Soft Comput (2006) 10: 891–901
DOI 10.1007/s00500-005-0015-9

FOCUS

Naoyuki Kubota · Yusuke Nojima
Fumio Kojima · Toshio Fukuda

Multiple fuzzy state-value functions for human evaluation through interactive trajectory planning of a partner robot

Published online: 12 October 2005

Abstract The purpose of this study is to develop partner robots that can obtain and accumulate human-friendly behaviors. To achieve this purpose, the entire architecture of the robot is designed, based on a concept of structured learning which emphasizes the importance of interactive learning of several modules through interaction with its environment. This paper deals with a trajectory planning method for generating hand-to-hand behaviors of a partner robot by using multiple fuzzy state-value functions, a self-organizing map, and an interactive genetic algorithm. A trajectory for the behavior is generated by an interactive genetic algorithm using human evaluation. In order to reduce human load, human evaluation is estimated by using the fuzzy state-value function. Furthermore, to cope with various situations, a self-organizing map is used for clustering a given task dependent on a human hand position. And then, a fuzzy state-value function is assigned to each output unit of the self-organizing map. The robot can easily obtain and accumulate human-friendly trajectories using a fuzzy state-value function and a knowledge database corresponding to the unit selected in the self-organizing map. Finally, multiple fuzzy state-value functions can estimate a human evaluation model for the hand-to-hand behaviors. Several experimental results show the effectiveness of the proposed method.

Keywords Fuzzy Modeling · Partner Robot · Trajectory Generation · Interactive Genetic Algorithm · Self-Organizing Map

N. Kubota (✉)
Department of System Design, Tokyo Metropolitan University,
PREST, Japan Science and Technology Agency
1-1 Minami-Ohsawa, Hachioji, Tokyo 192-0397, Japan
E-mail: kubota@comp.metro-u.ac.jp
Tel.: +81-426-772728
Fax: +81-426-772728

Y. Nojima
Department of Computer Science and Intelligent Systems,
Osaka Prefecture University,
1-1 Gakuen-cho, Sakai, Osaka 599-8531, Japan

F. Kojima
Department of Mechanical and Systems Engineering,
Graduate School of Kobe University,
1-1 Rokkodai-cho, Nada-ku, Kobe 657-8501, Japan

T. Fukuda
Department of Micro System Engineering,
Graduate School of Nagoya University,
1 Furo-cho, Chigusa-ku, Nagoya 464-8603, Japan

1 Introduction

Various human-friendly robots such as pet robots, humanoid robots, and partner robots, have been developed so far. These kinds of robots can perform many complicated behaviors, especially, dynamic walking and dancing, but it seems to be difficult to realize social communication with a human. Soft computing, which was proposed by Zadeh [1], is a new concept for information processing, and its objective is to realize a new approach for analyzing and creating flexible information processing with human beings such as sensing, understanding, learning, recognizing and thinking [1–4]. Soft computing including fuzzy, neural, and evolutionary computing has been applied successfully to motion planning and motion control of various robots in unknown or dynamic environments [5–8]. For example, fuzzy controllers and neural controllers are used for action systems representing the complicated relationship between sensory inputs and motion outputs in unknown or dynamic environments. Furthermore, neural networks and fuzzy inference rules are used for a perceptual system such as clustering and classification, and used for analyzing human behaviors. On the other hand, evolutionary optimization methods have been applied for parameter tuning, motion planning, and behavior acquisition of various robots. The researches of behavior acquisition based on evolutionary computing are well known as evolutionary robotics [9, 10]. Furthermore, interactive genetic algorithm, reinforcement learning, and organizational learning have been applied to various complicated problems [11–16]. Interactive genetic algorithm (IGA) is used for various design problems based on human evaluation [11, 12]. Reinforcement learning has been applied for learning multi-stage or sequential actions of a

robot in unknown environments. The reinforcement learning can estimate a value function, and generate its corresponding actions in a Markov decision process [13, 14]. Organizational learning is a new learning method based on multiple agents like a human society [15, 16]. In this way, various methods based on human intelligence and life have been proposed to build intelligent robots and artificial agents. Robotic intelligence is deeply related with the control architecture of a robot.

As a methodology for robotic control, a subsumption architecture was proposed by Brooks [17]. The concept of the subsumption architecture has led to one stream of behavior-based robotics. Behavior-based robotics emphasizes the importance of the interaction of an embodied robot with its environment [17–19]. According to cognitive psychology, the embodiment indicating an agent has not only a physical body, but also experience. Experience of a robot leads to the lifetime learning of perceptual system, action system, and communication system interacting with an environment and a human. The above discussion indicates that the importance is how to accumulate behaviors rather than how to obtain a brand-new behavior, i.e., the robot should generate and accumulate its behaviors, and should grow up by itself. Furthermore, a human-friendly robot must learn its behaviors by considering human factors. Especially, human-friendly physical expression using a robotic body is very important to realize social communication with a human, but it is very difficult to incorporate a human model into the evaluation function of a robot beforehand. Therefore, we discuss a mechanism for obtaining and accumulating robotic behaviors through human evaluation from the viewpoint of constructivism.

The mechanism of a robot becomes complicated and large increasingly as intelligent capabilities are added gradually to the robot. Moreover, its resulting behavioral patterns depend strongly on its information processing related to intelligence. Therefore, we should consider an entire structure of intelligence for processing information flowing over the hardware and software of a robot, not a single intelligent capability. Accordingly, we have proposed the concept of structured intelligence [8, 20] and structured learning [21]. The structured intelligence emphasizes the coupling of intuitive inference, logical inference, and self-consciousness. The structured learning emphasizes the importance of interactive learning of several modules through the interaction with its environment. In this paper, we propose a method for obtaining and accumulating a hand-to-hand behavior of a partner robot based on the structured learning. A hand-to-hand behavior is one of fundamental behaviors required for a partner robot interacting with a human. We have proposed a method of interactive genetic algorithm using human evaluation [21–23]. In order to obtain a hand-to-hand behavior, a genetic algorithm is used for generating a trajectory enable to maximize human evaluation. Because the human cannot evaluate all trajectory candidates, the robot must have a human evaluation model as an internal model. However, it is very difficult to identify an exact model of human evaluation, and furthermore the obtained exact model might not be able to be reused well because human evaluation is very vague and very changeable according to spatial and temporal conditions. For that reason, the robot should adaptively approximate the human evaluation model actually in displaying a trajectory candidate to a human. The estimated human evaluation model is used to evaluate trajectory candidates of the genetic algorithm using internal simulator of the robotic behaviors using kinematics. In [22], we applied a simple discrete state-value function to model the human evaluation through an interaction with the human. However, the discrete state-value function needed much interaction to learn the state-value function used for generating robotic trajectories. To solve this problem, we used a fuzzy state-value function [23]. In fact, the memory size of the proposed fuzzy state-value function used to represent a human evaluation model in the method [23] is 343 ($7 \times 7 \times 7$) where the number of membership functions used in each axis is 7, while that of the discrete state-value function is 27,000 ($30 \times 30 \times 30$) in the method [22]. In this way, we can reduce the memory size used in estimating a human evaluation model by using fuzzy partition, and furthermore, the number of human evaluation times can also be reduced [23]. These previous works succeeded to obtain human evaluation models through executing a single task. However, the obtained human evaluation model might not be able to be applied well for a new different task, and furthermore, the new task might make the previously obtained human evaluation model forgotten. In this paper, therefore, we apply mulitple fuzzy state-value functions to deal with various tasks composed of different human hand positions (target points). In this case, the structure of human evaluation is not static, but changeable according to a current task or an objective of hand-to-hand behavior, but an evaluation function should be designed systematically and statistically. If the human evaluation can be used systematically and statistically, the obtained structure of the human evaluation might be more reliable. Therefore, we divide the state of input space into several substates and try to estimate the structure of human evaluation. Accordingly, a self-organizing map (SOM) is used for clustering hand-to-hand behaviors according to the human hand position as the target points. Furthermore, multiple fuzzy state-value functions are applied for estimating the structure of human evaluation according to the clustering result of SOM. Here a fuzzy state-value function corresponding to an output unit of SOM is selected. In addition, an output unit has the best trajectory of the previous search. Therefore, initial trajectory candidates of an interactive genetic algorithm for a current target point are generated by using the previous best trajectory. Therefore, the proposed method can distributively estimate the structures of human evaluation according to various hand-to-hand behaviors, and can generate human-friendly trajectories for hand-to-hand behaviors.

This paper is organized as follows. Section 2 explains the concept of communication and structured learning for a hand-to-hand behavior using human evaluation, SOM for clustering target positions, multiple fuzzy state-value functions for estimating human evaluation, and IGA for trajectory

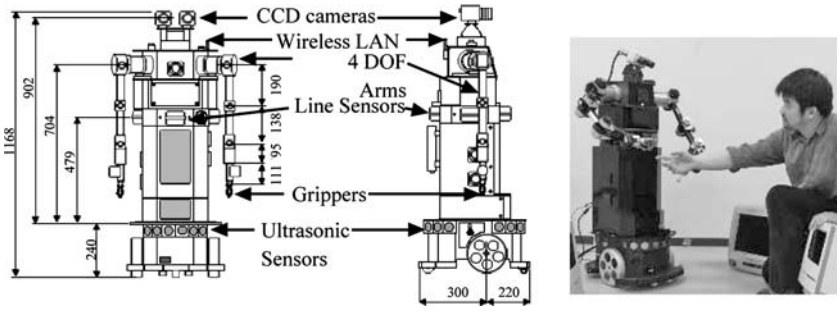

Fig. 1 A human-like partner robot, Hubot

planning. Section 3 shows several experiment results and comparison results with the previous methods.

2 A human-friendly partner robot

2.1 Communication and learning

Human-friendly robots require several intelligent capabilities such as perceiving, acting, communicating, and surviving like animals or humans. The capability to communicate is essential in building a relationship or even a friendship between a robot and its human owner or operator. We consider an example of a child playing with a pet robot. When a child begins to play with a pet robot, the child would try to have contact with the robot in various manners. The child will search for causal relationship between his or her contacting pattern and its reaction of the robot. The contacting pattern and its reaction of the child correspond to sensory inputs to the robot and motion outputs of the robot, respectively. A human can gradually find the boundary or structure of difference in the action patterns of the robot, and also the robot should learn a specific human contacting pattern and its corresponding actions. This interactive or mutual learning plays a very important role in their communication, because the causality of contact and reaction is useful for predicting future behaviors of each other. In many robots, their behavioral patterns and communication forms are designed beforehand, but we think the mechanism to enrich the relationship between a human and a robot is the architecture for learning the interrelation between the human and the robot. Consequently, the communication of a robot with a human requires the continuous interaction with the human, because the human tries to find out the causal relationship between human contact and its resulting robotic behavior, and furthermore, the human tries to find more complicated relationship according to the found relationships. Therefore, the robot needs to accumulate its behaviors through interacting with the human step by step.

2.2 Hand-to-hand behavior of a partner robot

We developed a human-like partner robot called Hubot in order to aim to realize the social communication (Fig. 1). This robot is composed of a mobile robot, body, two arms with grippers, and head with pan and tilt. The robot has various sensors such as two CCD cameras, four line sensors (infrared sensors), microphone, ultrasonic sensors, touch sensors as external sensors in order to perceive its environment. Each CCD camera can capture an image between the range of -30° and 30° in front of the robot. Furthermore, many encoders are equipped with the robot. Two CPUs are used for sensing, motion controlling, and communicating. In previous researches, we proposed a human detection method using a series of images from the CCD camera and a simple trajectory planning method for a hand-to-hand behavior [21]. In this paper, we focus on trajectory planning and learning methods for various hand-to-hand behaviors of the partner robot shown in Fig. 2.

Trajectory planning is one of the most important and essential task required by robot manipulators [24–28]. In general, a robot manipulator is composed of a gripper and an arm. To achieve a given task, the robot manipulator generally performs the following subtasks: (1) finding obstacles or modeling an environment (perception), (2) generating a collision-free trajectory (decision making), and (3) tracing the trajectory actually (action). First, the robot detects a human by using visual perception, and then, a tentative target position of the end-effector is decided according to the position of the detected human hand position. Next, the robot generates a reference trajectory based on the surrounding state built from sensory information. Various trajectory planning methods have been proposed to solve motion planning problems [24–28]. Basically, two main approaches have been proposed to generate collision-free trajectories. The first one is artificial potential field methods. A robot manipulator moves are based on the attractive force from the goal point and the repulsive force from the obstacles in the work space. The other is a configuration space (C-space) method. The C-space is transformed into an internal state space from a three-dimensional space of an environment, and therefore, the dimensions of the C-space are equal to the degrees of freedom (DOF) of the robot manipulator. In this paper, we focus on the control of the robot arm. Accordingly, a trajectory planning problem for a hand-to-hand behavior can result in a path planning problem on the C-space from an initial configuration to a final configuration corresponding to a target point of the detected human hand. Here a configuration θ is expressed by a set of

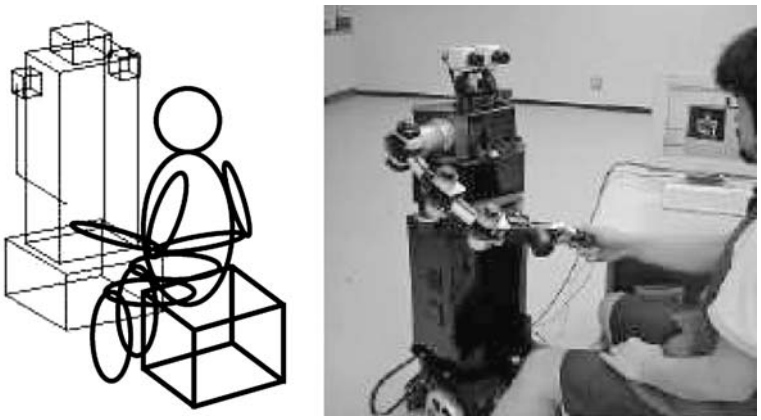

Fig. 2 An example of hand-to-hand behavior

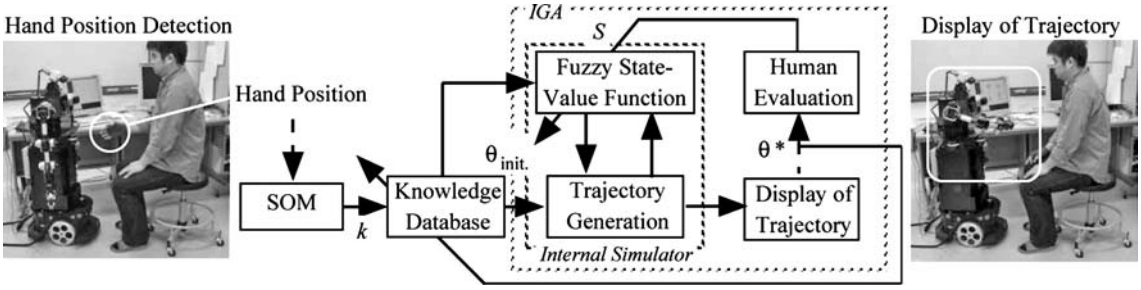

Fig. 3 Architecture of structured learning for obtaining a hand-to-hand behavior of a partner robot

joint angles, because all joints are revolute,

$$\theta = (\theta_1, \theta_2, \ldots, \theta_n)^T \in R^n \tag{1}$$

where n denotes the DOF of a robot arm. The number of DOF of the partner robot shown in Fig. 1 is 4 ($n = 4$). In addition, the position of the end-effector (robot hand or gripper), $P = (p_x, p_y, p_z)^T$, on the base frame is defined as follows:

$$P = {}^0T_1{}^1T_2, \ldots, {}^{n-1}T_n{}^nX = f(\theta) \tag{2}$$

where $P = (p_x, p_y, p_z, 1)^T$; ${}^{i-1}T_i$ denotes a homogeneous transformation matrix from a frame $i-1$ to a frame i; ${}^nX$ denotes a position of the nth joint on the frame n. Because a trajectory can be represented by a series of m intermediate configurations, the trajectory planning problem is to generate a collision-free and human-friendly trajectory combining several intermediate configurations.

Figure 3 shows a total architecture of generating a trajectory for a hand-to-hand behavior of a partner robot. Here one trial is defined as a process trajectory planning and learning for one hand-to-hand behavior of the robot. First of all, the robot detects the human hand as the target point at the tth trial. By using the target point as an input vector to SOM, the kth output unit that minimizes the distance from the input vector, is selected. And then, SOM outputs its corresponding kth fuzzy state-value function and the best trajectory θ_k^* by referring to the knowledge database stored. The trajectory is used for generating initial trajectory candidates θ_{init} as an initial population of IGA. Next, IGA generates candidate trajectories for a hand-to-hand behavior. After several iterations of an internal simulation in the robot, the best trajectory θ^* in the current population is displayed to the human. According to the hth human evaluation score $S(t, h)$ at the tth trial ($h = 1, 2, ..., H$), the fuzzy state-value function is updated. Furthermore, a next trajectory candidate is generated according to the updated fuzzy state-value function. And finally, the generated best trajectories and fuzzy state-value function are stored in the knowledge database linking with SOM.

2.3 Self-organizing map for clustering hand-to-hand behaviors

Various unsupervised learning methods have been proposed so far [2,29,30,32]. In a case of batch learning, a set of all data is required, but incremental learning can update design parameters when new data are given to the learning system. Here a human hand position used as the target point of the final configuration in a hand-to-hand behavior is changeable dependent not only on the posture and the distance against the robot (spatial conditions), but also on the meeting time and day (temporal conditions). Furthermore, the area including hand positions used as target points might be very specific to the human interacting with the partner robot. Therefore, we apply SOM for clustering target points sequentially. As one of unsupervised learning methods, SOM is often used for extracting a relationship among inputs data, since SOM can learn the hidden topological structure from the learning data [29–31].

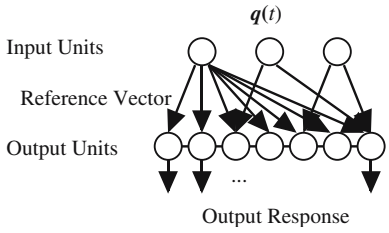

Fig. 4 A self-organizing map

In general, the Euclidian distance is defined as follows:

$$d_i = \|q(t) - c_i(t)\| \tag{3}$$

where $q(t) = (q_1(t), q_2(t), q_3(t))$ and $c_i(t) = (c_1, c_2, c_3)$ are the coordinates of the target point used as the input vector and the ith reference vector at the tth trial, respectively. We can obtain the kth output unit that minimizes the distance d_i by

$$k = \arg\min_i \{\|q(t) - c_i(t)\|\}. \tag{4}$$

Furthermore, the reference vector is trained by

$$c_i(t+1) = c_i(t) + \xi \cdot \zeta_{ki} \cdot \|q(t) - c_i(t)\| \tag{5}$$

where ζ_{ki} and ξ are the neighborhood function and learning rate respectively. The number of output units is the number of fuzzy state-value functions. Accordingly, the state-value function corresponding to the output unit nearest the input vector is selected. Here three-dimensional structure is used to represent the neighboring relationship among output units in SOM.

2.4 Fuzzy state-value functions for estimating human evaluation

A fuzzy theory is applied for estimating human evaluation. As mentioned in the previous subsection, the kth set of fuzzy rules corresponding to the kth output unit of SOM is selected as a human model of the current human hand position. A fuzzy rule is described as follows:

If x_2 is $A_{1,j}$ and $\cdots$ and x_N is $A_{N,j}$ then v is w_j

where x_i is the ith input; $A_{i,j}$ is a membership function for the ith input at the jth rule; v is an estimated human evaluation value; w_j is a singleton for the output of the jth rule; N is the number of inputs. Here we use Gaussian membership functions as follows,

$$\mu_{Ai,j}(x_i(r)) = \exp\left(-\frac{\left(x_i(r) - a_{i,j}\right)^2}{b_{i,j}}\right) \tag{6}$$

where $a_{i,j}$ and $b_{i,j}$ are the center and the width of a membership function, respectively. Next, we obtain the output at rth configuration by the following weighted average,

$$\mu_j = \prod_{i=1}^{N} \mu_{Ai,j}(x_i(r)) \tag{7}$$

$$v = \frac{\sum_{j=1}^{M} w_j \cdot \mu_j(r)}{\sum_{j=1}^{M} \mu_j(r)} \tag{8}$$

where μ_j denotes the firing strength of the jth rule; M is the number of fuzzy rules. Because a human evaluation model is used in the trajectory planning for a hand-to-hand behavior, the inputs to above fuzzy inference are coordinates $P(r)$ of the robot hand at the rth intermediate configuration ($r = 1, 2, ..., m$) in the trajectory for a hand-to-hand behavior, and therefore, the number of inputs is 3 (=N). Here seven linguistic values of *negative big*, *negative medium*, *negative small*, *zero*, *positive small*, *positive medium* and *positive big*, are used for fuzzifying input values. In this way, the fuzzy rules can map the state space of the robot hand position into the value space of the human evaluation. Therefore, we call the set of fuzzy rules, a fuzzy state-value function. The fuzzy state-value function is trained by using the human evaluation along the trajectory of a robot hand. The updating scheme for the jth rule of the fuzzy state-value function is as follows,

$$w_j \leftarrow w_j + \eta \cdot \left(\frac{S(t,h)}{10} - v\right) \frac{\mu_j}{\sum_{i=1}^{M} \mu_i} \tag{9}$$

where η denotes a learning rate satisfying $0 < \eta < 1.0$; $S(t,h)$ denotes the hth human evaluation score at the tth trial with $S(t,h) \in [0, 9]$. Here the human evaluation score 0 indicates excellent.

2.5 Interactive genetic algorithm for trajectory planning

Interactive genetic algorithm (IGA) is often applied to an optimization problem based on a fitness function including human evaluation. Various interactive optimization methods have been proposed so far to obtain good solutions based on human evaluation. However the problem is how to generate a next candidate solution to be displayed to the human, since the derivative information or searching direction for generating a next solution is not exactly included in the current solution. In the case of IGA, a population of candidate solutions might implicitly include possible searching directions, because genetic diversity to generate possible good candidate solutions is maintained in a population. Therefore, IGA can heuristically generate a next candidate trajectory by combining current candidate solutions. To reduce the human evaluation times, we use a fuzzy state-value function instead of actual human evaluation in a search with internal simulator (Fig. 3). Furthermore, the fuzzy state-value function is updated by using the human evaluation during the search of IGA. The procedure of the IGA for trajectory generation is shown as follows:

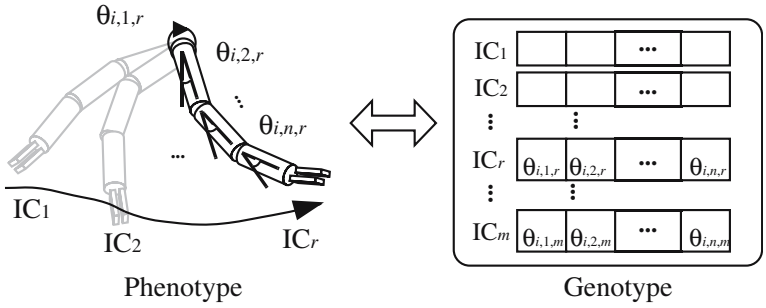

Fig. 5 The representation of the ith trajectory candidate composed of r intermediate configurations

```
begin
    Initialization
    repeat
        repeat
            Delete Least Fitness Selection
            Elite Crossover
            Adaptive Mutation
        until local_termination_condition is True
        Human Evaluation
        Update a fuzzy state-value function
    until termination_condition is True
end.
```

A trajectory candidate is composed of all joint variables of intermediate configurations (Fig. 5). Initialization generates an initial population based on the previous best trajectory stored in the knowledge database linked with SOM. The jth joint angle of the rth intermediate configuration in the ith trajectory candidate $\theta_{i,j,r}$, which is represented as a real number, is generated as follows ($i = 1, 2, ..., gn$; $j = 1, 2, ..., n$; $r = 1, 2, ..., m$):

$$\theta_{i,j,r} \leftarrow \theta^*_{k,j,r} + \gamma_j \cdot N(0, 1) \tag{10}$$

where $\theta^*_{k,j,r}$ is the kth trajectory stored in the knowledge database corresponding to the selected kth output unit of SOM; γ_j is a coefficient for the jth joint angle; $N(0, 1)$ is a Gaussian random variable with mean 0 and standard deviation 1. A fitness value is assigned to each trajectory candidate. The objective is to generate a trajectory realizing the possibly short distance from the initial configuration to the final configuration while realizing good evaluation. To achieve the objectives, we use a following multi-objective fitness function,

$$f_i = w_1 f_p + w_2 f_d + w_3 f_V \tag{11}$$

where w_1, w_2, and w_3 are weight coefficients. The first term, f_p, denotes the penalty about the distance between the hand position and the target point.

Figure 6 shows the penalty zone generated by using a sphere with a center on the line connecting the initial hand position and the target point. The radius of a sphere is decreased as the hand position approaches to the target point. The penalty zone is outside the sphere and used in each intermediate configuration. This factor simply restrains strange trajectories such that a human cannot touch a robot hand. The second term, f_d, denotes the squared sum of the difference of each joint angle between two configurations. This

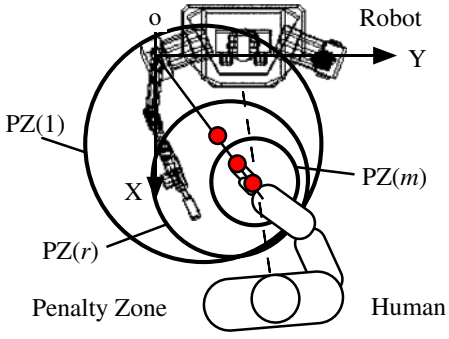

Fig. 6 Penalty zones defined as the outside of a sphere with the center on the line connecting the initial hand position and the target point

term is used to minimize motions of the manipulator [20]. But w_2 is set to a relatively small value, because the minimal motion is not always the best for a human. The last term, f_V, denotes the sum of the estimated human evaluation scores using the kth fuzzy state-value function. Therefore, this trajectory planning problem can result in a minimization problem.

"Delete least fitness" (DLF) is used as selection scheme, which removes the worst individual from the current population. This kind of selection scheme is called a continuous generation model or a steady-state genetic algorithm [4]. Next, one individual is randomly selected from the population. Here we use an elite crossover incorporating some genetic information from the best individual. Consequently, the worst individual is replaced with the individual generated by the elite crossover. Furthermore, we use the following adaptive mutation of the ith individual,

$$\theta_{i,j,r} \leftarrow \theta_{i,j,r} + \left(\alpha_j \frac{f_i - f_{\min}}{f_{\max} - f_{\min}} + \beta_j\right) N(0, 1) \tag{12}$$

where f_i is the fitness value of the ith individual, $f_{\max}$ and $f_{\min}$ are the maximum and minimum of fitness values in the population, respectively α_j and β_j are the coefficient and offset, respectively. The searching processes using the internal simulator are repeated until the local termination condition is satisfied. Here we use the maximal times of internal evaluations (T) as a local termination condition. After the search with the internal simulator (every T times), the best trajectory is displayed to the human. And then, the human evaluates

the trajectory by using keyboard and scores a value ($S(t, h)$) between 0 and 9, and 0 is excellent. Next, the kth fuzzy state-value function is updated according to this human evaluation score. If the human evaluation is excellent ($S(t, h) = 0$), the trajectory planning is stopped. Therefore, the human evaluation and the maximal human evaluation times (H) are used as the termination condition. And finally, the best trajectory obtained is stored in the knowledge database.

3 Experiments

This section shows experimental results of learning hand-to-hand behaviors using Hubot (see Figs. 1 and 2). Tables 1 and 2 show parameters used in this experiment.

First of all, we show learning results of SOM. The reference vectors of SOM used in front of the robot's body are initialized by small random values (Fig. 7). After 300 trials, the positions of the reference vectors were well updated to classify human hand positions (Fig. 8). The accuracy of this classification is confirmed by comparing with actual human hand positions showed in Fig. 9.

Here, the hand positions can be divided into two large classes; one is upper area, and the other is lower area. The upper area in the learning result corresponds to the hand positions at human standing posture. The lower area corresponds to hand positions when the human sits on a chair. These figures show that the robot can classify specific human hand positions through these experiments. Next, Fig. 10 shows the history of output units selected in SOM according to the human hand positions. Until 100 trials, the specific output units (state-value functions) were frequently used due to the transient state in the learning, because the topological distribution of output units is not suitable to the distribution of various hand positions of the human. For that, only a few output units tried to cover their neighboring hand positions. However, because the SOM has the simultaneous learning capability of the neighboring output units of the selected output units, the output units were crowded around the area used frequently as target hand positions of the human, and then the role of each state-value function was specialized to specific human hand positions after 200 trials.

Next, we discuss on the trajectory planning by IGA with human evaluation. The maximal times of evaluations in the internal simulator is 1000 at most in each trial (calculated by the product of IG and H), and the maximal human evaluation times in IGA is five in each trial. The total number of trials is 300, because we try to realize the life-time learning of a partner robot. As a result, the experiment was conducted for several days, although the internal simulation needs a short time and one actual display of trajectory spends a couple of minutes. Therefore, the evaluation and the objective of the human might change according to the emotional state of the human. Figures 11 and 12 show the average distance from the position of the ninth output unit to its neighboring output units in SOM and the history of human evaluations using the

Table 1 Parameters used in IGA

Parameter	Value
Chromosome length (nDOF × mIC)	24 (4 × 6)
Population size (gn)	200
Times of internal evaluations (IG)	200
Crossover rate	0.2
Mutation rate	1.0
Maximal human evaluation times for one trial (H)	5

Table 2 Parameters used in SOM and fuzzy state-value functions (FSVFs)

Parameter	Value
Input units for SOM	3
Output units of SOM (= number of FSVFs)	36 (3 × 4 × 3)
Learning rate for SOM (ξ)	0.02 ∼ 0.3
Learning rate for FSVFs (η)	0.05
Number of membership functions	7

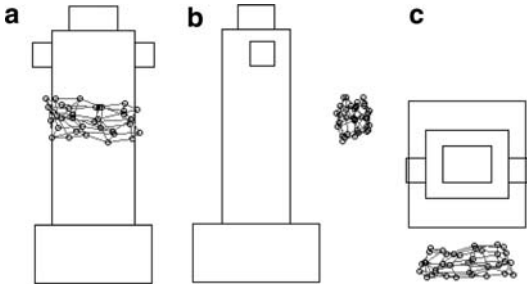

Fig. 7 Distribution of reference vectors of Self-organizing map (SOM) at initial state. **a** Front view (Y-Z). **b** Side view (X-Z). **c** Top view (Y-X)

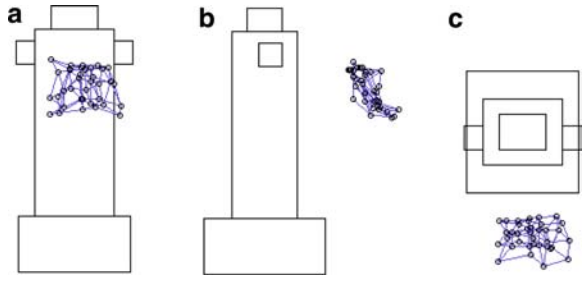

Fig. 8 Distribution of reference vectors of SOM after 300 trials. **a** Front view (Y-Z). **b** Side view (X-Z). **c** Top view (Y-X)

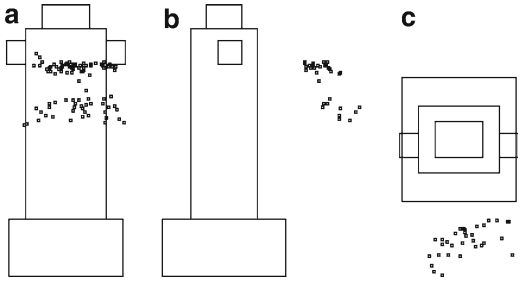

Fig. 9 Distribution of target points corresponding to human hand positions. **a** Front view (Y-Z). **b** Side view (X-Z). **c** Top view (Y-X)

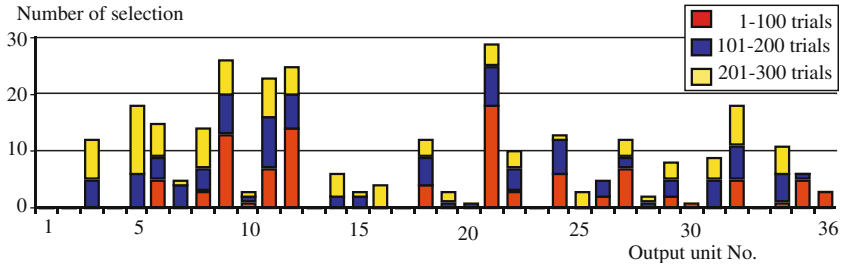

Fig. 10 History of output units selected in SOM to classify human hand positions

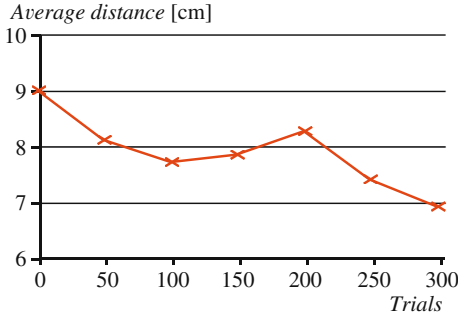

Fig. 11 Average distance from the position of ninth output unit to its neighboring outputs units in SOM

state-value function corresponding to the ninth output unit, respectively.

Here the ninth output unit is the frequently selected unit in SOM. And then, its neighboring output units (6th, 8th, 11th and 21st output units) are also frequently selected. The position of the ninth output unit as well as its neighboring output units were gradually updated according to human hand positions through 300 trials (see Fig. 11.). The change of human evaluation in Fig. 12 indicates that the robot was able to generate good trajectories trial by trial. Figure 13 shows the distribution of output values of the fuzzy state-value function corresponding to the ninth output unit and snapshots of the displayed best trajectory at the final trial. The size of the box indicates the degree of goodness. This figure shows the robot generates a trajectory passing through the area including high evaluation scores of the estimated human evaluation model. The region with good values is constructed along the robot's body and in front of robot (see Fig.13. a–c). In Fig.13 d, the robot moves its hand along the body until the height of shoulder, and then reaches out it to the human.

To compare the proposed method with previous method [22,23], we conducted a trajectory generation experiment without SOM. The robot has a single fuzzy state-value function to estimate the human evaluation model. Figures 14 and 15 show the average score and average times of human evaluations in every 100 trials, respectively. Here, average score means the average evaluation score of 100 trials where one trial includes five human evaluation times at maximal. Average times indicates the average number of evaluation times ($1 \leq$ Average times ≤ 5) until the robot obtain the best human evaluation in each trial. The average score and average times of the robot with multiple fuzzy state-value functions decrease as the increase of trials, while those of the robot without SOM don't decrease as the increase of trials. This indicates the robot with multiple fuzzy state-value functions and SOM-based clustering mechanism can estimate the human evaluation structure much better in case of learning in various hand positions, and IGA was able to generate trajectories satisfying human evaluation with less evaluation times.

Figure 16 shows a part of the history of human evaluations when only a single fuzzy state-value function was employed to the robot, in order to compare the previous method with the effectiveness of SOM clustering mechanism and multiple

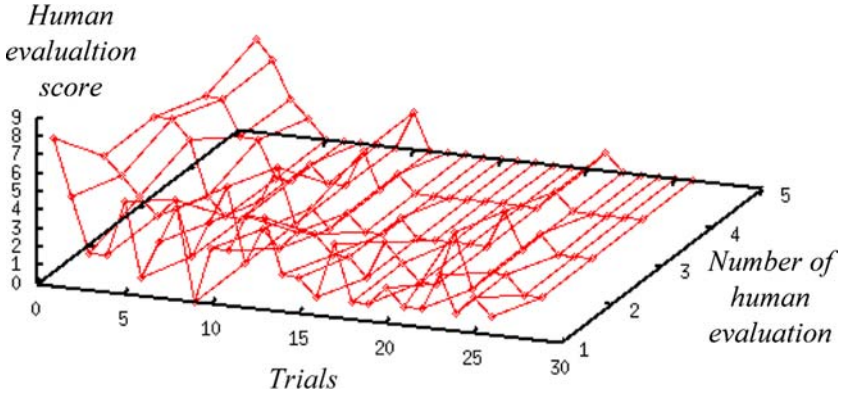

Fig. 12 History of human evaluations at each trial using state-value function for ninth output unit in SOM

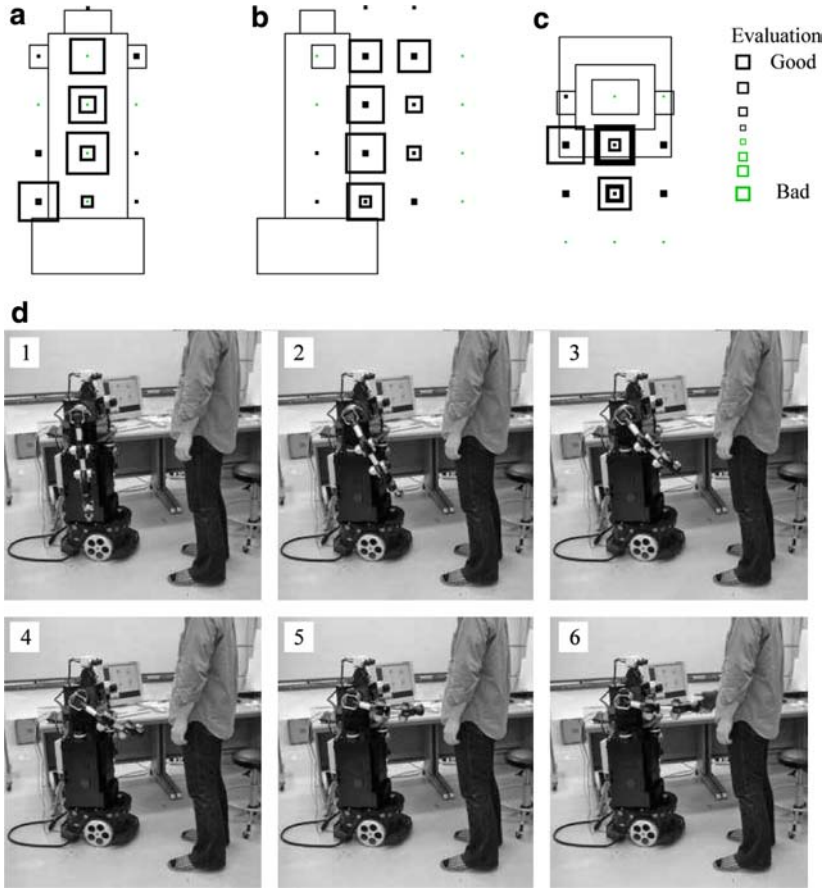

Fig. 13 A human-friendly trajectory obtained by interactive genetic algorithm (IGA) using human evaluation model of unit 9 in SOM. **a** Front view (Y-Z). **b** Side view (X-Z). **c** Top view (Y-X). **d** Snapshots of hand-to-hand behaviour

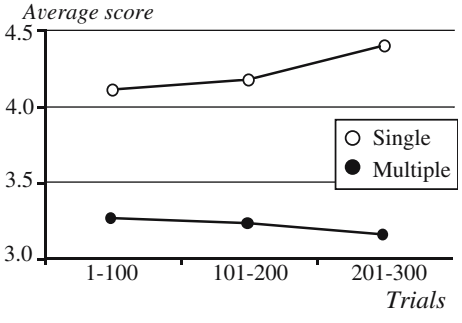

Fig. 14 Average score of human evaluations in every 100 trials

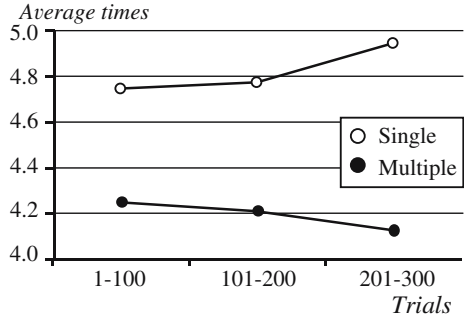

Fig. 15 Average times of human evaluations in every 100 trials

fuzzy state-value functions proposed in this paper. Since good human evaluation scores were not obtained all over the history in the experiment of various tasks, we showed the history of evaluations in the last 30 trials by using the previous method. A single fuzzy state-value function was not able to estimate human evaluation, because it was used for trajectory planning of various target points. Furthermore, the stored previous best trajectory and the state-value function used in other target hand position were not suitable for searching a new trajectory. As a result, the robot could not obtain good scores at the end (see Fig. 16). Since good scores were not obtained all over the history in the experiment of various tasks, we showed the history of evaluations in the last 30 trials.

In general, it is hard to verify the accuracy of the human evaluation model obtained by the proposed method, because the evaluation is very changeable every moment due to a little difference of trajectory, moving speed, distance, and other reasons. However, the important point for human-friendly

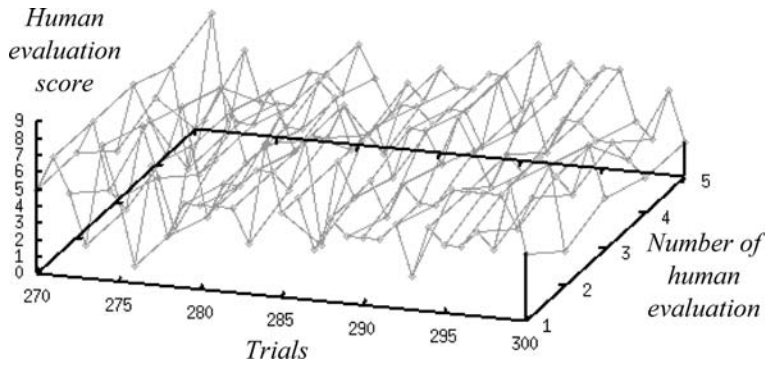

Fig. 16 History of human evaluations at each trial using a single state-value function

robots are the mutual adaptability, not the optimality in the behavior learning of only a robot, because the human can also learn the behavioral patterns of the robot. Therefore, the relationship between a human and a partner robot should be co-constructed like among humans. Furthermore, the reusability of the obtained evaluation structure and trajectories is also important in the life-time learning of a partner robot. The above experimental results show the proposed method realizes both the adaptability and reusability.

4 Concluding remarks

This paper proposed a trajectory planning and learning method for various hand-to-hand behaviors based on human evaluations to aim to build a human-like partner robot. In general, it is very difficult to design evaluation function for human-friendly robotic behaviors beforehand. Consequently, the robot should obtain a human evaluation model to realize human-friendly motion, but the human evaluation structure is much complicated. Therefore, we proposed multiple fuzzy state-value functions and applied SOM for clustering human hand positions used as target points of hand-to-hand behaviors. In this way, a fuzzy state-value function estimates human evaluation structure of a clustered human hand position. According to the fuzzy state-value function, the robot generates a trajectory by using interactive genetic algorithm with internal simulator. Actually, the robot obtains a human evaluation by displaying the current best trajectory in internal simulator to the human. The fuzzy state-value function is updated according to the actual human evaluation. And finally, the best trajectory is stored in a knowledge database. This knowledge database is used for next trajectory planning. In this way, the robot can generate various hand-to-hand behaviors by using human evaluations.

In general, it is very difficult to justify the most effective factor in many components to improve a system, as a system becomes complicated. However, we must consider the complication of the system such as human-like partner robots composed of many sensors and actuators. For that, the important point is the robot can obtain suitable behaviors by using less information like “good” or “bad”, instead of detailed instructions. And then, the clustering of behaviors and tasks is also important for partner robots to reuse these behaviors efficiently. Therefore, a partner robot should have an entire architecture of intelligent capabilities rather than a single intelligent capability, because the robot needs to obtain perceptual systems and action systems in unknown or dynamic environments including humans through the interaction with its environment. To realize this, the structured learning plays the important role in the life-time learning of the robot.

The robot should extract human evaluation through the actual interaction with the human, although this proposed method requires explicit human input from the keyboard to obtain a score of human evaluation. As a future work, we intend to incorporate the method for extracting human evaluation proposed in our previous works [21,31] for the human-like partner robot. As another future work, we will introduce a visual system using two CCD cameras for getting more essential information in communication with the human. Furthermore, we must discuss the relationship among communication and learning through interaction with a human.

References

1. Zadeh LA (1965) Fuzzy sets. J Inform Control 8:338–353
2. Jang J-SR, Sun C-T, Mizutani E (1997) Neuro-fuzzy and soft computing, Prentice-Hall, Inc.
3. Fogel DB (1995) Evolutionary computation, IEEE Press, New York
4. Syswerda G (1991) A study of reproduction in generational and steady-state genetic algorithms, In foundations of genetic algorithms, Morgan Kaufmann Publishers, San Mateo
5. Tani J (1996) Model-based learning for mobile robot navigation from the dynamical systems perspective. IEEE Trans Syst Man Cybern B 26(3):421–436
6. Wolpert DM, Kawato M (1998) Multiple paired forward and inverse models for motor control. Neural Netw 11:1317–1329
7. Xiao J, Michalewicz Z, Zhang L, Trojanowski K (1998) Adaptive evolutionary planner/navigator for mobile robots. IEEE Trans Evol Comput 1(1):18–28
8. Fukuda T, Kubota N (1999) An intelligent robotic system based on a fuzzy approach. Proc IEEE 87(9):1448–1470
9. Cliff D, Harvey I, Husbands P (1993) Explorations in evolutionary robotics. Adapt Behav 2:73–110
10. Nolfi S, Floreano D (2000) Evolutionary robotics: the biology, intelligence, and technology of self-organizing machines. MIT, Cambridge, USA

11. Takagi H (2001) Interactive evolutionary computation: fusion of the capabilities of EC optimization and human evaluation. Proc IEEE 89(9):1275–1296
12. Katagami D, Yamada S (2003) Teacher's load and timing of teaching based on interactive evolutionary robotics. In: Proceedings of the IEEE International Symposium on Computational Intelligence in Robotics and Automation, pp 1096–1101
13. Bertsekas DP, Tsitsiklis JN (1996) Neuro-dynamic programming. Athena Scientific, Belmont, USA
14. Sutton RS, Barto AG (1998) Reinforcement learning. MIT, Cambridge, USA
15. Takadama K, Nakasuka S, Shimohara K (2002) Robustness in organizational-learning oriented classifier system. J Soft Comput 6:229–239
16. Inoue H, Takadama K, Shimohara K, Katai O (2003) Acquisition of a specialty in multi-agent learning - approach from learning classifier system. In: Proceedings of the IEEE International Symposium on Computational Intelligence in Robotics and Automation, pp 306–311
17. Brooks RA (1999) Cambrian intelligence. MIT, Cambridge, USA
18. Pfeifer R, Scheier C (1999) Understanding intelligence. MIT, Cambridge, USA
19. Arkin RC (1998) Behavior-based robotics. MIT, Cambridge, USA
20. Kubota N, Arakawa T, Fukuda T (1998) Trajectory planning and learning of a redundant manipulator with structured intelligence. J Brazilian Comput Soc 4(3):14–26
21. Kubota N (2003) Intelligent structured learning for a robot based on percieving-acting cycle. Proceedings of the 12 yale workshop on adaptive and learning systems, pp 199–206
22. Kubota N, Indra AS, Kojima F (2002) Interactive genetic algorithm for trajectory generation of a robot manipulator. In: Proceedings of the 4th Asia-Pacific conference on simulated evolution and learning, pp 146–150
23. Nojima Y, Kojima F, Kubota N (2003) Trajectory generation for human-friendly behavior of partner robot using fuzzy evaluating interactive genetic algorithm. In: Proceedings of the IEEE international symposium on computational intelligence in robotics and automation, pp 306-311
24. Paul RP (1981) Robot manipulators; mathematics, programming, and control. MIT, Cambridge, USA
25. Davidor Y (1991) A genetic algorithm applied to robot trajectory generation. In: Handbook of genetic algorithms, Van Nostrand, Reinhold, pp 144–165
26. Walter JA, Martinetz TM, Schulten KJ (1991) Industrial robot learns visuo-motor coordination by means of "neural-gas" network. Artif Neural Netw pp 357–364
27. Latombe H-L (1991) Robot motion planning. Kluwer Academic Publishers, Dordercht
28. Canny JF (1988) The complexity of robot motion planning. MIT, Cambridge, USA
29. Kohonen T (1982) Self-organized formation of topologically correct feature maps. Biol Cybern 43:59–69
30. Kohonen T (1984) Self-organization and associative memory. Springer, Berlin Heidelberg New York
31. Kubota N, Hisajima D, Kojima F, Fukuda T (2003) Fuzzy and neural computing for communication of a partner robot. J Multiple-Valued Logic Soft-Comput 9(2):221–239
32. Hastie T, Tibshirani R, Friedman J (2001) The elements of statistical learning. Springer, Berlin Heidelberg New York