appl.*, 106, pp. 69–84, 1985.

[7] T. Inamura, I. Toshima, Y. Nakamura: Acquisition and Embodiment of Motion Elements in Closed Mimesis Loop, In *Proc. of the 2001 IEEE International Conference on Robotics and Automation*, pp. 1539–1544, 2002.

[8] R. Pfefer and C. Scheier.: Understanding Intelligence, MIT Press, Cambrridge, Massachusetts, 1999.

[9] S. Schaal., "Is imitation learning the route to humanoid robots?," *Trends in Cognitive Science*, pp. 233–242, 1999.

[10] V. Gallese and A. Goldman., Mirror neurons and the simulation theory of mind-reading, *Trends in Cognitive Science*, Vol. 2, No. 12, pp. 493-501. 1998.