Two main issues arise in practical imitation learning by humanoid robots observing human behavior—the first is segmenting and recognizing motion demonstrated naturally by a human being and the second is utilizing the demonstrated motion for imitation learning. Specifically, the first involves motion segmentation and recognition based on the humanoid robot motion repertoire for imitation learning and the second introduces learning bias based on demonstrated motion in the humanoid robot's imitation learning to walk. We show the validity of our motion segmentation and recognition in a practical way and report the results of our investigation in the influence of learning bias in humanoid robot simulations.

Keywords: motion segmentation and recognition, imitation learning, learning bias, humanoid robot, via-point representation

1. Introduction

Humanoid robots made commercially show many kinds of behavior, mostly through graphical user interfaces used to create new motion based on via-point representation. A user defines a set of via-points, each representing a set of joint angles defining the posture of a humanoid robot. The robot follows via-points one by one to generate a specific motion. The process of designing and modifying motion using via-point representation is, unfortunately, difficult and time-consuming, so imitation learning has draw attention [1–5]. These enable users to add and design new robot behavior. Most imitation learning for these robots focuses on mimicking the trajectory of joint angles for realizing the demonstrated behavior. This is insufficient, however, for imitating demonstrated motion by following the joint angle trajectory of demonstrated motion precisely because robot link structure and dynamics differ from those of a human demonstrator. Robot walking needs stability to keep the robot upright while following the demonstrated trajectory of joint angles. Overcoming the imitation gap between robots and human demonstrators requires additional learning to adapt imitated motion to robot body dynamics. Miyamoto and Kawato [1] show imitation learning through via-point representation for a tennis serve and upswing of robot arm. Via-point representation is widely used in many robots because it needs smaller parameters than other representations representing complex motion.

Reinforcement learning has been studied widely in robotics [6–8] because it enables a robot to acquire purposeful behavior through trial and error. Applying simple, naive reinforcement learning to a robot, however, requires long learning time and high cost making it desirable to introduce imitation into reinforcement learning to reduce learning time and unnecessary trial and error. Takahashi et al. [9] showed that a wheeled soccer robot enhances behavior acquisition and recognition based on the interaction between behavioral learning and observation. A robot learns several ways, including trial and error and reading rewards of observed teammate behavior. Hama-hata et al. [10] proposed designing reward system based on observing a human demonstration and applying it to learning on an inverted pendulum robot. These studies use reinforcement learning based on estimated state values. Value-based reinforcement learning, however, may involve issues such as the curse of dimensionality and relatively long learning time.

Policy gradient methods without state value estimation have the advantages of handling continuous state space, a high degree of freedom and a short learning time, although they do not have guarantee of convergence to the optimal policy. Kohl and Stone [11] shows the validity of the policy gradient method with a real 4-legged robot. Conventional approaches based on policy gradient based
reinforcement learning for imitation use a demonstrated motion trajectory to initialize the motion parameters to be learned. It is important to find good initial parameters when setting an optimal policy within reasonable learning time. Takahashi et al. [12, 13] show that an inverted-pendulum mobile robot acquires a good set of initial parameters from observing a demonstration by a human being for a robot learning a kicking motion. Miyamoto and Kawato [1] used a demonstrated trajectory as an initial parameter set of via-points for learning, but it is not clear how a demonstrated motion affected learning performance if it is used as a bias of reinforcement learning as imitation learning rather than as initial learning parameters.

A further point in imitation learning is available training data. Most imitation learning studies [2–5] assume that demonstrator’s joint angles are available as training data. Optical or mechanical motion capture is necessary for measuring joint angles of a demonstrator. This tends to be expensive and time-consuming preparation, however. Powerful inexpensive motion capture systems are available, such as KINECT from Microsoft Corporation. The KINECT device measures the positions of human joints but does not provide joint angle information directly. It is difficult to solve inverse kinematics in calculating joint angles from joint positions because of redundancy and differences in body link structure between human demonstrators and humanoid robots. Imitation learning should avoid unnecessary cost for solving inverse kinematics to acquire joint-angle trajectories from demonstrations. Imitation learning in a practical situation does not require a human demonstrator to declare the start and the end of a particular motion and the robot must segment and recognize motion itself.

This paper proposes two methods for practical imitation learning: (1) motion segmentation and recognition based on the motion repertoire of the humanoid robot for imitation learning, and (2) imitation learning based on a reasonable bias for a humanoid robot with policy gradient based reinforcement learning. (1) recognizes both the motion that the demonstrator shows and the cycle period and phase if a human demonstrator shows a periodic motion. (2) required that we investigate learning performance based on bias of motion automatically segmented in (1). A humanoid robot observes a human walking with motion capture and acquires the joint trajectory. Motion data based on the motion repertoire of the humanoid robot extracts only a walking sequence for imitation learning. Learning bias is designed with body link posture trajectories calculated by joint trajectory. Learning performance is investigated by strictly checking the weight of learning bias during imitation learning. Results of experiments showed that motion segmentation operates appropriately and indicates that learning bias contributes to better learning parameter exploration and motion stability during learning.

2. Overview

Figure 1 shows the imitation learning flow. The humanoid robot first observes a human demonstration with a motion capture device, such as the KINECT sensor. The device records a sequence of a human demonstrator’s joint positions in camera coordinates. Because it is difficult to estimate joint angles uniquely from information on the joint positions of the body, link postures are used to compare human and robot motion.

The recorded sequence of such link postures includes motions such as standing, walking, bowing, and hand waving. This sequence is divided into motions based on robot motion repertoires in which individual motions are defined using the via-point representation of joint angles. Motion segmentation is based on difference in link posture sequences between the human being and the robot. Motion that shows the smallest difference is labeled in the sequence of the human motion demonstrated.

The segmented sequence of human link postures is introduced in motion learning as a bias for imitating behavior. Basically, reinforcement learning is applied to improve motion using the body dynamics of the robot itself by updating via-points of joint motion angles. The learning bias derived from the segmented human motion sequence is used to improve learning performance similar to the way in which a trainer shows a model motion to a trainee.

3. Motion Recognition and Segmentation

A learning robot must recognize and segment observed motion demonstrated by a human being to choose which motions to learn as shown in Fig. 2 overview. We assume that our humanoid robot has a set of basic motions to be improved by imitation learning. The robot has 5 motions – standing, walking, bowing, waving the right hand, and
waving the left hand – each of which is described using via-point representation, and the sequence of link postures during motion is recorded beforehand. One of the modules shown in Fig. 2 is assigned to one motion and its link posture sequence.

The motion capture system is used to observe the human demonstration, measures human joint positions, and converts position data to the robot coordinate system, making the link posture sequence available. The link posture sequence is compared to that of each robot motion and the similarity of link posture sequence is calculated. The motion with the biggest similarity is regarded as the motion demonstrated by the human being.

3.1. Link Posture Estimation

It is difficult to determine joint angles from the sequence of observed joint positions because of redundancy in degrees of freedom. It takes time to find a reasonable joint angle sequence because the exploration space is huge. To avoid the issue, we use link posture estimation instead of joint angle estimation because link posture is calculated uniquely without the redundancy problem.

Figure 3 shows joints detected by KINECT and links whose postures are calculated with joints. Our KINECT-based motion capture system detects 16 joints. Links are defined and conducted using adjacent joints as shown in Fig. 3.

3.2. Similarity Based on Link Posture

The link postures sequence of human demonstrator $h\phi_i(t)$ is recorded by the motion capture system. Robot $r\phi_i(t)$ is recorded while the robot is actually executing the motion based on the joint trajectory represented by viapoints. The similarity of link postures $h\phi_i(t)$ and $r\phi_i(t)$ is defined as the inner product of these postures, that is, $h\phi_i(t) \cdot r\phi_i(t)$. The similarity of the two link posture sequences of the human demonstrator and the robot is defined as follows.

$$S = \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} h\phi_i(t) \cdot r\phi_i(t)$$

where $N$ and $T$ indicate the number of links and size of the sequence.
3.3. Periodic Motion Detection

The human demonstrator often shows periodic motion such as, walking, or waving hands. A sequence of motion data must be segmented into one cycle and the phase and period should be defined. The autocorrelation function reasonably finds the demonstrated motion period. The autocorrelation function in demonstrated motion period should be defined. The autocorrelation function data must be segmented into one cycle and the phase and period of demonstrated motion as shown in Fig. 5(b) shows an example of sequences of one cycle of observed motion demonstrated by (a) a human being and (b) autocorrelation function G(τ) of sequences.

\[
G(\tau) = \frac{1}{NT} \sum_{t=0}^{T} \sum_{i=0}^{N-1} \overline{\phi_i(t)} \phi_i(t + \tau) \quad \cdots \quad (3)
\]

\(\tau\) indicates the shift time, \(T\) the length of the window for calculating the autocorrelation, and \(N\) the number of links. Fig. 5(a) shows a schematic example of the sequence of the observed motion demonstrated by a human being. The black line indicates the original sequence of the observed posture of a link. The grey line shows the sequence shifted with time \(\tau\). The autocorrelation function shows a local maximum for \(\tau\) that is the period of the demonstrated motion as shown in Fig. 5(b). \(\tau\) showing the first local maximum of \(G(\tau)\) excluding \(\tau = 0\) is defined as the cycle period of demonstrated motion \(\tau_m\).

The phase is important for recognizing the state of motion correctly. Phase \(\alpha\) is calculated using the correlation between human and robot motions with time shift \(\alpha\). The correlation function is shown in Eq. (4).

\[
S(\alpha) = \frac{1}{NT} \sum_{t=0}^{T} \sum_{i=0}^{N-1} \overline{\phi_i(t)} \phi_i(t + \frac{\tau}{P} + \alpha) \quad \cdots \quad (4)
\]

\(P\) indicates the period of one-cycle robot motion. Fig. 6(a) shows an example of sequences of one cycle of observed human motion data and robot motion data. Similarity function \(S(\alpha)\) is maximum if phase \(\alpha\) matches the two motion data as shown in Fig. 6(b). Phase \(\alpha\) showing maximum \(S(\alpha)\) is regarded as the demonstrated motion phase \(\alpha_m\).

3.4. Trajectory Based on Via-Point Representation

The data size for representing the trajectory of raw joint angles of one motion is too huge to learn, so via-point representation is widely used to reduce learning parameter size [1]. Fortunately, many commercially available humanoid robots, including Robovie-PC produced by Vstone Co., Ltd. [14] which we use in our experiments, use via-point representation to design motion. Fig. 7 shows an example of via-point representation for walking. Each number in the figure indicates the index of a via-point. Humanoid skeletons adjacent to numbers indicate robot postures at via-points. A set of via points represents points on the trajectory of motion. A point between via-points is interpolated with, for example, spline interpolation or cubic curve interpolation. We use cubic curve interpolation in this paper because our robot uses it.

One motion is defined by a set of via-points. Via-point \(\theta_i = (\theta_{i1}, \theta_{i2}, \ldots)\) indicates a set of joint angles of the body at time \(t\). The walking we design for our experiments, for example, has 15 via-points. Time \(t\) is defined for each via-point to control motion speed.

Fig. 8 shows a joint angle trajectory represented by via-points. The trajectory is the actual angle of the right waist joint of the humanoid robot while it walks two steps. Black dots indicate via-points and the solid line indicates interpolated joint angles.
4. Imitation Learning

We use a simple policy gradient method introduced by Kohl and Stone [11] to improve walking by updating via-points that represents walking. It has an initial set of via-points representing its original walking and prepares test sets of similar via-points by adding small disturbances. It evaluates the test set of via-points one by one and estimates the gradient of the evaluation of walking based on via-points, then it updates via-points based on the estimated gradient of the evaluation. The learning system repeats this procedure and updates motion parameters to reach the local maximum of evaluation.

4.1. Policy Gradient Method

A set of via-points for the current motion is \( \Theta = (\theta_1, \theta_2, \ldots) = (\theta_{1}^{1}, \theta_{1}^{2}, \ldots, \theta_{M}^{1}, \ldots) \), where the \( k \) and \( j \) indicate indices of via-points and joints. The learning system prepares \( M \) motion sets of similar via-points \( 1^{\Theta}, 2^{\Theta}, \ldots, M^{\Theta} \) by adding small disturbances \( \epsilon_{1}^{k}, 0, -\epsilon_{1}^{k} \) to the current angle of motion via-point \( \Theta \).

\[
m^{\Theta} = ("m\theta_1^{1}, m\theta_2^{1}, \ldots) = ("m\theta_1^{1}, m\theta_2^{1}, \ldots) \ldots (8)
\]

\[
m\theta_{1}^{j} = \theta_{1}^{j} + r\epsilon_{1}^{k} \quad \text{where} \ r \in (-1, 0, 1) \ldots (9)
\]

It evaluates sets of via-points \( m^{\Theta} \) one by one after the robot generates walking based on the set of via-points.

Evaluation averages for disturbance \( \epsilon_{1}^{k} \) are estimated as follows:

- \( A_{k,+\epsilon}^{l} \) is average of evaluation of policies which parameter \( m\theta_{1}^{j} \) is \( \theta_{1}^{j} + \epsilon_{1}^{k} \)
- \( A_{k,0}^{l} \) is average of evaluation of policies which parameter \( m\theta_{1}^{j} \) is \( \theta_{1}^{j} \)
- \( A_{k,-\epsilon}^{l} \) is average of evaluation of policies which parameter \( m\theta_{1}^{j} \) is \( \theta_{1}^{j} - \epsilon_{1}^{k} \)

The set of via-points for walking \( \Theta \) based on the averages as follows:

\[
\theta_{1}^{l} \left\{ \begin{array}{ll}
\theta_{1}^{l} + \eta_{1}^{l} & A_{k,+\epsilon}^{l} > A_{k,0}^{l} \text{ and } A_{k,-\epsilon}^{l} \\
\theta_{1}^{l} - \eta_{1}^{l} & A_{k,-\epsilon}^{l} > A_{k,0}^{l} \text{ and } A_{k,+\epsilon}^{l} \\
\theta_{1}^{l} & \text{else}
\end{array} \right. \quad (10)
\]

\( \eta_{1}^{l} \) is the update step size for parameter \( \theta_{1}^{l} \). The learning system repeats this procedure and updates the set of via-points for motion to reach the local maximum of evaluation.

4.2. Evaluation of Motion

Evaluation of learning walking is designed based on three points – how fast the robot walks, the stability of walking, and how similar the walking is to demonstrated walking. The similarity to demonstrated walking \( S_{m} \) is defined based on the similarity \( S \) in Eq. (2) with the cyclic period and the phase of demonstrated motion \( \tau_{m} \) and \( \alpha_{m} \).

Walking distance \( D \) and similarity \( S_{m} \) are measured at each trial. Evaluation for learning \( E_{L} \) is defined as follows:

\[
E_{L} = w_{S}S_{m} + w_{D}D - P_{L} \ldots (11)
\]
where $w_S$ and $w_D$ are weights for similarity and walking distance. We set a constraint on weights so that $w_S + w_D = 1$. $P_L$ is a penalty against falling down during walking and zero is for when the robot does not fall down in the trial. Similarity $S_m$ works as a bias for imitation learning and the degree of bias is controlled by weight $w_S$.

Evaluation for learning $E_L$ includes similarity to demonstrated walking as a bias of imitation learning. To evaluate the performance of the learned walking itself, another evaluation $E_M$ is defined as follows:

$$E_M = w_D D - P_M$$  \hspace{1cm} (12)

$P_M$ is a penalty against falling during walking. Evaluation $E_M$ shows the pure performance of learned walking that does not consider similarity between robot walking and demonstrated walking.

5. Experiments

We use the humanoid robot (Robovie-PC) simulator that we developed instead of a real robot for exhaustive exploration based on the weight of learning bias during imitation learning.

5.1. Motion Recognition

Figure 9 shows recognition of the cyclic period and phase of observed human walking motion. A human demonstrator shows walking at various speeds, walking fast at the beginning and slowing down every 10 seconds, first (Fig. 9(a)). The human demonstrator then shows slow walking and speeds up every 10 seconds (Fig. 9(b)). Recognition results show that the period of cycle $\tau_m$ and phase $\alpha_m$ of walking is detected well enough to imitate.

Figure 10 shows an example of motion recognition and segmentation while the human demonstrator stands, waives his right hand, walks, bows, and stands again. Human motion is recognized successfully.

5.2. Imitation Learning

A human demonstrator shows walking at normal speed to the humanoid robot. The learning robot segments the demonstrated motion. Walking is selected for imitation learning, here. The humanoid robot starts learning walking on its own based on the policy gradient method in Section 4.1 with evaluation $E_L$ defined in Section 4.2 including learning bias based on demonstrated walking. Initial via-point parameters for walking are original ones of Robovie-PC walking. To investigate how learning bias effects learning, the weight for similarity $w_S$ is varied from 0.0 to 1.0. If $w_S = 0.0$, the robot ignores the learning bias of imitation and tries to maximize walking distance. If $w_S = 1.0$, the robot ignores walking distance and tries to imitate demonstrated walking.

The robot walks 10 steps in one trial. The size of parameter sets $M$ is 100 to update parameters of via-points for walking in this experiment. The number of updates is empirically set to 100 so that learning converges.

A conventional method is conducted to evaluate the proposed method. The comparative conventional method is conducted so that it uses motion demonstrated by a human being to set initial parameters of the via-points for walking, then updates parameters through trial and error as follows: the best joint parameters of each via-point for mimicking human demonstrated walking are searched for through trial and error which takes 10,000 trials. It then starts to learn walking based on the same learning method as our proposed method but without the learning bias of imitation $w_S = 0.0$.

Figures 11 to 14 show results of experiments. Each learning curve is an average of 5 learnings. Fig. 11 shows the average walking distance during learning with different weights of similarity $w_S$ and the conventional method.
All learning curves except the those with $w_S = 1.0$ and the conventional method are similar. Learning curves with $0.0 < w_S < 1.0$ shows better performance than curves $w_S = 0.0$ or $w_S = 1.0$. $w_S = 1.0$ indicates that $w_D = 0.0$ and the learning robot tries to imitate demonstrated walking ignoring walking distance. It shows a shorter walking distance than the others, but still shows better performance than at the beginning of learning. $w_S = 0.0$ indicates that the robot ignores the learning bias of imitation and just tries to maximize walking distance. Nonetheless, it shows worse performance than others with $w_S$ greater than 0. The weight of similarity $w_S = 0.6$ shows the best learning performance in experiments. The conventional method actually shows faster walking than the proposed method. Note, however, that learning started after the best parameters for mimicking human demonstrated walking are found through 10,000 trials.

Figure 12 shows the cumulative number of falls during learning. Fig. 12(a) shows results with the conventional and proposed methods. The conventional method shows a much bigger number of falls during learning, which means that the conventional method shows an unstable walking pattern during learning. The proposed method shows many fewer falls during learning, which means that walking during learning is stable. This is one of the big advantages of our proposed method.

Figure 12(b) shows results without the conventional method. If the learning robot ignores the bias of imitation with motion similarity $S_m$, it walks unstably and often falls. It actually shows the worst stability during learning if it ignores the bias of imitation by $w_S = 0.0$. The learning robot with weight of similarity $w_S = 1.0$ shows the best stability performance, rarely falling during learning. If the weight of similarity increases, stability tends to be high. One exception, however, $w_S = 0.9$ shows worse performance than $w_S = 0.6$. The weight of similarity $w_S = 0.6$
Figure 13 shows the evaluation of motion without the weight of similarity, $E_M$, defined by Eq. (12). Weight of similarity $w_S = 0.6$ shows a stable learning curve whereas $w_S = 0.0$ shows frequent falls during learning because it is often given a penalty for falling.

Figure 14 shows walking similarity $S$ defined by Eq. (2) during learning. All similarity curves except $w_S = 1.0$ decrease during learning, indicating that walking tends to be affected by the evaluation of walking distance rather than similarity of walking. Fig. 11 shows that learning bias based on walking similarity leads to better walking parameters. Fig. 12 shows that bias contributes to walking stability during learning, so learning bias based on walking similarity contributes to better parameter search and motion stability during learning, although the evaluation of motion similarity decreases in learning.

5.3. Real Robot Experiments in Imitation Learning

We conducted real robot experiments to show the validity of our proposed imitation learning method. Figs. 15 and 16 show typical sequence of real robot walking before and after imitation learning. Walking before learning is stable but slower (Fig. 15). Walking learned based on walking distance and imitation similarity, $w_S = 0.9$, shows faster walking while also being slightly unstable.

Figure 17 shows a sequence of real robot walking learned based on distance walking alone without the imitation of human walking. In general, this walking is too unstable to continue, although walking distance is longer than that of others. The upper body tends to swing back and forth using dynamics for faster walking but it almost falls down.

Figure 18 shows the sequence of real robot walking learned based on only similarity of imitation in which the walking distance is not considered. Despite this, it shows faster walking than the original while showing stable walking. Walking speed is slightly slower than that learned with walking distance and imitation similarity in real robot experiments. Simulation results show a significant difference between them, possibly due to the difference in friction between the foot and floor and/or the different characteristic of servo-motor dynamics of joints.

6. Conclusions and Future Work

This paper has proposed two methods for humanoid robot imitation learning with via-point representation. One involves motion recognition and segmentation for imitation learning and recognizes both motion itself and the period and phase of the periodic motion cycle. The other involves imitation learning that introduces the bias of the similarity of motion demonstrated by a human being. We investigated learning performance based on the weight of bias. Results of experiments showed that the first method successfully recognized and segmented the
Fig. 16. Real robot walking pattern sequence after learning based on the evaluation of both walking distance and imitation similarity.

Fig. 17. Real robot walking pattern sequence after learning based on the evaluation of walking distance alone.

demonstrated motion with period and phase information on the periodic motion cycle and that learning bias contributes to better parameter search and learning motion stability during learning.

Future work should extend the proposed method to the imitation learning of new motion even though the method
Motion Segmentation and Imitation Learning of Humanoid Robot

we proposed in this paper focuses on motion of the robot as one of motion repertory.

References:
Name: Yasutake Takahashi
Affiliation: Department of Human and Artificial Intelligent Systems, Graduate School of Engineering, University of Fukui
Address: 3-9-1 Bunkyo, Fukui, Fukui 910-8507, Japan
Brief Biographical History:
2000-2009 Assistant Professor of Department of Adaptive Machine Systems, Graduate School of Engineering, Osaka University
2006-2007 Visiting researcher at the Fraunhofer IAIS
2009-2012 Senior Assistant Professor of Department of Human and Artificial Intelligent Systems, Graduate School of Engineering, University of Fukui
2012-Associate Professor of Department of Human and Artificial Intelligent Systems, Graduate School of Engineering, University of Fukui
Main Works:
Membership in Academic Societies:
- The Robotics Society of Japan (RSJ)
- Japan Society for Fuzzy Theory and Intelligent Informatics (SOFT)
- The Institute of Electrical and Electronics Engineers (IEEE)

Name: Hiroki Hatano
Affiliation: Department of Human and Artificial Intelligent Systems, Graduate School of Engineering, University of Fukui
Address: 3-9-1 Bunkyo, Fukui, Fukui 910-8507, Japan
Brief Biographical History:
2010-2014 Faculty of Engineering, University of Fukui
2014-Graduate School of Engineering, University of Fukui
Membership in Academic Societies:
- Japan Society for Fuzzy Theory and Intelligent Informatics (SOFT)

Name: Kazuyuki Usui
Affiliation: Department of Human and Artificial Intelligent Systems, Graduate School of Engineering, University of Fukui
Address: 3-9-1 Bunkyo, Fukui, Fukui 910-8507, Japan
Brief Biographical History:
2009-2013 Faculty of Engineering, University of Fukui
2013- Hokusin Co., Ltd.
Name: Yoichiro Maeda

Affiliation: Department of Manufacturing Technologists, Institute of Technologists

Address: 333 Maeya, Gyoda City, Saitama 361-0038, Japan

Brief Biographical History:
1983- Researcher, Central Research Lab., Mitsubishi Electric Corp.
1989-1992 Senior Researcher, Laboratory for International Fuzzy Engineering Research (LIFE)
1995- Associate Professor, Osaka Electro-Communication University
1999-2000 Visiting Researcher, University of British Columbia (UBC), Canada
2002- Associate Professor, Faculty of Engineering, University of Fukui
2007- Professor, Graduate School of Engineering, University of Fukui
2013- Professor, Faculty of Engineering, Osaka Institute of Technology
2015- Professor, Department of Manufacturing Technologists, Institute of Technologists

Main Works:

Membership in Academic Societies:
- The Society of Instrument and Control Engineers (SICE)
- The Robotics Society of Japan (RSJ)
- Japan Society for Fuzzy Theory and Intelligent Informatics (SOFT)
- The Japanese Society for Artificial Intelligence (JSAI)
- Japan Society of Kansei Engineering (JSKE)