iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: V Month of publication: May 2025

DOI: https://doi.org/10.22214/ijraset.2025.70478

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

# Artificial Intelligence Based Realtime Face Tracking Gimbal for Live Video Session

Samiksha Kale[1], Mrs.M.M.Gudadhe[2], Sejal Ghagare[3], Samiksha Jungari[4], Ishika Bhalavi[5], Vartika Wasekar[6]
*Department of Information Technology, Rashtrasant Tukadoji Maharaj Nagpur University*

***Abstract: The Real-Time Face Tracking Gimbal for Live Video Session is a dynamic system that enhances video streaming by automatically tracking and cantering a subject's face in the camera frame. Utilizing advanced algorithms for face detection and tracking, this system ensures smooth camera movements across pan, tilt, and roll axes. It leverages a combination of image processing and embedded systems to achieve real-time performance. The gimbal's camera captures video, processed by a microcontroller integrated with a pre-trained model, which detects the subject's face and calculates its position within the frame. Any deviation prompts the microcontroller to adjust the gimbal's motors, ensuring the subject remains cantered. This closed-loop feedback system makes it ideal for live streaming, vlogging, and professional videography. The system's design focuses on portability, scalability, and ease of use, incorporating hardware such as servo motors, motor drivers, and a USB (Universal Serial Bus) camera. It also supports software tools for live streaming and algorithm customization. Applications span entertainment, education, and virtual meetings. By automating face tracking, the gimbal reduces manual intervention, improves video quality, and ensures a seamless viewing experience for audiences. This design reduces the need for manual camera handling, offering a more professional and user-friendly experience. Its compact, portable structure makes it suitable for live streaming, vlogging, video conferencing, and surveillance applications. Furthermore, the system is highly scalable and cost-effective, allowing for integration into various industries requiring real-time tracking.***

## I. INTRODUCTION

In today's digital age, the quality of video content plays a pivotal role in how effective information is communicated, especially in educational and live streaming contexts. As more educators, trainers, and content creators leverage video to engage audiences, the demand for seamless and professional video presentations has never been higher. However, achieving this level of quality often presents challenges, particularly when it comes to camera stabilization and the space, maintaining proper framing and focus becomes a significant hurdle. As a result, both the content quality and viewer engagement may suffer, limiting the potential impact of the educational material being presented. To address these issues, this project proposes the development of an innovative system that integrates artificial intelligence (AI) with gimbal technology to automate camera adjustments.

The integration of AI enables real-time processing and decision-making, allowing the system to adapt swiftly to various movements and angles. This results in a smoother viewing experience, eliminating the interruptions caused by manual gimbal adjustments. Moreover, the automated system frees trainers to focus on delivering high-quality content without the distraction of managing the camera setup. Ultimately, this project aims to elevate the standards of video production in educational and live streaming environments. By ensuring that the trainer is always perfectly framed, the proposed system enhances audience engagement and comprehension, leading to more effective learning outcomes. In a world where digital content is paramount, leveraging technology to streamline the presentation process promises to make a meaningful difference in how knowledge is shared and absorbed.

The system's hardware includes servo motors, motor drivers, and an Arduino Uno microcontroller, which plays a pivotal role in coordinating motor movements. Using stepper motors controlled via the Arduino Uno, the gimbal achieves precise adjustments that align the camera dynamically with the subject's movements. The Arduino IDE (Integrated Development Environment) serves as the programming environment, enabling customization of motor control algorithms and seamless integration with the face-tracking software. This design ensures that the system is not only efficient but also accessible for modifications, making it suitable for a wide range of applications. By automating face tracking, this gimbal system significantly reduces the need for manual intervention, allowing users to focus entirely on content delivery. Its compact and portable design makes it ideal for live streaming, vlogging, video conferencing, and even surveillance applications.

## II. LITERATURE REVIEW

Recent advancements in object and face tracking technologies have significantly improved Unmanned Aerial Vehicle (UAV) and robotic systems' performance, particularly through the integration of 3-axis gimbals and real-time visual tracking. Several approaches utilize active tracking mechanisms to reduce motion blur, geometric distortions, and re-projection errors during fast UAV movements, enhancing pose estimation accuracy. Real-time tracking systems have been validated through simulations for both 3D(Three-Dimensions) and 6D(Six-Dimensions) object tracking, showing promising results for improved stability and target localization. Many studies incorporate deep learning-based detectors and visual tracking algorithms to maintain targets within the camera's field of view, even in dynamic scenarios.

Researchers have developed gimbal control strategies that rely on image measurements and sensor fusion, using GPS (Global Positioning System) and IMU (Inertial Measurement Unit) data for accurate camera alignment and pointing. While fixed-wing UAVs face challenges like video transmission lag and vibration, algorithms are being optimized to mitigate these effects. Several systems also leverage classic techniques like Haar cascades and the KLT(Kanade-Lucas-Tomasi) algorithm for human face detection and tracking in real time. These methods demonstrate robust performance across various environments and user conditions, though challenges persist under harsher lighting or with occlusions.

In video surveillance and human-robot interaction domains, combining facial recognition with alignment and sparse representation methods has enhanced detection accuracy, even under variable expressions and orientations. The increasing size of datasets and training samples has led to improved recognition accuracy, emphasizing the role of data in refining detection systems. Future research directions across studies include improving multi-target tracking, incorporating laser ranging, handling lens distortions, and adopting advanced deep learning algorithms for real-world, complex environments.

## III. OBJECTIVE

The main goal of this project is to create a new AI-based gimbal that overcomes the limitations of mechanical gimbals. This next-generation system will:

Automatically follow subjects: Utilize robust computer vision algorithms to precisely identify and follow human faces or other specified objects in real-time, irrespective of lighting, background detail, or subject motion.

Provide superior image stabilization: Employ sophisticated control algorithms to suppress camera shake and deliver smooth, cinematic footage even in demanding conditions.

Adapt to dynamic scenarios: Continuously adapt and learn to changing conditions, including fluctuations in lighting, occlusion, and fast-moving subjects, to ensure optimal tracking and stabilization performance.

Provide an easy-to-use interface: Offer an intuitive interface for simple operation and adjustment of settings, enabling users of all levels to take professional-quality video.

Be light and portable: Create a lightweight and compact gimbal that is simple to transport and use in a variety of environments, from outdoor excursions to studio shoots.

Integrate advanced features: Investigate the integration of other features like object tracking, face detection, and auto-framing to improve the user experience and enhance the capabilities of the gimbal. With these goals, this project will transform how filmmakers, content creators, and hobbyists get amazing visual content, enabling them to bring their stories to life with unprecedented creativity and accuracy.

## IV. METHEDOLOGY

The system operates by continuously detecting a face in the camera's field of view and keeping it cantered using feedback control. The working principle can be broken down into the following steps:

Step [1]: Face Detection- The first step involves detecting the face in each video frame:

[I]. Use a pre-trained face detection model, such as:

Haar Cascades or HOG (Histogram of Oriented Gradients)-based detectors for traditional approaches.

Deep learning models like YOLO (You Only Look Once) or SSD (Single Shot Detector) for robust and fast detection.

[II]. Extract a bounding box representing the face's location ( x, y, width, height).

Step [2]: Face Tracking- Once the face is detected, tracking ensures smooth transitions between frames:

[I]. Simple tracking uses the centroid of the bounding box as the face's position.

[II]. Advanced tracking may involve Kalman filters or optical flow algorithms to predict the face's next position, improving robustness under temporary occlusions or motion blur.

Step [3]: Position Error Calculation
[I]. The field of view of the camera is partitioned into a grid, with the middle point being the ideal placement for the face.
[II]. The algorithm computes the error between the face's current location and the center of the grid in both horizontal (x-axis) and vertical (y-axis) directions.
Step [4]: Feedback Control
[I]. A Proportional-Integral-Derivative (PID) Controller computes control signals to minimize the position error.
Proportional (P): Reacts to the current error.
Integral (I): Accounts for accumulated past errors.
Derivative (D): Predicts future error based on current trends.
[II].These control signals determine the required adjustments in the gimbal's pan and tilt angles.
Step [5]: Motor Control
The computed adjustments are translated into motor commands.
Brushless DC motors adjust the gimbal's orientation to re-center the face in the frame.
Step [6]: Stabilization
The IMU compensates for unintended camera movements (e.g., hand jitters or external vibrations). Gyroscopic and accelerometric data are used to stabilize the camera dynamically.

## V. APPLICATIONS

Online Learning: Having an AI-powered tracking gimbal in virtual classrooms enables instructors to roam freely while enjoying flawless framing. This increases participation by ensuring the instructor is always within sight, thus helping students pay greater attention and retain information well.
Corporate Training: Organizations can use this technology to carry out remote training sessions that preserve visual attention on trainers. This results in a richer experience, where workers can track with demonstrations or presentations without losing visual contact with the speaker.
Fitness and Wellness Instruction: Fitness trainers can teach classes in which their movements are monitored in real time. This makes it possible for participants to see clearly demonstrations of exercises, enhancing their form and performance while doing the same.
Content Creation: Content producers enjoy high-definition video without the inconvenience of manual adjustments. The AI gimbal provides smooth shots, raising the production level and maintaining viewer interest throughout the content.
Event Coverage: At concerts or theatre performances, keeping stable and well-framed shots improves the watching experience. The audience can become fully engaged in the performance and capture the emotional subtlety and energy.
Broadcast Journalism: Reporters are able to utilize real-time tracking in order to remain framed while they move about dynamic settings. This function ensures continued viewer interest, offering a professional appearance during breaking news coverage.

## VI. LIMITATIONS

Processing Latency: Real-time face detection and gimbal control can experience slight delays due to the high computational demand of AI models, leading to a noticeable lag in fast-paced scenarios.
Battery Consumption: Real-time processing and continuous motor control can quickly drain battery power, limiting usage duration in portable setups.
Overheating Hazards: Ongoing high-load computing has the risk of overheating, especially with smaller hardware.
Lighting Sensitivity: Face detection accuracy can be substantially reduced in low-light or high-glare conditions, impacting tracking performance.

## VII. CONCLUSIONS

The AI-based real-time face tracking gimbal offers a significant improvement in live video quality. Its applications extend beyond education to include areas like intelligent cinematography and robotics. Future work includes optimizing the system for more complex scenarios and integrating additional features like multi-person tracking. This project demonstrates the potential of AI and robotics in transforming live video technology, with broad implications for various industries. The AI-based real-time face tracking gimbal represents a significant advancement in live video quality, addressing the common challenges faced in dynamic recording environments. By leveraging cutting-edge technologies in artificial intelligence and robotics, our system ensures that the trainer's face remains centred and clearly visible, enhancing viewer engagement and experience.

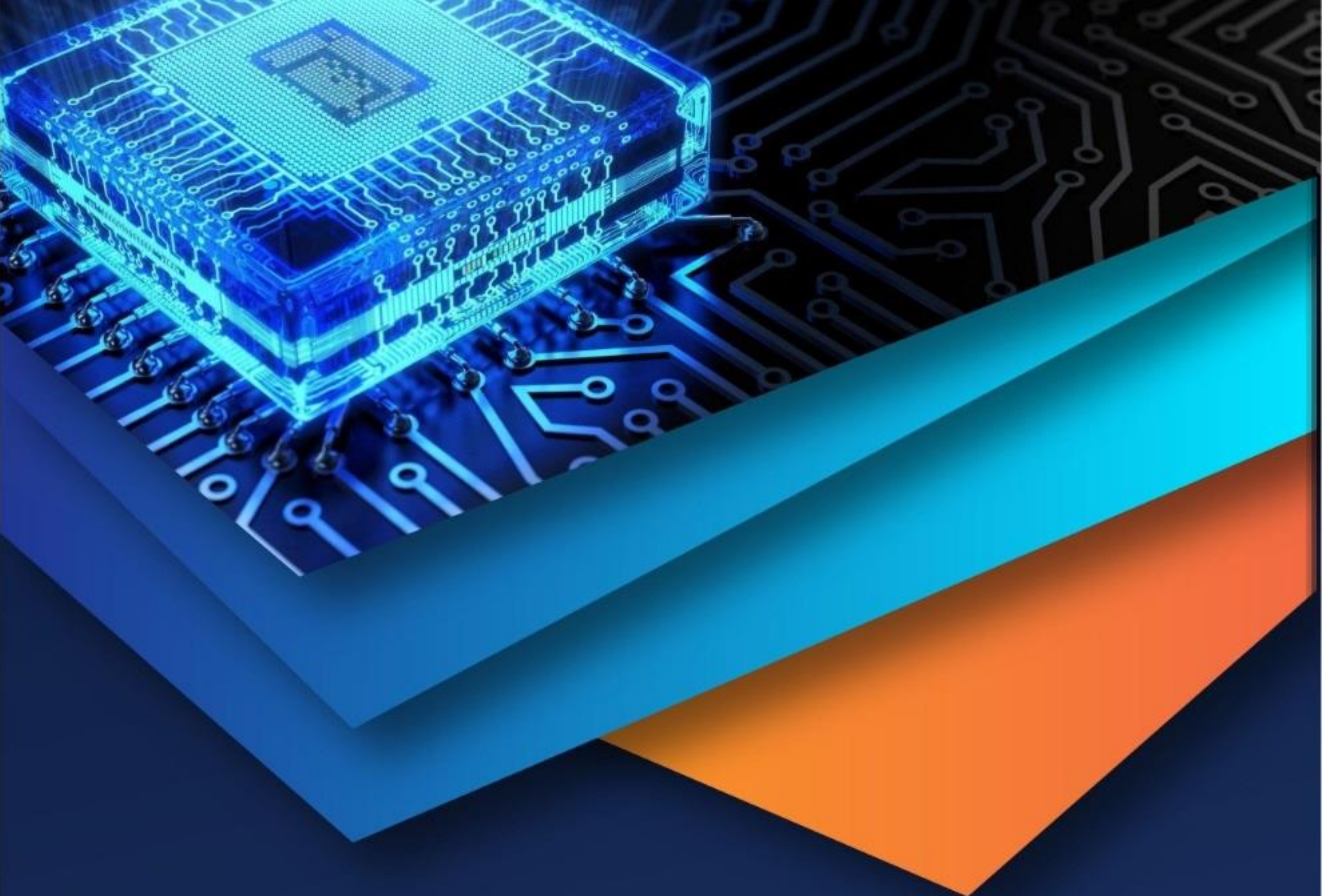
This innovation not only improves the stability and clarity of live streams but also showcases the potential of automated solutions in diverse applications. Beyond the educational sector, the implications of this technology are vast. Intelligent cinematography can benefit from automated face tracking, allowing filmmakers to focus on storytelling rather than technical camera adjustments. Similarly, in robotics, this system can facilitate more intuitive human-robot interactions, where the robot can maintain focus on a specific individual, enhancing communication and functionality.

## VIII. ACKNOWLEDGEMENT

We would like to put on record our sincere gratitude to everyone who assisted with the project. We would like to thank our partners for help and assistance thought out the project. We would also like to acknowledge our professer Dr. M.M. Gudadhe , Department of Information Technology, Priyadarshini College of Engineering, Nagpur, for her strong encouragement, leadership, and help in this work throughout.

## REFERENCES

[1] Hansen, J. G., & de Figueiredo, R. P. (2024). Active Object Detection and Tracking Using Gimbal Mechanisms for Autonomous Drone Applications. Drones, 8(2), 55.
[2] Cunha, R., Malaca, M., Sampaio, V., Guerreiro, B., Nousi, P., Mademlis, I., ... & Pitas, I. (2019, September). Gimbal control for vision-based target tracking. In Proceedings of the European Conference on Signal Processing.
[3] Abdullah, D. B., & Alnuaimy, M. (2022). Real-time Face Tracking for Service-Robot.
[4] 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.
[5] 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.
[6] Kumar, Shivaji & Kumar, Sanket & poddar, parag. (2022). Face Detection and Recognition in live video - Technical Report.
[7] 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.
[8] Mendoza Torres, G., & Ródriguez Gómez, G. (2021). A Fuzzy Control MIMO for Stabilization and attitude of a Cube Sat. Technium: Romanian Journal of Applied Sciences and Technology, 3(1), 12–22. Retrieved from https://techniumscience.com/index.php/technium/article/view/2621 G. B. N. (2020). IRAD
[9] Xavier Fontaine, Radhakrishna Achanta, Sabine Susstrunk. Face recognition in real-world images In Proceedings of IEEE CVPR 97, pages 495–501,2016.
[10] Bozcan, I.; Kayacan, E. Au-air: A multi-modal unmanned aerial vehicle dataset for low altitude traffic surveillance. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 8504–8510




10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429


# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY