# Co-Operative Strategy for an Interactive Robot Soccer System by Reinforcement Learning Method

**Hyoung-Rock Kim, Jung-Hoon Hwang, and Dong-Soo Kwon***

**Abstract:** This paper presents a cooperation strategy between a human operator and autonomous robots for an interactive robot soccer game. The interactive robot soccer game has been developed to allow humans to join into the game dynamically and reinforce entertainment characteristics. In order to make these games more interesting, a cooperation strategy between humans and autonomous robots on a team is very important. Strategies can be pre-programmed or learned by robots themselves with learning or evolving algorithms. Since the robot soccer system is hard to model and its environment changes dynamically, it is very difficult to pre-program cooperation strategies between robot agents. Q-learning - one of the most representative reinforcement learning methods - is shown to be effective for solving problems dynamically without explicit knowledge of the system. Therefore, in our research, a Q-learning based learning method has been utilized. Prior to utilizing Q-learning, state variables describing the game situation and actions' sets of robots have been defined. After the learning process, the human operator could play the game more easily. To evaluate the usefulness of the proposed strategy, some simulations and games have been carried out.

**Keywords:** Robot soccer, reinforcement learning, human-robot cooperation, entertainment robots.

## 1. INTRODUCTION

Robot soccer games like MIROSOT [2] and Robocup [3] have fascinated many people and increased public interest in the field of robotics. But, for people to participate in these games, they need to have technical knowledge about robot hardware, software, vision systems, etc. For this reason, people lacking expert knowledge can only remain as spectators. If more people can participate in the robot soccer games easily, the games will be more challenging. In addition, controlling the robot directly would make the game more dynamic and amusing.

An Interactive Robot Soccer System has been developed to supplement existing robot soccer games with the new functions mentioned above (Fig. 1). A joystick interface has been adopted for easy control of the robots. Infrared communication and a robot position/orientation detecting system using electromagnetic field effects have been developed so the system can be installed and used in arcades, where the environment cannot be regulated. The Interactive Robot Soccer System is described in detail in section 2.

In the proposed system, a human operator can control a robot at any moment. The other autonomous robots should move according to their programmed strategies. Because robot soccer games are dynamic and complicated, it is very difficult to program strategies manually and much experience in robot soccer games is required.

Accordingly, many researchers have proposed strategy-learning algorithms with a reinforcement

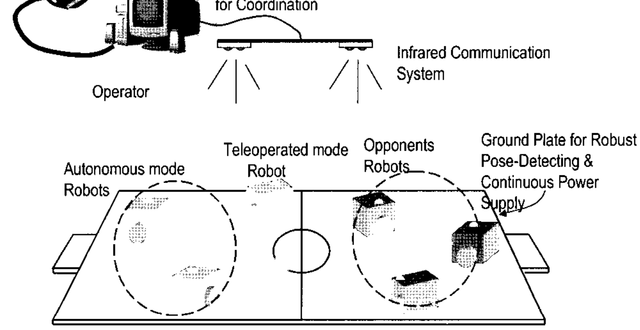


Fig. 1. Schematic of an interactive robot soccer system.

Manuscript received Febuary 28, 2002; accepted June 24, 2002. This work is supported in part by the University Research Program supported by the Ministry of Information & Communication in the Republic of Korea and Roccer Co.

Hyoung-Rock Kim and Jung-Hoon Hwang are with the Department of Mechanical Engineering, KAIST, 373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701, Korea. (e-mail: {kimhr, hwangjh} @robot.kaist.ac.kr).

Dong-Soo Kwon is with the Department of Mechanical Engineering, KAIST, 373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701, Korea. (e-mail: kwonds@mail.kaist.ac.kr).

*Corresponding author.

learning method [5-7]. Because robot soccer games cannot be modeled exactly, reinforcement learning, an unsupervised learning method, is used to develop robot soccer strategies. In reinforcement learning, when the state of the environment changes according to an agent's action, reward or punishment for that action is given and that agent learns through the results of his action progressively (Fig. 2).

H. T. Chu and B. R. Hong categorized robot actions into eight groups [5]. A robot selects his actions according to the proposed ASPL (Action Selection Priority Level). In their research, robots learn action priorities by reinforcement learning. State information about the current game situation is represented by a summation of two-dimensional vectors. This method simplifies the environment information and allows for fast learning.

Y. Takahashi and M. Asada have proposed an algorithm with layered Q-learning [6]. It divides the learning process into several steps and makes learning more efficient. Contrary to Takahashi and Asada's research, K. Kostiadis and H. Hu divided the robot's roles and applied slightly different learning modules to each robot [7]. This method converges learned strategies into a sub-optimal solution quickly.

In previous studies about strategy learning of autonomous robots, researchers have focused on accelerating learning speed. However, a human operator is included in the proposed Interactive Robot Soccer System and autonomous robots should be able to play the game around the human-operated robot. Therefore, our research objective is to make autonomous robots cooperate with a human operator rather than propose a more efficient learning algorithm.

Simple Q-learning has been used in our research and state variables and a robot action set have been selected in consideration of their relation with the human operated robot. Detailed description about the proposed strategy structure is provided in section 3. Simulation and game results with human operators are presented in Section 4.

## 2. INTERACTIVE ROBOT SOCCER SYSTEM

The Interactive Robot Soccer system is a robot soccer game, in which some human operators can participate in real-time and it is designed for the entertainment field [1].

For entertainment, a robot soccer system requires several features: It must be easy to operate for a wide range of users, interactive for more fun, evolvable to maintain user interest, and manageable for business markets.

A user interface with a joystick and buttons is proposed to make the robot soccer system interactive and easily operable. The interface allows a person to take part in the control of one robot while the

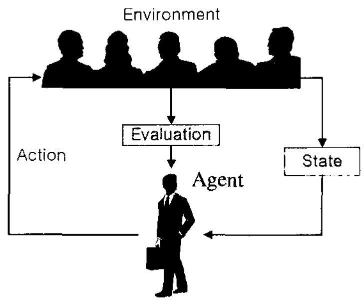


Fig. 2. Reinforcement learning.

other robots run autonomously via a programmed strategy. and easily operable. The interface allows a person to take part in the control of one robot while the other robots run autonomously via a programmed strategy. The operator can select the teleoperated robot arbitrarily and change the programmed strategies on-line with the user buttons

A teleoperation method with respect to (w.r.t.) ground-based coordinates is developed for easy maneuverability of the soccer robot. To achieve teleoperation w.r.t. ground-based coordinates, a robust pose-detecting system is developed.

A camera vision system is one of the more popular devices for position/orientation detection because many off-the-shelf products exist. However, camera vision systems have the following limitations:

- Limited sampling rate
- Intensive computation
- Sensitiveness to environmental illumination

To overcome these limitations, a new pose-detecting system is proposed. This system applies the magnetic field effect. The robots have two magnetic field generators (MFGs), one on the right side and one on the left. The host computer organizes the on-off order and activates only one MFG through wireless communication at a given instant. The sensing wires are buried under the playing ground. Those wires are connected to the main computer through an analog-to-digital conversion module. The sensing wires show a certain voltage level according to the location of the MFG and the robot's position and orientation can be recognized from the positions of the two MFGs.

To complete the robot soccer system so that it can be practically used for entertainment, additional technologies have been developed which include the following. An infrared communication system and a continuous power supply system are implemented for wireless teleoperation and continuous operation without recharging. The schematic of the system is shown in Fig. 1, and the system developed in this study is outlined in Fig. 3.

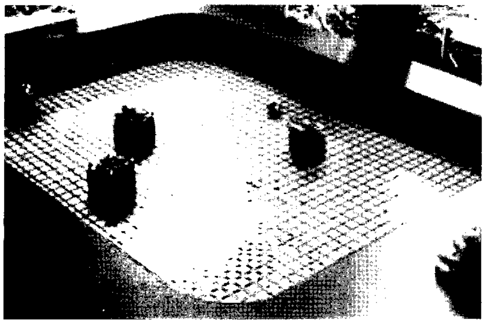

Fig. 3. The interactive robot soccer system.

## 3. ARCHITECTURE OF LEARNING SYSTEM

To facilitate cooperation between the human operator and autonomous robots in the game, the robots should know the operator's intentions exactly [4]. The intentions of the operator constitute the operator's strategic plans. However, these intentions cannot be transmitted to other robots in a team directly during the game.

After a series of test games with many people, human operators were found to aggressively dash to the opponent's goal, and generally, they wanted to handle the ball as much as possible. These kinds of human operator tendencies reduced the autonomous robots' chances of holding on to the ball. Therefore, the object of a cooperation strategy in the Interactive Robot Soccer Game is to aid the human-operated robot in attacking and defending smoothly. In addition, autonomous robots should be able to quickly adapt to changes in the operator's strategy. Otherwise, a human operator would have to participate in thousands of games until the robots adapted to his strategies.

### 3.1. Strategy structure

The Interactive Robot Soccer system has been designed for 3 robots versus 3 robots per game. Our research is only concerned with situations in which human operates one robot and the other two robots move autonomously. The two autonomous robots' roles are divided into attacker and defender. The robot more closely located to the opponent's goal is assigned the role of attacker, and the other is assigned the role of defender. This kind of division of roles reduces the search area of the learning system and the time required to search local minima [7].

Strategy refers to the policy that attains the most profitable action in a current situation. According to the roles of the robots, they have a decision-making rule structure as presented in Fig. 4 and 5.

When the ball is in front of the attacker and the path to the opponent's goal is clear, the attacker tries to shoot. On the other hand, when the ball is in front of the attacker and the path to the opponent's goal is not clear, the attacker tries to pass the ball to the human-

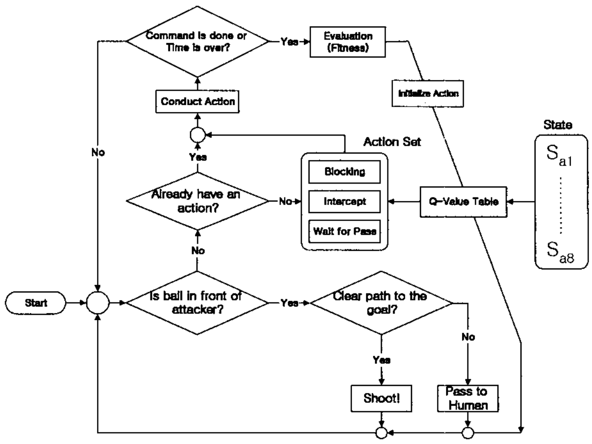


Fig. 4. Strategy structure of attacker.

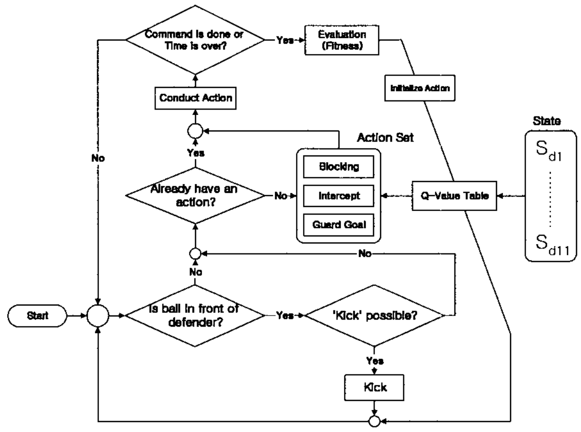


Fig. 5. Strategy structure of defender.

operated robot. When the ball is far from the attacker, it selects actions among pre-defined actions according to the current situation. After a selection is made, the attacker robot keeps trying the selected action until the completion condition of each action is satisfied. If the action completion condition is satisfied or the time limit of each action is over, the attacker robot evaluates its action. Even if the above conditions are not met, when the ball is in front of an attacker, the attacker robot evaluates its action and decides whether to shoot or pass.

The decision-making structure of the defender robot is similar to that of the attacker robot. In our research, the robots learn and memorize by a Q-learning method which action would be profitable in assisting the human-operated robot in the current situation when shooting, passing or kicking is impossible.

In Q-learning, each pair of current state and action has a Q-value. If the Q-values are high, the probability of selecting that action in the given situation will be high. Q-values are updated with Equation (1) after completion of each action.

$$Q_{n+1}(i,a) = (1-\eta_n(i,a))Q_n(i,a) + \eta_n(i,a) \\ [g(i,a,j) + \gamma J_n(j)]. \tag{1}$$

Table 1. Possible actions of each robot.

| | Attacker Robot | Defender Robot |
|---|---|---|
| Possible Actions | Blocking | Blocking |
| | Wait for Pass | Intercept |
| | Intercept | Guard the Goal |

Table 2. State variables of each robot.

| | Variables | Discretization |
|---|---|---|
| Attacker | Distance between(btw.) Attacker and Ball | F, M, C |
| | Angle btw. Attacker and Ball | Fr, Le, Ri, Be |
| | Distance to opponent's goal | F, M, C |
| | Distance btw. Attacker and $R_t$ | F, M, C |
| | Angle bwt. Attacker and $R_t$ | Fr, Le, Ri, Be |
| | Clearance to the opponent's goal | Yes, No |
| | Distance btw. $R_t$ and $O_t$ | F, M, C |
| | Angle bts. $R_t$ and $O_t$ | Fr, Le, Ri, Be |
| Defender | Distance btw. Defender and Ball | F, M, C |
| | Angle btw. Defender and Ball | Fr, Le, Ri, Be |
| | Distance to the Own Goal | F, M, C |
| | Defender's Lateral Position | LW, Mid, RW |
| | Distance btw. the Own Goal and $O_b$ | F, M, C |
| | Lateral Position of $O_b$ | LW, Mid, RW |
| | Distance btw. the Defender and $O_b$ | F, M, C |
| | Angle btw. the Defender and $O_b$ | Fr, Le, Ri, Be |
| | Distance btw. the Own Goal and Rt | F, M, C |
| | Lateral Position of Rt | LW, Mid, RW |
| | Moving Direction of the Ball | Forward, Backward |

where,

$Q_n(i,a)$ : Current Q-value

$Q_{n+1}(i,a)$ : Updated Q-value

$\eta(i,a)$ : Learning Rate

$g(i,a,j)$ : Reward Value

$\gamma$ : Discounting Factor

$J_n(j) = \min_{b \in A_j} Q_n(j,b)$ : Expected Reward

As stated above, for the possible actions of autonomous robots, necessary information about the current situation and fitness evaluation should be defined first.

3.2. State variables and action set

The necessary actions of autonomous robots are reduced in comparison with a conventional autonomous robot soccer game. Autonomous robots' actions are listed in Table 1.

In Table 1, 'Blocking' means that a robot blocks the way of an opponent robot. 'Intercept' means that a robot chases and gets the ball. 'Wait for Pass' indicates when a robot moves somewhere and receives a ball. Finally, 'Guard the Goal' is when a robot guards its own team's goal against an opponent robot's shot.

After defining the possible actions of the robots, the environmental state variables should also be defined. In a robot soccer game, environment information includes the position and orientation of the robots, the velocity of the robots, the position and velocity of the ball, and the location of each team's goal. Moreover, these variables should be expressed as discrete variables.

Rt : Human operated robot
Ob : Closest Opponent to the Ball
Ot : Closest Opponent to Rt
F : Far M : Middle C : Close
LW : Left Wing
Mid : Midfield RW : Right Wing
Fr : Front Le : Left Ri : Right Be : Behind

If all the variables describing environmental information are to be used and discretized narrowly, huge memory will be necessary in storing Q-values [5]. In addition, this would make learning time much longer. Therefore, it is necessary to reduce the variables used in learning. When a human operator is included in the game, the situation around human-operated robot is sufficient information for an autonomous robot to assist a human-operated robot. This is because human intelligence is used, and the human operator plays the leading role in the game.

All the variables used in learning have been selected empirically and discretized into certain levels. Variables related with distance have been discretized into three levels, Far, Middle, and Close. Similarly, variables related with angle have been discretized into four levels, such as Front, Left, Right, and Behind. In addition, variables related with the robot's lateral position in the playground have also been discretized into three levels, Left Wing, Midfield, and Right Wing. These variables are listed in Table 2.

3.3. Fitness evaluation

When a point is scored, the action conducted by a team's autonomous robots in relation to the current situation is given the maximum fitness value, 1. On the other hand, when the opposing team scores a point, -1 is given to the action. When no point is lost or made, 0 is given.

## 4. SIMULATION RESULTS

4.1. Learning in regulated situations

With the proposed actions and variables, some simple simulations in regulated situations have been conducted. Simulations of two cases are as follows:

**Case 1:** In a 2 robots vs. 2 robots situation, all 4 robots are positioned as in Fig. 6. It is difficult for the human operator to shoot by himself because the opponent robot near the goal tries to block the robot and take the ball. Therefore, it makes scoring a point easier if the attacker receives the ball and shoots.

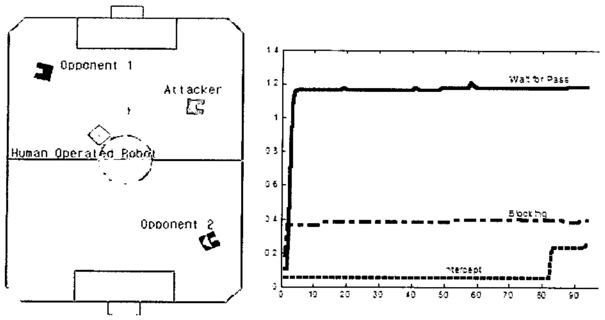


Fig. 6. Simulation condition and result of case 1.

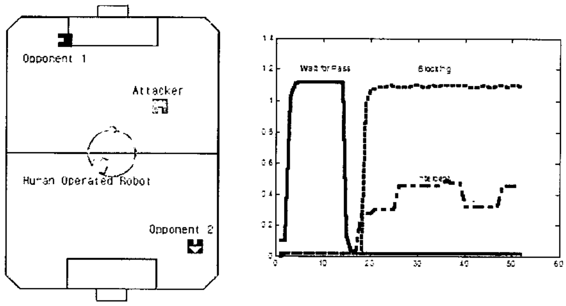


Fig. 7. Simulation condition and result of case 2.

**Case 2:** In a similar situation to Case 1(Fig. 7), a human operator tries to pass the ball to the attacker robot first. After fifteen trials, the human operator tries to shoot by himself or herself. Because the angle between the opponent goal and the human-operated robot is narrower in this case, the human operator can to shoot by himself.

These simulations have been conducted to confirm that autonomous robots can learn to cooperate with the human-operator and cope with the changes in the human operator's intentions. In the series of simulations, it was confirmed that the Q-value of a necessary action in the given situations has increased and the probability of selecting that action also has increased.

Fig. 6 shows the changes of Q-values in Case 1. The attacker learns that waiting for a pass from the human-operated robot is most profitable in such a situation. In Fig. 7, it can be identified that the attacker changes its tropism from a 'wait for pass' action to a 'blocking' action. This shows that the proposed learning mechanism is able to cope with changes in the intentions of human players.

4.2. Learning with simulated human operator

A 3 robots versus 3 robots simulation game also has been conducted to confirm the usefulness of the proposed strategy. In this simulation, the human operator should be modeled. In general, the learning process takes too long , and the human operator cannot participate in the learning process from the beginning to end. The human operator in our simulation has been modeled as follows.

1. He/she selects and operates the robot nearest to the ball.
2. When he/she receives the ball, he/she dribbles the ball toward the opponent's goal
3. When the path to the opponent's goal is not clear, he/she passes the ball to the nearest team mate robot.

The strategy of the opponent team has been pre-programmed. Each game takes two minutes and 200 games were conducted. Fig. 8 and Fig. 9 shows the scores from 200 ally and opponent team games. The score of the ally team increases continuously and that of the opponent team decreases slightly over the course of these games.

Fig. 10 and Fig. 11 show the changes in the period of the ball that remains in each team's region. The period of the ball staying time in the ally region increases in the initial phase of the learning process and maintains an even level. On the other hand, the period of the ball staying time in the opponent region decreases.

According to the changes in scoring in a game and period of ball staying time, it is verified that the performance of the team with a modeled human operator is improved.

4.3. Game with two human operators

After the simulation with a modeled human operator, a simulation game with two human operators has been conducted. In these games, each human operator controls a robot from each team. One team used a pre-programmed strategy and the other team used the proposed strategy structure. At every 10th game, the human operators changed their team with each other to account for any difference in the control ability between two operators so as not to affect the learning process. Each game took 2 minutes, and 60 games were conducted. Fig. 12 and Fig. 13 show the changes in score in a game, and Fig. 14 and Fig. 15 show the period of ball staying time for each team. The goals scored in a game by the team to which the learning strategy is applied increases slightly as in the simulation results in Section 4.2. On the contrary, the goals scored in a game by the team to which the pre-programmed strategy is applied decreases slightly. The period of ball staying time shows the results similar to those of the simulation in Section 4.2.

4.4. Analysis of simulation results

From the series of simulation results, it is verified that the proposed strategy structure can help a human operator to play the game more easily. The human operator should have the willingness to play a game, cooperating with autonomous robots on the same team for the proposed strategy structure. However, the precise way of playing the game varies from operator to operator. Therefore, more games with many human operators and many pre-programmed opponent strategies are necessary.

In addition, the proposed state variables still require too much memory. To reduce the time required for learning, the state variables should be reduced within a limit not affecting the performance of the learned strategy.

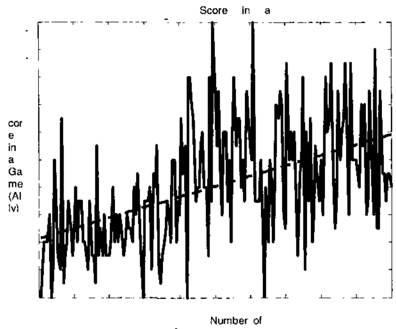


Fig. 8. Change of score of the ally team with simulated human-operator.

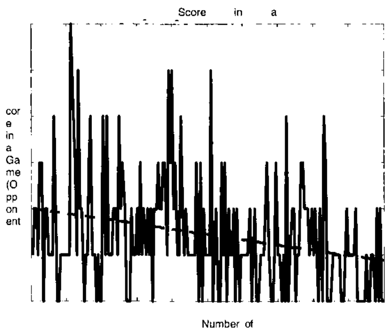


Fig. 9. Change of score of the opponent team.

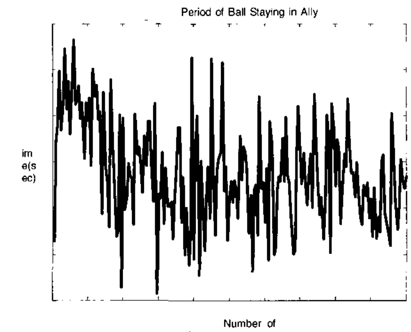


Fig. 10. Change of period of ball staying time in ally region.

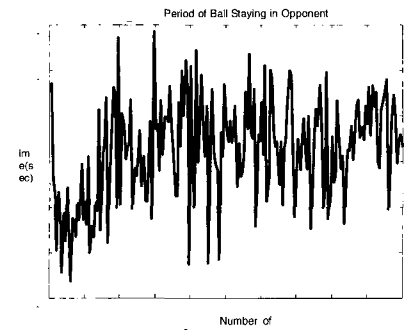


Fig. 11. Change of period of ball staying time in opponent region.

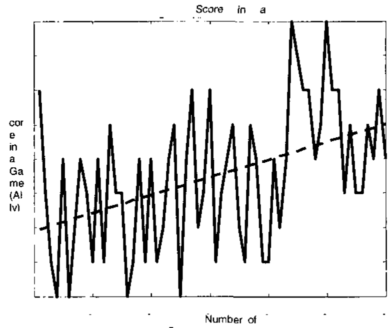


Fig. 12. Change of score in a game of ally team.

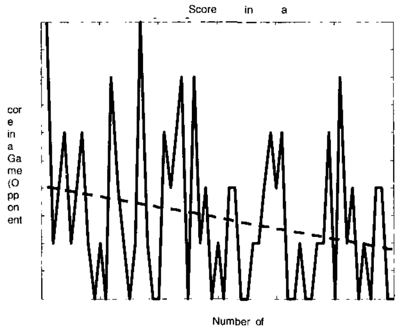


Fig. 13. Change of score in a game of opponent team.

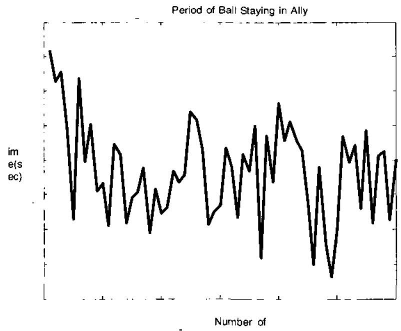


Fig. 14. Change of period of ball staying time in ally region.

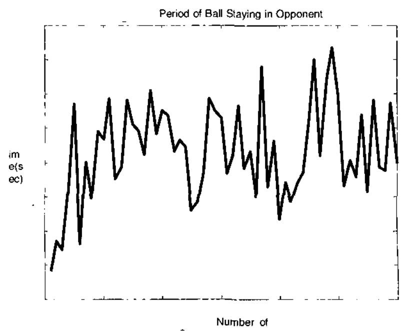


Fig. 15. Change of period of ball staying time in opponent region.

## 5. CONCLUSION

The framework for the strategy of an Interactive Robot Soccer System has been proposed. The proposed strategies have simple structure and do not need any hard manual coding. The possibility of autonomous robot cooperation with a human operator is identified by using the proposed strategy structure. Although a series of simulations have been conducted with restrictive opponent strategies and limited players, the reinforcement learning algorithm of the autonomous robots shows the capability to cooperate with rough and inconsistent human operators. In the future, more adjustments in the state variables and modifications in the learning algorithm will be considered to make the learning strategy structure more efficient.

## REFERENCES

[1] D. S. Kwon et al., "A novel interactive robot soccer system," *Proc. of IEEE/ASME Int. Conf. on Advanced Intelligent Mechatronics*, pp. 1082-1086, 2001

[2] J. H. Kim, H. S. Shim, H. S. Kim, M. J. Jung, I. H. Choi, and J. O. Kim, "A cooperative multi-agent system and its real time application to robot soccer," *Proc. of the 1997 IEEE Int. Conf. on Robotics and Automation*, Albuquerque, New Mexico, pp. 638-643, April 1997.

[3] H. Kitano, M. Asada, Y. Kuniyoshi, I. Noda, and E. Osawa, "RoboCup: The robot world cup initiative," *Proc. of the IJCAI-95 Workshop on Entertainment and AI/Alife*, 1995.

[4] P. M. Jones and J. L. Jacobs, "Cooperative problem solving in human-machine systems: theory, models, and intelligent associate systems," *IEEE Trans. On Systems, Man, and Cybernetics-Part C: Applications and Reviews*, vol. 30, no. 4, pp. 397-407, November 2000.

[5] H. T. Chu and B. R. Hong, "Cooperative behavior acquisition in multi-robots environment by reinforcement learning based on actions selection level," *Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 1397-1402, 2000.

[6] Y. Takahashi and M. Asada, "Vision-guided behavior acquisition of a mobile robot by multi-layered reinforcement learning," *Proc. of the 2000 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 395-402, 2000.

[7] K. Kostiadis and H. Hu, "Reinforcement learning and co-operation in a simulated multi-agent system," *Proc. of the 1999 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 990-995, 1999.

[8] S. Haykin, Neural Networks – A Comprehensive Foundation, Prentice Hall, New Jersey, 1999.

**Hyoung-Rock Kim** received the B.S. degree in mechanical engineering from Seoul National University, Korea in 2000, and the M.S. degree in mechanical engineering from KAIST, Korea in 2002. He is currently pursuing the Ph.D. degree in the department of mechanical engineering at KAIST. His research interests are learning algorithm, intelligent robotics, and multi-agent system.

**Jung-Hoon Hwang** received the B.S. degree in mechanical engineering from Yonsei University, Korea in 1997, and the M.S. degree in mechanical engineering from KAIST, Korea in 1999. He is currently pursuing the Ph.D. degree in the department of mechanical engineering at KAIST. His research interests are multi-agent system, behavior robotics, robot control, human-robot haptic interface, and teleoperation.

**Dong-Soo Kwon** received the B.S. degree in mechanical engineering from Seoul National University, Korea in 1980, the M.S. degree in mechanical engineering from KAIST, Korea in 1982, and Ph.D. degree in mechanical engineering from Georgia Institute of Technology, Atlanta in 1991. From 1991 to 1995, he was a research staff at Oak Ridge National Laboratory. He is currently an associate professor of mechanical engineering at KAIST, Daejon, Korea. His current research interests include teleoperation, Human-robot haptic interface, medical robot, robust control, and intelligent robotics. He is a member of IEEE, ICASE and KS.