Linguistic and Numerical Heterogeneous Data Integration with Reinforcement Learning for Humanoid Robots

Changjiu Zhou

Department of Electrical and Electronic Engineering, Singapore Polytechnic
500 Dover Road, Singapore 139651
zhoucj@sp.edu.sg, changjiu.zhou@ieee.org

Abstract. Autonomous humanoid robots have to be intelligent to determine their own motions in unknown environments based on both sensory data (numerical) and expert knowledge (linguistic), that is, the linguistic-numerical heterogeneous data (LNHD). This paper looks at the recent development in the area of fuzzy data fusion approach in conjunction with the concept of reinforcement learning (RL) for intelligent control of humanoid robots. To make the most use of the information available for humanoid robots, we classify the LNHD into four categories: direct data (DD), direct rules (DR), indirect data (ID), and indirect rules (IR). Based on fuzzy data fusion theory, the LNHD can be directly built into the RL model as a starting configuration so as to speed up its learning. We demonstrate that, even for complex tasks like humanoid robot control, the inclusion of available LNHD can have a significant effect on learning speed. The results are illustrated with a simple biped robot. It shows that the initial gait can be formed through the integration of the LNHD such as linguistic rules obtained from human intuitive walking knowledge and numerical data received from humanoids experiments. The fuzzy RL starts synthesizing humanoids gait from this initial state and thus can quickly improve its behavior.

1 Introduction

Research in humanoid robotics has uncovered a variety of new problems and a few solutions to classical problems in robotics, artificial intelligence, and control theory [2, 8, 9, 16]. Humans display complex dynamic behaviors in real time, and learn various motor skills throughout life. As an example, a human arm has 7 degrees of freedom (DOF), the hand 23, and the control of an actuated human spine is still far beyond current considerations. Humanoid robot is a highly redundant control system, involving sensory and effector limitations and various forms of uncertainties. These uncertainties are primarily due to sensor imprecision and unpredictability of the environment, that is, lack of full knowledge of the environment characteristics and dynamics. Humans, on the other hand, seem to cope very well with uncertain and unpredictable environments, often relying on approximate or qualitative (fuzzy) data and reasoning to make decisions and to accomplish their objectives. To meet such challenges, it is important to develop a humanoid robot control system that can integrate various types of sensory information and human knowledge in order to carry out tasks with or without need for human intervention. However, most of information obtained for humanoids control are hybrid, that is, their components are not homogeneous but a blend of numerical and linguistic, direct and indirect, and etc. We call the data of above types or vectors whose components are any of this, Linguistic-Numerical Heterogeneous Data (LNHD).

It is unlikely that one technique or architecture can provide a uniformly superior solution for humanoid robot control. Neural networks (NN) and conventional control theory are useful if sufficient numerical data is available or measurable, while fuzzy logic (FL) and artificial intelligence (AI) are appropriate if sufficient linguistic knowledge is available. Therefore, there is a need to use hybrid intelligent techniques [6, 35] to integrate the LNHD for humanoid robot control.

There exists voluminous literature on the subject of making use of either sensory information (numerical data) or expert knowledge (linguistic rules) for robotic planning and control [3, 12, 13, 26, 28]. However, very few papers are found that report the linguistic and numerical heterogeneous data (LNHD) integration for intelligent robotic control, and humanoids control in particular.

Several successful applications have been reported on the sensor fusion approach in conjunction with fuzzy behavior concept for intelligent robotic systems. Most of the results on data fusion for robot planning and control were obtained under some assumptions, such as the complete knowledge of the working environment is usually assumed. However, one of the greatest challenges facing the research community is to extend the

domains of application of intelligent robotic control to the general class of unstructured environments, that is, the environments that are generally dynamic, not fully known *a priori*, and typically unpredictable. Therefore, the real challenge in this area is to integrate both sensory information and human knowledge, that is, the LNHD, for intelligent robotic systems, and humanoid robot in particular in this paper. We will show that the autonomous humanoid robot can acquire basic perceptual and motor skills through the LNHD integration, however, it still lacks adaptive behavior. Hence, a learning approach is required. Most of learning methods need a supervisor that shows desired output values for each input variable. If supervised learning (SL) can be used in control, then it has been shown to be more efficient than reinforcement learning (RL) [29]. However, for some real-world control applications, such detailed and precise training data is usually difficult or even impossible to obtain. In such a case, the RL techniques are more appropriate than the SL for practical systems, such as humanoid robots with complex mechanisms in this paper. This paper focuses on how humanoid robot can set up its initial behavior by the LNHD integration, and then it can learn from this initial state rather than from scratch, so as to speed up its learning to acquire basic perceptual and motor skills.

Most of the existing RL methods focus on numerical evaluation [3, 4, 5, 20, 31], *i.e.,* allow systems to learn from *scalar* reward signals only. However, using the numerical (scalar) critical signals for the RL is surely not the biologically plausible. As an example, for human "humanoids", we usually use *fuzzy* in stead of *numerical* critical signal, such as "near fall down", "almost success", "slower", "faster", and so on, to evaluate the walking behavior. In this case, using fuzzy evaluative feedback is much closer to the learning environment in the real world. Therefore, there is a need to study the reinforcement learning with fuzzy evaluative feedback [19, 37], which is called fuzzy RL (FRL), and its applications to humanoid robot control. Most of RL applied to robots has relied on a coarse discretization of the control surface, but the FRL architecture proposed in this paper is based on a form of continuous evaluation. During the learning process, both structure learning and parameter learning are performed simultaneously. The proposed FRL method constructs an intelligent control system automatically and dynamically through the reward/penalty signals or fuzzy evaluative feedback signals.

In the following section, we introduce the information available in humanoid control and the representation of LNHD. In the third section, some information integration methods and its application to LNHD integration by means of hybrid intelligent techniques are described [6]. In the fourth section, the LNHD integration with RL is proposed for a humanoid robot. We then demonstrate how this method can be used to improve the humanoid robot behavior. In the fifth section, a simple humanoid biped robot is used to illustrate the effectiveness of the proposed method. Some simulation results of the gait synthesis by LNHD and LNHD + RL are described and compared. This is followed by some concluding remarks.

2 Linguistic-Numerical Heterogeneous Data (LNHD) in Humanoid Robot Control

2.1 The Information Available for Humanoid systems

The intelligence of a robotic system can be characterized by three functional abilities. First, the robotic system should be controlled directly as a task level; that is, it should take task-level commands directly, without any planning type decomposition to joint level commands. Second, the control systems of robots should be designed for a large class of tasks rather than a specific task. Finally, the robotic system should be able to handle some unexpected or uncertain events. Intelligent robotic control depends on a large extent of the capability of the robotic system to acquire, to process, and to utilize both *sensory information* and *human knowledge* to plan and execute actions in the presence of various changing or uncertain events in the robot-working environment.

Figure 1 shows a functional architecture for an intelligent robotic controller with an interface to an autonomous humanoid robot involving sensing and actuation, and an interface to humans and other systems [25]. The organizer provides for the supervision of lower level functions and for managing the interface to the humans. The coordinator provides for tuning, scheduling, supervision, and redesign of the control algorithms; crisis management; planning and learning capabilities for the coordination of the execution-level tasks; and higher-level symbolic decision making. The controller has low-level numeric signal processing and control algorithms. Based on the principle of increasing precision with decreasing intelligence (IPDI) [25], the highest level is assigned with the highest machine intelligence and smallest complexity (size of database) and the lowest level with the lowest machine intelligence and largest complexity. As an example, for a humanoid robot [36], at the organization level, the system reasons symbolically for strategic plans or schedules of the humanoids motion using the knowledge base. This level may also include the common sense knowledge for the biped motion. At

the organizational level, a human operator mainly gives the knowledge in a top-down manner. There are different kinds of information available for the coordination level, such as the sampled data from the experiments (numerical); the intuitive knowledge (linguistic); and the walking knowledge which has been obtained from the biomechanics studies of human walking (linguistic + numerical). At the control level, the information is mainly given by measuring instruments in a bottom-up manner (numerical). Moreover, the perceptual and motor skills' knowledge is also useful for the joint controller design and tuning. Unfortunately, each of above sets of information alone is usually incomplete. Therefore, there is a need to integrate both expert knowledge and sensory information for control and learning of humanoid robots.

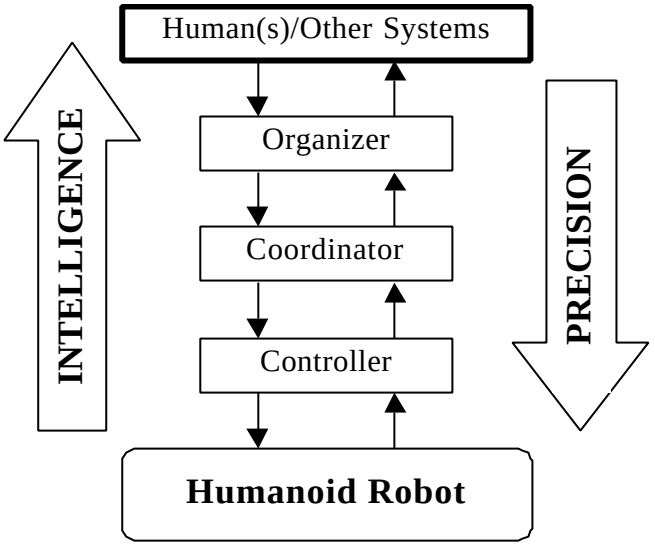

Fig. 1. Hierarchical intelligent robotic control system for a humanoid robot

2.2 Linguistic-Numerical Heterogeneous Data (LNHD)

From the above discussion we found that the information available for the humanoid robot control systems can be classified into two types of data in general: numerical data received from sensor measurements, and linguistic data obtained from human operators and domain experts. We call the data of above types or vectors whose components are any of this, Linguistic-Numerical Heterogeneous Data (LNHD).

In the real-world situations, the numerical data may be noisy, inconsistent, and incomplete, and the linguistic information is usually imprecise and vague. Therefore, a fuzzy set is a naturally good choice to integrate the LNHD. As an example, let's consider the speed measurement of a humanoid robot (see Figure 2). The speed can be expressed as a real number if the sensors can precisely measure it. If the speed is evaluated non-numerically by a human observer, it may be interpreted as the linguistic term, such as the speed is "fast." It can also be represented as a single interval. Uncalibrated instruments could lead to this situation.

In [15, 23], Pedrycz *et al.* proposed two heterogeneous fuzzy data (HFD) models: parametric and non-parametric models, and their applications to the pattern recognition. Let $\Re$ be the real numbers, $I(\Re) = I$ be all real closed intervals such as $[a,b]$, and $F(\Re) = F$ be the real fuzzy subsets of $\Re$. An element of F is a membership function $m: \Re \alpha\ [0,1]$ whose values $\{m(x) : x \in \Re\}$ are the grades of membership of each x in the fuzzy set m. Every real number and interval have crisp membership functions. Thus the HFD (real numbers, real intervals, and fuzzy sets) are all the elements of F. In this way, the most general form of the HFD becomes a collection of n $p-$tuples $M = \{m_1, m_2, \Lambda\ , m_n\} \subset F^p$, where p is the number of features observed in the chosen representation of the HFD. The above method assumes all the sensors and experts have the same importance, that is, the same degree of truth. However, for some applications, each sensor and expert may not have the same contribution to the final fusion decision. On the other hand, for humanoid robot control systems, there are some special problems to be considered.

As an example of the LNHD, we consider a humanoid robot control system that consists of at least a planner and a controller. For controller design, there are two types of expert knowledge. One is the *robot knowledge* that describes the behavior of the robots; another is *control knowledge* that can be described as fuzzy IF-THEN rules that state in which situations what control actions should be taken. The control knowledge can be *directly* used to design the controller, while the robot knowledge must be used in an *indirect* way. The control knowledge may also be classified into two categories: *conscious* knowledge and *subconscious* knowledge. By conscious knowledge we mean the knowledge that can be explicitly expressed in words, therefore, we can simply ask the *human* experts to express it in terms of fuzzy IF-THEN rules. By subconscious knowledge we refer to the situations where the human experts know what to do but cannot express exactly in words how to do it. In a common sense, we all are experts in humanoid control. Everyday, we perform different humanoid tasks, however, we almost have no conscious knowledge to describe this process. What we can do is to ask the human experts to *demonstrate*, and collect a set of input-output numerical data pairs. Research on learning from imitation and demonstration [21, 27] may help us to acquire this kind of subconscious knowledge.

Based on the above consideration, in order to make the most use of the LNHD available for humanoid robot control, we classify the LNHD into four categories [36]: direct data (DD), direct rules (DR), indirect data (ID), and indirect rules (IR). The DD is used to describe the subconscious knowledge of the humanoid control. The DR is the experts' linguistic description on how to control the humanoid robots. The ID is collected from humanoid demonstration, and the IR is the experts' description on the humanoid behavior rather than how to control it.

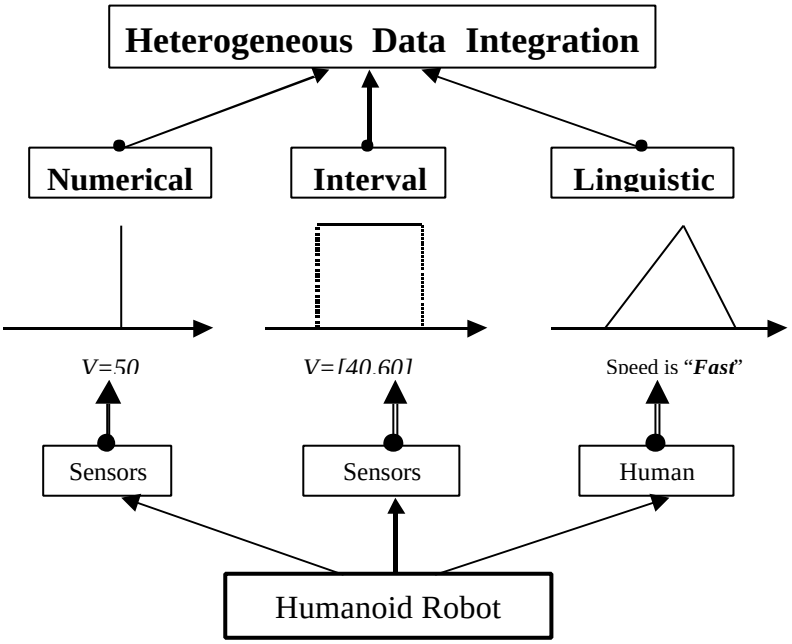

Fig. 2. Example: heterogeneous data integration for the speed measurement of a humanoid robot.

2.3 Representation of the LNHD

In this paper, we consider a single-output fuzzy rule based system in the *n*-dimensional input space $[0,1]^n$. Let us assume that human experts give the following fuzzy IF-THEN rules:

$$R^{(l)}: IF \quad x_1 \ \ is \ \ F_1^l \ \ and \Lambda \ and \ \ x_n \ \ is \ \ F_n^l \qquad THEN \ \ y \ \ is \ \ G^l \tag{1}$$

Where F_i^l ($i = 1,2,\Lambda\ ,n$) and G^l are fuzzy sets, $l = 1,2,\Lambda\ ,M$, and M is the number of rules in the fuzzy rule base. $X = (x_1,\Lambda\ , x_n)^T \in U \subset \Re^n$ and $y \in V \subset \Re$ are input and output linguistic variables, respectively.

We also assume that the following input-output pairs are given as numerical data from measuring instruments:

$$(x_1^p, x_2^p, \Lambda\ , x_n^p; y^p) \tag{2}$$

Where $p = 1,2,\Lambda\ ,P$, and P is the number of input-output data pairs. $X_p = (x_1^p, x_2^p, \Lambda\ , x_n^p)$ is the input vector of the p th input-output pair and y^p is the corresponding output.

As an example, we assume that a robot has only one state $x(k)$ and one input $u(k)$, the control objective is to design a controller such that the robot state $x(k)$ can follow a desired trajectory $x_d(k)$ as closely as possible. If we assume that the order of the model is given, and all the state variables are measurable, and the inverse dynamics of the model exist, then we have

$$\begin{aligned} x(k+n) &= F(x(k), u(k)) \\ u(k) &= G(x(k), x(k+n)) \end{aligned} \tag{3}$$

For the above robot control system, the LNHD can be described as following.

- Direct Rules (DR): $IF \quad x(k) \quad AND \quad x(k+n) \quad THEN \quad u(k)$
- Indirect Rules (IR): $IF \quad x(k) \quad AND \quad u(k) \quad THEN \quad x(k+n)$
- Direct Data (DD): $(x(k), x(k+n); u(k))$
- Indirect Data (ID): $(x(k), u(k); x(k+n))$

3 Linguistic and Numerical Heterogeneous Data (LNHD) Integration

3.1 Information integration

In recent years, multi-sensor data fusion technology has been rapidly evolved. Numerous prototype systems have been developed. Robot designers can choose from a large number of sensor types and sensing modules. These are sometimes complementary, sometimes redundant, and there also exist architectures where sensors are used in both fashions.

Data fusion operators have been traditionally defined in a numerical setting. Most numerical data fusion operations are those based on the weighted mean and the ordered weighted average. However, in many applications, such as the humanoid robot systems, the information to be integrated is not the pure qualitative or ordinal. In such cases, the usual technique is to map the quantitative values and weights into a numerical scale, and then perform the integration of those numerical values, and optionally go back to a value in an ordinal scale, if needed.

When pieces of information issued from several sources have to be integrated, each of them represented as a degree of belief in a given event, these degrees are combined in the form $F(x_1, x_2, \Lambda\ , x_n)$, where x_i denotes the representation of information issued from sensor i. A large variety of information combination operators F for data fusion have been proposed [5, 7, 10].

Let us consider a function F acting only two pieces of information x and y. Where x and y denote two real variables representing the degree of belief to be combined, they take values into the interval $I \subset [0,1]$. Under closure property assumption, $F(x, y)$ also has values in the interval I. Based on the definitions for fuzzy operators, the fusion operator F can be classified into three categories:

- F is conjunctive if $F(x, y) \le \min(x, y)$.
- F is disjunctive if $F(x, y) \ge \max(x, y)$.

- F behave like a compromise if $\min(x, y) \le F(x, y) \le \max(x, y)$.

In the framework of fuzzy sets and possibility theory, triangular norms (T-norms), triangular conorms (T-conorms), and mean operators are typical examples of conjunctive, disjunctive, and compromise fusion operators respectively. Therefore, we may choose some fuzzy operators for information integration. The following are some most common used fusion operators.

- T-norm operators: $\min(x, y)$, xy, $\max(0, x + y - 1)$.
- T-conorm operators: $\max(x, y)$, $x + y - xy$, $\min(1, x + y)$.
- Mean operators: the median operators $m_a(x, y) = med_a(x, y, a)^2$, the harmonic mean $2xy/(x+y)$, the geometrical mean $\sqrt{xy}$, the arithmetical mean $(x+y)/2$, the quadratic mean $\sqrt{(x^2 + y^2)/2}$, the ordered weighted operators.

The choice of the fusion operators depends on the aim of the fusion. Bloch [5] proposed a classification of fusion operators. The Context Independent Constant Behavior operators (CICB) have the same behavior whatever the values of the information to be combined. It can be proved that the T-norms, T-conorms, and mean operators discussed above are all the CICB operators [5]. The Context Independent Variable Behavior operators (CIVB) are more or less severe, depending on the degrees of belief. The Context Dependent operators (CD) are computed not only from the degrees of belief but also depend on a global knowledge or measure on the sources to be fused, such as the degree of conflict, source reliability, special constraints. Most of the above fusion operators assume that all the information sources including sensors and expert have the same importance or degree of belief. However, for heterogeneous data fusion, each sensor or expert does not have the same contribution for the final fusion decision.

3.2 The LNHD Integration

There are various approaches to integrate the linguistic rules and numerical data. In [36] we studied different kinds of methods to integrate the ID, IR, DD and DR for controller design and learning. As an example, the comparison of the above LNHD fusion methods for the integration of the direct rules (DR) and indirect data (ID) is shown in Figure 3.

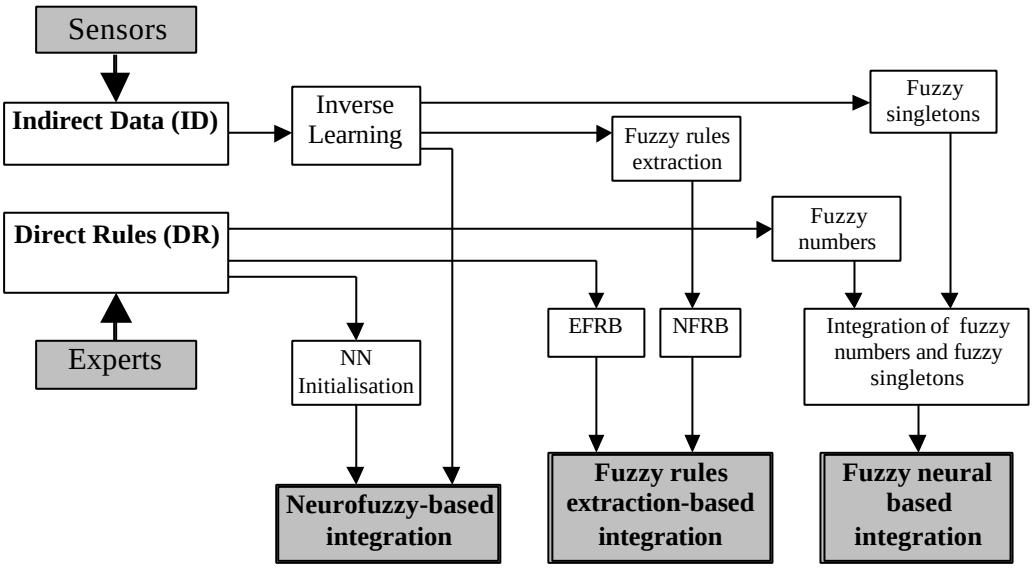

Fig. 3. The comparison of the LNHD integration methods

The straightest way is to construct a neural-network-based fuzzy system using the fuzzy rules obtained from experts, and then fine-tune its parameters by the numerical data. This is the most common method used for

neuro-fuzzy systems [6, 35, 36]. Another approach is to develop a fuzzy neural system with fuzzy supervised learning to process the integration of numerical and fuzzy information [18]. The inputs, outputs, and weights of a neural network, which is the connectionist realization of a fuzzy inference system, can be fuzzy numbers of any shape. The training data can be hybrid of fuzzy numbers and numerical numbers through the use of fuzzy singletons. In order to handle linguistic information in NN, the fuzzy data are viewed as convex and normal fuzzy sets represented by α-level sets. Since the α-level sets of fuzzy numbers are intervals, the operations of real numbers in the traditional NN are extended to closed intervals. A notable deficiency in this approach is that the computations become complex (e.g. multiplication of interval) and time-consuming. Another method is fuzzy rules extraction-based integration method. In this approach, fuzzy IF-THEN rules are first generated from input-output numerical data pairs [37]. We then construct the numerical data based fuzzy rule base (NFRB) with the degree of belief (DOB) using direct and heuristic matching method. We acquire experts-based fuzzy rules (EFRB) with the DOB from experts. The final uniform fuzzy rule base (UFRB) can be constructed through the integration of NFRB with EFRB. After normalizing NFRB and EFRB to ensure the completeness and consistency of the fuzzy rule bases, most data fusion operators can be used to integrate the LNHD through fusing the DOB.

3.3 Example: Fuzzy Rules Extraction-Based LNHD Integration

There are many methods to extract NFRB from the numerical data obtained from the sensors [37]. If the numerical database has the format as $(x_1^p, x_2^p; y^p)$, $p = 1,2,\Lambda\ ,P$, here P is the number of the data pairs, then we can construct the NFRB with the following fuzzy rules:

$$R_i^N : IF\ \ x_1\ \ is\ \ A_i\ \ AND\ \ x_2\ \ is\ \ B_i\ \ THEN\ \ y\ \ is\ \ C_i^N \quad (w_i^N) \tag{4}$$

where w_i^N is the DOB of the ith rule in the NFRB. The next task is to construct an expert knowledge-based fuzzy rule base (EFRB). Expert knowledge acquisition is one of the greatest challenges facing the AI community. For robot control knowledge acquisition, the most common problem is that the human experts know how to control the robot but cannot express in words how to do it. Thus, transferring human empirical knowledge to a controller may turn out to be a difficult task. We refer this as *subconscious* knowledge. In this case, we may extract fuzzy rules from the action of an experienced manual operator. Modeling characteristics of the human operator (HO) as an element in control systems has been the subject of much research efforts since the late 1940's [34]. Early studies have focused on identifying linear transfer functions. More recently, neural networks [22] and fuzzy logic [34] have been applied in this field. Conventional models such as transfer functions and even neural networks, although useful to reproduce human control actions, often fail to give insight in actual strategy adopted by the operator. Fuzzy IF-THEN rules can help visualizes, in a more natural way, the logical relations between a set of cause-effect data. We consider two situations for the expert knowledge acquisition. If the control expert can explicitly express the knowledge (conscious), we can acquire the knowledge in terms of fuzzy IF-THEN rules. For the subconscious knowledge, we can extract the rules from experimental human operator data.

Using above knowledge acquisition methods, we can construct an expert knowledge-based fuzzy rule base (EFRB) with the DOB. For two inputs one output controller, the fuzzy rule has the following forms:

$$R_i^E : IF\ \ x_1\ \ is\ \ A_i\ \ AND\ \ x_2\ \ is\ \ B_i\ \ THEN\ \ y\ \ is\ \ C_i^E \quad (w_i^E) \tag{5}$$

where w_i^E is the DOB of the ith rule in the EFRB.

Let us assume the fuzzy IF-THEN rules in the UFRB has the following forms:

$$R_i : IF\ \ x_1\ \ is\ \ A_i\ \ AND\ \ x_2\ \ is\ \ B_i\ \ THEN\ \ y\ \ is\ \ C_i \quad (w_i) \tag{6}$$

where w_i is the DOB of the ith rule in the UFRB.

If the fuzzy rules are consistent, that is, the THEN parts are same for the same IF condition, set $C_i = C_i^E$ or C_i^N. The DOB of UFRB can be calculated by different ways using fusion operators [5].

$$C_i = C_i^E \quad or \quad C_i^N \tag{7}$$
$$w_i = FUSION(w_i^E, w_i^N)$$

where *FUSION* is a data fusion function. For example, we can use a T co-norm operator to choose the maximum DOB, that is, $w_i = \max(w_i^E, w_i^N)$. If a global measure is needed, we can choose a mean operator, for example, a weighted averaging operator, $w_i = a w_i^E + (1-a) w_i^N$. Where α is a weighted factor. If experts knowledge plays an important role in the integrated information, we can choose a bigger α, otherwise, choose a smaller one. As an example, if only numerical information from measuring instruments is used, set $a = 0$, otherwise, if we only use linguistic rules obtained from experts, let $a = 1$. We can set $a = 0.5$ for processing the LNHD if the expert's knowledge is as important as the measuring numerical data.

By the above LNHD integration method, we can integrate both expert knowledge and sensory numerical data into a uniform fuzzy rule base (UFRB). The UFRB will be directly built into the RL model as a starting configuration so as to speed up the robot learning.

4 The LNHD Integration with Reinforcement Learning

In humans, the capability of performing a variety of motor actions by interacting with the unpredictable environment, and without any particular conscious effort, is the result of the combination of several sophisticated brain sub-systems [30]. It is widely accepted that thinking behavior (high-level component) and motor behavior (low-level component) in the brain can be identified as two different entities able to co-ordinate when necessary, even if probably residing in different regions of the brain [1]. For both aspects of human behavior, *learning* is one of the basic issues for the acquisition of motor co-ordination schemes.

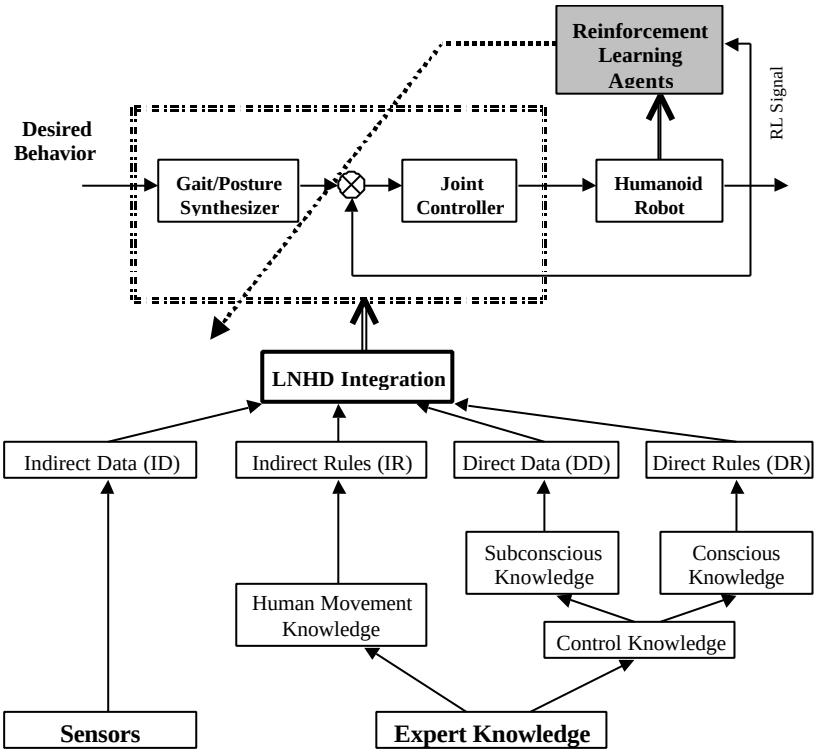

Fig. 4. LNHD integration with RL learning for a humanoid robot

Most of learning methods need a supervisor that shows desired output values for each input variable. However, such detailed and precise training data is usually difficult or even impossible to obtain for robot learning, especially, for a humanoid robot system with complex mechanisms. This section will discuss how humanoid robots can learn from both *a priori* knowledge and sensory information in order to acquire perceptual and motor skills by means of LNHD integration with reinforcement learning (RL).

For the conventional neural networks implementation of the RL, both the action-state evaluation network (AEN) and the action selection network (ASN) are initialized *randomly*, so they need a very large number of trials for learning. In this paper, the fuzzy reinforcement learning (FRL) agent is introduced, so the LNHD can be built into the FRL agent for representing the heuristic knowledge for both action selection and state evaluation. By the LNHD integration with RL, we want to explore if the corporation of a priori knowledge and sensory information can lead to substantial reduction in learning time for some complex system like humanoid robots.

The proposed intelligent control scheme for a humanoid robot is shown in Figure 4. The humanoid robot can establish its initial perceptual and motor skills through the LNHD integration described above. Then its performance can be improved by the RL further.

4.1 RL with Numerical Evaluative Feedback

The RL usually requires an unambiguous representation of states and actions of systems and existence of a *scalar* reward function. Learning proceeds by trying actions in a particular state and based on the received, possibly delayed reward, updating an evaluation function that assigns expected rewards to possible actions. After learning, the action with the highest expected value in each state is chosen to achieve the task goal. The RL applications imply the mapping of situations to actions in a huge situation-action space. On the other hand, the duration of the learning phase must be as short as possible to reduce engineering costs of development. Therefore, generalization is very important for RL. Neural networks (NN) implementation of RL is a naturally good choice because of its good generalization capability. There are several representative NN architectures for RL [4, 29, 31].

For RL problems, most of the existing learning methods of neural networks focus on *numerical* evaluative information. In this paper, the RL agent is based on a modified GARIC fuzzy-neural architecture [4]. It has three components as shown in Figure 5: the action selection network (ASN), the action evaluation network (AEN), and the stochastic action modifier (SAM). The ASN maps a state vector into a recommended action F using fuzzy inference. The AEN maps a state vector and a failure signal into a scalar score that indicates state goodness. It is also used to produce internal reinforcement $\hat{r}$. The SAM uses both F and $\hat{r}$ to produce a desired control action.

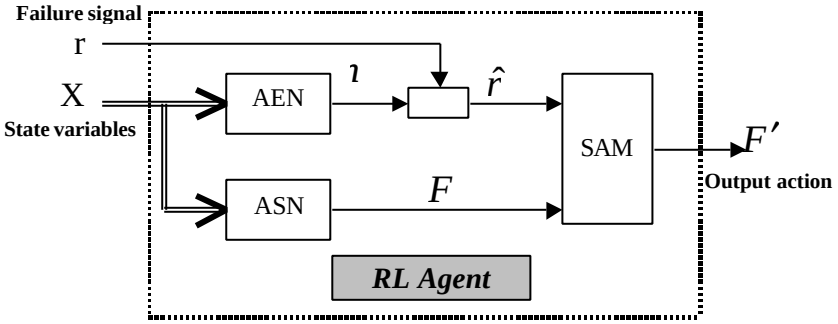

Fig. 5. The architecture of the RL agent

The learning mechanism of the AEN is similar to a reward/punishment scheme for neural networks [2].

$$\hat{r}[t+1] = \begin{cases} 0 & start \quad state \\ r[t+1] - v[t,t] & failure \quad state \\ r[t+1] + gv[t,t+1] - v[t,t] & otherwise \end{cases} \tag{8}$$

where $\hat{r}[t+1]$ is the internal reinforcement at time $t+1$, it plays the role of an error measure in the update of the output weights (since it is impossible to measure error directly). $0 \le g \le 1$ is the discount rate. $v[t,t]$ is an evaluation of the state (a prediction of the external reinforcement signal). $r[t+1]$ is an external failure signal. The following are some most common used numerical reinforcement signals.

4.2 RL with Fuzzy Evaluative Feedback

Using the numerical (scalar) critical signal for RL is surely not the biologically plausible. In stead of using scalar critical signal $r(t)$, we consider the reinforcement signal as a fuzzy number $R(t)$. We also assume that $R(t)$ is the fuzzy signal available at time step t and caused by the input and action chosen at time step $t-1$ or even affected by earlier inputs and actions. Let's assume the fuzzy reinforcement signal is expressed by a fuzzy number such that

$$R(t) \in \{R_1, R_2, \Lambda\ , R_n\} \tag{9}$$

It should satisfy the following inequality relation:

$$-1 \le defuzzifier(R_1) \le defuzzifier(R_2) \le \Lambda\ \le defuzzifier(R_n) \le 1 \tag{10}$$

where $defuzzifier(R_i)$ represents the discrete degree of reward or penalty.

Based on a modified architecture of the RL agent with numerical evaluative feedback as shown in Figure 5, we propose a FRL agent architecture given in Figure 6. It can generate the failure signal using *fuzzy* evaluation rule base by means of a fuzzy inference system similar to fuzzy controller.

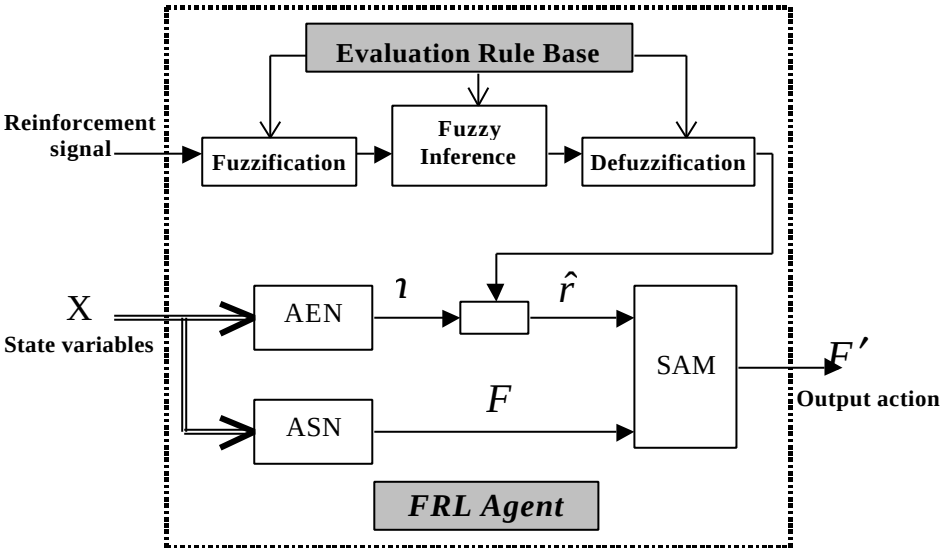

Fig. 6. The architecture of the FRL agent

The proposed FRL algorithm with fuzzy evaluative feedback can be summarized as following.

Step 1: Construct the AEN (2-layer).
Step 2: Using fuzzy control rule base to construct the ASN and set its initial parameters.
Step 3: Using the ASN to generate the recommended control action F.
Step 4: Using fuzzy evaluation rule base to generate reinforcement signals $r(t-1)$ by the Fuzzy Critical Signal Processor as shown in Figure 6.

Step 5: Calculate the internal reinforcement $\hat{r}(t-1)$.

Step 6: Using F and $\hat{r}(t-1)$ to calculate F' through the SAM.

Step 7: Update the AEN and the ASN using $\hat{r}(t-1)$.

Step 8: If reach to the maximum number of trials, then STOP; otherwise, goto Step 3.

5 Case Study: Humanoid Biped Gait Synthesis Using LNHD Integration with RL

In this section, experiment and simulation are presented to illustrate the proposed LNHD with RL and its application to gait synthesis for a simple humanoid biped as shown in Figure 7.

Humanoid biped control is a challenging problem due to its high order nonlinear dynamics and uncertainty. In the following, we demonstrate how to use the LNHD integration with RL to improve the gait of a human biped. The dynamic equation of the biped robot in workspace can be derived from the Lagrangian approach as [32].

$$D(q)\ddot{q} + C(q,\dot{q})\dot{q} + G(q) = t \tag{11}$$

where $t \in \Re^n$ is the vector of generalized torque, $q \in \Re^n$ is the vector of the joint angular position, t_i is corresponding to the variable q_i, $D(q)$ is the positive symmetric $n \times n$ inertia matrix, $C(q,\dot{q})$ is the $n \times n$ matrix which includes terms from the centrifugal and Coriolis toques, and $G(q)$ is the n-dimensional vector which includes the gravitational torque.

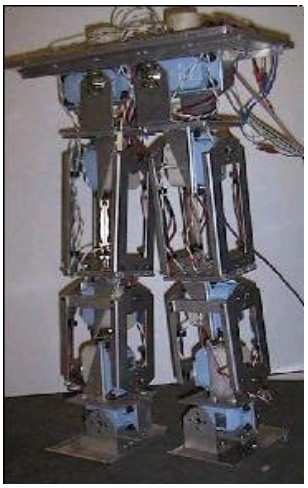

Fig. 7. A simple humanoid biped

To synthesize the biped walking motion, it is required to take a workspace variable $\boldsymbol{p}$ from an initial position p_i at time t_i to a final position p_f at time t_f. The motion trajectory for $p(t)$ can be obtained by solving an optimization problem. To achieve the dynamic walking, the zero moment point (ZMP) [32] is usually used as a criterion, therefore, we can determine the biped motion to minimize the following performance index

$$Minimize \quad \int_{t_i}^{t_f} \left\| P_{zmp}(t) - P_{zmp}^{d}(t) \right\|^2 dt \tag{12}$$

subject to the boundary conditions of both $p(t)$ and $\dot{p}(t)$ at time t_i and t_f, where P_{zmp} is the actual ZMP, and P_{zmp}^{d} is the desired ZMP position. The control objective of the gait synthesis for biped dynamic walking can be described as

$$P_{zmp} = (x_{zmp}, y_{zmp}, 0) \in S \quad \textbf{(13)}$$

where ($x_{zmp}, y_{zmp}, 0$) is the coordinate of the ZMP with respect to O-XYZ. S is the domain of the supporting area.

For the RL with numerical evaluative feedback, the reinforcement signal is generated according to the evaluation of the actual ZMP, $r(t) = \begin{cases} 1 & p_{zmp} \in S \\ -1 & failure \end{cases}$.

For the RL with fuzzy evaluative feedback, the generation of the reinforcement signal is based on the experts' evaluation. It can be described as some fuzzy IF-THEN rules. As an example, the following fuzzy rules can be used to evaluate the biped dynamic balancing in the sagittal plane.

IF $|\Delta x_{zmp}(t)|$ *is* **Big** *THEN* $r_x(t)$ *is* **Bad**

IF $|\Delta x_{zmp}(t)|$ *is* **Medium** *THEN* $r_x(t)$ *is* **Good**

IF $|\Delta x_{zmp}(t)|$ *is* **Small** *THEN* $r_x(t)$ *is* **Excellent**

where $r_x(t)$ is the reinforcement signal for training the FRL in segittal plane. $|x_{zmp}(t)|$ is the displacement of the ZMP (zero moment point), which is usually used as a criterion to evaluate the dynamic walking of the biped, in the sagittal plane.

The proposed LNHD integration with RL based gait synthesizer is shown in Figure 9. For the simulation set up, please refer to [39].

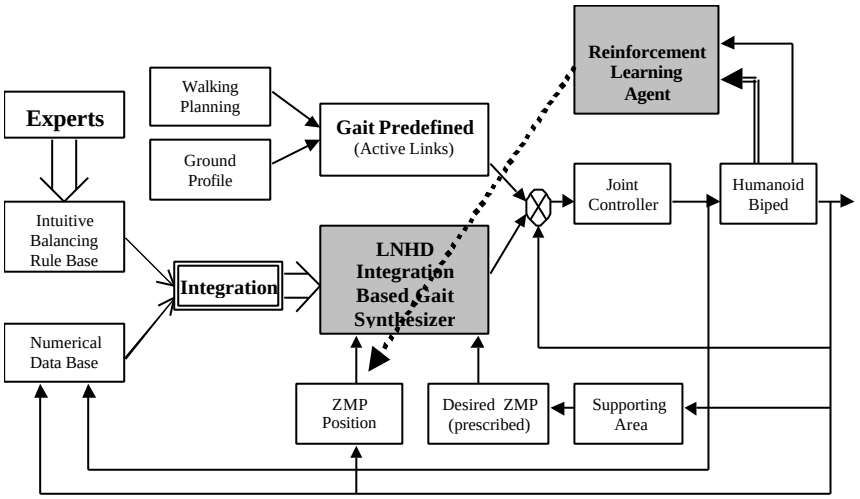

Fig. 8. Humanoid biped gait synthesizing system based on LNHD integration with RL

Based on intuitive balancing knowledge, 25 gait synthesis fuzzy rules for sagittal plane and another 25 rules for the frontal plane are obtained [36, 39]. As an example, the following fuzzy rule can be used to balance the biped in the sagittal plane,

IF *X* *is* **Positive_Medium**
AND *DX* *is* **Negative_Small**
THEN *swing in the sagittal plane* *is* **Positive_Small**

Where $X(t) = x_{zmp}(t) - x^{d}_{zmp}(t)$, $DX(t) = X(t) - X(t-T)$. x_{zmp} and y_{zmp} are actual ZMP in the sagittal and frontal plane respectively. x^{d}_{zmp} and y^{d}_{zmp} are desired ZMP.

Figure 9 shows the gait synthesizing result by the LNHD integration based on of the EFRB obtained from human balancing knowledge, and the NFRB which is extracted from the biped control experiments. We found that the actual ZMP can roughly track the desired trajectory, however, the ZMP tracking error is very big. Based on the initial gait established by LNHD integration, we conduct an experiment by using the RL agent with

numerical evaluative feedback shown in Figure 5. The result is given in Figure 10. It can be seen that the gait has been improved. From Figure 11, we can see that the ZMP is much closer to the desired trajectory by using the RL agent with fuzzy evaluative feedback as shown in Figure 6.

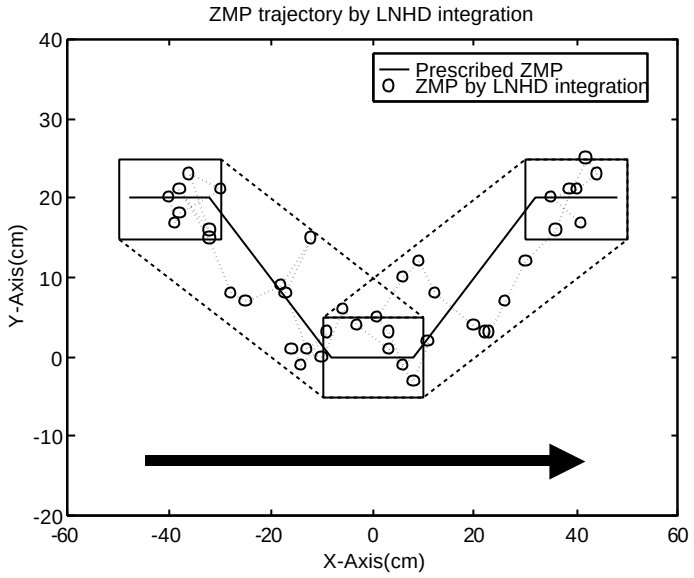

Fig. 9. Initial gait by the LNHD integration

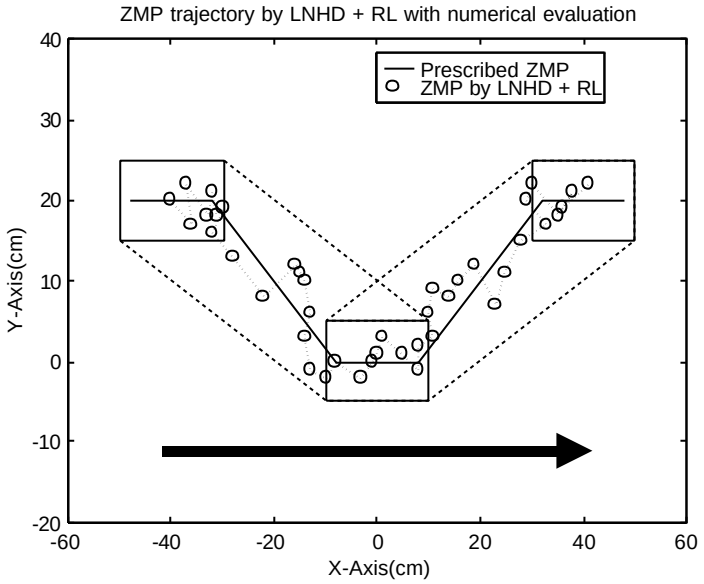

Fig. 10. The ZMP tracing result by RL with numerical evaluative feedback

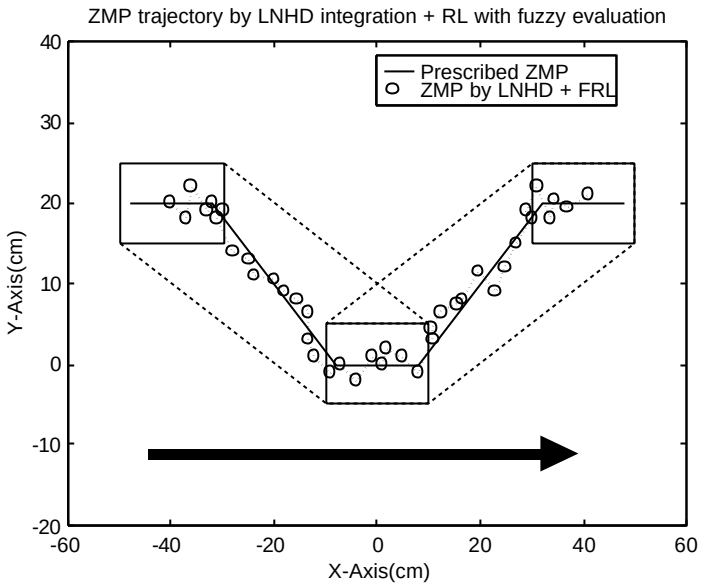

Fig. 11. The gait improved by the RL with fuzzy evaluative feedback

6 Conclusions

In this paper we look at the recent development in the area of fuzzy data fusion approach in conjunction with the concept of reinforcement learning (RL) for humanoid robots. Then we present the results of a research aiming at making the most use of the information available to improve the humanoids' behavior through the LNHD integration with RL. We demonstrate how the incorporation of the LNHD could lead to substantial reduction in learning time for humanoids. The results show that it is possible for the humanoid robots to start with heuristic knowledge and sensory information such as the LNHD and learn to refine it through experiments using RL. Following are some further points:

1) The hybrid information for humanoid robot control usually has different forms, symbolic and numerical, precise and fuzzy, direct and indirect, and etc., however, the proposed method can only process a few simple forms of hybrid information. Therefore, there is a need to develop a more general LNHD integration method for humanoid robot.
2) The proposed method may allow the integration of fuzzy rules extracted from different experts and numerical databases. Research should be conducted on whether and when the LNHD integration yields better results than simply averaging the different kinds of information.
3) The existence of the inverse of the plant or controller is not valid in general. Some results from adaptive inverse control [33] should be useful for the general fusion of the LNHD with inverse mapping properties.
4) For humanoid robots with many degree of freedoms (DOFs), there is an exponential explosion in the number of actions that can be taken in every state. It is impossible to search a huge state-action space, hence, it is necessary to either find a more compact state-action representation, or to focus learning on those parts of the state-action space that are actually relevant for the movement task at hand. Some research works show that both topics can be approached in the framework of imitation learning [27].
5) The RL approach used in this paper is with the most general actor-critic architecture. Its main drawback is that it usually suffers from the local minima problem in the network learning due to the use of gradient descent learning method. Because genetic algorithm (GA) doesn't require derivative information, and it is a general purpose optimization algorithm. Some research shows that the integration of the actor-critic architecture with the GA can solve the local minima problem by using the global optimization capability of the GA.

The RL is too slow for the real time control of the humanoid robots. In contrast, learning by imitation is particularly appealing because it allows the designer to specify entire behaviors by demonstration, instead of using low-level programming or trial and error by the robots [21]. On the other hand, the FRL agent can house available expert knowledge and sensory information to speed up its learning. How to use the LNHD integration

as a starting configuration for the RL, imitation learning, and other machine learning methods to speed up humanoid robot learning will be our future research.

Acknowledgments

I would like to thank Dr. D. Ruan at the Belgian Nuclear Research Center for his helpful comments and suggestions on the LNHD integration. I would also like to thank Dr. J. Kanniah, Mr. S. H. Ker, Miss S. C. Oh, and my students Adrian Kong and Y.W. Tham at the Singapore Robotic Games (SRG) Lab of Singapore Polytechnic for their support in biped hardware development. The research described in this paper was made possible by support of Singapore Tote Fund.

References

1. Albus, J.S.: Brains, Behavior and Robotics. Byte Books, Mc Graw Hill, Peterborough, N.H. (1981)
2. Bekey, G.A.: On autonomous robots. Knowledge Engineering Review **13** (1998) 143-146
3. Beom, H. K., and Cho, H. S.: A sensor-based navigation for a mobile robot using fuzzy logic and reinforcement learning. IEEE Trans. Systems Man Cybernetics **25** (1995) 464-477
4. Berenji, H. R., and Khedkar, K.: Learning and tuning fuzzy logic controllers through reinforcements. IEEE Trans. Neural Networks **3** (1992) 724-740
5. Bloch, I.: Information combination operators for data fusion: a comparative review with classification. IEEE Trans. on Systems, Man, and Cybernetics **26** (1996) 52-67
6. Bonissone, P.P., Chen, Y.-T., Goebel, K. and Khedkar, P.S.: Hybrid soft computing systems: industrial and commercial applications. Proc. IEEE **87** (1999) 1641-1667
7. Bouchon-Meunier, B. (ed.): Aggregation and Fusion of Imperfect Information. Physica-Verlag (1998)
8. Brooks, R.A., Breazeal, C., Marjanovic, M., Scassellati, B., and Willimson, M.M.: The cog project: building a humanoid robot.
9. Brooks, R.A.: Prospects for human level intelligence for humanoid robots. Proc. 1st Int. Symposium on Humanoid Robots (1996) 17-24
10. Dubois, D. and Prade, H.: Combination of fuzzy information in the framework of possibility theory. Data Fusion and Machine Intelligence. Academic Press (1992) 481-505
11. Filippidis, A..: Data fusion using sensor data and a priori information. Control Eng. Practice **4** (1996) 43-53
12. Ghosh, B. K., Xi, N., and Tarn, T. J. (eds): Control in Robotics and Automation: Sensor-Based Integrattion. Academic Press (1999)
13. Goodridge, S.G., Kay, M.G., and Luo, R.C.: Multilayered fuzzy behavior fusion for real-time reactive control of systems with multiple sensors. IEEE Trans. Industrial Electronics **43** (1996) 387-394
14. Harris, G.F. and Smith, P.A.(eds): Human Motion Analysis: Current Applications and Future Directions. IEEE Press (1996)
15. Hathaway, H.J., Bezdek, J.C., and Pedrycz, W.: A parametric model for fusing heterogeneous fuzzy data. IEEE Trans. on Fuzzy Systems **4** (1996) 270-281
16. Hirai, K., Hirose, M., Haikawa, Y., and Takenaka, T.: The development of the Honda humanoid robot. Proc. IEEE Int. Conf. On Robotics and Automation (1998)
17. Hong, L. and Wang, G.-J.: Centralised integration of multisensor noisy and fuzzy data. IEE Proc. Control Theory Appl. **142** (1995) 459-465
18. Lin, C.-T. and Lu, Y.C.: A neural fuzzy system with fuzzy supervised learning. IEEE Trans. Systems, Man, and Cybernetics **26** (1996) 744-763
19. Lin, C.-T. and Kam, M.-C.: Adaptive fuzzy command acquisition with reinforcement learning. IEEE Trans. on Fuzzy Systems **6** (1998) 102-121
20. Mataric, M. J.: Reinforcement learning in the multi-robot domain. Autonomous Robots **4** (1997) 73-83
21. Mataric, M. J.: Getting humanoids to move and imitate. IEEE Intelligent Systems (2000)
22. Nechyba, M.C. and Xu, Y.: Stochastic similarity for validating human control strategy models. IEEE Trans. Robot Automation **14** (1998) 437-451
23. Pedrycz, W., Bezdek, J.C., and Hathaway, R.J.: Two nonparametric models for fusing heterogeneous fuzzy data. IEEE Trans. on Fuzzy Systems **6** (1998) 411-425
24. Pratt, J. and Pratt, G.: Intuitive control of a planar bipedal walking robot. Proc. IEEE Int. Conf. Robotics and Automation (1998) 2014-2021.

25. Saridis, G.N.: Intelligent robotic control. IEEE Trans. Automatic Control **28** (1983) 547-557
26. Shi, X., Lever, P.J.A., and Wang, F.-Y.: Fuzzy behavior integration and action fusion for robotic excavation. IEEE Trans. Industrial Electronics **43** (1996) 395-402
27. Schaal, S.: Is imitation learning the route to humanoid robots? Trends in Cognitive Science **3** (1999) 233-242
28. Stover, J.A. and Gibson, R.E.: A fuzzy-logic architecture for autonomous multisensor data fusion. IEEE Trans. Industrial Electronics **43** (1996) 403-1996
29. Sutton R. S. and Barto A. G.: Reinforcement Learning: An Introduction. MIT Press (1998)
30. Taddecucci, D., et al: Model and implementation of an anthropomorphic system for sensory-motor perception. Proc. IEEE/RSJ Int. Conf. On Intelligent Robots and Systems (1998) 1962-1967
31. Touzet C. F.: Neural reinforcement learning for behavior synthesis. Robotics and Autonomous Systems **22** (1997) 251-281
32. Vokobratovic, M., Borovac, B., Surla, D., and Stokic, D.: Biped Locomotion: Dynamics, Stability, Control and Application. Springer-Verlag, (1990)
33. Widraw, B. and Walach, E.: Adaptive Inverse Control. Prentice-Hall, Englewood Cliffs, NJ (1996)
34. Zapata, G.O.A., Galvao, R.K.H., and Yoneyama, T.: Extraction fuzzy control rules from experimental human operator data. IEEE Trans. Systems, Man, and Cybernetics **B29** (1999) 398-406
35. Zhou, C., Ruan, D. and Zhu, S.: Hybrid intelligent fuzzy control for nonlinear systems. J. of Communication & Cognition and AI (CCAI) **14** (1997) 163-189
36. Zhou, C. and Ruan, D.: Integration of linguistic and numerical information for biped control. Robotics and Autonomous Systems **28** (1999) 53-70
37. Zhou, C. and Meng, Q.: Fuzy rules extraction-based integration of linguistic and numerical information for hybrid intelligent systems. Lecturer Notes in Artificial Intelligence **1531** (1998) 282-293
38. Zhou C. and Meng Q.: Reinforcement learning with fuzzy evaluative feedback for a biped robot. Proc. IEEE Int. Conf. Robotics and Automation (2000)
39. Zhou, C.: Neuro-fuzzy gait synthesis with reinforcement learning for a biped walking robot. J. Soft Computing (2000) (in press)