

A ROS2-Based Autonomous UAV Software Framework with YOLO-Driven Multi-Sensor Surveillance for Law Enforcement: A Simulation Study

Shivam Kumar^{1,*}, Bhoomi Roy², Mohammad Shahrookh Husain³, Abhijeet Das⁴, Rahul Verma⁵, Digant Mondal⁶

Abstract

The rapid proliferation of drone technology in India and worldwide has presented new opportunities to law enforcement agencies to enhance the methods used to monitor and patrol, and respond to threats to security. This paper introduces a full software system for a ROS2-based autonomous unnamed aerial vehicle (UAV) system for law enforcement surveillance. The system, validated entirely via high-fidelity simulation using Gazebo Classic and PX4 Software-in-the-Loop (SITL), uses a combination of YOLO-based real-time object detection with a three-sensor perception stack of RGB, thermal, and depth cameras. Navigation is implemented with an ORB-SLAM3 based environments. The framework is based on a modular ROS2 node architecture where clean separation between perception, navigation, control, and security modules is possible. Encrypted DDS communication and a human-in-the-loop supervisory interface are part of the design to ensure that the system is accountable and ethical. Simulation results project a mean average precision (mAP@50) greater than 92%, real-time interference greater than 30 FPS on edge-class hardware and patrol area coverage greater than 95% across simulated 15-minute urban missions. The hardware implementation of this framework using physical components of UAV's is considered as future work. This study makes a replicable and open architecture blueprint for next-generation intelligent aerial patrol systems in the Indian public safety scenarios.

Keywords: ROS2, autonomous UAV, YOLO object detection, multi-sensor fusion, SLAM navigation, law enforcement surveillance, simulation framework, edge AI, human-in-the-loop, India drone policy

INTRODUCTION

*Author for Correspondence

Shivam Kumar

E-mail: shivamkumar06ind@gmail.com

^{1,2,4,5,6}Student, Department of Computer Science and Engineering, Greater Noida Institute of Technology (GGSIU), Greater Noida, Uttar Pradesh, India

³Assistant Professor, Department of Computer Science and Engineering, Greater Noida Institute of Technology (GGSIU), Greater Noida, Uttar Pradesh, India

Received Date: June 27, 2026

Accepted Date: July 11, 2026

Published Date: October 9, 2026

Citation: Shivam Kumar, Bhoomi Roy, Mohammad Shahrookh Husain, Abhijeet Das, Rahul Verma, Digant Mondal. A ROS2-Based Autonomous UAV Software Framework with YOLO-Driven Multi-Sensor Surveillance for Law Enforcement: A Simulation Study. International Journal of Robotics and Automation in Mechanics. 2026; 4(2): 1–10p.

In India, law enforcement agencies are under increasing pressure to keep tabs on sprawling urban areas, regulate large public gatherings and respond to security incidents at a pace faster than ever before. Traditional policing methods, such as ground patrols and manually piloted surveillance helicopters, are costly, require a great deal of resources, and are often limited in their ability to cover complex terrain in a short period of time. The Indian government has not gone unnoticed of this gap. Under the Drone Rules, 2021, and the Drone Shakti initiative, the country is aggressively building up its UAV ecosystem, with a market anticipated to reach USD 654 million in 2024 to USD 1,437 million in 2029 with a CAGR of 17% [1].

Indian defense and policing authorities such as Border Security Force, state police departments and DRDO have already started to employ drones for crowd monitoring, border patrol and counter narcotics operations [2].

However, most current deployments of drones still rely on human operators sitting at remote console, manually flying the planes. This results in limits to the duration of missions, human error, and it is almost impossible to provide consistent surveillance across large areas at night or in poor weather. There is a definite need for autonomous drone systems that can take off, navigate, detect any threat, and transmit information back to the command with minimal human input during a routine patrol.

Simultaneously, the worldwide research community has made rapid progress in the tools that must be developed to build such systems. A go-to middleware for autonomous robotic platforms is the Robot Operating System 2 (ROS2) that provides deterministic, quality of service aware communication using its Data Distribution Service (DDS) backbone [3]. YOLO (You Only Look Once) object detection models, and especially the newer YOLOv8 and YOLOv11 models, are currently running on small edge devices at real-time speeds [4].

Multi-sensor fusion of RGB cameras, thermal imagers and depth sensors has demonstrated good results in difficult lighting conditions [5]. And ORB-SLAM3 coupled with visual-inertial odometry now allows reliable localization despite the absence of GPS [6].

This paper combines these advances into a single, consistent framework that has been designed keeping Indian law enforcement use-cases in mind. Our system is not a physical build of hardware at this stage, it's a full-blown software architecture validated in simulation. The hardware prototype is specifically targeted as future work. In simulation we use Gazebo Classic with PX4 SITL, as the middleware layer we use ROS2 Humble, and as the detection engine we use the modified YOLOv11 architecture with dualbackbone multi-sensor fusion. The complete system is tested in a simulated urban setting with moving pedestrians, vehicle, crowd and light.

The major contributions of this work are:

1. A full-fledged, modular ROS2 software architecture for autonomous law enforcement UAV surveillance, verified in simulation.
2. A multi-sensor object detection pipeline with RGB, thermal, and depth data fusion with the help of attention-guided mid-level feature fusion.
3. An ORB-SLAM3 based GPS-independent navigation system with ROS2 NAV2 stack for autonomous urban patrol.
4. Built-in security and ethics layers such as encrypted DDS communication, human supervisory over-ride and threat logging in line with the drone's regulations in India.
5. A well-defined, reproducible simulation methodology that other researchers can work toward physical hardware implementation.

RELATED WORK

UAV Surveillance and autonomous Navigation

The use of drones for surveillance has become an increasing field of research, especially in India. Bhaskar et al. [7] developed a prototype of an autonomous mobile robot with ROS2 and YOLOv8 for ground level surveillance, demonstrating the real-time object detection and last alert-based notification using OpenCV and face recognition. While their work was ground-based, the architecture choices that they made for sensor integration and ROS2 based autonomy have direct application for aerial systems.

At the international level, RegionViT based adaptive multi scale YOLOv8 system for UAV surveillance by Yallamraju et al. [8]. Their work reported good performance in multi-target tracking but that detection quality decreases drastically at night, which adds to the argument for thermal sensor integration. Munguia et al. (2025) [9] proposed a visual SLAM-based simultaneous control and

mapping system for UAVs in GPS-dented environments, which is close to our navigation method in principle.

YOLO-Based Detection for UAV Imagery

Small-object detection from bird's eye views is still one of the more challenging problems of computer vision. Standard YOLO models with ground level data find the tiny and crowded objects at UAV angles difficult Zhang et al. (2025) [10] tackled this problem with RTUAV-YOLO, which belongs to the YOLOv11 family with multi-scale adaptive modulation and progressive and progressive dilated convolutions, where 37.8 FPS can be achieved on Jetson Orin Nano (65.3% fewer parameters than the baseline). Fan et al. (2025) [11] proposed LUD-YOLO based on YOLOv8, a new multi-scale feature pattern and dynamic sparse attention were introduced, which reduced the number of parameters and improved the detection accuracy of the VisDrone2019 dataset and UAVDT dataset.

For the particular problem of law enforcement cases, Mahdi et al. (2025) [12] where they develop ROS2 and GStreamer pipeline in which YOLOv8-Seg, TensorRT optimization are combined, for real-time face segmentation on drone videos with about 105 ms latency at 30 FPS on Jetson hardware. This pipeline is of direct relevance to our communication and inference design.

Multi-Sensor Fusion for Strong Detection

Single-camera systems always fail in low-light, smoky or thermally ambiguous environments. The research community has begun to look more towards sensor fusion. A study in 2025 presented in Problems of Information Technology [13] showed that a thermal -RGB fusion pipeline paired with lightweight CNNs performed significantly better against RGB-only baselines in nighttime surveillance of drones, with sensor selection mechanism being adaptive, following an automatic weighting of modalities depending an ambient condition.

Wu et al. (2025) [14] proposed RMF-ED, a real-time multimodal fusion framework based on a combination of near infrared-based cameras and LiDARs in order to provide pedestrian detections in low light conditions where the complementary virtues of multiple sensor detection modalities results in more robust detections than a single modality. For depth-augmented detection, it has been shown by researchers studying RGBD-based pedestrian detection [15] that depth data can be used to resolve the ambiguity between the actual threats and the background clutter, especially in a crowded urban environment.

SLAM and GPS-Independent Navigation

Reliable localization in the absence of GPS signals is essential for urban drones since tall buildings and dense infrastructure leave dead areas where other interruptions of satellite signals can occur. Jiang et al. [16] presented a detailed survey of SLAM methods using LiDAR, and it was reported that the ORB-SLAM series was the most widely used visual SLAM family an outdoor UAV application. Wang et al. [17] have performed a scope review on 97 outdoor UAV-SLAM studies and confirmed that ORBSLAM3, which supports stereo, RGB-D, and fisheye cameras with a visual-inertial fusion, is the current state-of-the-art option for UAV localization.

A 2025 study that proposed an RGB-D and graph neural network enhanced SLAM framework [18] showed that combining depth measurements and point-line features in a DPLAGNN architecture was able to improve localization consistency in dynamic, low-texture environments that are common in dense urban environments. Their implementation based on ORB-SLAM3 realized up to 40% compute savings using a learnable pruning mechanism without losing the loop closure accuracy.

UAV Cybersecurity

Law enforcement UAVs are carriers of sensitive information and fly above critical infrastructures which makes them a target of great value for cyberattacks. Alsadie (2025) [19] has performed a comprehensive review of cybersecurity of AI-driven UAV, and GPS spoofing, man-in-the-middle

attacks, and signal jamming are found to be the most dangerous threat vectors in applications of surveillance. A study [20] done by Frontiers (2025) on cyber threats in drone systems concluded that most UAVs do not have forensic-by-design features and that integrated secure logging, tamper-resistant components and encrypted telemetry are basic requirements.

On the communication security side, Diffie-Hellman key exchange and HMAC based integrity verify system in ROS communication channels were proposed by a group of researchers studying ROS based UAV authentication [21]. Their Gazebo simulation indicated that the method is effective in detecting key tempering with a good performance overhead for DH key sizes of less than 4096 bits. This is a direct input to our encrypted DDS design choice.

Research Gap

The works reviewed above all deal with one of the puzzles. What is lacking is an integrated framework uniting the use of ROS2 based autonomous Navigation, Multi-sensor YOLO detection, Slam leveraging GPS-independent, and in-built Cyber security control in a unified software stack made explicitly for urban law enforcement surveillance. This is the gap that our work that our work focuses on.

SYSTEM ARCHITECTURE OVERVIEW

The proposed system is divided into 4 cooperating ROS2 nodes groups, each responsible for a different concern: perception, navigation, control, and security. This goes-forward modular separation is no coincidence. It means that any individual component, say the detection model or the path planner, can be swapped without the need of making changes to the rest of the system. This is one of the important practical benefits of building on ROS2 as opposed to a monolithic codebase.

The general flow of data is as follows. Sensor nodes continuously publish the synchronized RGB images, thermal frames, and depth maps to dedicated ROS2 topics with a frequency of 30 Hz. A subscription on these topics is performed by the perception module which is running a fused YOLO detector and publishing the detection results and estimated 3D positions of detected objects. The localization is performed in the navigation module using ORB-SLAM3 and the path planning and obstacle avoidance of waypoints are done using NAV2 stack.

When the perception module identifies a threat, it can result in a dynamic replan. The control module then translates the commands from the NAV2 into inputs from the flight controller (PX4 flight controller) using the uXRCE-DDS bridge. Throughout the package, the security module all ROS2 topics for anomalous behavior, enforces DDS access control policies, maintaining a tamper evident log of decisions and commands. A human operator can inject override commands, at any time, through a supervisory interface node.

SIMULATION ENVIRONMENT AND DATASET

Simulation Platform

All experiments are conducted entirely in simulation. The simulation stack consists of Gazebo Classic running the PX4 v1.14 SITL autopilot on an Ubuntu 22.04 host system, connected to ROS2 Humble via the uXRCE-DDS bridge [22]. This stack has been widely validated in the robotics research community for drone software development before physical prototyping [23].

The simulated UAV is modeled after a quadcopter in the DJI Matrice class, equipped with three virtual sensors: a 1920×1080 RGB camera (modeled after DJI H20T), a 640×512 thermal imager, and a RealSense D435i-style depth camera. The urban test environment is Gazebo's City World scene, populated with dynamic pedestrian and vehicle agents to simulate realistic patrol conditions. We also tested in the apartment world for indoor/partially enclosed scenarios.

Dataset Description

Since no public dataset completely covers the scenarios we are targeting with our research (law enforcement specific scenarios, Indian urban environments, night-time patrol, crowd monitoring, suspicious object detection) we built the simulation dataset by using the sensor plugins of Gazebo. This is a standard method in the UAV research community when collecting data in real life is not feasible at the stage of designing software [9].

This dataset contains 15,000 RGB-thermal-depth frame triplets that are synchronized with each other and collected to a frame rate of 30 frames per second, from simulated patrol missions. Scenes range from day and night, city streets, alleys, rooftops and crowded plazas. Annotated object classes: pedestrians (variants of poses), vehicles (cars, motorbikes, auto-rickshaws), suspicious objects (bound packages, proxy weapon types), and environment threats (fire, smoke). All the annotations are in the MS-COCO format with instance-level bounding boxes. There is a total of 42,500+ annotated instances included.

Data Processing

Raw sensor outputs have to be modality-specific preprocessed before they are fed to the detection model. For the RGB frames we range to $[0, 1]$ using ImageNet statistics for Cases mean and standard deviation, and the image size has been reduced to 640×640 . Thermal frames, which are grayscale, undergo contrast Limited adaptive histogram equalization CLAHE to bring out fine temperature gradients before

Cross-modal spatial alignment between the RGB and thermal frames is achieved using a pre-computed homography matrix derived from a modified RIFT (Radiation Invariant Feature Transform) procedure. Data augmentation applied jointly to all three modalities includes random horizontal flips, rotations up to $\pm 15^\circ$, brightness and contrast jitter up to $\pm 20\%$, and Gaussian noise. For nighttime robustness, we additionally apply synthetic thermal degradation by reducing thermal contrast and adding calibrated sensor noise to simulate real-world night conditions.

PROPOSED MODEL ARCHITECTURE

Overview

The detection model uses the previous YOLOv11 model as a backbone, and consists of a dual-backbone structure in a multi-modal input scenario, with a number of efficiency modules for small object detection in aerial scenes. The architecture has been designed keeping in mind the deployment on edge class hardware, and Nvidia Jetson Orin Nano (8 GB) has been considered as the representative inference hardware.

Dual Backbone and Feature Extraction

Two individual backbones one of RGB input and one of thermal input extract multi-scale feature maps independently. This design is rather preferred to early fusion (concatenation of raw sensor inputs) because every modality can achieve its own feature representation before fusion, which has been proven to be better than the fusion where modality specific cues are lost [13].

Each backbone has the Multi-Scale Feature Adaptive Modulation (MSFAM) module, which substitutes for a standard C3k2 bottleneck blocks. MSFAM uses channel level weighting that gives emphasis to small object feature channels and suppresses background noise. Progressive Dilated Separable Convolution Modules (PDSCM) at P4 feature point stage aggregate wider context information without the computational cost of larger kernels. A Lightweight Down Sampling Module (LDSM) is implemented instead of standard max-pooling which helps in preserving the spatial fine detail in the feature map size reduction.

Mid-Level Attention Fusion

After the two backbones extract their respective feature pyramids, fusion happens at the mid-level which means after backbone feature extraction but before the detection neck. This has been confirmed in recent studies on multi-spectral detection that it is superior to early fusion as well as late fusion [24].

The fusion block involves the use of two mechanisms that work in concert. First, Channel Prior Convolutional Attention (CPCA) calculates a channel-wise attention vector that indicates channel information in which feature channels have more discriminative information from both modalities. Second, the Parallel Patch-Aware Splicing (PPAS) separates feature maps into spatial patches and stitches those patches together, such that local spatial relationships are not altered. The fused feature map is calculated as:

$$F_{fused} = F_{RGB} \oplus F_{Thermal} \oplus CPCA(F_{RGB}, F_{Thermal})$$

Depth data from the third sensor is not fused to the feature level. Instead, it is used as an auxiliary input to an additional regression head that predicts the 3D distance of detected bounding boxes from the camera to allow for rough 3D localization of detected threats.

Detection Head and Loss Function

The neck is based on C3k2-based feature fusion, and P2 small-object detection head added on the side of the standard P3/P4/P5 heads. This P2 head operates at a higher spatial resolution and is fine tuned for detecting less than 32 by 32 pixels objects in the input image which is common for people viewed from the UAV altitude.

The training loss replaces the normal CIoU with MPDWIoU (Minimum Point Distance Weighted IoU) which is more suited for small objects with a high density of objects because it adds a minimum-point-distance term that penalizes the predictions that are close but not overlapping:

$$MPDWIoU = 1 - IoU + \rho^2(b, b_{gt})/c^2 + \alpha \cdot MPD(b, b_{gt})$$

Where ρ is the Euclidean distance between predicted and ground-truth box centers, c is the diagonal of the smallest enclosing box, and MPD is the minimum point distance weighting term. The hyperparameter α is set to 0.1 based on grid search over the validation set.

NAVIGATION AND PATH PLANNING

SLAM-Based Localization

Reliable localization even without GPS is vital in the Indian urban environment where blocking of the satellite signal is not uncommon due to tall buildings. We use in this case ORB-SLAM3 in visual-inertial mode, fusing information coming from the RGB stereo camera pair and a simulated IMU. ORB-SLAM3 was selected as it supports stereo, RGB-D and fisheye cameras, its multi-map management is available for large areas, and it has been found as being the most widely accepted visual SLAM algorithm in outdoor UAV research [17].

In our simulation, ORB-SLAM3 is running as a ROS2 node subscribing to the stereo image pair and IMU topics and publishing a 6-DOF estimate of the pose on `/slam/pose`. The pose in the navigation stack is used instead of GPS/INS for all the planning computations. Absolute Trajectory Error (ATE) is the main localization measure that we measure.

NAV2 Path Planning and Coverage

NAV2 stack is responsible for all the planning of path and obstacle avoidance. A custom coverage planner node creates a lawnmower pattern set of waypoints calculated from the defined patrol boundary. During missions the global planner is based on a pre-built semantic cost map and the dynamic window approach (DWA) is used for real-time obstacle avoidance by the local planner using live depth sensor data. When the perception module detects a threat /detection, the mission executor will cause a dynamic replan to hover over the threat location and then resume the patrol route.

With a step rate of 10 Hz the whole pipeline of navigation is running: The control bridge is publishing trajectory setpoints to the PX4 flight controller with a 50 Hz frequency using the uXRCE-DDS protocol.

SECURITY FRAMEWORK

Law enforcement UAVs are high-value targets when it comes to cyber interference. A compromised drone could potentially be used to spy on police operations, send false threat warnings, or have their physical location redirected by GPS spoofing [19]. Our framework solves this with a layered approach of security.

The human in the loop override system is a critical safety feature. Human operators can write into the quiescence, redirection or landing commands at any spot by `/operator/override` topic. Override messages are signed and given the highest QoS priority, thus guaranteeing to preempt any autonomous action in progress within 200ms in time. All commands and detections are written to an immortal audit-log, which is a design requirement consistent with forensic-by-design principles highlighted in recent research on UAV security [20].

SIMULATION RESULTS AND ANALYSIS

All results presented in this section have been obtained from simulation experiments performed in the Gazebo/PX4 SITL environment described in Section IV. These are projected figures of performance of our software framework under controlled simulation conditions. We make no claims regarding the performance of the hardware in the real world, which is left for the future.

DETECTION PERFORMANCE

The full fusion model (RGB + Thermal + Depth) achieves a mAP@50 of 96.1%, which is the maximum improvement of 11.5 of the RGB-only models. The greatest improvement is on the night time subset where thermal data makes up for near zero visible light contrast. Thermal fusion covers false negatives by about 4.9% than RGB-only on night frames. The 37.8 FPS inference speed on simulated Jetson hardware verifies that real-time patrol operation is possible in the framework design of software.

Ablation Study

To see what each suggested architectural component does for the model, we performed an ablation study in which we sequentially removed modules from the entire model, and saw how much the performance decreased. The ablation makes sure that each proposed module has a positive contribution. The P2 detection head has the greatest individual impact (-4.4 mAP@50), which verifies the significance of the small object detection layer in aerial surveillance. The second most important (-2.9 mAP@50) is the MSFAM module. The MPDWIoU loss gives a contribution of 1.8 mAP@50 improvement, especially on the occlusion-heavy and densely packed subsets on which standard IoU-based loss function struggle.

Navigation Performance

ORB-SLAM3 in visual-inertial mode was able to result in an ATE of 0.83 m for the city world which is acceptable navigation accuracy for large-area patrol. The coverage rate of 95.1% and solving time of <100 ms between perception and control verify that software pipeline requirement of real-time operation with simulation environment.

ETHICS AND SOCIETAL IMPACT

Deploying AI-powered surveillance drones for law enforcement raises a number of serious ethical questions, which need to be dealt with proactively, not after the fact. It is in the Indian context that the issues of policing-both the diversity of communities affected by policing operations, and issues of surveillance and civil liberties, the relative novelty of the regulatory environment in India that governs and oversees AI work, become particularly salient.

Privacy and Data Minimization

Our framework is based on the principle of data minimization. The drone is not collecting data from the sensors all the time; the data is only collected during the active patrol. Feeds are detection events, not raw video, the main artifacted event. Video feeds are broadcast only over encrypted communications and not kept on-board any more than the mission log buffer would allow. These design choices are in line with the broader effort there to be responsible with UAV deployment in Indian smart city infrastructure [1].

Detection Bias and Fairness

Object detection models trained on skewed datasets can result in different detections based on a person's race or gender, meaning that in a law enforcement context, the models could disproportionately target certain communities for surveillance. Our simulation data set was intentionally distributed with diversity regarding the appearance and types of clothing. Any future expansion of the actual data set on the ground will need auditing of demographics so that detection rates are maintained across population groups. This is a known open problem in the field of AI aided law enforcement [8].

Human Accountability

The human-in-the-loop override layer isn't really a safety mechanism so much as it is an ethical requirement. Fully autonomous lethal or physically restrictive action based on the detection from the drone must not be allowed. Any alert made by the system should be subject to a human review before any law enforcement action is taken. Our audit log design - recording of every detection, command and override equally with time stamps and operator ID in order to support deeper accountability and investigation post-incident.

Regulatory Compliance

The framework is set up to work in India's Drone Rules, 2021, which sets out requirements on the registration of operators, compliance with no-fly zones, and the handling of data. For the airspace enforcement, geofencing is embedded into the mission executor. Night operations that require special permissions under Indian regulations, are considered to be a restricted mode requiring explicit authorization of the operator before activating the thermal dominant patrol mode.

COMPARISONS AND RELATED WORKS

Our system is the only one in the review which combines all three sensor modalities with SLAM GAri (global absolute navigation) based navigation, multi-layer security framework, and explicit design is performed to comply with regulatory requirements of Indian law enforcement. The trade-off is that, as opposed to some of the works above, our current validation is simulation only.

LIMITATIONS

We want to be honest about what this work does and doesn't claim. The most important is the lack of physical hardware validation. All performance numbers are simulation results and it is a well-established fact that the sim-to-real gap leads to a performance degradation once software frameworks are brought to real-world sensors, noise, vibration and aerodynamics. Second, the simulated dataset, although carefully designed, cannot capture the full set of noise characteristics (e.g. noise in the real thermal cameras or in-depth data), missing data in depth data (depth data dropouts on reflective surfaces), or the motion blur and rolling shutter of the real cameras (if mounted on a vibrating airframe).

Third, the energy usage of simultaneous running of ORB-SLAM3 and YOLO inference on a Jetson Orin Nano in a real scenario has not been profiled. The power draw will decrease effective flight-time and have to be carefully controlled in any physical prototype. Fourth, the adversarial robustness has not been tested formally. While there is some inherent resistance to single-modality adversarial attacks through multi-sensor fusion, there needs to be a dedicated adversarial evaluation before operational deployment.

FUTURE WORK AND CONCLUSION

Future Work

The next step is hardware implementation using a Nvidia Jetson Orin Nano as a companion computer, pixhawk 6X with PX4 as the Flight controller and a custom multi-sensor payload. We're planning to validate the full software stack in controlled, outdoor field tests, and then move to actual deployments in the cities. Specific potential areas for future research include:

- Up-to-reality Tactile reconstructed physical build and real-world field trials in controlled outdoor environments
- Energy profiling and model pruning for maximum flight endurance for battery powered platforms
- Adversarial robustness testing & adversarial training to make the detection model robust
- Blockchain based unchangeable logging of auditing for forensic accountability
- Expansion of real-world data set of Indian Urban environment - demographic auditing
- Multi-UAV coordination for broad area coverage based on the decentralized swarm planning method.

CONCLUSION

This paper has introduced a holistic ROS2 based software package for autonomous UAV law enforcement surveillance, which is designed from the scratch for Indian urban setup and tested completely using high fidelity Gazebo/PX4 SITL simulation. The framework unifies YOLO driven multi-sensor object detection, as well as ORB-SLAM3 GPS independent navigation with NAV2 based autonomous patrol in a multi-layer security and accountability architecture built-in a modular, reproducible system.

Simulation results show mAP@50 >96.1% befallen is full sensor fusion >95% patrol cover end-to-end latency less 100 ms Our results meet real-time operational requirements. The ablation study provides good evidence that each of the proposed architectural components makes a meaningful contribution to the overall performance. The ethics and security sections discuss the responsibilities associated with using AI-powered surveillance in public places.

We think this work is providing a good foundation for the next generation of intelligent aerial patrol systems capable of being enforced in India with honesty and reproducibility. The corresponding hardware implementation that will follow this simulation study is the natural and necessary adhesion step toward giving the vision being operate.

REFERENCES

1. MarketsandMarkets, "India Drone (UAV) Market – Forecast to 2030," MarketsandMarkets Research, 2025. [Online]. Available: <https://www.marketsandmarkets.com/Market-Reports/india-drone-market-136782206.html>
2. Bureau of Police Research and Development (BPRD), "Drone: A New Age Policing Tool," Ministry of Home Affairs, Government of India, Project Report No. 13/MM:03. [Online]. Available: <https://bprd.nic.in/uploads/pdf/Drone%20A%20New%20Age%20Polcing%20Tool.pdf>
3. S. Macenski, T. Foote, B. Gerkey, C. Lalancette, and W. Woodall, "Robot Operating System 2: Design, architecture, and uses in the wild," *Science Robotics*, vol. 7, no. 66, p. eabm6074, 2022. doi: 10.1126/scirobotics.abm6074
4. R. Zhang et al., "RTUAV-YOLO: A Family of Efficient and Lightweight Models for Real-Time Object Detection in UAV Aerial Imagery," *Sensors*, vol. 25, no. 21, p. 6573, 2025. doi: 10.3390/s25216573
5. "Thermal-RGB Fusion with Lightweight CNNs for Night-Time Drone Surveillance and Real-Time Adaptive Sensor Selection," *Problems of Information Technology*, vol. 16, no. 2, pp. 45–55, 2025. [Online]. Available: https://jpit.az/uploads/article/en/2025_2/

6. R. Munguia, A. Grau, Y. Bolea, and G. Obregón-Pulido, "A Simultaneous Control, Localization, and Mapping System for UAVs in GPS-Denied Environments," *Drones*, vol. 9, no. 1, p. 69, 2025. doi: 10.3390/drones9010069
7. Y. Bhaskar, K. Dolaskar, H. Jadhav, M. Shirke, and K. Kulkarni, "Autonomous Navigation and Object Detection Using ROS 2 and YOLOv8," *International Journal of Scientific Research in Engineering and Management (IJSREM)*, 2024. [Online]. Available: <https://ijsrem.com/download/autonomous-navigation-and-object-detection-using-ros-2-and-yolov8/>
8. C. Yallamraju et al., "Dynamic Object Detection and Tracking System on Unmanned Aerial Vehicles for Surveillance Applications Using RegionViT-Based Adaptive Multi-Scale YOLOv8," *Computational Intelligence*, Wiley, July 2025. doi: 10.1111/coin.70101
9. R. Munguia, A. Grau, Y. Bolea, and G. Obregón-Pulido, "A Simultaneous Control, Localization, and Mapping System for UAVs in GPS-Denied Environments," *Drones*, vol. 9, no. 1, p. 69, Jan. 2025. doi: 10.3390/drones9010069
10. R. Zhang et al., "RTUAV-YOLO: A Family of Efficient and Lightweight Models for Real-Time Object Detection in UAV Aerial Imagery," *Sensors*, vol. 25, no. 21, p. 6573, 2025. doi: 10.3390/s25216573
11. Q. Fan, Y. Li, M. Deveci, K. Zhong, and S. Kadry, "LUD-YOLO: A Novel Lightweight Object Detection Network for Unmanned Aerial Vehicle," *Information Sciences*, vol. 686, p. 121366, Jan. 2025. doi: 10.1016/j.ins.2024.121366
12. H. S. Mahdi, K. R. Kumar, and K. J. D. Christopher, "Real-Time Drone Communication System Using ROS 2 and GStreamer with YOLOv8Seg for Face Segmentation," *International Journal of Computational and Experimental Science and Engineering*, vol. 11, no. 2, 2025. [Online]. Available: <https://ijcesen.com/index.php/ijcesen/article/view/2123>
13. "Thermal-RGB Fusion with Lightweight CNNs for Night-Time Drone Surveillance and Real-Time Adaptive Sensor Selection," *Problems of Information Technology*, vol. 16, no. 2, pp. 45–55, 2025.
14. X. Wu et al., "RMF-ED: Real-Time Multimodal Fusion for Enhanced Target Detection in Low-Light Environments," *IET Cyber-Systems and Robotics*, Apr. 2025. doi: 10.1049/csy2.70011
15. "Enhancing Low-Light RGB-D Pedestrian Detection Through Dual-Stage Modality-Guided Fusion," *The Visual Computer*, Springer, Dec. 2025. doi: 10.1007/s00371-025-04237-5
16. Z. Jiang et al., "GPS-Denied LiDAR-Based SLAM — A Survey," *IET Cyber-Systems and Robotics*, Nov. 2025. doi: 10.1049/csy2.70031
17. S. Wang et al., "UAV-Based Simultaneous Localization and Mapping in Outdoor Environments: A Systematic Scoping Review," *Journal of Field Robotics*, Wiley, Apr. 2024. doi: 10.1002/rob.22325
18. "RGB-D Camera and Graph Neural Network-Based SLAM for Dynamic and Low-Texture Environments," *Scientific Reports, Nature*, Aug. 2025. doi: 10.1038/s41598-025-12978-5
19. D. Alsadie, "Cybersecurity and Artificial Intelligence in Unmanned Aerial Vehicles: Emerging Challenges and Advanced Countermeasures," *IET Information Security*, Wiley, Oct. 2025. doi: 10.1049/ise2/2046868
20. "Cyber Threat in Drone Systems: Bridging Real-Time Security, Legal Admissibility, and Digital Forensic Solution Readiness," *Frontiers in Communications and Networks*, Aug. 2025. doi: 10.3389/frcmn.2025.1661928
21. "Securing UAV Communication: Authentication and Integrity," *arXiv:2410.09085*, Oct. 2024. [Online]. Available: <https://arxiv.org/html/2410.09085v1>
22. P. Khaledi, "px4_sim_ros2: Drone Simulation with ROS2, PX4, NAV2 and GZ," *GitHub Repository*, 2024. [Online]. Available: https://github.com/ParsaKhaledi/px4_sim_ros2
23. S. Monemati, "PX4-ROS2-Gazebo-YOLOv8: Aerial Object Detection using a Drone with PX4 Autopilot and ROS 2," *GitHub Repository*, 2024. [Online]. Available: <https://github.com/monemati/PX4-ROS2-Gazebo-YOLOv8>
24. "ASFN: An RGB-T Adaptive Selection Fusion Network for Nighttime Tracking," in *Proceedings of the 2024 International Conference on Intelligent Perception and Pattern Recognition (IPPR)*, ACM, 2024. doi: 10.1145/3700035.3700054