

Development of Image-Based Visual Servoing for Tracking by a Quadrotor

Shabnam Mostafa Gharebaghi¹, Abdol Majid Khoshnood²

¹PhD candidate, Department of Aerospace Engineering, K. N. Toosi University of Technology, Tehran, Iran.

²Assoc. Prof., Department of Aerospace Engineering, K. N. Toosi University of Technology, Tehran, Iran.

*Corresponding author: khoshnood@kntu.ac.ir

Received: 08/06/2025 Revised: 12/28/2025 Accepted: 03/03/2026

Abstract

Target tracking is considered a key and advanced application of Unmanned Aerial Vehicles (UAVs). In the past, this task was primarily carried out using radar systems, which, due to their limited flexibility, could only cover a specific area. However, advancements in image processing and machine vision technologies have enabled the use of lighter and smaller cameras that can be easily mounted on UAVs. The primary objective of this paper is to develop an Image-Based Visual Servoing (IBVS) method for tracking stationary targets using a quadrotor. This method utilizes visual feedback to generate appropriate control inputs to maintain the target position at the center of the drone's field of view. In conventional IBVS approaches, the optimal design and tuning of controller gains to ensure satisfactory performance under various conditions has always been a fundamental challenge. The main innovation of this research is the application of the Q-learning reinforcement learning algorithm for the optimal adjustment of these gains.

Keywords: Visual tracking, Fixed target, image based visual servoing, quadrotor.

1. Introduction

Quadrotors have become increasingly popular tools for a wide range of applications due to their autonomous capabilities, high maneuverability, and adaptability to diverse conditions. Among the most significant of these applications is target tracking, which is employed in both military and civilian domains.

In this context, visual signals offer several advantages over other tracking methods. Compared to techniques requiring multiple and costly sensors, vision-based technologies enable continuous target detection and tracking by processing video frames in real time using lightweight computers, less space, and lower energy consumption. Ultimately, with recent advancements in image processing and machine learning, vision-based approaches have gained particular importance in the design of tracking systems.

There are two principal approaches in vision-based control: (1) Position-Based Visual Servoing (PBVS), which reconstructs the target position from visual information for use in control laws, and (2) Image-Based Visual Servoing (IBVS), where control computations are performed directly in the image plane [1]. Achieving precise estimates of reconstruction parameters—such as depth, object model, intrinsic parameters, and 3D geometry—in PBVS can be challenging [2,3]. Consequently, many researchers prefer IBVS methods due to their relatively simple control formulation and robustness against camera calibration errors and depth estimation uncertainty [4].

The IBVS approach utilizes image features to finely

adjust the quadrotor's position and distance relative to the target. Such technologies ultimately enhance UAV autonomy and enable more complex and sensitive mission execution. Image-based control is a technique that uses visual feedback from a camera to guide a robot toward a desired position. This method proves highly effective in applications requiring precise positioning and orientation, such as robotic manipulation and inspection.

The main objective of this paper is to develop a vision-based control method for tracking static targets using a quadrotor, in which the controller gains are optimized by a Q-learning algorithm. Accordingly, the dynamic model of the quadrotor is first introduced, followed by the fundamentals of the IBVS method, and finally, the analysis results are presented.

2. Dynamics

For the simulation of the quadrotor, the reference equations from [5] have been utilized.

$$\begin{bmatrix} \ddot{x} \\ \ddot{y} \\ \ddot{z} \end{bmatrix} = -g \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + T/m \begin{bmatrix} S_\theta C_\phi C_\psi + S_\phi S_\psi \\ S_\theta C_\phi S_\psi - S_\phi C_\psi \\ C_\theta C_\phi \end{bmatrix} - \frac{1}{m} \begin{bmatrix} A_x & 0 & 0 \\ 0 & A_y & 0 \\ 0 & 0 & A_z \end{bmatrix} \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{bmatrix} \quad (1)$$

$$\begin{bmatrix} \dot{p} \\ \dot{q} \\ \dot{r} \end{bmatrix} = \begin{bmatrix} (I_{yy} - I_{zz})qr/I_{xx} \\ (I_{zz} - I_{xx})pr/I_{yy} \\ (I_{xx} - I_{yy})pq/I_{zz} \end{bmatrix} - I_r \begin{bmatrix} q/I_{xx} \\ -p/I_{yy} \\ 0 \end{bmatrix} + \begin{bmatrix} \tau_\phi/I_{xx} \\ \tau_\theta/I_{yy} \\ \tau_\psi/I_{zz} \end{bmatrix} \quad (2)$$

$$\begin{bmatrix} \dot{\phi} \\ \dot{\theta} \\ \dot{\psi} \end{bmatrix} = \begin{bmatrix} 1 & S_\theta T_\theta & C_\theta T_\theta \\ 0 & C_\theta & -S_\theta \\ 0 & S_\theta/C_\theta & C_\theta/C_\theta \end{bmatrix} \begin{bmatrix} p \\ q \\ r \end{bmatrix} \quad (3)$$

Ax,y,z represent the drag coefficients, and wr is the summation of the quadcopter's angular velocities.

3. Vision-based Control

Vision-based control is an image feedback-based control framework encompassing both Image-Based Visual Servoing (IBVS) and Position-Based Visual Servoing (PBVS). This research utilizes the IBVS approach, where the two-dimensional target image information is directly incorporated into the closed-loop control system.

The objective of all vision-based control schemes is to minimize the error $e(t)$, which is typically defined as follows:

$$e(t) = s(m(t), a) - s^* \quad (4)$$

The parameters in the above relationship are defined as follows. The vector $m(t)$ represents the set of image coordinates of the target points. The set a comprises parameters that indicate potential additional information about the system (such as approximate intrinsic camera parameters or 3D object models). The vector s^* includes the desired values for the features.

Once s is selected, the design of the control system becomes relatively straightforward. One direct approach is designing a velocity controller. For this purpose, a relationship between the time derivative of s and the camera velocity is required. The camera velocity is denoted by $V_c = (v_c, \omega_c)$, where v_c is the linear velocity of the camera center and ω_c is the angular velocity of the camera frame.

$$\dot{s} = L_s V_c \quad (5)$$

The matrix L_s has dimensions $k \times 6$ and is known as the interaction matrix or feature Jacobian, where k represents the number of feature points. By utilizing relationships (4) and (5), the relationship between the camera velocity and the temporal variation of the error is immediately derived:

$$\dot{e} = L_s V_c \quad (6)$$

Considering V_c as the input to the robot controller, if the goal is the exponential and independent reduction of the error, in other words, $\dot{e} = -\lambda e$, we conclude:

$$V_c = -\lambda L_s^+ e \quad (7)$$

In (7), L_s^+ is a $6 \times k$ matrix that acts as the Moore-Penrose pseudoinverse of the matrix L_s .

4. Results and Discussion

For the implementation of the vision-based control approach, at least four marked feature points are required. Subsequently, based on the aforementioned methodology, the simulation is carried out in the Simulink environment, and the obtained results are presented in the following section.

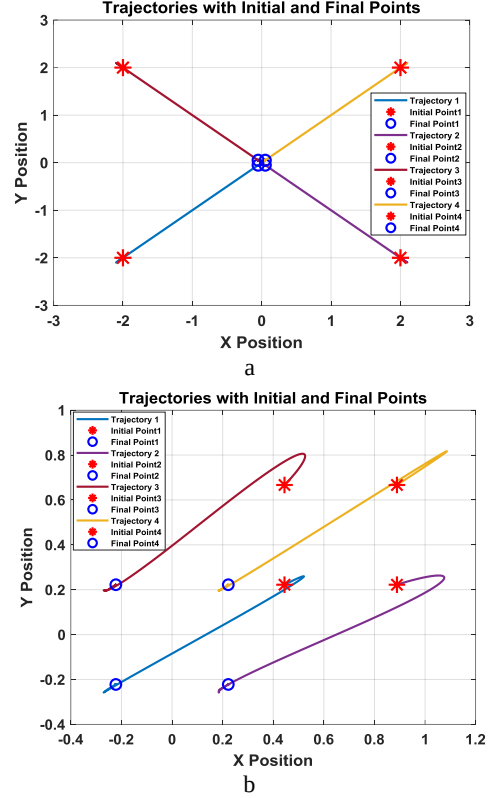


Figure 1. displacement of target points in image plane a) horizontal motion b) vertical motion

In Figure 1 (a) it is observed that the quadrotor is increasing its altitude to reach the reference point. Moreover, Figure (b) illustrates the displacement of the marked feature points during the horizontal motion.

5. Conclusions

In this study, target tracking technology is examined as one of the fundamental and vital technologies whose significance in practical applications is steadily increasing. Traditionally, this process relied primarily on radar-based systems, which, due to inherent limitations such as low flexibility, were only capable of covering specific areas.

In this paper, a novel target tracking approach based on vision-based control is proposed. In the proposed method, the pixel displacement is considered as the input to the controller, which is subsequently converted into the desired velocity and ultimately into the target position and orientation. The controller parameters are tuned using the Q-learning reinforcement learning algorithm.

According to the obtained results, the designed controller demonstrates satisfactory performance and is capable of consistently maintaining the target within the camera's field of view. However, one of the main limitations of this research lies in the system's sensitivity to visual disturbances; that is, when obstacles appear in the camera's field of view or the target temporarily disappears, the tracking process fails.

Despite this limitation, the proposed approach

exhibits strong potential for further development and improvement in future research.

6. References

- [1] Janabi-Sharifi, F., Deng, L. and Wilson, W.J., 2010. Comparison of basic visual servoing methods. *IEEE/ASME Transactions on Mechatronics*, 16(5), pp.967-983.
- [2] Wu, A.D., Johnson, E.N. and Proctor, A.A., 2005. Vision-aided inertial navigation for flight control. *Journal of Aerospace Computing, Information, and Communication*, 2(9), pp.348-360.
- [3] Metni, N., Hamel, T. and Derkx, F., 2005, December. Visual tracking control of aerial robotic systems with adaptive depth estimation. In *Proceedings of the 44th IEEE Conference on Decision and Control* (pp. 6078-6084). IEEE.
- [4] Hamel, T. and Mahony, R., 2002. Visual servoing of an under-actuated dynamic rigid-body system: an image-based approach. *IEEE Transactions on Robotics and Automation*, 18(2), pp.187-198.
- [5] Luukkonen, T., 2011. Modelling and control of quadcopter. Independent research project in applied mathematics, *Espoo*, 22(22), pp.1-24.