A Learning Control System for Rapid Position Control with Obstacle Consideration for Aerial Hovering Vehicles

Syafiq Fauzi Kamarulzaman[1], Takeshi Shibuya[2], and Seiji Yasunobu[3]
[123]Dept. of Intelligent Interaction Technologies, University of Tsukuba, Japan
[1]Faculty of Computer Systems and Software Engineering, Universiti Malaysia Pahang, Malaysia
Email: syafiq@fz.iit.tsukuba.ac.jp, shibuya@iit.tsukuba.ac.jp, yasunobu@iit.tsukuba.ac.jp

***Abstract*—Aerial hovering vehicles, like helicopters, has the ability of remaining stationary while on air, which is useful for maintaining its position at a certain location for a period of time. Operating a rapid hovering maneuvers while avoiding obstacles is difficult and mostly depends on the skills of an expert operator. In this research, we develop a control system that learns to form a control knowledge that can perform rapid position control while considering obstacles. The system works by manipulating angular orientation of the vehicles for performing a rapid movement, while considering existing obstacles in an operation environment. The effectiveness of the system was evaluated and proven in series of simulations.**

I. Introduction

Aerial hovering vehicles, such as helicopters, has been useful in various applications. One of the useful application of such vehicles is in operations that needs the vehicle to remain stationary on air, hovering and maintaining its position at a certain location over a period of time. UAV that has such characteristic has been beneficial in many operation, particularly in case of observation, due to its smaller size.

Operating rapid position control with hovering maneuvers around obstacles is difficult without certain interruption when making control decision. The effectiveness of such control decisions are mostly depending on the skill of an expert operator. An expert operator develops his own control knowledge from training and experience to operate such maneuvers around existing obstacles [1]. If such control knowledge can be learned by a control system autonomously, dependency on a skilled human operators can be reduce, while having possible application on developing an autonomous UAV.

Previously, we have developed a learning-based control system that can perform rapid position control by manipulating a hovering UAV's angular orientation while hovering at a certain altitude, and applied on an inverted pendulum system [2]. In this research, we develop a learning-based control system that learns the necessary control knowledge for aerial hovering vehicles, that can learn to operate the rapid position control, around obstacles, towards an assigned target position. Reinforcement Learning is applied to create control knowledges for manipulating angular orientation in order to perform a rapid position control around obstacles towards a target position. Simulations were constructed to confirm the effectiveness of the system.

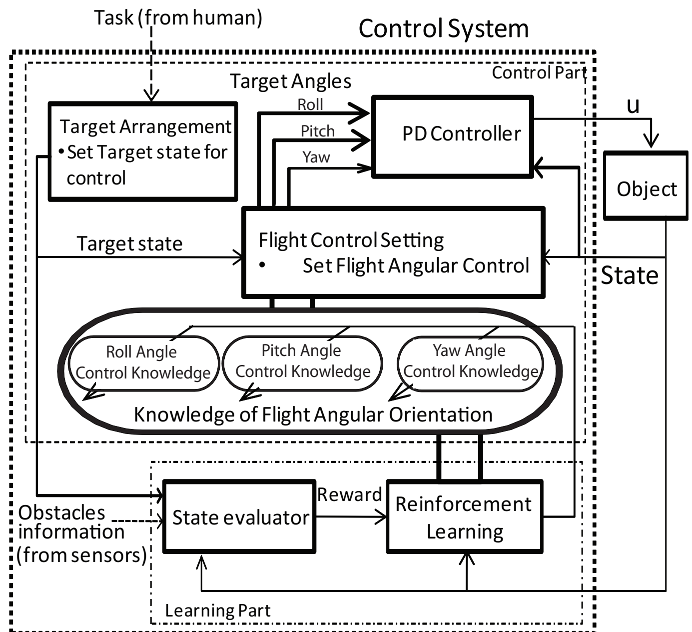

Fig. 1. The structure of the system.

In the next chapter, we discuss the structure of the control system, including the dynamics of aerial hovering vehicles that for performing rapid position control and the learning process that make the vehicle able to consider any obstacle while performing a rapid position control. Later, the structure of simulations done to confirm the effectiveness of the system, and finally, the results of the simulations.

II. The Structure of the Control System

Position control for aerial hovering vehicles requires the operator to configure the angular orientation of the vehicle to stabilize while performing transition between two positions. An expert operator are capable of analyzing and react to a situation in case of any obstacles appear in the vehicle control path. The operator decides the manipulation techniques for the angular orientation of the vehicle, to reconfigure its operation path around the obstacles safely with optimum results. Operating the vehicle angular orientation is difficult

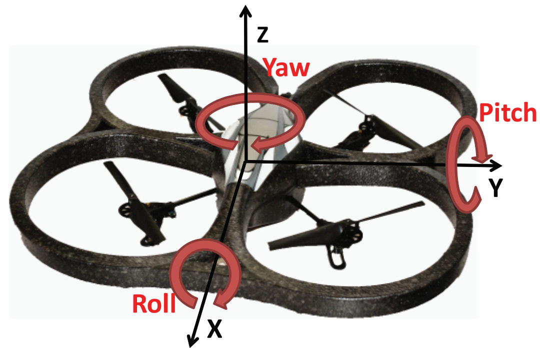

Fig. 2. The angular dynamics of a hovering UAV (ArDrone by Parrot)

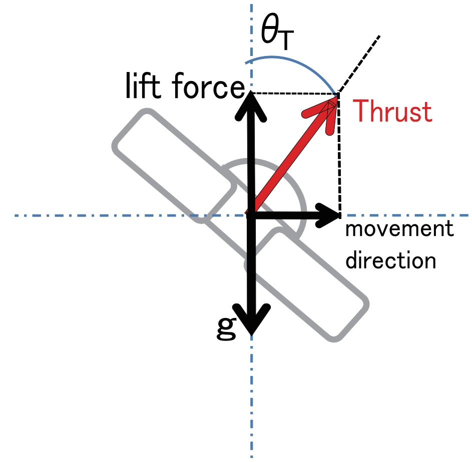

Fig. 3. The trajectory of a hovering UAV when manipulating target angle

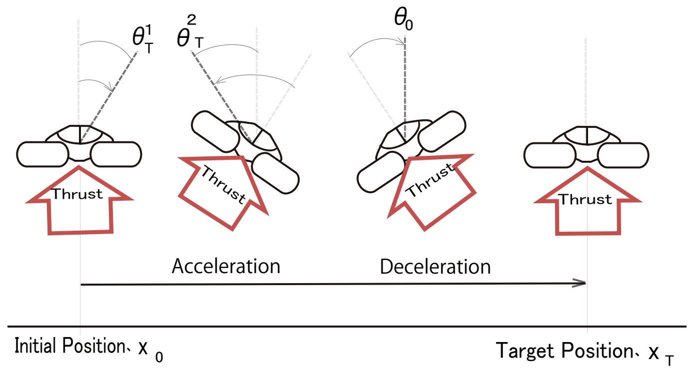

Fig. 4. The maneuver for position control through target angle.

due to constraints in selecting the best angular orientation since it is also required to stabilized the vehicles on air. Therefore, an adequate control knowledge are needed for performing an optimum angular transition that is required for position transition.

The structure of the system developed in this research is shown in Fig. 1. This system consisted two main part; the Control Part and the Learning Part. The Control Part controls the controlled aerial hovering vehicle for performing rapid position control using three angular parameters; roll, pitch and yaw. The Learning Part rewrites the control knowledge depending on the successful and failed control attempts in particular operation environment, therefore a successful operation can be found and maintained in a specific environment.

In this research, we applied the ability to learn the knowledge of the expert operator into a control system based on a hovering UAV shown in Fig. 2. This hovering UAV consisted three parameters of angular orientation in three dimensions; roll, pitch and yaw. These three angular parameters is manipulated in our system for operating position control while considering any existing obstacles.

A. Rapid Position Control of Aerial Hovering Vehicles

The control part obtained a target state best on task instructed by a human operator. This part then selects the best combination of angular orientation that can perform a position control. In our previous research [2], We understood that it was possible to manipulate angular orientation to perform a rapid position control. This is because, manipulating the angular orientation changes the direction of the thrust. In order to maintain the altitude, the force of the thrust will be increased and that creates a horizontal force as shown in Fig. 3. This effect produces acceleration for performing horizontal movements.

Target angle θ_T is used for changing the direction of the thrust to create a horizontal force for horizontal movement while hovering at a constant altitude. Fig. 3 shows the direction of the thrust according to target angle θ_T that makes the horizontal movement possible. Fig. 4 shows by manipulating target angle at a constant altitude, acceleration and deceleration can be produced for position control. Therefore, optimum combination of target angles are needed for performing a precise and optimum rapid position control.

Here, three angular parameters are needed to be manipulated for an effective rapid position control around an existing obstacles. This function is done by the learning part, that learns the optimum maneuver around the obstacles by stores the knowledge for manipulating these angular parameters.

B. Learning Rapid Position Control with Obstacle Consideration

The Learning Part rewrites the control knowledge depending on successful and failure attempts in a control operation. In an unknown environment, it is difficult to perform a successful and optimum control operation due to availability of obstacles and other constraints. Here, Reinforcement Learning is applied to rewrite the control knowledge by determining the favorable state s; location and velocity, for an action a, which is the optimum target angular orientation θ_T for rapid position control. The control knowledge, Q is rewritten using Q-learning as (1) and (2), which is

$$Q(s, \theta_T) = (1 - \alpha)Q(s, \theta_T) + \alpha[rew + \gamma Q_{max}], \quad (1)$$

$$Q_{max} = \max_{\theta'_T} Q(s', \theta'_T) \quad (2)$$

Where s and s' denotes state and future state, α is Learning Rate, γ is the discount rate and r is the reward.

However, as shown in Fig. 2, the hovering UAV does have three parameters of angular orientation, therefore, 3 optimum target angle must be learned in order to perform a rapid position control. Plus, effective combination of three target angles may help perform an optimum rapid position control around obstacles. Therefore, target angle θ_T is a set of three target angles from the three parameters of angular orientation, as

$$\mathbf{\Theta_T} = \{\theta_{roll}, \theta_{pitch}, \theta_{yaw}\}.$$

From above, a set of 3 independent control knowledge Q is created for each target angle, as

$$\mathbf{Q} = \{Q_{roll}, Q_{pitch}, Q_{yaw}\}.$$

Since there will be three sets of independent control knowledge will be used in the system based on three dimensional angular orientation, state s were prepared to be three dimensional coordinates and velocities. State s consisted location $\mathbf{r}$, where

$$\mathbf{r} = \{x, y, z\},$$

and velocity according to each axis, $\mathbf{v}$, where

$$\mathbf{v} = \{v_x, v_y, v_z\}.$$

Therefore, state s is denoted as

$$s = \{\mathbf{r}, \mathbf{v}\}.$$

The reward rew used to update the control knowledge $\mathbf{Q}$ is based on (3),

$$rew = \frac{d_s - d_{s'} + 1}{d_{s'}} \quad (3)$$

Where d_s is the distance between the control object at state s and the target location, and $d_{s'}$ is the distance between the control object at state s' and the target location, as shown in Fig. 5.

Reward r in (3) is used for two reasons; To have the control object travel a large distance between two states, and To have the control object distinguish the favorability of states that are closer to target position. This is because, larger travel distance between two states represent higher acceleration that is needed for performing rapid position control to reach the target state at a faster rate.

Beside (3), reward rew is a constant when the system failed to reach the target state within the designated simulation time, and when the control object exceed the designated movement range for the simulation. The details of the simulation is explained in the next chapter.

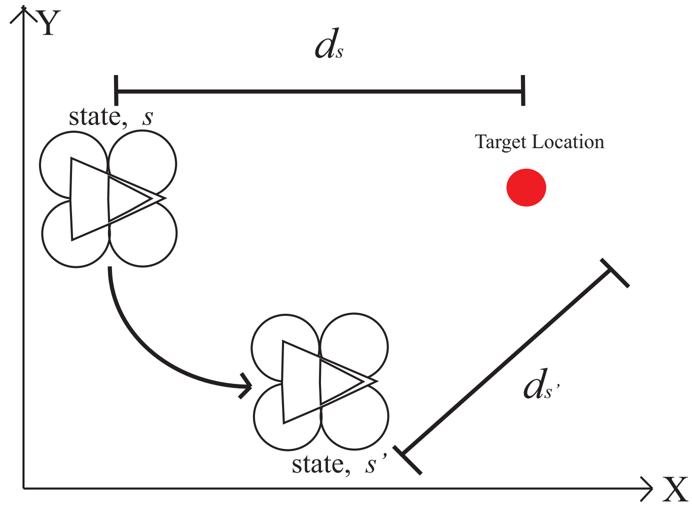

Fig. 5. Parameters for determining rewards.

III. Structure of Simulation

In this research, the simulation was done in MATLAB Simulink based on the parameters of the hovering UAV shown in Fig.2. These parameters is shown in Table I. A series of simulations which consisted different target position was created to confirm the effectiveness of the system. Obstacles was also included in the simulation to confirm that the system is able to operate through obstacles as intended. The target states and obstacles were placed as shown in Fig. 6.

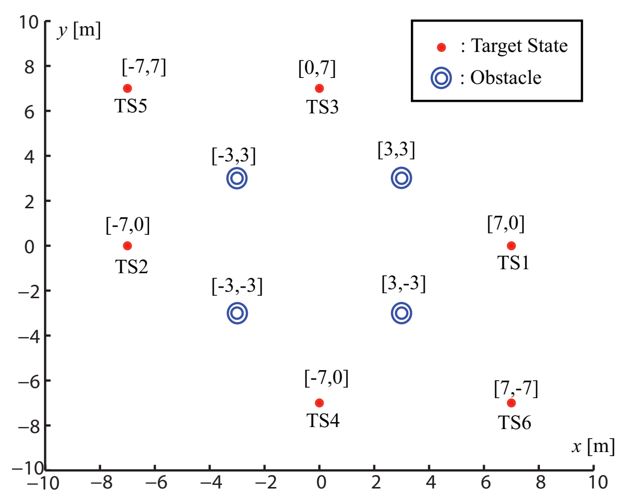

Fig. 6. Obstacles and target location assigned in simulations.

The parameters for Q-learning is as shown in Table II. These parameters was selected pre-simulation. The position control only applied on horizontal movements with constant altitude, within a movement range assigned.

There are several properties designated into the simulations before the simulation was conducted. For each simulations with different target state, the properties is as follows.

TABLE I
SPECIFICATIONS OF THE SIMULATED AERIAL HOVERING VEHICLE

Parameters	Value
Weight	0.42 [kg]
Size:	
Length	0.53 [m]
Width	0.52 [m]
Height	0.1 [m]

TABLE II
Q-LEARNING PARAMETERS

	Parameters	Range	Intervals
State	Location,r[m]	$-10 < r(x, y) < 10$ $r(z) = 1$	2
	Velocity, v [m/s]	$-10 < v < 10$	2
Action	Target Angle, θ_T[rad]	$-0.25 < \theta_T < 0.25$	0.05

Learning rate, α	0.5	Discount rate, γ	0.3

- Simulation runs six times with different target state assigned with each having 4 four permanent cylindrical obstacles with diameter of 1[m].
- Simulation end at 3000 episodes of trials.
- 30 second operation time for each episode.
- Action is evaluated for reward and target angles were renewed every 1 second.
- $\varepsilon-$greedy selection of each target angles
- $rew = -2$ when the action leads to out of range or obstacles.
- Due to large intervals on states, the controller for states within 1[m] around the target state will be switched to PD control.

The results from the simulations is determine by the accumulated rewards through the simulations and the successful attempts on reaching the target position by operating with and without obstacles. The results continues in the next chapter.

IV. RESULTS OF THE SIMULATION AND DISCUSSION

At the end of the simulation, the result of the trials for each episode was collected and analyze to confirm the reliability of the system. The results should provided the information on the control path for each target state assigned. This includes position transition and angular transition which is important for distinguish the reliability of the system, with or without obstacles in its operation's environment. The results also provide information regarding the improvement that happens in the control knowledge. This can be understood by viewing the accumulated rewards in the simulations, since more rewards accumulated leads to more successful operation were attempted through the simulation.

Therefore, The results of the simulation is viewed in two aspects. The first aspect is the characteristic of rapid position control operation that successfully operates within an environment. The results from the first aspect also compares an operation without obstacles with an operation with obstacles. The second aspect is the improvement of control knowledge that is used to perform the rapid position control. The accumulated reward from the trials throughout the simulation is monitored to confirm the effectiveness of the learning process.

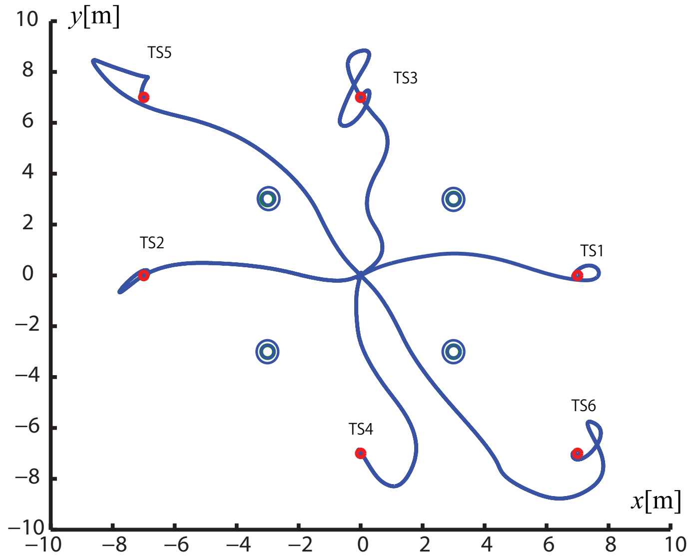

Fig. 7. Successful control operation for the simulation with assigned target state.

A. Successful Control Operations towards designated target states

This results confirms the reliability of the system for performing successful control operation that is required to reach the assigned target state. There are 6 target states were assigned with the same initial starting position. Control attempts for each target states that was learned by the system during the simulation is shown in Fig. 7.

Fig. 7 shows the control operation that was accomplished at the final, 3000th episode of the simulation for each target angle assigned. During this episode, the most successful control maneuvers in an environment for a target state should have been learn through trial and error in the previous episode.

The results shows that the control system was able to control the control object towards each designated target states. Simulation for target 1 to 2 shows that direct movement from start position was able to achieved, when the movement path is not obstructed by any obstacles.

However, for target 3 and 4, the movement path was not so smooth compared to target 1 and 2. This is because, the system learns the most effective maneuvers, and in case for target 3 and 4, the optimum maneuvers that was learned here were not as smooth as for target 1 and 2, in Fig. 7.

For target 5 and 6, the control system bent the movement path so that the control object can avoid the obstacles, but still reaches the assigned target state.

1) Successful Control Operation without obstacles in direct path.: This results explains the movement path of the control object that was operated by the system towards reaching target

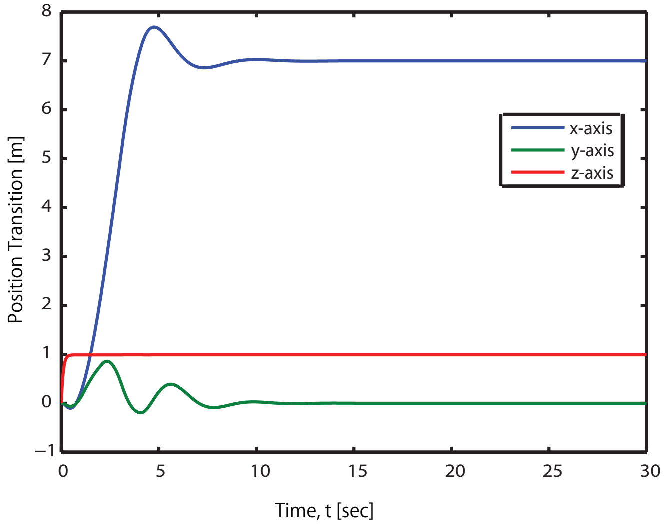

Fig. 8. Position transition during successful control operation without obstacles in direct path (Target State 1)

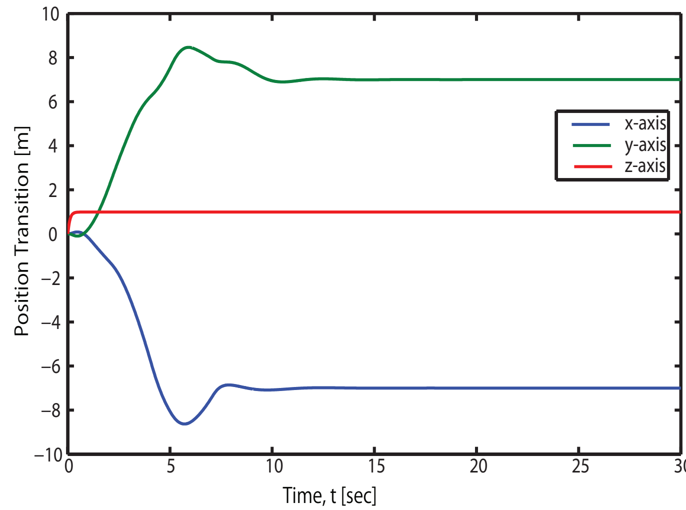

Fig. 10. Position transition during successful control operation with obstacles in direct path (Target State 6)

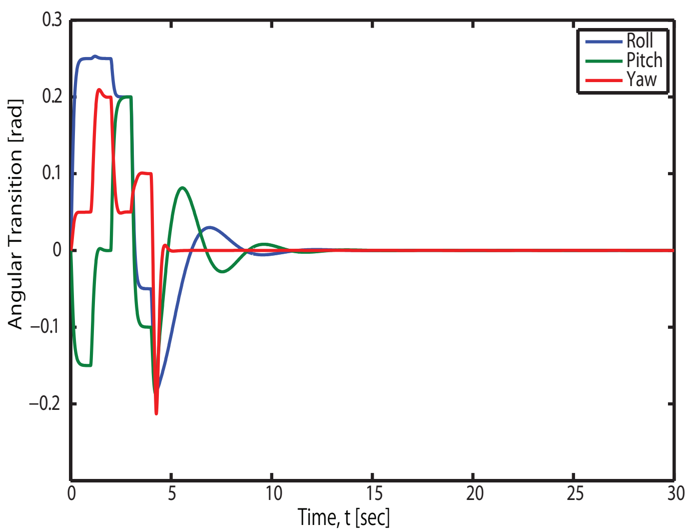

Fig. 9. Angles transition during successful control operation without obstacles in direct path (Target State 1)

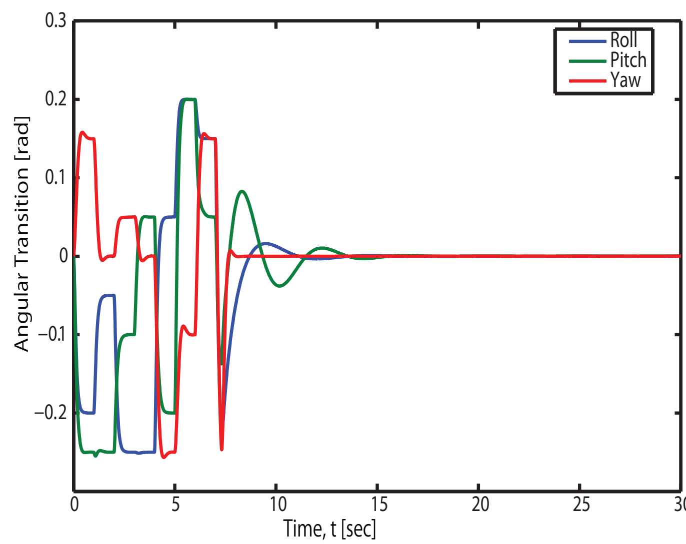

Fig. 11. Angles transition during successful control operation with obstacles in direct path (Target State 6)

state 1. The direct path towards target state 1 is unblocked by any obstacles but the system are needed to be careful of the obstacles at the side of the direct path. The details of the control operation for reaching target state 1 is shown in Fig. 8 and Fig.9.

Fig.8 shows the position transition of the control object in each 3 axis, during the final episode of simulation for Target State 1. Here, the system selects the optimum position transition for achieving the target state, with less unnecessary movements according to each axis.

Fig.9 shows the transition of angular orientation based on roll, pitch and yaw during the final episode of simulation for Target State 1. Here, the manipulation of angle can be seen to influence the position transition in Fig.8.

It is understood from Fig.7 that the same effect are also happening in simulation with target state 2, 3 and 4, since there are no obstacle obstructing the direct path towards the target state. Therefore, it is understood from the simulation that the most optimum combination of target angles was selected to construct the most optimum position maneuvers needed to achieve the target state.

2) Successful Control Operation with obstacles in direct path.: This results explains the movement path of the control object that was operated by the system towards reaching target

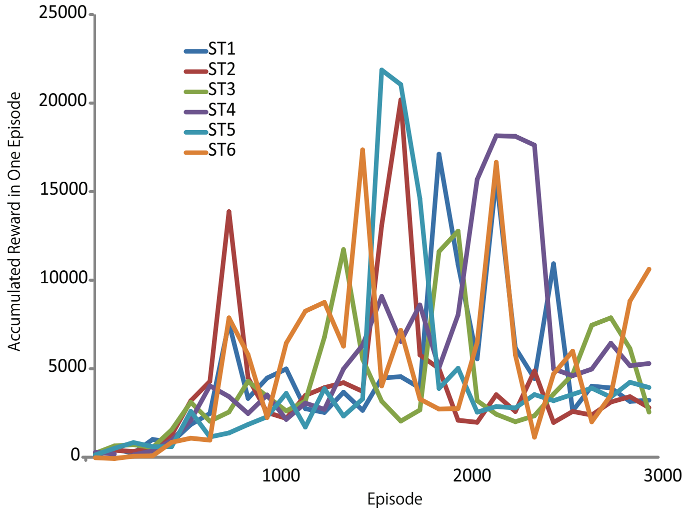

Fig. 12. Accumulation of reward during simulation.

state 6. The direct path towards target state 6 is blocked by an obstacle and the system are needed to consider this obstacle when performing control operation to reach target state 6. The details of the control operation for reaching target state 1 is shown in Fig. 10 and Fig.11.

Fig.10 shows the position transition of the control object in each 3 axis, during the final episode of simulation for Target State 6. Here, the system selects the optimum position transition for achieving the target state, with necessary movements according to each axis, needed to avoid the obstacles place in the environment.

Fig.11 shows the transition of angular orientation based on roll, pitch and yaw during the final episode of simulation for Target State 6. Here, the manipulation of angle can be seen to influence the position transition in Fig.10 for taking necessary movements to avoid the assigned obstacle.

It is understood from Fig.7 that the same effect are also happening in simulation with target state 5, since there are obstacle obstructing the direct path towards the target state. Comparing the successful control operation without obstacles, target angles assigned in the control operation with obstacles are more frequent. This is because, the angular orientation applied are necessary to help the control object avoid the obstacles that exist in the direct path towards the target state. Therefore, it is understood from the simulation that the most optimum combination of target angles was selected to construct the most optimum position maneuvers needed to achieve the target state while considering existing obstacles in the environment.

B. Control Knowledge Improvements during Control Operations towards Designated Target States

This result explains the improvement that occurred during the simulation. For each episode, knowledge has been updated to satisfy the environment where the control operation will be performed. Therefore, we can understand that the increasing number of accumulated rewards represent the more successful a control operation was. This explains that the system learned the best control operation needed by attempting the operation that leads to most reward in each episode. The results of the accumulated reward is shown in Fig.12

Fig.12 shows the accumulated reward during the simulation, according to the assigned target state. Here, we understood that the amount of accumulated reward increases with more episodes of trials. The amount of accumulated rewards was low at the starting episode, but increases towards the end. Therefore, it is proven that the successful control attempts were learned during the simulation that leads to more reward accumulated through more episodes.

V. Conclusion

In this research, we develop a learning-based control system that can perform rapid position control by manipulating a hovering UAV's angular orientation while considering obstacles. Three independent control knowledge for three parameters of angular dynamics were constructed and updated during operation in order to obtain an optimum knowledge depending on environment situation.

Simulations that involved control operations in an environment with different target state assigned were created in order to confirm the effectiveness of the system. Obstacles was placed in the environment that make some target states are easy to achieve and some require the system to make some consideration on the operation maneuver for avoiding the designated obstacles.

Results shows that all the target states were achieved and operation maneuvers were successful in avoiding the obstacles. The results also shows that the time taken to achieved the target states were not influenced much by the obstacles, and around the same as the time taken to achieved the target states without any obstruction. Therefore, the learning control system for rapid position control with obstacle consideration for aerial hovering vehicles was achieved.

References

[1] T.Matsubara and S.Yasunobu, *An Intelligent Control Based on Fuzzy Target and Its Application to Car Like Vehicle*. SICE Annual Conference, 2004.

[2] Syafiq F. K., T. Shibuya and S. Yasunobu, *A Learning Control System for Rapid Position Control of Aerial Hovering Vehicles*. SICE Annual Conference, pp.1546-1551, 2012.

[3] Richard S. Sutton and Andrew G. Barto, *Reinforcement Learning An Introduction*. MIT Press, 1998.

[4] J. Valasek, J. Doebbler M.D. Tandale and A.J. Meade, *Improved Adaptive-Reinforcement Learning Control for Morphing Unmanned Air Vehicles*. IEEE SMC, pp.1014-1020,2008.

[5] Y. Nakamura, S. Ohnishi, K. Ohkura and Kanji Ueda, *Instance-Based Reinforcement Learning for Robot Path Finding in Continuous Space*. IEEE SMC, pp.1229-1234,1997.