Archives available a t journals.mriindia.com

**International Journal on Advanced Computer Engineering and Communication Technology**

ISSN: 2278-5140
Volume 14 Issue 01, 2025

# Gimbal Control for Vision-based Target Tracking

Prof. Jayshree Borkar[1], Mr. Yatharth Maharwade[2], Mr. Suyesh Bansod[3], Mr. Rajat Kathane[4], Mr. Sahil Kawale[5]
[1] *Assistant Professor Suryodaya college of Engineering and Technology/ Computer Engineering, Nagpur, India*
[2-3] *Suryodaya college of engineering/Computer Enginerring, Nagpur, India*
[1]*jayuborkar.coet@gmail.com,* [2] *yathhu2004@gmail.com,* [3]*bansodsuyesh@gmail.com,*
[4]*rajatkathane27@gmail.com,* [5]*sahilkawale13@gmail.com*

**Peer Review Information**

*Submission: 05 Feb 2025*
*Revision: 17 Mar 2025*
*Acceptance: 18 April 2025*

**Keywords**

*Gimbal Control*
*Vision-Based Tracking*
*Target Tracking*
*Electro-Optical Tracking*

**Abstract**

This paper addresses the problem of controlling the orientation of a 3-axis gimbal that is carrying a cinematography camera, using image measurements for feedback. The control objective is to keep a moving target of interest at the center of the image plane. A Region-of-Interest (ROI) that encloses the target's image is generated through the combination of a visual object detector and a visual object tracker based on Convolutional Neural Networks. The proposed system uses a fast, computationally efficient CNN-based object detector and tracker to generate a Region-of-Interest (ROI) around a moving target, computes the attitude error as a rotation matrix, and employs an attitude controller to ensure the target remains centered in the image plane, with experimental validation demonstrated using a human face.

## INTRODUCTION

Gimbals are widely utilized in aerial vehicles to stabilize cameras and minimize vibrations caused by angular motion and external disturbances through a process known as inertial stabilization [1]–[3]. Beyond stabilization, gimbals play a crucial role in expanding the camera's Field of View (FOV) by enabling angular adjustments that can either counteract or complement the vehicle's translational motion to achieve specific filming objectives. This capability becomes particularly essential when tracking moving targets, where maintaining a consistent and accurate view is critical [4].

Autonomous target tracking using a gimbal-camera system is a fundamental feature for various vision-enabled robotic applications, such as autonomous cinematography and intelligent shooting systems [5]–[8]. Achieving precise target tracking requires two primary tasks: visual object detection and tracking, and gimbal control. The first task involves obtaining image measurements of the target, which serve as inputs for the second task — controlling the gimbal to ensure the camera's optical axis remains directed towards the target.

## LITERATURE SURVEY

[2] Prashant Kumar1, Sarvesh Sonkar2 (2020, June).[2] intro Visual surveillance has a wide field of applicability, such as mapping, surveillance, object identification, rescue missions, tracking etc. Targets are detected using some characteristics features size, shape or color, the character of movement using a set of algorithms on a live video feed. Small aerial vehicles provide an excellent platform for different payloads like offer electro-optical (EO) and infrared (IR) cameras. Most of the small and miniature air vehicles presently in operation are

equipped with an EO or IR camera. . [1]. Assembling is supposedly stiff, so there is no vibrant connection between the camera's position and size and the UAV position and direction. The camera head steering algorithm is focused on the geometry of camera relations and tracked object locations. Similarly, in [2] [3]Liu, X., Yang, Y., Ma, C., Li, J., & Zhang, S. (2020). Because of the COVID-19 pandemic, many people are prohibited to go out for exercise. Exercising or working out at home thus becomes popular and important. For working out or fitness, people used to go to a gymnasium and got advice from a fitness coach, in order to make fitness more effective and avoid injury. Therefore, working out at home without professional advice but with the risk of injury becomes a problem. A real-time fitness posture evaluation system is thus demanded. In this work, we propose a posture evaluation system specially for fitness videos. Two common fitness movements are studied, i.e., Dumbbell Lateral Raise and Biceps Curl.

[4] P. Nousi, E. Patsiouras, A. Tefas, and I. Pitas (2018)"Convolutional Neural Networks for Visual Information Analysis With Limited Computing Resources," it's essential to understand the context of the challenges and goals that drive this area of research. Here's a breakdown of the typical introduction to such a topic. Convolutional Neural Networks (CNNs**)** CNNs have revolutionized visual information analysis, achieving state-of-the-art results in

[7]

tasks like image classification, object detection, and image segmentation.Their ability to automatically learn hierarchical features from images makes them incredibly powerful.

[5] Bhavani, K., Dhanaraj, V., & Siddesh, N. V. (2017).Face detection can be defined as a creative process that deals with scanning an object’s image and taking the resultant datasets. It takes a high level of coding using an algorithm that It requires a high level of coding using an algorithm that can detect moving images from a running video stream and capture the different poses that form it’s dataset[7]. It is probably one of the most popular areas of research in image processing and has a wide range of real-world applications including surveillance, access control, identity authentication, and photo based image detection and recognition

[6] O. Zachariadis, V. Mygdalis, I. Mademlis, N. Nikolaidis, and In the (2017) . Unmanned Aerial Vehicles (“UAVs”, or “drones”), are a recent addition to the cinematographer’s arsenal. By exploiting their agility and ability to fly, drones are potentially able to capture video streams that would be impossible otherwise.One of the most demanding media production applications is the aerial coverage of live outdoor (e.g., sports) events, sinceit mainly involves filming multiple moving targets (e.g., athletes, boats, cars etc.). The drones must capture stable, noncluttered and visually pleasing AV streams, which requiresdemanding drone piloting skills or impressive drone intelligence.

**LITERATURE SURVEY TABLE**

| **Author** | **Title** | **Year** | **Key Focus** | **Methodology** | **Results/Findings** |
|---|---|---|---|---|---|
| [2]Prashant Kumar1 , Sarvesh Sonkar2 | Real-time visionbased tracking of a moving terrain target from Light Weight Fixed Wing UAV using gimbal control | 2020 | Real-time Object Tracking and Location from a Fixed-Wing UAV | The algorithm focuses on controlling the camera's azimuth and elevation angles to keep the target within the field of view.. | The simulation and implementation results demonstrate the applicability of the proposed gimbal control algorithm for object tracking from a fixedwing UAV. |
| [3]Liu, X., Yang, Y., Ma, C., Li, J., & Zhang, S. | A Posture Evaluation System for Fitness Videos based on Recurrent Neural Network | 2020 | The primary focus is developing a system to evaluate the correctness of fitness postures in videos, especially for home workouts | The video is segmented into clips, with each clip containing a single exercise movement, based on the evolution of joint positions. | The LSTM-based method outperformed baseline methods, demonstrating the effectiveness of considering the temporal evolution of joints. |
| [4] P. Nousi, E. | Convolutional | 2018 | The primary focus is on developing and | Reducing the size of | The implied result is that by the research conducted, |

| | | | | | |
|---|---|---|---|---|---|
| Patsiouras, A. Tefas , and I. Pitas | Neural Networks for Visual Information Analysis With Limited Computing Resources | | utilizing CNNs for visual information analysis in situations where computing resources are limited. | CNN models through techniques like pruning, quantization, and knowledge distillation. | that it would be possible to run complex CNNs on devices that have limited computing resources. |
| [5] Bhavani, K., Dhanaraj, V., & Siddesh, N. V. | Real time Face Detection and Recognition in Video Surveillance | 2017 | the potential of extending the system for automated attendance tracking and person location tracking. | the system uses an algorithm to detect faces from image data, including moving images from video streams.It captures different poses of faces to build a dataset. | The paper claims that the proposed approach can nearly double the recognition rate while halving the computational runtime compared to existing methods. |
| [6] O. Zachariadis, V. Mygdalis, I. Mademlis, N. Nikolaidis | 2D VISUAL TRACKING FOR SPORTS UAV CINEMATOGRAPHY APPLICATIONS | 2017 | the paper is dedicated to analyzing and optimizing 2D visual tracking methods for UAVbased target | 2D coordinates of the tracked target are used to estimate the relative 3D position. | The research identifies an effective 2D visual tracking implementation for UAVs, combining findings from multiple methods. |

**OBJECTIVE**

The objective of this system is to develop a robust and efficient vision-based target tracking framework using a 3-axis gimbal control system. The system aims to achieve the following specific goals:

1. **Accurate Target Detection:** Implement deep learning-based object detection algorithms, such as SSD and YOLO, to accurately detect and identify moving targets within the camera's Field of View (FOV), even under varying lighting conditions and diverse orientations.
2. **Real-Time Target Tracking:** Utilize advanced visual tracking algorithms, including correlation filters and CNN-based trackers like SiamFC, to ensure continuous and precise tracking of the target, regardless of changes in position, scale, or partial occlusions.
3. **Attitude Control on SO(3):** Formulate an attitude tracking controller based on rotation matrices within the Special Orthogonal Group SO(3), allowing effective computation of orientation errors and precise control of the gimbal's orientation.
4. **Elimination of Depth Estimation:** Design a control system that constructs error rotation matrices directly from image measurements, thereby eliminating the need for depth estimation and enhancing the system's robustness and simplicity.
5. **Maintaining Horizontal Alignment:** Integrate accelerometer measurements to ensure that the camera maintains horizontal alignment during tracking, thereby improving stability and reliability in dynamic environments.
6. **Integration with Gimbal Control System:** Develop a control mechanism that generates appropriate angular rate commands to adjust the gimbal orientation, ensuring the target remains centered in the camera's image plane.
7. **System Evaluation and Optimization:** Test the proposed system's performance under various scenarios involving dynamic targets and evaluate its robustness, accuracy, and realtime capabilities, followed by optimization of the control algorithms and hardware integration.

**PROPOSED WORK**

The proposed work aims to design and implement a real-time vision-based target tracking system using a 3-axis gimbal control mechanism. The objective is to maintain the target of interest centered in the camera's image plane by employing a robust attitude tracking controller on the Special Orthogonal Group SO(3). The proposed framework is divided into the following key components:

**1. Visual Object Detection and Tracking:** The visual processing module will combine a deep learning-based object detector with a visual tracker to accurately detect and track moving targets in real-time. Object Detection Single Shot Detector (SSD) with MobileNet v1 as the base feature extractor will be employed due to its efficiency in providing a favorable balance between speed and accuracy, especially when operating on low-power devices like NVIDIA Jetson Tegra X2. Object Tracking a lightweight version of the fully convolutional Siamese network (SiamFC) will be developed, optimized using a depth factor to enhance processing speed while maintaining acceptable tracking precision.

**2. Gimbal Control System Design:** The control mechanism of the gimbal is divided into an inner loop and an outer loop. Inner Loop (Low-Level Control) a high-bandwidth control system responsible for rejecting disturbances and tracking angular rate commands derived from rate gyroscope measurements. Outer Loop (High-Level Control) a vision-based control system designed to maintain the target's position at the center of the camera's image plane. This is achieved by generating angular rate commands from image measurements and feeding them to the inner loop.

**3. Attitude Control on SO(3):**The core of the proposed control strategy involves formulating the gimbal control problem as an attitude tracking problem on the Special Orthogonal Group SO(3).
The orientation of the gimbal is represented by a rotation matrix RC∈SO(3)R_C \in SO(3)RC ∈SO(3), which describes the transformation from the camera frame to the world frame. A desired rotation matrix RC∗R_C^*RC∗ is defined based on the relative position between the target and the camera.

**4. Error Matrix Construction Without Depth Estimation:** Unlike traditional methods that require depth estimation, the proposed approach constructs the error matrix ReR_eRe directly from image measurements. This is achieved by computing the relative position between the target and camera using intrinsic camera parameters and accelerometer data.

**5. Horizontal Alignment:** To ensure horizontal alignment, accelerometer measurements are incorporated to maintain the camera's orientation concerning the gravitational vector, enhancing stability and robustness.

**6. System Evaluation and Testing:** The proposed system will be evaluated through experimental testing involving real-time tracking of dynamic targets (e.g., human faces). The performance of the controller will be assessed by Measuring the accuracy of target detection and tracking.Evaluating the robustness of the control system under various conditions (e.g., different lighting, occlusions, and camera motion). Comparing the proposed control law's performance with traditional methods that rely on depth estimation.

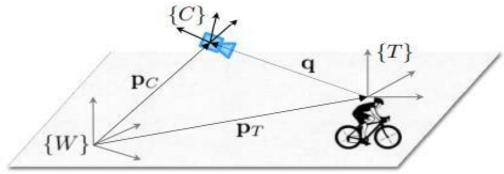


*Figure1. Problem setup and notation*

**METHODOLOGY**

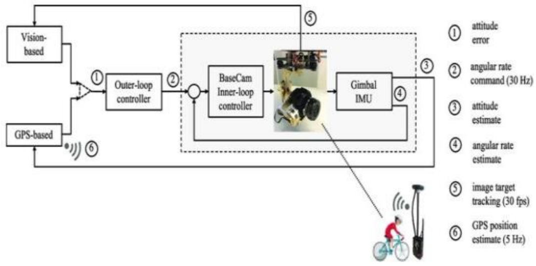


*Figure2. Diagram showing the proposed methodology for Gimbal Control.*

1. **Data Collection:** Capture real-time video data using a camera mounted on a 3-axis gimbal.
   Record video sequences involving dynamic targets, such as human faces or other moving objects, under various conditions (e.g., different lighting, angles, and speeds). Store the recorded data with appropriate labels for effective processing and analysis.

2. **Visual Object Detection:** Implement Single Shot Detector (SSD). Use SSD with MobileNet v1 as the base feature extractor to detect targets within the camera's Field of View (FOV). Pretrain the model on large annotated datasets like COCO and fine-tune it to accurately identify specific targets. Detection Refinemen continuously refine the detector by adjusting network parameters to improve detection accuracy and speed.

3. **Object Tracking:** Implement SiamFC Lite Tracker deploy a lightweight version of the fully convolutional Siamese network (SiamFC) to perform real-time tracking of detected targets. Apply a depth factor to reduce computational complexity and enhance processing speed without compromising accuracy. Tracking Initialization and Update. Initialize the tracker using bounding boxes generated by the SSD detector. Continuously update the tracker with new image frames to maintain accurate tracking.

4. **Error Calculation (Rotation Matrix Construction):** Formulate Attitude Control on SO(3) Represent the gimbal's orientation as a rotation matrix RC∈SO(3)R_C \in SO(3)RC∈SO(3) describing the transformation from the camera frame to the world frame. Define Desired Orientation construct the desired rotation matrix compute Orientation Error

5. **Control Law Design (Gimbal Control System):** Inner-Loop Control (Low-Level Control. Receive angular rate commands from the gyroscope and reject disturbances to maintain stability.Outer-Loop Control (High-Level Control). Generate angular rate commands based on image measurements to ensure accurate pointing and tracking of the target.

6. **Horizontal Alignment Control:** Utilize accelerometer measurements from the gimbal's Inertial Measurement Unit (IMU) to maintain horizontal alignment. Compensate for gravitational effects to ensure the camera's orientation remains stable and aligned with the target.

7. **System Evaluation and Testing:** Test the system's performance through real-time experiments involving dynamic targets.Evaluate Performance Metrics. Accuracy: The ability to keep the target centered in the image plane. Robustness: The system's resilience to varying conditions, including occlusions and rapid movements. Response Time: The speed at which the system reacts to changes in the target's position. Compare the proposed approach with conventional methods relying on depth estimation for gimbal control.

**RESULT**

*Figure2. Gimbal Control for Vision-based Target Tracking*

Our approach gives a good overall recognition rate, the recognition rate entirely depends on the camera resolution. We did our experiments using the integrated web-cam in the laptop. For a dataset of 30 images of a person, the recognition rate is around 75-80% for the captured frontal face image orientation. If there exists is a different orientation other than the image stored in the dataset, the face is not recognized. Fig3. Recognized Faces in the live stream Fig.3 shows how the faces get recognized from a live video stream. It encircles the recognized faces within rectangular frames along with the values assigned to them.

**CONCLUSION**

In this paper, we addressed the problem of controlling the orientation a gimbal-mounted camera to point at a target of interest. The proposed solution combines a fast and reliable deep learning visual object detector and tracker, suited for low computational power implementation, with an attitude controller that is based on image accelerometer measurements and guarantees convergence of the target image to origin of the image plane. Experimental results have shown the effectiveness of the proposed solution, involving human face detection and tracking. Future work will focus on more dynamic scenarios, involving different targets and more aggressive camera and target motions.

**References**


Kumar, Shivaji & Kumar, Sanket & poddar, parag. (2022). Face Detection and Recognition in live video - Technical Report.

Kumar, P., Sonkar, S., Ghosh, A. K., & Philip, D. (2020, June). Real-time vision-based tracking of a moving terrain target from Light Weight Fixed Wing UAV using gimbal control. In *2020 7th international conference on control, decision and information technologies (CoDIT)* (Vol. 1, pp. 154-159). IEEE.

Liu, X., Yang, Y., Ma, C., Li, J., & Zhang, S. (2020). Real-time visual tracking of moving targets using a low-cost unmanned aerial vehicle with a 3-axis stabilized gimbal system. *Applied Sciences*, *10*(15), 5064.

P. Nousi, E. Patsiouras, A. Tefas, and I. Pitas, “Convolutional neural networks for visual information analysis with limited computing resources,” in Proceedings of the IEEE International Conference on Image Processing (ICIP), 2018.

Bhavani, K., Dhanaraj, V., & Siddesh, N. V. (2017). Real time face detection and recognition in video surveillance. *Int. Res. J. Eng. Technol*, *4*(6), 1562-1565.

O. Zachariadis, V. Mygdalis, I. Mademlis, N. Nikolaidis, and In the 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Pitas presented research on 2D visual tracking techniques designed for UAV cinematography in sports applications.

P. Casau, R. Cunha, R. G. Sanfelice, and C. Silvestre, “Hybrid feedback for global asymptotic stabilization on a compact manifold,” in 2017 IEEE 56th Annual Conference on Decision and Control (CDC), Dec. 2017, pp. 2384–2389.

J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, and Guadarrama's work, presented at the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), explores the balance between speed and accuracy in modern convolutional object detectors.

L. Bertinetto, J. Valmadre, J.F. Henriques, A. Vedaldi, and P.H.S. Torr, “Fully-convolutional siamese networks for object tracking,” in Proceedings of the European Conference on Computer Vision (ECCV). Springer, 2016, pp. 850–865