SICE Annual Conference 2012
August 20-23, 2012, Akita University, Akita, Japan

A Learning Control System for Rapid Position Control of Aerial Hovering Vehicles

Syafiq Fauzi Kamarulzaman[1] , Takeshi Shibuya[2] and Seiji Yasunobu[3]

[123]Department of Intelligent Interaction Technologies, University of Tsukuba, Ibaraki, Japan
[1]Faculty of Computer Systems and Software Engineering, Universiti Malaysia Pahang, Malaysia
(E-mail: syafiq@fz.iit.tsukuba.ac.jp, shibuya@iit.tsukuba.ac.jp, yasunobu@iit.tsukuba.ac.jp)

Abstract: A learning based control system for rapid position control of aerial hovering vehicles is proposed. An aerial hovering vehicle uses angular orientation to accelerate during position control where target angle is preserve to produce acceleration. By arranging target angle, the acceleration and deceleration during position control can be configured, thus able to produce a rapid position control. In order to create a learning based control system for rapid position control, the characteristic of the position control by target angle of an aerial hovering vehicle is simplified an applied to an inverted pendulum system. The effectiveness was confirmed on the inverted pendulum system through simulations where several target position was assigned to be achieved.

Keywords: Reinforcement Learning, Unmanned Aerial Vehicle, Inverted Pendulum

1. INTRODUCTION

Unmanned Aerial Vehicles (UAV) is becoming ubiquitous in recon and observation operation. Conventional UAV includes remotely controlled aircraft that requires a human operator to pilot from a certain location. The distance of the remote pilot from the UAV can affect the control condition of the UAV that mainly caused by delays in control command. Therefore, the UAV dependency on human operator should be reduced, towards making the UAV itself autonomous.

Rotorcraft UAV as in Fig. 1 is an aerial hovering vehicles that beneficial compared to any aeroplane-shaped UAV because of the capability of performing hovering movements since hovering movements can provide close range observation.

Fig. 1 A rotorcraft UAV propelled by four rotors.

In order to perform position changing operation, an aerial hovering vehicle is required to change its angular orientation (attitude) for position control. The angular orientation of a rotorcraft vehicle is controlled by a human operator using cyclic, similar to the joystick of a conventional aircraft.

An expert human operator of an aerial hovering vehicle is capable of performing accurate position control rapidly by configuring the angular orientation. To perform a rapid position control as an expert operator would require skills that are hard to develop for any aerial hovering vehicles.

In this paper, we developed a control system that can learn a knowledge that uses target angle to produce acceleration and deceleration for performing rapid position control. The system learns to acquire the knowledge by repeatedly try to arrange the period of applying certain target angles for performing an operation to achieve an assigned target position. The target angle based position control method of an aerial hovering vehicle can be simplify and applied on the inverted control system. In order to construct the learning control system for rapid position control, a simpler experiment device as the inverted pendulum system is used, and its effectiveness is evaluated through simulation.

2. RAPID POSITION CONTROL OF AN AERIAL HOVERING VEHICLE

In order to create the learning control system for the rapid position control, the dynamics of an aerial hovering vehicle is studied. The problem of the dynamics, which consist of position control by referring target angle, is simplified and later emulated on an inverted pendulum system.

2.1 The dynamics of an aerial hovering vehicle

The system is developed to learn the best coordination of target angle θ_T that can perform a rapid position control. Target angle θ_T is used to change the direction of the thrust to create horizontal force that can create a horizontal movement while airborne. Fig. 2 shows the direction of the thrust according to target angle θ_T that makes the horizontal movement possible.

However, by configuring the target angle θ_T, the thrust must be increased to preserve the leaning angle against gravity. This means that when the preservation period

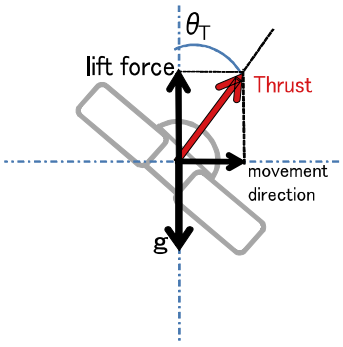

Fig. 2 Configuration of an aerial hovering vehicle movement by angular orientation.

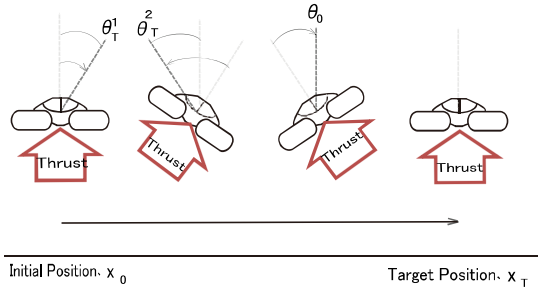

Fig. 3 A position control of an aerial hovering vehicle using target angle θ_T as reference.

of the leaning angle increased, the horizontal velocity of the aerial hovering vehicle will be increased. Therefore, certain configuration strategy of the target angle θ_T is needed to provide acceleration and deceleration for a precise position control.

Fig. 3 shows the angular orientation of the aerial hovering vehicle during a position control. A target angle θ_T^1 is set to provide a force for acceleration while another θ_T^2 is to provide a force for deceleration before returning to its initial angle θ_0. Fig.4 shows the transition of the target angle θ_T during operation.

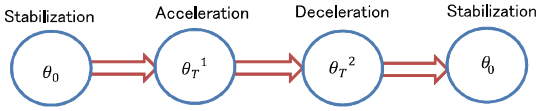

Fig. 4 Transition of target angle θ_T during operation.

Therefore, by configuring the preservation period at a certain angular orientation, optimum acceleration and deceleration strategy can be produced that can have the aerial hovering vehicle perform a precise position control. A rapid position control means that the system must be able to perform a position control within short amount of time based on any assigned target position x_T. In order to perform a rapid position control, the best combination of target angle θ_T with respect to acceleration and deceleration must be learned by the system.

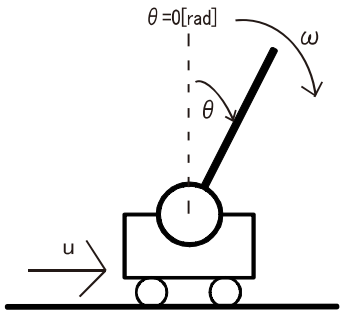

Fig. 5 The stabilization control of inverted pendulum.

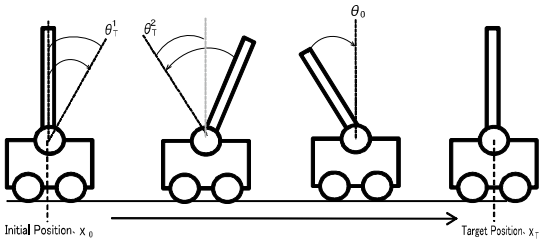

Fig. 6 The inverted pendulum position control using target angle θ_T as reference.

2.2 Emulating an aerial hovering vehicle on inverted pendulum system

Emulating the position control of an aerial hovering vehicle, the inverted pendulum position control applies position control through angular orientation. The inverted pendulum is controlled using combinations of target angle θ_T of the pendulum as reference to achieve a target position x_T.

The inverted pendulum position control is as shown in Fig. 5. Force is applied on the cart to change its position while at the same time maintaining the pendulum pointing upwards as the position and angular control of an aerial hovering vehicle. The force applied to the cart is referred to the target angle of the pendulum. Hence, the force is applied to maintain the pendulum at a certain angle. In this research, the position control of the inverted pendulum which refers to target angle θ_T was constructed using PD control.

Fig. 6 shows the position control of the inverted pendulum, based on the position control of an aerial hovering vehicle. By changing the target angle θ_T, certain force is applied to the cart that moves the cart and the pendulum towards a certain position. Combination of target angle that can provide horizontal acceleration and braking for the position control of the inverted pendulum can be learned through the system.

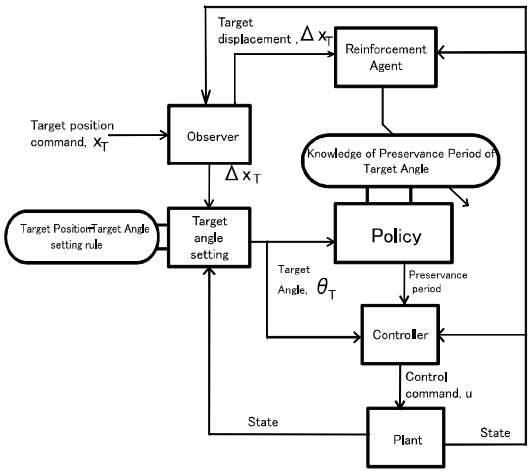

Fig. 7 The structure of the system.

3. THE STRUCTURE OF THE LEARNING CONTROL SYSTEM

The learning control system for above purpose is created based on the structure shown in Fig. 7. Reinforcement learning is used to acquire the knowledge of preservation period of the target angle with respect to the target position x_T.

Target position x_T is given as the control objective for the system. Target angle θ_T is selected based on a rule provided relating to the assigned target position x_T, while the learning process only consists of configuring the preserve period of the target angle.

The learning process updates the knowledge using rewards based on the achievement of the operation. When the system reaches the target position, reward is applied while none is given when the system fails.

4. SIMULATION

Fig. 8 shows the device where the simulation for this research was based. This research was done in simulations since the learning process is time consuming that could cause durability issues on the device if operated for a long time since the cart movement is limited to a smaller range.

The properties of the experiment device were applied in the simulation to have the system be applicable in real world environment. The properties of the device are as shown in Table 1.

Table 1 The parameter properties of the experiment device used in simulations

Parameters	Range
Pendulum Mass, m[kg]	0.08
Pendulum Length, l[m]	0.4
Cart Mass, M[kg]	3.117
Cart Motor Output, V_{out}[V]	-9.9∼9.9
Cart Movement Range, x[m]	-1.0∼1.0

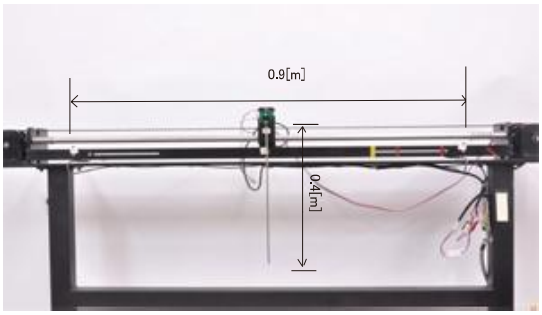

Fig. 8 Experimental device (Japan E.M. Co. Ltd.) on which the simulation were based

4.1 Preparation for simulation

Before conducting the simulation, a series of test is done to confirm that the inverted pendulum control system controls using target angle θ_T. It is known that a UAV horizontal acceleration increases when the leaning angle increased, therefore the same conclusion must be confirmed in the pendulum control system before the simulation.

Table 2 shows the result of the test that uses a set of three target angle θ_T. From this table it is confirmed that acceleration increases as target angle θ_T increase. Fig. 9 shows the motor output at every sampling pulse for each tested target angle θ_T.

Based on Fig. 9, a certain amount of output, total output V_{out} is produced for each target angle θ_T at a certain time. Total output V_{out} helps calculate the period of maintaining the target angle therefore is used in the learning algorithm to help produce the learning control

Table 2 Pre-experimental results for determining the output required

Target Angle,θ_T [rad]	0.02	0.05	0.1
Total output, V_{out}[V] needed to maintain θ_T for 3[sec] (Sampling time:0.01[sec])	184.5	460.0	919.6
Distant covered, x[m] in 3[sec]	0.80	2.00	4.02
Total output, V_{out}[V] needed to maintain θ_T for 5 [sec] (Sampling time:0.01[sec])	504.2	1259.8	2522.9
Distant covered, x[m] in 5[sec]	2.28	5.71	11.45
Average amount of output per distant covered, V_{out}[V/m]	225.89	225.28	224.55
Acceleration, a [ms^{-2}]	0.13	0.24	0.48

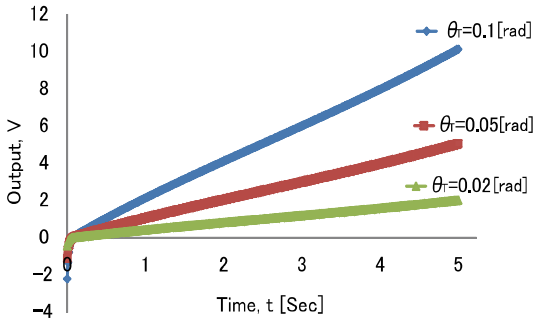

Fig. 9 The reference data used to calculate the period of output for each target angle.

system. Total output V_{out} is used instead of time t as the unit of period is for lesser burden in computation during the simulation. Since the operation able cart position range is limited, the range of the total output V_{out} were limited up to 100 [V].

Based on these results, the angular orientation is known to be related to the acceleration of the cart. This confirmed that the system have the dynamics similar to an aerial hovering vehicle.

4.2 Method for updating knowledge

In this paper, learning is defined as knowledge update, where the knowledge is updated to produce an optimum knowledge needed to perform a successful operation. The method used in this research is reinforcement learning, therefore Q-learning is used to produce a value function $Q(\theta_T, V_{out})$ that will define the best combination of target angle θ_T and total output V_{out}. Q-learning algorithm updates the value function $Q(\theta_T, V_{out})$ using reward r for producing a better knowledge.

The algorithm is written as

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

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

where θ_T denotes continuing target angle θ_T and V_{out} denotes the total output V_{out} of the continuing target angle. α is defined as learning rate while γ is defined as discount rate.

The parameters of the Q-learning algorithm used in the simulation are as shown in Table 3. These state and

Table 3 Q-learning parameters

	Parameters	Range	Intervals
State	Target Angle, θ_T [rad]	-1 ∼ 1	0.05
Action	Total Output, V_{out} [V]	0 ∼ 100	20

Learning rate, α	0.5	Discount rate, γ	0.3

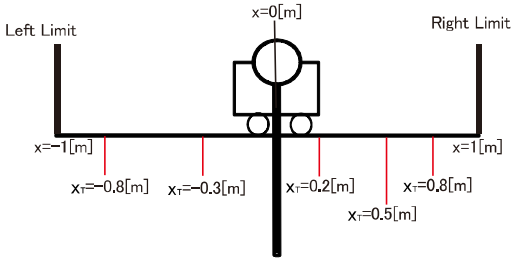

Fig. 10 Target position assigned

action parameters range was selected depending on the properties of the control object device. The intervals of these parameters were randomly selected before the simulation. The learning rate α and discount rate γ was also randomly selected before the simulation was conducted.

4.3 Simulation properties

The simulation was conducted by using five target cart position as shown in Fig. 10. The objective of this simulation is to have the system learns the best control operation for achieving the target cart position assigned. All those targets were selected randomly before the simulations. Five target positions were selected to confirm that the system was able to learn a rapid position control at any direction and distance.

The simulation properties were selected before conducting the simulation. These properties are used for all five target positions assigned previously. An operation is the process of attempting position control of the inverted pendulum towards the target position within 10 seconds. Each operation done is counted as trials. The other simulation properties are as follows.

- Simulation runs five times with different target position assigned.
- Simulation end at 550 trials.
- 10 second operation time for each trial.
- $\varepsilon-$greedy selection of output
- Reward is given after operation ends.
- Full reward, $r = 1$ is given to acceleration target angle, θ_T^1 if successfully achieve target position x_T at the end of an operation
- Half reward, $r = 0.5$ is given to deceleration target angle, θ_T^2 if successfully achieve target position x_T at the end of an operation
- zero reward, $r = 0$ is given to both target angle θ_T^1 and θ_T^2 if it fails to achieve target position x_T at the end of an operation.

The results were collected and analyzed after the simulation is finished with 550 trials for each five assigned target position x_T.

5. RESULTS

The result for the simulation is separated into two categories. The first shows improvement achieved through the learning process while the second shows the success-

ful operation achieved at the end of the simulation. The improvement achieved in the first result confirms the validity that the learning process was able to create a better knowledge through the simulation that can lead to a successful operation. The operation shown in the second result confirms that the successful operation operates the position control towards the target position x_T.

5.1 Knowledge improvement through learning process

At the beginning of the simulation, the value function $Q(\theta_T, V_{out})$ is at zeros, where any operation using this knowledge will less likely to be successful as no particular optimum combination of angle orientation can be detected from the knowledge. At the end of the simulation, the optimum combination has become clear due to the update by Q-learning algorithm. Therefore, the target position can be achieved during operation at the end of the simulation.

Fig. 11 shows the final cart position at the end of every operation trials. The final cart position at the beginning of the simulation is scattered around the cart movement range. However, the final cart positions are focused to the target position at the end of the simulation. At the end of the simulation, a successful operation that can achieve the target position is obtained.

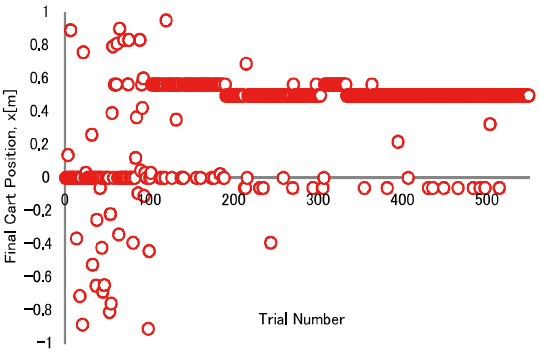

Fig. 11 Improvement of final cart position with respect to the number of trials.(Target position, x_T=0.5[m])

5.2 Success operation learned through simulation

When the simulation ended, an optimum knowledge that can lead to a successful operation was obtained. The result below shows the operation result that uses the knowledge after 550 trials.

Fig. 12 shows the pendulum angular trajectory during an operation that uses the knowledge obtained after 550 trials. The pendulum trajectory varies depending on each target position assigned. However, it can be seen that the pendulum angle stabilized at $\theta = 0[rad]$ around 5 seconds.

Fig. 13 shows the cart trajectory during the operation that uses the knowledge obtained after 550 trials. The cart trajectory can be seen to be moving towards the target position and stabilizes near the target position with an

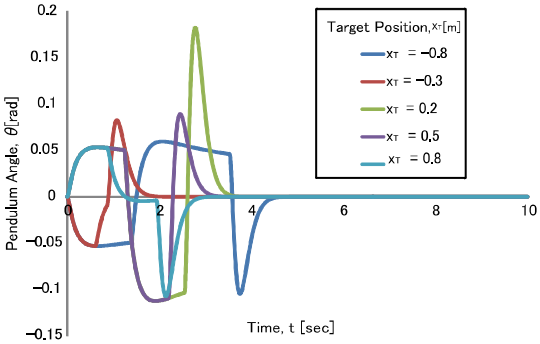

Fig. 12 Pendulum angular trajectory during an operation that uses the knowledge obtained after 550 trials

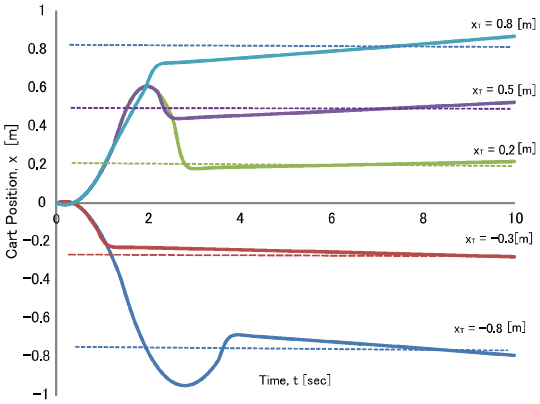

Fig. 13 Cart trajectory during an operation that uses the knowledge obtained after 550 trials

error margin around ± 0.1[m].

The details of the successful operation can be seen in Table 4. Based on Table 4, for each cart position, particular acceleration angle θ_T^1 and deceleration angle θ_T^2 was selected to complete the operation at certain amount of output V_{out}.

The target angles θ_T^1 and θ_T^2 that were selected during the operation shows a certain pattern. Acceleration angle θ_T^1 is facing the direction of the target position x_T 's direction. However, deceleration angle are either facing the opposite of the target position x_T 's direction or zero. This shows that the system learns that deceleration angle θ_T^2 is selected to decelerate for attempting to stop at the target position x_T. Therefore, any target angle, θ_T that can helps decelerate the cart movement is relevant to be selected.

The total output V_{out} varies according to the target angle θ_T depending on necessity upon achieving the target position x_T.

Based on the result, the system was able to learn the best combination of target angle θ_T to produce a rapid

Table 4 Time required to complete a position control during a successful operation

Target Position, x_T	-0.8	-0.3	0.2	0.5	0.8
Acceleration angle,θ_T^1 [rad]	-0.05	-0.05	0.05	0.05	0.05
Deceleration angle,θ_T^2 [rad]	0.05	0.05	-0.1	-0.1	0.0
Acceleration output,V_{out} [rad]	100	80	80	80	40
Deceleration output,V_{out} [rad]	80	40	60	20	100
Time until achieved stabilization, t[sec]	3.8	1.2	2.8	2.5	2.3

position control toward target position. The rapid position control can be seen from the usage of target angle θ_T to produce high acceleration for particular target position x_T. For example, further target position would require more acceleration for position transition at a shorter time; hence bigger target angle is required. Therefore, it is understood that the system are able to perform a rapid position control.

6. CONCLUSION

In this paper, we developed a control system that can learn knowledge for performing rapid position control for aerial hovering vehicles. The system learns to acquire the knowledge by repeatedly updating the knowledge after assembling and evaluating the required configuration of target angle for acceleration and deceleration based on an assigned target position.

This system was applied in a cart-pendulum inverted control that emulates the position control using target angle by an aerial hovering vehicle. The result shows that the system was able to achieve the target position after several trials. For each target position, the configuration of target angle is different in order to have shorter operation time. The result also shows that the transition of positions finishes at the same time with each transition process requires different configuration of target angle. This shows that the system were able to perform rapid position control by configuring the angular orientation to achieve the assigned target position.

From the simulation on the inverted pendulum system, the proposed learning control system is capable on learning the knowledge needed for performing rapid position control according to assigned target position. This system will be applied on a rotorcraft UAV in future works.

REFERENCES

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

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

[3] Syafiq F. K., T. Shibuya and S. Yasunobu, "A Substitute Target Learning-based Inverted Pendulum Swing-up Control System", *Joint 5th SCIS & 11th ISIS International Conference*, SA-D3-3, pp.1-6, 2010.

[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.