2009 IEEE/ASME International Conference on Advanced Intelligent Mechatronics
Suntec Convention and Exhibition Center
Singapore, July 14-17, 2009

Concurrent Learning of Task and Attention Control in the Decision Space

Maryam S. Mirian, Hadi Firouzi, Majid Nili Ahmadabadi, Babak N. Araabi

***Abstract*— learning attention control is a real need specifically when a robot tries to learn a sequential decision-making-type task. This is even more critical when learning directly in the perceptual space is not feasible mainly due to the high dimensionality thus non-homogeneity. Therefore, two learning problems are raised to be solved at the same time. In this paper, a novel approach with three learning phases is proposed to facilitate learning of these two coupled problems: 1) learning how to divide attention among multiple dimensions of robots perceptual space and also how to shift it efficiently inside one modality from one spatial part to another and 2) learning the main task. The main task is considered "driving in a simulated road using a miniature mobile robot" in order to demonstrate the necessity of attention control. An important new feature of the proposed learning method is that the attention is learned in the decision space rather than the original perceptual space and this brings some discussed advantages. Obtained results justify practicability and usefulness of learning attention control in the proposed alternate space.**

I. Introduction

Shaping the complex desired behavior of a fully autonomous robot is an everlasting challenge. The expression robot shaping [1] denotes the use of learning as a means to translate suggestions coming from an external trainer into an effective control strategy that allows a robot to do a predefined task. Shaping a complex behavior such as driving, naturally bares a huge multi-dimensional sensory space. Selective attention is thought to be necessary because there are too many things in the environment to perceive and respond to at once considering the demand for fast response in face of limited computational resources in the driving task. Lack of experience in control of spatial attention is one of the major problems that affects even human beings' driving [2]. Attention control in fact solves the information bottleneck problem and makes a manageable input sensory space out of a rather distracting one. The great necessity of attention control is in fact for reduction of probable confusion among multiple dimensions of the perceptual space and also to acquire faster response time.

According to the conceptual framework proposed for understanding the role of selective attention in driving [3], there are two ways that a selection process might work: automatic and controlled. There are a large number of papers on rule-based and hard coded attention shift and sensor selection, however; here we argue that attention shift should be learned for performing complicated tasks –like driving- as its dynamics is not fully known at the design time and changes as well. In this paper, we opt for learning attention control in multimodal space; which results in deliberative and controlled attention shift.

Demand for learning attention shift brings forth a new optimization problem in addition to the problem of optimizing motor actions in face of constraints and costs. Therefore, as in [4], we couple motor actions with those that are performed solely for the change of attention focus. It means attention shift is learned in concert with motor actions in a unified problem. In addition, we use Reinforcement Learning (RL) for formulating and solving the optimization problem. In this problem, the goal of the driving robot is to maximize its expected reward by proper selection of focus of attention –which is called perceptual action- and motor actions.

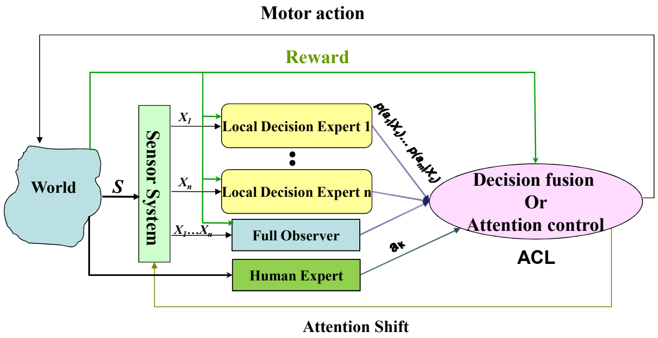

Fig. 1. Local Decision Experts (LDEs) and the full observer agent learn in parallel. Their output is their belief about appropriateness of actions based on their observation. A human expert drives the system at the initial stages of learning or suggests corrective actions in the next stages. His actions get directly through when present. Attention Control Learner (ACL) is inactive in the first stage and human actions are executed. A fixed decision fuser (uses max operator) is used instead of ACL in the second stage and the human expert is removed. In the last stage the full observer agent is also removed and ACL learns decision fusion and attention.

Driving demands dealing with multimodal and continuous perceptual space. It means that ordinary RL methods that discretize the perceptual space do not fit into the problem because of their inefficient learning. Therefore, an efficient

Maryam S. Mirian is in Robolab, ECE dept. Faculty of Engineering, University of Tehran, Iran; e-mail: mmirian@ut.ac.ir

Hadi Firouzi is in Robolab, ECE dept. Faculty of Engineering, University of Tehran, Iran; e-mail: h.firouzi@ece.ut.ac.ir

Majid Nili Ahmadabadi is in Robolab, ECE dept. Faculty of Engineering, University of Tehran, Iran, and School of Cognitive Science, IPM; e-mail: mnili@ut.ac.ir

Babak N. Araabi is in Robolab, ECE dept. Faculty of Engineering, University of Tehran, Iran, and School of Cognitive Science, IPM; e-mail: araabi@ut.ac.ir

978-1-4244-2853-3/09/$25.00 ©2009 IEEE

continuous RL algorithm based on Bayesian framework is used; see [5] and [6][1]. In addition, our robot learns the driving task and the attention shift in three phases to expedite its learning; see Fig. 1. Details are given in the next sections.

We have two options for state representation of the attention learning system. The first one is sensory data itself. The second one is the belief of the local agents about the appropriateness of motor actions. Here we choose the second alternative; see [10] for details. It is worth-mentioning that the driving[2] task is used here just as a testbed to reveal the necessity of attention control in a demanding task.

The paper contains one review section with the primary focus on driving preferably with ideas of attention and fusion. Then, the proposed approach is presented and after that the simulator as well as the acquired results is given. Finally, the conclusions and future works are discussed.

II. Related Works

In this section, we review those research mostly focused on fusion based attention control for robot navigation. In [11] a sensor fusion framework for mapping unknown environments for mobile robots is presented which enables on-line selection of the most reliable logical sensors and the most suitable fusion algorithm. The main aptitude of the framework is in its ability to select the simple sensor fusion algorithm whenever possible and the advanced algorithm when needed. This results in efficient resource utilization thus can be considered valuable from attention point of view. Our approach for attention control is somehow similar to the approach in this paper in the way that our approach is also selection of those more beneficial beliefs of LDEs trained in local sensor spaces and fusing them in an appropriate way. In [12] the cooperative localization of a heterogeneous group of road vehicles is proposed. Every member of the group maintains an estimation of the state of its environment and transmits it to its neighbors. The global state of the environment is obtained by fusing the environment states of the vehicles. This approach is almost similar to our proposed approach in the sense that we have also tried to learn the state of the environment based on the partial beliefs of the LDEs. But, we also try to map this state with the best action of the robot and shape its behavior accordingly.

III. Our Approach

As it is depicted in Fig. 1, in our approach, the input sensory space is partitioned and given to different agents learning and acting in parallel; called **L**ocal **D**ecision **E**xperts (LDEs). The world is partially observable for these agents. There are two other entities in our framework; namely the advisor and the full observer agent. The advisor acts as a driving instructor or an expert driver when an LDE agent is sitting next to it. The full observer agent has access to the full sensory space and learns in parallel with the LDEs. It simulates the situation when the driving agent is driving slowly thus carefully or drives in a controlled environment and has sufficient resources to look at all of its sensory information. Existence of such a full observer speeds up the learning process of LDEs. Note that the number of the learning agents does not change the number of required learning trials. It is because all agents observe the state in parallel, sense the executed action concurrently, get the reinforcement signal equally and finally learn in parallel.

In the **first phase** of learning, there is no attention control and all LDEs and the full observer agent learn based on the actions of the advisor and the reward function. It means that these agents learn the value of the expert's action in their sensory space.

In the **second phase,** the advisor is removed and a fixed decision fuser based on *max* operator is used to fuse all agents' decisions. The learning in this phase is just based on the reward given to the motor decision found based on the fusion. Although the fusion strategy is so simple, it is merely to help the training of the LDEs in order to make them ready and as expert as possible for the next phase.

In the **third phase**, , the full observer agent is also removed and attention shift and decision fusion are learned as a coupled action by **A**ttention **C**ontrol **L**earner (ACL) agent while learning in LDEs still continues. Learning in this stage is based on the reward the agent gets for its motor actions and cost of its attention shifts. In other words, ACL learns to which LDEs it should attend and how to fuse their beliefs in order to maximize its expected reward. The state of ACL is shaped by concatenating the belief vector of attended LDEs; which forms a multidimensional continuous space. The dimension of this space is equal to the driver agent's number of actions multiplied by the number of LDEs. The same continuous Bayesian RL method [5] used for training LDEs (in phase 1, 2) is also employed here (in phase 3).

In this paper, six LDEs are trained over different non-overlapping spatial areas of robot's visual field and one is trained on the infra-red sensor space of the robot. Each distinct part of the following areas is assigned to an LDE to explore and learn the driving task (which will be clearly defined in section IV):

- Middle-Near part of the scene (MN)
- Left-Near part of the scene (LN)
- Right-Near part of the scene (RN)
- Middle-Far part of the scene (MF)
- Left-Far part of the scene (LF)
- Right-Fat part of the scene (RF)
- Eight Infra-red sensors around the robot body (IR)

[1] The idea of this paper is inherently independent from the employed continuous RL algorithm. Every other continuous RL algorithm such as approaches in [7], [8], [9] can be employed to handle continuous space. Therefore, this algorithm is briefly explained in the appendix and interested users are referred to [5].

[2] It is tried to keep the main concerns of driving however; we have simplified them to be implementable on a miniature driving simulator.

A. Phase 1, 2: Training Local Decision Experts in the Perceptual Space

The agents participate in these two phases observe their own partial spatial area of the perceptual space.

In the first phase, the action selection and also doing the action is the human advisor's responsibility. The LDEs just sense the action and the feedback from the environment. The association is then reinforced in their minds. This phase is considered just for helping the LDEs not to start from the scratch and also in order to make the learning more analogous to the driving learning of a human.

In the second phase, LDEs have more important roles. In addition to partially observing the state, they propose their decisions at each state to the fusion center. The final decision is made based on a measure of *maximum* expertness[3] of LDEs on that state. Then, the final decision is applied to the environment. The corresponding reinforcement signal is received and used to update the learning of all participating LDEs. As soon as their learning is converged, the full observer is removed and the other partial observers turn out to be our local decision experts[4]. As mentioned before, LDEs and ACL learn by a Bayesian continuous RL algorithm [5]. In this algorithm, the suitability of the i_{th} action in the state $X^s(t)$ can be encoded as the probability $P(action_i \mid X(t))$. These distributions are learned to shape LDE's beliefs through the mentioned continuous RL algorithm (at the second phase) and proposed to ACL in an on-demand manner (in the third phase). More about the computation of this term can be found in Appendix. For LDEs and the full observer, the state space *S* and action space *A* are defined below and the *reward function* which is correlated to the driving task is defined in section IV:

$$A = Actions_{motor} \tag{1}$$

$$S_{LDE_{1,2,3,4,5,6}} = \{s \mid s \in [O_1, O_2, ..., O_c]\} \tag{2}$$

$$S_{FullObserver} = \{s \mid s \in [S_{LDE_1}, S_{LDE_2}, ..., S_{LDE_6}]\} \tag{3}$$

$$S_{LDE_7} = \{s \mid s \in [ir_1, ir_2, ..., ir_8]\} \tag{4}$$

Where O_i is the amount of dominant $Object_i$ or $color_i$ existing in the corresponding spatial part of LDEs of vision at each state, *C* is the number of dominant objects (colors) of interest and ir_i is the value of i^th IR sensor. More details will come in section IV.

B. Phase 3: Learning Attention Control

The A*ttention Control Learner* (ACL) agent marked bold in Fig. 2, at each step, has two types of actions to decide between: either consult an LDE (perform a perceptual action) or to perform a motor action. The ACL agent's state is changed either mentally[5], when it consults an expert or physically if a motor action is performed. In fact, its mental state is transferred from an initial coarse view[6] of the scene to a completely clear form when it consecutively decides to attend to the decision of all LDEs. Here, consulting a local expert means paying attention to the corresponding sensory input, then processing it and finally making decision accordingly. For ACL agent, the state space *S* and action space *A* are defined below:

$$A = \{Actions_{perceptual} \cup Actions_{motor}\} \tag{5}$$

$$S_{ACL} = \{s \mid s \in \bigcup_{i=1}^{L} (D_{LDE_i} \parallel null) \tag{6}$$

Where *L* is the number of LDEs and D_{LDE_i} is a pdf constructed according to the i^th local expert's decision. When an LDE is not attended, null is considered in the ACL's state instead. D_{LDE_i} is defined as:

$$D_{LDE_i} = [P(action_i \mid \hat{X})]_{i=1}^{N} \tag{7}$$

where *N* is the number of motor actions and $\hat{X}$ is the partial observation space available to the *LDE_i*. This is the mechanism for belief composition which in fact defines the state of our learner.

The *reward function* in this step has two parts: motory and perceptual. The first motory part (which is the same for all three phases of the framework) as well as the second perceptual part is defined in section IV (where the testbed is introduced.)

In order to make the ACL agent to learn the decision-making with minimum possible cost -the main goal of attention control- each consultation with an LDE bares a cost. Based on the complexity of the situation the robot faces with, it consults a proper number of experts.

IV. Testbed

The simulation is done in a 3D environment[7] created in Webots™ [13] containing an irregular shaped road (with 2

[3] Basically, the expertness of a reinforcement learning agent in a specific state can be characterized based on the following two criteria: 1) the agent should have visited the state enough times, 2) the agent should have received higher rewards, in accordance with number of visits. Therefore, we considered 'max' fusion operator to fuse the decision of LDEs and find the motor decision in phase 2.

[4] Actually, this is the time we can call them 'local decision experts' but for simplicity we call them LDE in entire paper.

[5] The robot keeps doing its last selected motor action until the next one is selected.

[6] This rough view is shaped by one default perceptual action (here the belief of the middle near LDE) that is performed at the start of each episode to help the ACL to find gist of the scene.

[7] At the first phase, the partial observers and also the full observer are trained 'by demonstration' of the human advisor's driving. In order to realize the first step of learning (i.e. demonstration-based learning by a human expert), we used the keyboard's arrow keys (↑ for 'go forward fast', ← for 'turn left', → for 'turn right') plus two other keys (**m** for 'go with medium speed' and **s** for 'slow down') to select one out of five possible motor actions as the best action according to the view that the robot perceives from the road scene.

lines) plus multiple miniature vehicles (two other miniature cars) of simulated E-puck 0 robots. The E-puck robot is facilitated with a color camera and eight infra-red sensors around it shown in **Fig. 3**.a. It is noticeable that in this simulation, the learning state of the robot is shaped in MATLAB® while its body is located in Webots™. Each of six LDEs can observe one of six available areas described in section III. One more LDE is trained over IR sensor space. ($L = 7$).

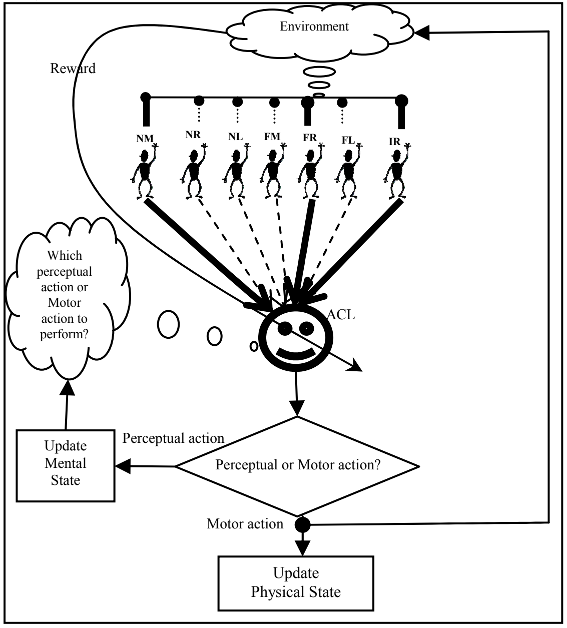

Fig. 2. Phase 3: Learning Attention Control. ACL has decided to consult NM, FR and IR LDEs one at a step from the beginning of the current learning episode till now. Maybe it prefers to consult more LDEs and maybe it stops by performing a motor action. Therefore, each of three LDEs has actively observed the environment then proposed own belief to ACL one at a time to make a mental state for ACL.

Dominant objects (C) of the scene which are depicted with distinct dominant colors (for simplifying object detection) shown in **Fig. 3**.b are:

- the road which is black
- the white lines of the road
- the red obstacles (other E-pucks are covered with red bounding box)
- the free outside road which is green
- the blue color of the boundary marker which is not considered in the robot's state.

Motor Reward:

The reward of the physical actions for the agents is defined according to the fact that whether it has collided to any other car or not, the amount of robot's bottom area which is inside the road boundary, appropriateness of its view to the road (according to the amount of each color observed in each part) and finally the appropriateness of its velocity (difference with the maximum allowed velocity). As seen, the maximum value of *r(t)* for motor actions is zero, i.e. the robot should try to minimize the punishment.

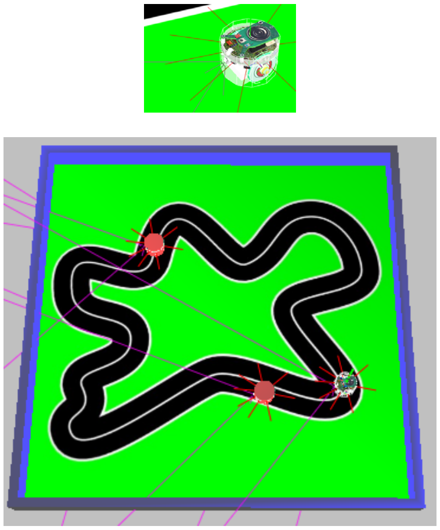

Fig. 3.

(a) simulated E-puck robot,

(b) One snapshot of the environment: containing two passive red e-pucks and our learner.

The reason behind such design of reward function is that we should prioritize the factors of a good driving behavior to help the robot find the effect of actions and not to become confused:

- If there is an accident, it should be definitely punished disregarding all other concerns.
- Otherwise, if there is no accident, it should be punished if it is out of the road boundary.
- Otherwise, if it is inside the road, it should be punished if it has not a good view to the road.
- Otherwise, if it is in road and has a good view, it should be punished for its inappropriate velocity.
- Otherwise, if none of the above penalty conditions are met, there will be no punishment.
- Otherwise, there is no reason for punishment.

Perceptual Reward:

If perceptual action is "Consulting IR-LDE" cost = -0.5 is considered. For perceptual action of "Consulting vision-LDEs of LN, RN, LF, RF" cost = -2 is given. Finally, for perceptual action of "consulting Vision-LDEs of MN, MF" cost = -1.5 is considered. Thus, average punishment unit[8] is equal to -2.

As seen, the maximum value of motor reward is zero, i.e. the robot should try to minimize the punishment. It is important to note that each perceptual action also bares a cost. This is to make the robot take as least perceptual step as possible according to the situation it is faced. The cost of processing spatial parts of vision is considered less than that of perceiving IR sensor values to make the robot to prefer utilization of least cost sensors. Moreover, the cost of processing middle parts (either far or near) is considered slightly less than those of four spatial areas. This is to push the robot to mostly prefer the front view.

[8] This is the average cost of perceptual actions that full observer performs and also that of the fusion method (in the second phase).

V. Simulation results

As shown in Fig. 4, Results express that the robot can learn attention control as well as its main job in a simultaneous manner in the decision space. The robot consults with those LDEs (on spatial visual areas or even on IR space) it found their opinion more helpful in a specific state and decides which motor action to perform to gain more rewards.

The results of the no-attention-control case (i.e. full observer) are also shown in Fig. 4. In this case, the robot has enough sources to fully consider all the decisions of LDEs and fuse them to find the best action. To have a fair comparison, for the full observer, the A*verage Punishment Unit* (-2) multiplied by the number of LDEs (7) is considered to adjust the total reward[9] shown in Fig. 4 (in dash dot line format).

The average number of perceptual steps the robot takes in each state is 2.5 using attention control while it is 7 in the case without attention control. This shows that the robot has learned to shift its attention among modalities of vision and IR and also has learned to divide its attention efficiently inside modality of vision. It is worth-noting that there are harsh conditions in which the robot decides to consult 5, 6 or even 7 number of experts, but this is on-demand and almost rare. Fig. 5 demonstrates the histogram of bins of number of perceptual steps that the ACL has taken before performing a motor action. As it is shown, when the learning progresses, the percentage of selecting less number of perceptual steps increases and the robot does not pay to look at every part. Rather, it selects more beneficial perceptual actions. This is a superior sign of learning which shows that the robot has learned which experts are wise to be consulted in a situation.

There are some interesting emerged attention behaviors by the ACL (i.e. robot): when the robot is faced with a free road (according to the rough initial gist) it regularly checks middle parts (both far and near or just one of them). When the scene seems to contain an obstacle very close to the robot, it initially checks the IR modality and then immediately checks the corresponding spatial part of the vision space according to the position the obstacle has been sensed. In the case the robot wants to turn and follow the road curvature, it only checks the corresponding spatial visual area and has nothing to do with IR space.[10]

VI. Conclusions, discussions and future works

It was shown that learning attention control in concert with motor action is crucial for learning to perform complicated tasks; like driving. The main outcome of the paper is to show that learning attention control is feasible in decision space when attention control is mainly spatial. Learning attention control in decision space benefits some interesting advantages over learning attention control in perceptual space. The major benefits are sharing one common space (decision space) among tiny decision agents, utilizing not-necessarily-similar learning algorithms for decision agents and finally making a more confident decision. The simulation results demonstrated that the agent could learn the task with average 2.5 attention shifts with the same level of reward and the behavior attained by a full observer agent. It means that the robot has learned to remove the unnecessary redundancy and make use of the most related (here the most rewarding) piece of information. It should be noted that solving information bottleneck is a big challenge and therefore learning attention control in concert with learning motor action is a must.

The learning core of our proposed approach works in continuous state space whereas most of existing interactive and unsupervised methods work with discrete representation of sensory space. This is a big merit that makes possible learning of different complicated tasks.

There are several extensions planned for the proposed approach. The most important theoretical one is to find a more compact yet meaningful decision space within which attention control is learned. This space may possibly bring the higher advantages such as faster learning speed, lower cost and maybe more robustness. The other one is to add a mechanism of object detection instead of current color-based detection policy. Another step which is more about the driving application is to make the robot extract (learn) some traffic rules. The only rule that the robot has already learned is to drive inside lines. It is planned to teach the robot more rules preferably in the first phase while it drives under the human supervision. This may need facilitating the robot with more powerful sensors such as range-finder and GPS to surpass the results taken solely by IR and vision.

Appendix

As mentioned before, we need a learning algorithm in the continuous space. We used the Bayesian RL framework proposed in [5]. The idea of this algorithm is based on online concept-based partitioning of the perceptual space initially proposed in [6]. It means that the perceptual space is partitioned based on the value of the agent's concepts (actions). The partitioning method employs a Bayesian formulation in order to handle inherent uncertainty in the environment and in the agent. In this method, the perceptual space is first considered as a single concept and the agent gradually and interactively partitions it using a series of multi-modal distribution functions. This partitioning is guided by the reinforcement signal. That is, the perceptual space is divided into some continuous parts where perceptual points in each part share the same action-value. In this algorithm, the suitability of the i_{th} action in the state $X^s(t)$ can be encoded as the probability $P\left(action_i \mid X(t)\right)$. The decision making problem in a continuous state space can be reduced to computing the posterior

[9] *Total_reward = motor_reward – Average_Punishment_Unit * 7*

[10] One short video is recorded from the driving behavior of the ACL robot in the road and uploaded to the conference management system.

probabilities $P(action_i \mid X(t))$.

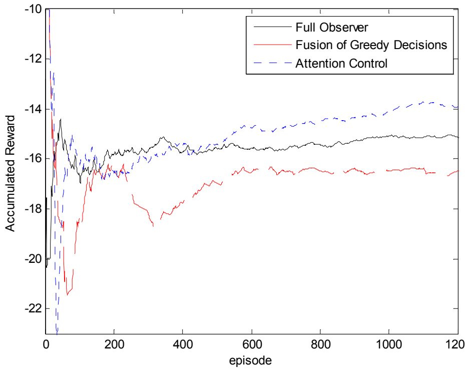

Fig. 4. The average reward acquired during learning of the driving task

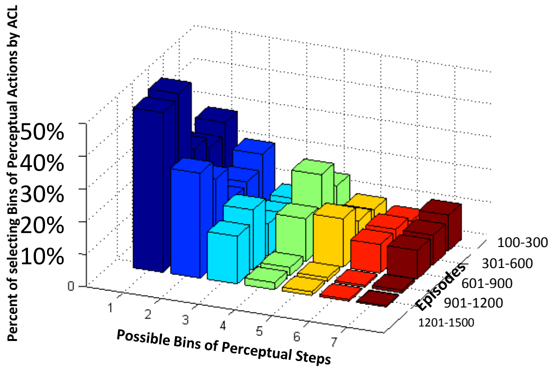

Fig. 5. The histogram of possible bins of taken perceptual steps during episodes of learning

On the other hand, by applying the Bayes rule, we have:

$$P(action_i \mid X(t)) = \eta.P(X(t) \mid action_i)P(action_i) \qquad (8)$$

where $P(X(t) \mid action_i)$ and $P(action_i)$ are the likelihood of $action_i$ and the prior probability of $action_i$, respectively. To model the probability distribution $P(X(t) \mid action_i)$, the mixture densities model is used. By substitution, we get:

$$P(action_i \mid X(t)) = \frac{P(action_i)\sum_{j=1}^{q} P(X(t) \mid M_j)P(M_j \mid action_i)}{\sum_{k=1}^{r}\left\{P(action_k)\sum_{j=1}^{q} P(X(t) \mid M_j)P(M_j \mid action_k)\right\}} \qquad (9)$$

According to equation (9), we can conclude that to compute $P(action_i \mid X(t))$, we should first estimate the probability distributions $P(action_i)$, $P(M_j \mid action_i)$ and $P(X(t) \mid M_j)$ which their computation detail is given in [5]. Describing the complete details of the Bayesian RL framework contains all the following aspects such as modeling approach, learning algorithm, method of estimating pdfs, policy of computing updating weights, deciding whether to update an existing component or just adding a new one, computing the TD-error thus updating values of states. It is out of the scope of this paper. Interested readers may refer to [5] to find the details of this framework.

ACKNOWLEDGMENT

This research was supported by University of Tehran and has been realized in close collaboration with the BACS project supported by EC-contract number FP6-IST-02´140, Action line: Cognitive Systems. The authors would like to highly appreciate the kind efforts of Mr. Mohammad Afshar (m.afshar@ece.ut.ac.ir) one of M.Sc. students of Robolab to develop the free powerful interface for Webots™ and MATLAB® which greatly facilitates the simulation.

REFERENCES

[1] M. Dorigo and M. Colombetti, Robot Shaping, 1997, The MIT Press.

[2] Attention: from theory to practice, editted by Arthur F. Kramer, Douglas Wiegmann, Alex Kirlik, Oxford university press, 2007

[3] Trick, L. & Enns, J.T. (2004). Driving and selective attention: a conceptual framework for understanding the role of selective attention in driving. In Gale A.G., (Ed.), Brown I. D., Haslegrave C. M. & Taylor S. P. (co-eds.): Vision in Vehicles X. (Amsterdam) Elsevier Science B.V., North-Holland

[4] H. F. Shariatpanahi, M. N. Ahmadabadi, Biologically Inspired Framework for Learning and Abstract Representation of Attention Control, Proceedings of International Workshop on Attention in Cognitive Systems, at IJCAI 2007, Hyderabad, India, pp. 63-80, 2007

[5] H. Firouzi, M. Nili, B. Araabi, "A Probabilistic Reinforcement-Based Approach to Conceptualization", International Journal of Intelligent Technology (IJIT), Volume 3, pp 48-55, 2008 ; available to be downloaded at: http://www.waset.org/ijist/v3/v3-1-10.pdf

[6] S. Amizadeh, M. N. Ahmadabadi, B. N. Araabi, R. Siegwart, A Bayesian Approach to Conceptualization Using Reinforcement Learning, IEEE/ASME International Conference on Advanced Intelligent Mechatronics, Zürich, Switzerland, Sep. 4-7, 2007.

[7] S.Whiteson, M. E. Taylor, and P. Stone, Adaptive Tile Coding for Value Function Approximation, AI Technical Report AI-TR-07-339, University of Texas at Austin, 2007.

[8] K. Doya, "Reinforcement learning in continuous time and space," *Neural Computation*, vol. 12, pp. 219–245, 2000.

[9] H. R. Berenji, "Fuzzy Q-Learning for generalization of reinforcement learning," In Proceedings of FUZZIEEE'96, New Orleans, US, September 1996.

[10] Maryam S. Mirian, Majid Nili Ahmadabadi, Babak N. Araabi, Ronald R. Siegwart, Comparing Learning Attention Control In Perceptual and Decision Space, LNAI 5395, pp. 242 – 256, 2009, Springer-Verlag Berlin Heidelberg 2009.

[11] O .Cohen and Y.Edan, A Sensor Fusion Framework for On-Line Sensor and Algorithm Selection, Proceedings of the 2005 IEEE International Conference on Robotics and Automation, pp 3166-3172, Barcelona, Spain, April 2005

[12] N. Karam, F. Chausse,R.Aufrere,R.Chapuis, Localization of a Group of Communicating Vehicles by State Exchange, Proceedings of 2006 IEEE/RSJ IROS'06, pp 519-524

[13] Professional Mobile robot Simulator, http://www.cyberbotics.com
E-puck, EPFL Education Robot, http://www.e-puck.org.