Goal Seeking of Mobile Robot Using Fuzzy Actor Critic Learning Algorithm

F. Lachekhab, M. Tadjine

***Abstract*— In this paper, we present a study of a basic behavior of mobile robot, which is goal seeking. Firstly, we use the heuristics approaches to develop a controller based on fuzzy logic to control the robot and bring it from an initial position to a final position. Secondly, we use the learning technique -Fuzzy actor critic learning algorithm (FACL)-. The reinforcement learning FACL has to select the actions available in each fuzzy rule. The main advantage of the proposed method is the automatic construction of these rules. So far, simulation has shown that the controller is able to perform successful navigation task in known environments, and it has smooth action and exceptionally good robustness.**

I. Introduction

Few decades ago, a special effort had been made in the fields of research and industry to build autonomous mobile robots with minimal human intervention [8].

However, when the environment becomes more complex, it is essential that the robot would be equipped with ability of decision-making to react to hazards that can thwart his movements.

The navigation resolution problem of a mobile robot has been largely a subject of research and results of artificial intelligence. We have chosen to use a reasoning based on the use of fuzzy logic, where imprecise information can be processed in similar way as it has been done by human. It allows avoiding the complex phase of mathematical modeling of systems to command. It provides more decision-making and information processing in much reduced computation times. Due to its performance, we chose to use this tool in the proposed approach module.

The work presented in this article focuses on the development of a control system for a mobile robot based on fuzzy logic. The overall goal is to make the robot learn a simple behavior as going towards a goal.

For this, in the first phase, the navigation module was made with a standard fuzzy inference system where parametric and structural characteristics have been determined by a human expert (heuristically). In the second phase, we use a reinforcement learning method which is based on the active exploration of the environment in order to discover the states causing the emission of rewards and punishments. In this context, we use the Fuzzy Actor-Critic Learning (FACL) algorithm based on the prediction method of temporary differences (TD). It allows the selection of the best action from a set of available actions in each fuzzy rule.

This paper is organized as follows: In Section II, the robot model is given. In Section III the heuristic method and the results are presented. In Section IV the reinforcement learning method is introduced then the function of this module (goal seeking) is verified by simulation. In Section V real experiments on the Pioneer 2P robot are given. Finally the paper is concluded in section VI.

II. Experimental platform

Pioneer 2P-DX is a small mobile robot shown in “Fig. 1”. It contains all the basic components for navigation in a real environment, mostly intended for use in indoor environments on hard and flat surfaces.

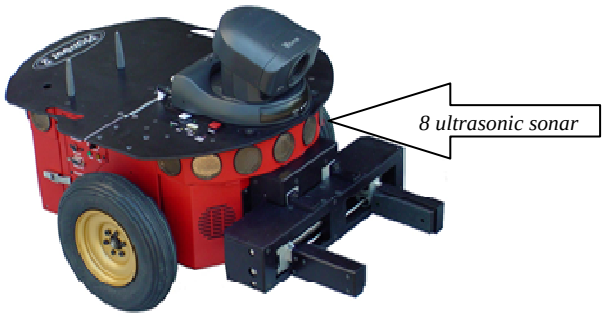


Figure 1. Image of the robot Pioneer 2P-DX.

Pioneer 2 has 8 ultrasonic sonar transducers to provide objects detection, each sensor Si (for i=1…...8) gives a distance di from the robot to the obstacle in its field of view of 20°. For obstacle detection and avoidance purposes, the sensors are divided into 4 sensor groups G_i (for i=1…3). Saphira software is a robot control system developed at the International Artificial Intelligence Laboratory SRI. It is the interface for robots Pioneer Amigobot, PeopleBot and some other kinds of robots [4].

From the version 8.x, several functions of Saphira have been moved completely to ARIA "ActivMedia Robot Interface for Applications" Saphira and Aria software is written in C ++.In our study, due to the simplicity offered by the toolbox Simulink of MATLAB and his ability of integration of the developed program in C ++, we use a block diagram to achieve our simulations.

F. Lachekhab is with the University of Mhamed Bougara, Boumerdes BP3500. Algeria. She is with Laboratory of Automatic and Applied Signal Processing. (e-mail: lachekhab_f@yahoo.fr).

M.Tadjine is with Polytechnic National School, Algiers, Algeria. He is with Process Control Laboratory (e-mail: tadjine@yahoo.fr).

III. HEURISTIC METHOD

In this section, we design the controller heuristically. Furthermore the membership functions of inputs and outputs and the fuzzy rules are determined by human expertise.

A. *Principle of control*

This behavior allows the action "convergence toward the goal" by the robot. For that the control uses the knowledge of current robot position and the definition of a position relative to achieve (the goal). The purpose is to control the movements of the robot so it can reach the destination. It is clearly understood that this behavior can only operate in uncongested environments where no obstacles impede the progress of the robot [7].

The definition of the target position is done through two input variables "θ_{Rb}" and "ρ_b", the polar coordinates of this point expressed in the coordinate system of the robot in "Fig. 2". the variable "θ_{Rb}" represents the angle between the robot velocity vector and the vector Goal-Robot, the variable "ρ_b" is the distance Robot-Goal. The two output variables are the rotation speed robot Vrot_AB and Vit_AB linear speed translation.

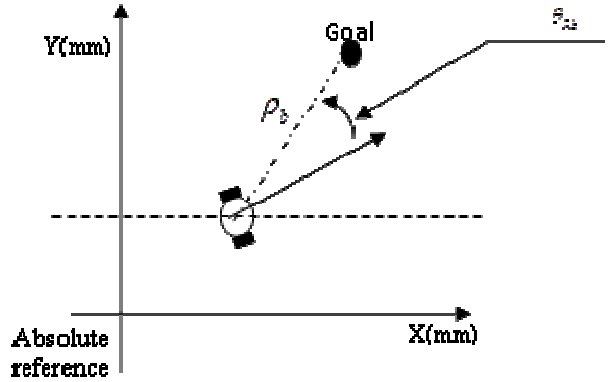


Figure 2. Representation of controller input variables.

B. *Definition of membership functions*

The universe of discourses of the input variable θ_{Rb} and ρ_b of the first fuzzy controller examined are respectively decomposed into three and two fuzzy sets. This strong fuzzy partition is simple and provides a concise rule-based and easy to interpret (06 fuzzy rules).

The robot-goal angle variable θ_{Rb} is partitioned into three fuzzy subsets "Fig. 3": N (Negative), ZE (Zero), P (Positive) while the variable ρ_b is partitioned into two fuzzy subsets N (near) and F (far). The output variables are Vit_AB as translational speed and Vrot_AB as rotation speed.

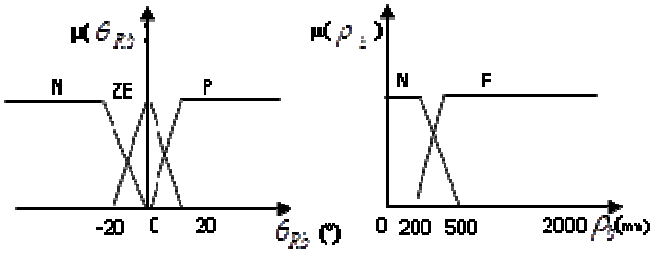


Figure 3. The controller input variables « convergence toward the goal».

The developed fuzzy controller is a zero order Sugeno type, each output variable is partitioned into three fuzzy sets. The numerical values of conclusions for the translational speed expressed in (mm/s): ME=150; Z=0; FA=300; and for the speed of rotation expressed in (°/s): TR=-15, ZE=0, TL=15.

The inference table 1, represents the fuzzy rule base for the speed of translation, and the inference table (2) is the basis of the rules of the rotational speed. Labels are: ME (medium), Z (zero), RA (fast) for the translational speed, and TR (turn right), ZE (zero), TL (turn left) for the rotational speed.

TABLE I. FUZZY RULE BASE FOR THE TRANSLATIONAL SPEED

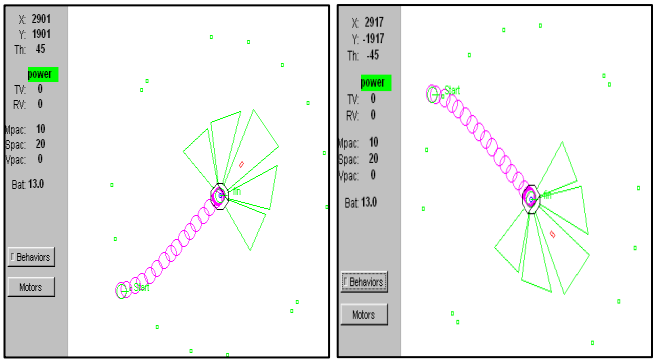
θ_{Rb} / ρ_b	**N**	**ZE**	**P**
N	ME	Z	ME
F	FA	FA	FA

TABLE II. FUZZY RULE BASE FOR THE SPEED OF ROTATION

θ_{Rb} / ρ_b	**N**	**ZE**	**P**
N	TR	ZE	TL
F	TR	ZE	TL

C. Results of the experiment

"Fig. 4", is a screen copy showing the path taked by the robot. The fuzzy controller with six fuzzy rules has given satisfactory results.



A B

Figure 4. Robot paths for the targets points A (2920, 1920) and B (2920 , -1920)

In the following we present a simulation of several sub goals with this same controller.

In this simulation we define sub goals. The robot starts from the initial position and joins sub goals. We note that the robot did not take the shortest way to go to the second sub-goal as illustrated in "Fig. 5".

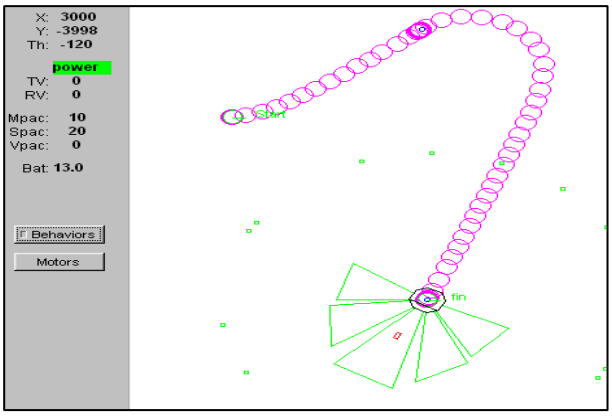


Figure 5. Robot path for the target point (2920, 1920) and (3000,-4000)

IV. REINFORCEMENT LEARNING METHOD

In this section, we use the learning method to develop a controller of behaviour convergence toward the goal. This learning method is based on the reinforcement using fuzzy inference system called Fuzzy-actor -critic-learning [5].

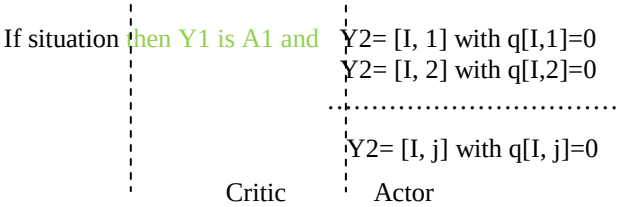
A. Fuzzy actor critic -learning

In the reinforcement learning paradigm an agent receives a scalar reward value called reinforcement from its environment. This can be Boolean (true, false) or fuzzy (bad, fair, very good ……..) etc…

A sequence of control actions is often executed before receiving any information on the whole sequence.

Therefore, it is difficult to evaluate the contribution of one individual action. This Credit Assignment problem has been widely studied since the pioneering work of Barto, Sutton[1], and Anderson the whole methodology is called Temporal difference Method [2] and contains a family of algorithms. Recently, Watkins proposed a new algorithm of this family; Actor critic learning, this algorithm, is a form of competitive learning which provides agents with the capability of learning to act optimally by evaluating the consequences of actions. Actor critic-learning keeps a function, which attempts to estimate the discounted future reinforcement for taking actions from given states. This function is a mapping from state-action pairs to predicted reinforcement. In the fuzzy actor critic –learning, the selected agent is a SIF of Takagi-Sugeno type of zero order because of its simplicity, its characteristic of universal approximate, its capacity of generalization and its applications in real time [6]. Its input variables membership functions are a triangular and trapezoidal, and it has two outputs.

The SIF is consists of N rules of the form:



In the algorithm FACL each rule has:

- A set of discrete actions U_i identical for all the rules.
- A vector of parameters q indicating the quality of the various discrete actions available and used in the definition of the current policy.
- The premise part is imposed by the operator (the input variables membership functions are fixed characteristics).
- The conclusions are chosen by reinforcement learning algorithm among a whole of discrete actions available for each fuzzy rule.

B. Description of the execution procedure

In this learning method the idea consists on using the matrix q to implement not only the local policies with the rules, but also to represent the evaluation function of the polic-the global t-optimal-. One obtains then the FACL, adaptation of AC-Learning for an apprentice of the type SIF.

The execution on a step of time can be divided into six principal stages [3]. The step of current time is t+1; the apprentice executed the elected action to the step of time previous and it received the primary education reinforcement for the transition from the state S_t to S_{t+1}. After the determination of the values of truth of the rules α R_i, the six stages are as follows:

1- The first stage of the algorithm consists to determine the function of t-optimal evaluation of the current state using the matrix q:

$$V_t(S_{t+1}) = v_t.\phi_{t+1}^T \quad (1)$$

And update of the traces of eligibility:

$$\bar{\phi}_t \leftarrow \bar{\phi}_t - \gamma\rho\phi_{t+1} \quad (2)$$

2- Determination of the error TD:

$$\tilde{\varepsilon}_{t+1} = r_{t+1} + \gamma.V_t(S_{t+1}) - V_t(S_t) \quad (3)$$

3- Update the rates of training corresponding to the parameters of the critic for all the rules in three stages:

$$\begin{aligned} &- \delta_t^i = \tilde{\varepsilon}_{t+1}.\bar{\phi}_t^i \\ &- \beta_{t+1}^i = \begin{cases} \beta_t^i + k & if \quad \bar{\delta}_{t-1}^i.\delta_t^i > 0, \\ \beta_t^i(1-\psi) & if \quad \bar{\delta}_{t-1}^i.\delta_t^i < 0, \\ \beta_t^i & else. \end{cases} \\ &- \bar{\delta}_t^i = (1-\psi)\delta_t^i + \psi\bar{\delta}_{t-1}^i \end{aligned} \quad (4)$$

Where in our case

- β_t^i is the rate of training of the critic for the rule with the step of time
- $\bar{\delta}_t^i = (1-\psi)\delta_t^i + \psi\bar{\delta}_{t-1}^i$ represent the exponential average.

4- The update of the matrix q and the vector v :

$$v_{t+1} = v_t + \tilde{\varepsilon}_{t+1}.\beta_{t+1}.\bar{\phi}_t^T \quad (5)$$

$$q_{t+1} = q_t + \tilde{\varepsilon}_{t+1}e_t \quad (6)$$

5- Recalculation of the evaluation functions of the current state by the critic, but this time with the newly updated parameters:

$$V_{t+1}(S_{t+1}) = v_{t+1}.\phi_{t+1}^T \quad (7)$$

This value will be used to calculate the error TD in the next step of time.

6- The action should now be chosen to be applied to the state St+1 .In the case of continuous actions; the crisp action is determinated from the various actions elected for each rule:

$$U_{t+1}(S_{t+1}) = \sum_{R_i \in A_{t+1}} Election_U(q_{t+1}^i).\alpha_{R_i}(S_{t+1}), \forall U \in U, \quad (8)$$

Where Election is defined by:

$$Election_U(q_{t+1}^i) = ArgMax_{U \in U}(q_{t+1}^i(U) + \eta^i(U) + \rho^i(U)), \quad (9)$$

Then we update traces of eligibility

$$\bar{\phi}_{t+1} = \phi_t + \gamma\lambda.\bar{\phi}_t,$$

$$e_{t+1}^i(U^i) = \begin{cases} \gamma\lambda'.e_t^i(U^i) + \phi_{t+1}^i, (U^i = U_{t+1}^i), \\ \gamma\lambda'.e_t^i(U^i), \text{otherwise} \end{cases} \quad (10)$$

C. Details of the experiment

The lack of the Need for a preliminary model of the process and integration expertise which characterize the heuristic fuzzy control, responds perfectly to our goals [9]. However, this first approach has also demonstrated the limits of human expertise, in particular the difficulty of assessing the interactions. Indeed, in the design of FIS a choice is carried out relatively crude and the development of a fuzzy controller can be long and difficult especially if the number of fuzzy rules is important [10]. Therefore, the fuzzy inference systems designed from only the human expertise often have very average performance.

Moreover, learning methods have theoretically and practically proven their ability to address these problems. For this we use a reinforcement learning algorithm in order to determine the control of translational speed and rotation robot speed (the conclusions of the FIS).

1. The apprentice

The Apprentice is a Sugeno fuzzy controller zero order in which the conclusions of the translational speed, the rotation speed and the conclusions of critical are initially set to zero. The inputs of the fuzzy controller "ρ_b" and "θ_{Rb}" are defined respectively by two and three membership functions "Fig. 3".

For the set of actions we consider a set of actions U common to all fuzzy rules, formed of seven values for the speed of rotation (a1= -15(°/s), a2 = -10(°/s), a3 = -5(°/s), a4 = 0(°/s), a5 = 5 (°/s), a6= 10(°/s), a7 = 15(°/s),) and three values for the translational speed (a'1 = 0(mm/s), a'2 = 100(mm/s), a'2 = 300 (mm/s)) . The fuzzy controller can generate continuous action between -15 and 15 °/s for the speed of rotation, and between 0 and 300 mm/s for the translational speed, by interpolation of these discrete actions.

The basic rule is as follows:

If (θ_{Rb} is N) and (ρ_b is P) then (Vrot is A1) (Vit is B1) (C_ Vrot is v1)

If (θ_{Rb} is N) and (ρ_b is L) then (Vrot is A2) (Vit is B2) (C_ Vrot is v2)

If (θ_{Rb} is ZE) and (ρ_b is P) then (Vrot is A3)) (Vit is B3) (C_ Vrot is v3)

If (θ_{Rb} is ZE) and (ρ_b is L) then (Vrot is A4) (Vit is B4) (C_ Vrot is v4)

If (θ_{Rb} is P) and (ρ_b is P) then (Vrot is A5) (Vit is B5)(C_ Vrot is v5)

If (θ_{Rb} is P) and (ρ_b is L) then (Vrot is A6) (Vit is B6) (C_ Vrot is v6).

2. Reinforcement function

The actor in our case consists of the pair of actions: rotation speed and translation speed. The reinforcement function is defined to allow the rotation of the robot in the direction of the object point and direct convergence to the goal.

- If the robot is far to the goal

  - 1 for ($\theta_{Rb} . \dot{\theta}_{Rb} < 0$)
  - 1 for (-1°< θ_{Rb} <+1°)
  - 0 for ($\theta_{Rb} . \dot{\theta}_{Rb} = 0$)
  - -1 else

- If the robot is near then the goal

  - 10 for ($\theta_{Rb} . \dot{\theta}_{Rb} < 0$) & if Vit=0
  - -1 else

3. Conditions of experimentation

The conditions of the experiment for the case of the algorithm FACL are as follows:

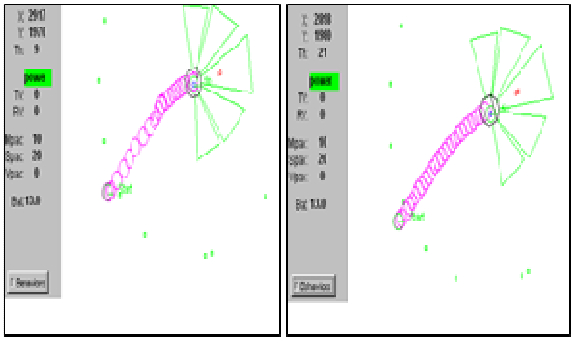
- An adaptive learning rate β.
- A discount rate γ (to evaluate the influence of the long-term prediction of the evaluation function)
- An exploration and exploitation SP, θ
- Traces of eligibility: proximity the critical factor λ .

Table 3 indicates the values of parameters of the algorithm FACL included in these simulations.

TABLE III. TABLE OF USED PARAMETERS.

γ	λ	λ'	ψ	φ	θ	Sp
0.9	0.7	0.5	0.8	0.6	10	0.1

"Fig. 6", shows trajectories of the robot during and after the learning phase. After an important exploration phase, the robot moves towards the target point.



Learning phase Validation phase

Figure 6. Robot trajectories before and after learning by FACL algorithm « point goal (2920, 1920) ».

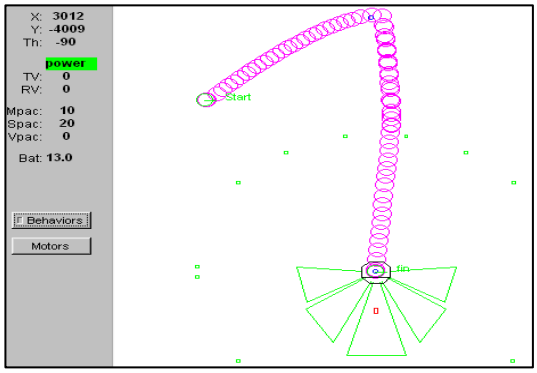


Figure 7. Robot path for the target point (2920, 1920) and (3000,-4000)

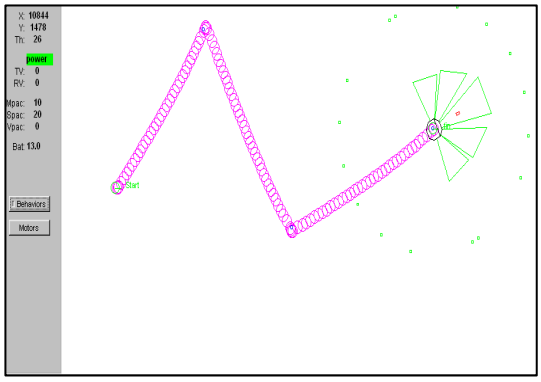


Figure 8. Robot path for several targets points.

We note that after learning the robot comes to properly reach sub-goals. "Fig. 8", shows trajectories of the robot.

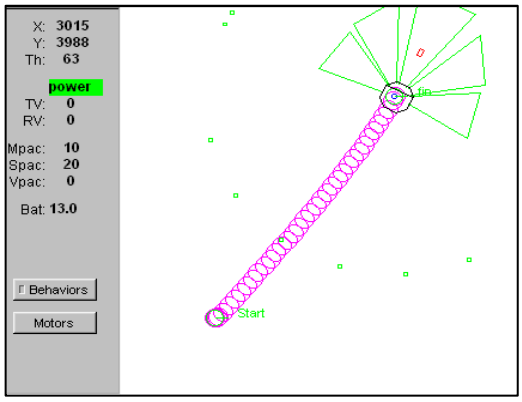


Figure 9. Robot path for the target point.

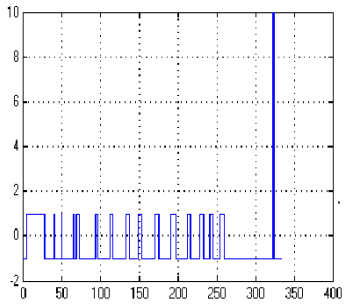


Figure 10. Rreinforcement function.

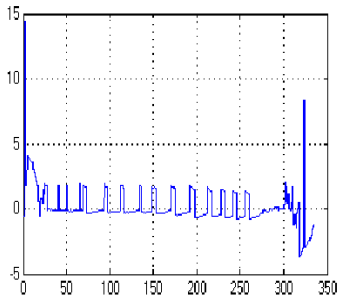
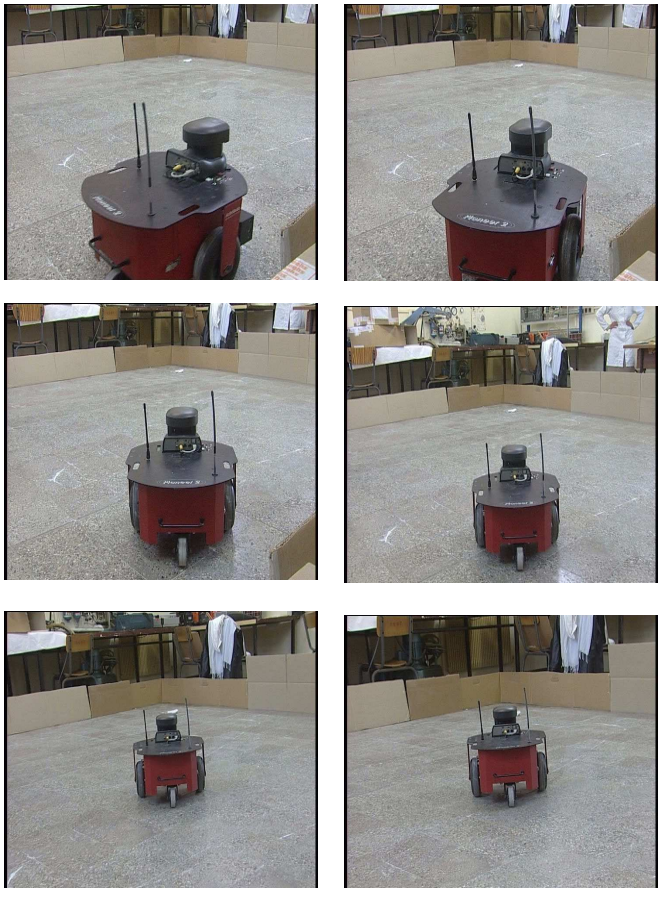


Figure 11. Error TD.

V. Real experiments on the Pioneer II robot

Using quality matrix values obtained by the algorithm FACL after validation, we performed a real test on the mobile robot Pioneer II. “Fig. 12”, represents the movements of the robot in an open environment. Starting from an initial position (0,0), the robot performs a rotation on site in the direction of an assigned goal (X_b, Y_b) = (3000.3000) and moves with a speed of 150 mm / s. Then the robot begins to decrease its speed once it approaches the point goal and finally stops on it.



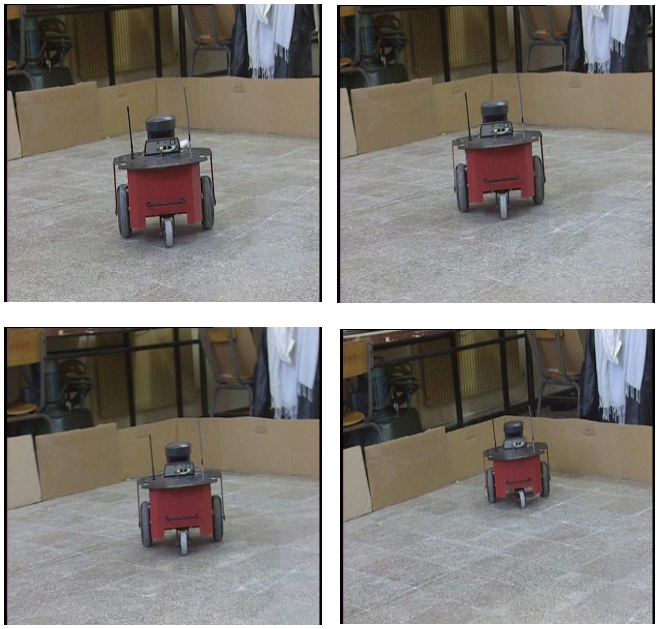


Figure 12. Behavior "convergence toward a goal" of the real robot

VI. Conclusion

In this paper, we have simulated the behavior of "convergence toward the goal" of a mobile robot using heuristic approaches and the reinforcement learning algorithm FACL. The experiments show that with a proper choice of parameters influencing learning (learning rate, discount factor, choice of reinforcement function, policy exploration and exploitation) the controller of six fuzzy rules obtain fast convergence towards the goal.

These results also show the generalization capacity offered by this learning algorithm unlike to the heuristics method of FIS developing.

Finally, the latest results show that FACL algorithm gives satisfactory results, however, non-adapted choice of the parameters of this algorithm results degradation in performance.

References

[1] R. S. Sutton, D. Mcallester, S. Singh and Y. Mansour, “Policy Gradient Methods for Reinforcement Learning with Function Approximation”, In Advances in Neural Information Processing Systems, Volume 12, 2000.

[2] P. Y. Glorennec, “Fuzzy Q-Learning and Dynamical Fuzzy Q-Learning”, *IEEE,* Proc of the Third International Conference on Fuzzy System, pp. 474-479, 1994

[3] L. Jouffe and P.Y. Glorennec, “Comparison Between Connectionist and Fuzzy Q-Learning”, Proc of Iizuka’96, Fourth International Conference on Soft computing, Lizuka, Fukuoka, Japan, September, pp. 557-560, 1996.

[4] K. G. Konolige, *Saphira Robot Control Architecture*, Edition SRI International, Avril, 2002

[5] C. Ye, N. H. C. Yung, D. Wang, “A Fuzzy Controller with Supervised Learning Assisted Reinforcement Learning Algorithm for Obstacle Avoidance”, IEEE, Transactions on Systems Man and Cybernetics -Part B Cybernetics, vol.33, NO.1, pp.17-27, February 2003.

[6] M,Zucker. and J. A,Bagnell. “Reinforcement planning: RL for optimal planners. In IEEE International Conference on Robotics and Automation (ICRA). 2012.

[7] R. S, Sutton and A.Koop, and D,Silver., “ the role of tracking in stationary environments”, In International Conference on Machine Learning (ICML).2007.

[8] C,Lakhmissi and M,Boumehraz : “Designing of Goal Seeking and Obstacle Avoidance Behaviors for a Mobile Robot Using fuzzy Techniques”,J. Automation & Systems Engineering 6-4.2012: 164-171.

[9] L. M. Zamstein, A. A. Arroyo, E. M. Schwartz, S. Keen, B. C. Sutton, and G. Gandhi , "Koolio : “ Path Planning using Reinforcement Learning on a Real Robot Platform FCRAR”, 19th Florida Conference on Recent Advances in Robotics, Miami, Florida, May 25-26, 2006.

[10] C,Lakhmissi and M,Boumehraz: “Using Q-Learning and Fuzzy Q-Learning Algorithms for Mobile Robot Navigation in Unknown Environment”, International Journal of Science and Engineering Investigations. vol. 1, issue 2, March 2012.