Acquiring Motion Elements for Bidirectional Computation of Motion Recognition and Generation

Tetsunari Inamura[12], Iwaki Toshima[3], and Yoshihiko Nakamura[21]

[1] CREST
Japan Science and Technology Corporation
Saitama, Japan
inamura,nakamura@ynl.t.u-tokyo.ac.jp
[2] Department of Mechano-Informatics
University of Tokyo
7-3-1, Hongo, Bunkyo-ku, Tokyo, Japan
[3] Nippon Telegraph and Telephone Corp
Kanagawa, Japan

Abstract. Mimesis theory is one of the primitive skill of imitative learning, which is regarded as an origin of human intelligence because imitation is fundamental function for communication and symbol manipulation. When the mimesis is adopted as learning method for humanoids, convenience for designing full body behavior would decrease because bottom-up learning approaches from robot side and top-down teaching approaches from user side involved each other. Therefore, we propose a behavior acquisition and understanding system for humanoids based on the mimesis theory. This system is able to abstract observed others' behaviors into symbols, to recognize others' behavior using the symbols, and to generate motion patterns using the symbols. In this paper, we mention the integration of mimesis loop, and confirmation of the feasibility on virtual humanoids.

1 Introduction

The research of humanoid robots has a long history and has accumulated a substantial amount of literature. The focus of early efforts was mostly on the dynamics and control of biped walk. Although it has not yet reached the level of complete solution with full of liability and adaptability, the hardware technology has been established for building autonomous humanoids.

Recently, the human behavioral science and the human intelligence have become conspicuous as a real research issue of robotics. Although the motivation of the artificial intelligence originated there, the physical limitations have forced or justified the researchers to carry on their research in a limited scope and scale of complexity. It ought to be the major challenge of contemporary robotics to study robotic behaviors and intelligence in the full scale of complexity mutually sharing research outcomes and hypotheses with the human behavioral science and human intelligence.

The discovery of mirror neurons[1] have been a notable topic of brain science on the background of such issues. Mirror neurons, which have been found in primates' brain and humans' brain, fire when the subject observes a specific behavior and also fire when the subject start to act the same behavior. The fact suggests that the

B. Siciliano and P. Dario (Eds.): Experimental Robotics VIII, STAR 5, pp. 372–381, 2003

behavior recognition process and behavior generation process are combined as same information processing scheme. We believe that a paradigm can be proposed taking advantage of the mirror neurons, with considerations of Deacon's contention[2] that the language and brain had evolved each other; and Donald's contention [3] that the symbol manipulation and communication ability are the remarkable feature of the primates.

So far, many researcher have tackled with the issues between the imitation learning for humanoids and human intelligence[4][5]. In feedback error learning proposed by Kawato et al[6], dynamics in observed behavior are acquired, however, consideration for symbolization is not well studied. There are some suggestions that module structure of basic motions is needed for the symbolization and representation of complex behavior, such as Schaal's work[4]. In Kuniyoshi's approach[7], robots can reproduce complex behaviors from observation of human demonstration with the abstraction and symbolization, however, it is difficult to be applied to general recognition and reproduction process because of lack of dynamics point of view. Samejima *et al* have proposed an imitation learning framework with symbolization modules[8]. In this case, a premise is set that sequence of symbol is given from others by communication, thus a certain representation model for dynamics of the whole body motion are needed.

We have proposed "mimesis model"[9] as a framework which can abstract whole body motion into symbol representations, generate motion patterns from the symbols, and recognize total behavior. In other words, it is embodiment of the mirror neurons and mimesis theory from an engineering point of view. However, the framework is formed on the assumption that appropriate motion elements exist in advance. In this paper, we discuss how to evolve the mimesis model. Appropriate motion elements are ought to be acquired during imitation process because the elements are fundamental factors for both behavior recognition and generation. First, we introduce the mimesis model in Sec.2, and propose the acquisition method in Sec 3.

2 Outline of mimesis model

2.1 Motion elements and HMMs

The *mimesis model* consists of two phases as shown in Fig.1. In the first half, observed motion patterns are analyzed into motion elements, and the dynamics in the sequence of the elements is abstracted using by symbol representations. We call the symbol representations as *proto-symbols*. Motion elements are fundamental pieces of motion, which hold joint angle $\boldsymbol{\theta}$ and angular velocity $\dot{\boldsymbol{\theta}}$ for short period of time as follows:

$$e_i = \begin{pmatrix} \boldsymbol{\theta} \\ \dot{\boldsymbol{\theta}} \end{pmatrix} \tag{1}$$

In the latter half, a sequence of motion elements is decoded from a proto-symbol. However, the generated motion patterns would be inappropriate for real humanoids.

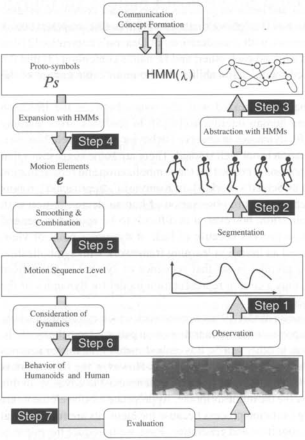

Fig. 1. An outline of proposed mimesis model

For the issue, we adopt an approach where motion elements are modified based on evaluation of generated behavior on a humanoid

For the transformation between motion patterns and the proto-symbols, Hidden Markov Models (HMMs) are adopted as mathematical backbone. HMMs are one of stochastic processes which takes time series data as an input, then outputs a probability that the data is generated by the HMM. In this framework, state transition is performed probabilistically and some labels $\boldsymbol{o}_i$ are outputted during the transition as shown in Fig.2. $\boldsymbol{\lambda}$ indicates the property of probabilistic dynamics of the HMM; which consists of state transitions probability matrix $\boldsymbol{A}$, output probability matrix $\boldsymbol{B}$, and initial condition $\boldsymbol{\pi}$. Here, we define the proto-symbols as the HMM parameter $\boldsymbol{\lambda}$, and also define the motion elements $\boldsymbol{e}_i$ as output labels $\boldsymbol{o}_i$.

The mimesis model can recognize others' motion and generate self-motion. Therefore, the use of HMM can be regarded as a bidirectional computational model

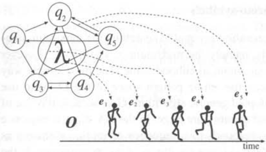

Fig. 2. Motion Elements and Hidden Markov Models

of mirror neurons. From now on, recognition, generation, and learning processes are explained.

2.2 Creating proto-symbols through observation

Observed behavior are transfered into sequence of the motion elements, with segmentation. After that, a parameter of a HMM ($\boldsymbol{\lambda} = \{\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi}\}$) which output the sequential elements plausibly is calculated and registered as a proto-symbol $\mathcal{P}_S$. Humanoids gather several motion patterns as a stock of observed data for the learning. In the case of an unknown motion is input, the robot creates a new HMM. $\boldsymbol{\lambda}$ can be calculated by Baum-Welch algorithm which is one of EM-algorithms[12].

2.3 Motion recognition using proto-symbols

To recognize others' motion, observed motions are transformed into sequence of motion elements $\boldsymbol{O} = (\boldsymbol{o}_1, \boldsymbol{o}_2, \ldots, \boldsymbol{o}_t)$, and a parameter $P(\boldsymbol{O}|\mathcal{P}_S)$ is calculated. This parameter indicates a probability that a motion pattern $\boldsymbol{O}$ is generated by a proto-symbol $\mathcal{P}_S$. This value is called as likelihood, calculated by forward algorithm[12].

Each proto-symbol is corresponded to each motion, thus, likelihood values of the input motion against each proto-symbol are calculated. A proto-symbol which corresponds to input pattern should indicate high likelihood, and other proto-symbols ought to indicate low likelihood. In order to distinguish these likelihoods, following criterion is introduced.

$$R = \frac{\max(P(\boldsymbol{O}|\mathcal{P}_{Si}))}{second(P(\boldsymbol{O}|\mathcal{P}_{Si}))} \tag{2}$$

Where $second(\mathbf{x})$ means that the second highest value in the components of $\mathbf{x}$. Mimesis model recognizes the input motion without any confusion when the R indicates high value. In this case, the recognition result becomes $\mathcal{P}_{Sj}$, where

$$j = \arg\max_i P(\boldsymbol{O}|\mathcal{P}_{Si}) \tag{3}$$

When the R indicated low value, the recognition was failed and Mimesis model tries to shift to proto-symbol creation phase.

2.4 Motion generation using proto-symbols

In mimesis framework, motion generation is equal to searching the maximum likelihood of motion using by HMMs, namely the maximum of $P(O|\mathcal{P}_S)$. However there is no algorithm to find the maximum likelihood value. The most simple way to search the best pattern is to scan the whole pattern space of the $\boldsymbol{O}$, or to use Monte-Carlo method. Here, we adopted generic algorithm (GA) because the size of the search space and computational quantity are huge. In the GA stage, a sequence of genes, that is, chromosome corresponds to a sequence of motion elements as shown in Fig.3, the likelihood $P(O|\mathcal{P}_S)$ is used as fitness of the chromosome. In the mutation phase, transposition process is adopted in order to let the searching speed high, because keeping a series of motion elements as a block leads to keeping of high fitness[9].

Finally, the sequential motion elements are transformed into continuous joint angle representations, and modified with *Dynamics Filter*[10], which considers the feasibility of the motion from dynamics point of view.

2.5 Problems on the usual mimesis model

Usual mimesis model mentioned above have a fatal problem that there are no method to decide the motion elements. The performance of motion recognition and generation is influenced by the motion elements, therefore, the humanoid ought to acquire motion elements with the consideration of embodiment through repetition of motion perception and generation.

We adopted an approach that the system searches the best motion elements with an evaluation criterion whether the generated motion corresponds with the purpose of original motion.

3 Acquisition of motion elements through repetition of motion observation and generation

3.1 Introduction of Continuous HMMs and applying to mimesis model

In this phase, We introduce continuous HMMs[12], which is a kind of HMMs which can treat continuous multi-dimensional data. The difference between normal discrete

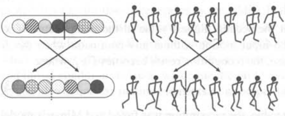

Fig. 3. Chromosome description of motion sequence and transposition of gene

HMMs (DHMMs) and continuous HMMs (CHMMs) is that the transition process outputs continuous multi-dimensional vectors. Therefore, output probability matrix $\boldsymbol{B}$ becomes probability density functions in CHMMs. Here, the density function is approximated with linear combination of Gaussian functions as follows.

$$P_i(\boldsymbol{o}) = \sum_{j=1}^{m} c_{ij}\mathcal{N}_{ij}(\boldsymbol{o}; \boldsymbol{\Sigma}, \boldsymbol{\mu}) \tag{4}$$

Where $P_i(\boldsymbol{o})$ is probability density function for output of continuous vector $\boldsymbol{o}$ at i-th state node, $\mathcal{N}(\boldsymbol{o}; \boldsymbol{\Sigma}, \boldsymbol{\mu})$ is the Gaussian function:

$$\mathcal{N}_{ij}(\boldsymbol{o}; \boldsymbol{\Sigma}, \boldsymbol{\mu}) = \frac{1}{\sqrt{(2\pi)^D |\boldsymbol{\Sigma}_{ij}|}} \exp\{-\frac{1}{2}(\boldsymbol{o} - \boldsymbol{\mu}_i)' \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{o} - \boldsymbol{\mu}_i)\} \tag{5}$$

$\boldsymbol{\Sigma}$ is covariance matrix, $\boldsymbol{\mu}$ is mean vector, m is the number of mixture Gaussian functions, c_{ij} is mixture coefficient, and D is the number of dimension of continuous vector $\boldsymbol{o}$.

The property of the CHHHs are decided by parameter $\{\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{c}, \boldsymbol{\Sigma}, \boldsymbol{\mu}\}$. These parameters are calculated using Baum-Welch algorithm .

Here, each mean vector of the Gaussian function is regarded as important representation of the observed motion. Therefore, we divide the parameters of CHMMs, and redefine the *motion elements* $\boldsymbol{e}$ as follows.

$$\boldsymbol{e} \stackrel{\text{def}}{=} \{\boldsymbol{\Sigma}, \boldsymbol{\mu}\} \tag{6}$$

In other words, the number of motion elements is as many as the number of mixture Gaussian components. An important issue is that the motion elements are automatically calculated by the Baum-Welch algorithm as well as mentioned in Sec.2.2. In previous works[9], the number of motion elements m is fixed, and entrusted to the developer, however, in CHMMs, it is easy to change the m according to complexity of the motion.

Motion elements could be regarded as filter between continuous motion representation and discrete motion representation. When continuous motion would be transferred into discrete motion, $\boldsymbol{e}_i$ where $i = \arg\max_j \mathcal{N}_j(\boldsymbol{o})$, is adopted as a typical motion element for each time period. When discrete motion would be transfered into continuous motion, sequence of $\boldsymbol{\mu}_i$ is used directly.

To sum up, mimesis system can take following advantages with CHMMs.

- Motion elements are able to express the whole body motion, therefore, various motion patterns are available easily.
- Parameters of motion elements are automatically calculated.

3.2 Hybrid Hidden Markov Model

Although many advantages are available, CHMMs have a disadvantage that huge computational quantity is needed. It should take much time for motion generation and

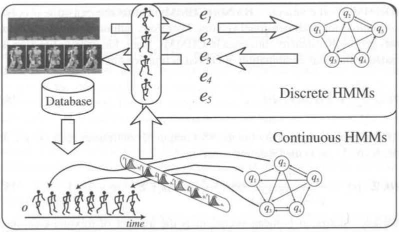

Fig. 4. A model of Hybrid Hidden Markov Model

Fig. 5. Motion Capturing System: step motion for learning data

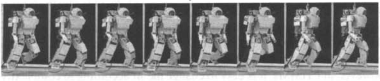

Fig. 6. A result of step motion generation on a humanoid robot

recognition. Therefore we propose a hybrid Hidden Markov Model which consists of CHMMs and DHMMs as shown in Fig.4.

In motion recognition and generation phase, DHMMs are used which computational quantity is little. In motion elements acquisition phase, CHMMs are used which computational quantity is large.

3.3 Closing the mimesis loop for embodiment

Mimesis models decide *motion elements* and *proto-symbols* according to the observed motions, therefore, when the physical condition of learner differs from the one of teacher, acquired motion elements are influenced by teacher's motion, and are unsuitable for learner's body. Here, we introduce loop structure of mimesis model in order to absorb the difference.

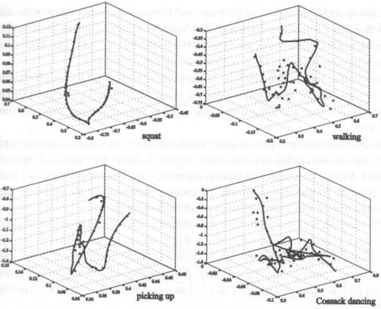

Fig. 7. A result of motion elements acquisition against four kinds of motion

Humanoids collect motion patterns during the observation, in addition to that, generated motions are added to the database when they are fit for the real humanoid body. Using the CHMMs and the database which is formed during observation-generation loop, suitable motion elements and whole body motions are able to be acquired with consideration of humanoids embodiment.

We use evaluation criterion for first condition as follows:

$$E_{\theta} = \frac{1}{T}\int_0^T |\boldsymbol{\theta}_{in}(t) - \boldsymbol{\theta}_{out}(t)| dt \tag{7}$$

Where $\boldsymbol{\theta}_{in}(t)$ and $\boldsymbol{\theta}_{out}(t)$ indicate the joint angle of an observed ideal motion and a generated motion. For second condition, we adopted following criterion:

$$E_R = R^{i+1} - R^i \tag{8}$$

where R^i indicates recognition ratio in i-th learning loop. When the sign of E_R is plus, the mimesis system judges that i-th motion data is suitable for recognition process, adds the motion data into database, and calculates the motion elements again.

4 Experiments of motion elements acquisition

We have confirmed the performance of our method by experiments where the mimesis model observes humans' motion and generate motions for a real humanoid. Four kinds of motions; walking, squat, picking up, and Cossack dancing, were recorded using by the *Behavior Capturing System* [13]. Using the system, joint angle data for 20 DOFs are directly observed because the DOFs of the humanoid is 20. The time period of each motion is about 2[sec] with sampling time 20[msec].

The dimension of motion elements is 20, however, it is difficult to describe such a complex vector in figure, therefore we put the motion elements in charge of three joint; hip(pitch axis), knee and ankle(pitch axis), that is, the dimension of the motion elements is three.

After the 50 observations, motion generation process is executed 50 times, and appropriate motions are added into database. Figure 7 shows the acquired motion elements. Each dot mark indicate the motion element, solid line indicate the original motion's trajectory. As the figure shows, the motion elements are located near the original motion, that is, our method shows good performance.

Using the method, we also performed the imitation learning experiment against humans' walking motions as shown in Fig5. Fig.6 shows experimental result on a humanoid robot. The generated motion is suitable for the real robot because the acquisition process considers the embodiment of the humanoid. The experiments are executed with its body dangled because any balance controller are not considered in current stage.

5 Conclusions

We proposed *mimesis model* using *motion elements* and *proto-symbols* in order to realize a computation model which leads to the integration of imitation learning and symbol emergence. In this framework, there was a problem that fundamental motion elements are embedded by developer in advance. In this paper, we introduced a method for acquisition of motion elements during motion observation and generation based on hybrid HMMs, that is integration of continuous HMMs and discrete HMMs. For the future works, a language development model in which humanoids try to make communications with others, and build a relationship representation between proto-symbols and linguistic symbols, is planned. We believe that this approach leads to build a intelligent system which connect humanoids intelligence and behavior science.

Acknowledgement.
This research was supported by the "Robot Brain Project" under the Core Research for Evolutional Science and Technology (CREST program) of the Japan Science and Technology Corporation. The authors would like to thank to Mr. K.Kurihara for the use of *Behavior Capturing System*.

References

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

2. Terrence W. Deacon. *The symbolic species*. W.W. Norton & Company. Inc., 1997.
3. Merlin Donald. *Origins of the Modern Mind*. Harvard University Press, Cambridge, 1991.
4. Stefan Schaal. Is imitation learning the way to humanoid robots? *Trends in Cognitive Sciences*, Vol. 3, No. 6, pp. 233–242, 1999.
5. Maja J Mataric. Getting humanoids to move and imitate. *IEEE Intelligent Systems*, pp. 18–24, 2000.
6. M. Kawato, K. Furukawa, and R. Suzuki. A hierarchical neural-network model for control and learning of voluntary movement. *Biological Cybernetics*, Vol. 57, No. ?, pp. 169–185, 1987.
7. Yasuo Kuniyoshi, Masayuki Inaba, and Hirochika Inoue. Learning by Watching: Extracting Reusable Task Knowledge from Visual Observation of Human Performance. *IEEE Transaction on Robotics and Automation*, Vol. 10, No. 6, pp. 799–822, 1994.
8. Kazuyuki Samejima, Ken'ichi Katagiri, Kenji Doya, and Mitsuo Kawato. Symbolization and imitation learning of motion sequence using competitive modules. *The trans. of the institute of electronics, information and communication engineers*, Vol. J85-D-II, No. 1, pp. 90–100, 2002.
9. Tetsunari Inamura, Yoshihiko Nakamura, Hideaki Ezaki, and Iwaki Toshima. Imitation and primitive symbol acquisition of humanoids by the integrated mimesis loop. *the Proc. of IEEE Int'l Conf. on Robotics & Automation*, pp. 4208–4213, 2001.
10. Katsu Yamane and Yoshihiko Namamura. Dynamics Filter–Concept and Implementation of On-Line Motion Generator for Human Figures. *the Proc. of IEEE Int'l Conf. on Robotics & Automation*, pp. 688–695, 2000.
11. Qiang Huang, Yoshihiko Nakamura, and Tetsunari Inamura. Humanoids walk with feedforward dynamic pattern and feedback sensory reflection. *Int'l Conf. on Robotics & Automation*, pp. 4220–4225, 2001.
12. Steve Young et al. *The HTK Book*. Microsoft Corporation, 2000.
13. Kazutaka Kurihara, Shin'ichiro Hoshino, Katsu Yamane, and Yoshihiko Nakamura. Optical motion capture system with pan-tilt camera tracking and realtime data processing. *the Proc. of IEEE Int'l Conf. on Robotics & Automation*, pp. 1241–1248, 2002.