IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 2, February 2026

DOI: 10.17148/IJARCCE.2026.15220

AimSense: A Real-Time AI-Assisted Threat Detection and Response System with Human-in-the-Loop Protocol

Amal Sankar M[1], Albin Varghese Mathew[2], Ajin Anil[3], Jishnu Jayakumar[4], Ancy Das Y R[5]

Department of Computer Science and Engineering, College of Engineering Kottarakkara

APJ Abdul Kalam Technological University, Kerala, India[1-5]

Abstract: In the domain of modern security and surveillance, the delay between threat detection and response is a critical vulnerability. This paper presents AimSense, a computer vision-based threat detection system currently under development. The system utilizes the You Only Look Once (YOLO) version 11 (YOLOv11) architecture to integrate object detection (weapon recognition) with pose estimation (human skeleton analysis), enabling accurate identification of active threats based on grasping interactions rather than mere object presence. An important component of the proposed system is the Human-in-the-Loop (HITL) interface, which ensures that all the engagement decisions are verified by a human operator prior to execution. This paper describes the prototype architecture, the sector-based localization algorithm, and the optimization techniques that enable real-time performance on standard hardware.

Keywords: Computer Vision, YOLOv11, Threat Detection, Human-in-the-loop, Pose estimation, Surveillance Systems.

I. INTRODUCTION

Automated surveillance systems have traditionally relied on motion detection or basic object classification techniques. With the advancement of Artificial Intelligence (AI), modern surveillance systems are increasingly capable of performing context-aware analysis and more accurate threat assessment. However, in high-stakes environments, false positives can still occur when harmless objects are misidentified as weapons or when inactive or holstered weapons are incorrectly classified as immediate threats, which can lead to serious operational consequences.

The AimSense project addresses these limitations by incorporating AI-drive context aware detection that evaluated human-object interaction rather than simply identifying the presence of a weapon. The system determines whether a detected weapon is actively being handled by a human subject, thereby improving threat assessment accuracy and reducing unnecessary alerts.

To mitigate the ethical and safety concerns associated with autonomous systems, AimSense employs a strict HITL framework. The system performs detection, locking, and tracking functions, while the final engagement decision remains under human supervision through a dedicated Graphical User Interface (GUI). This paper presents the current development stage of the AimSense prototype, focusing on its software architecture, decision-making logic, and planned transition to real-world hardware implementation.

II. LITERATURE REVIEW

Recent advancements in autonomous defense technologies have focused primarily on improving automated surveillance and fire-control capabilities. Early systems relied mainly on motion-tracking mechanisms. For instance, Louali et al. [1] developed a vision-based autonomous turret that used tracking-learning-detection methods to follow moving targets. Although effective for tracking single-targets, their study revealed limitations in managing multiple targets simultaneously under real-time conditions. Extending this line of research, Keswani et al. [2] introduced an AI-driven sentry turret designed for continuous area surveillance, demonstrating the importance of automated perimeter monitoring in modern security infrastructures.

High-performance target detection in such systems requires computationally efficient object detection frameworks. Sun et al. [3] proposed YOLO-E, a lightweight adaptation of the YOLO architecture optimized for military target recognition, demonstrating that architectural optimization can significantly reduce processing latency on edge computing hardware. Similarly, Zheng et al. [4] introduced LAM-YOLO, which enhances detection accuracy in visually challenging environments by addressing lighting occlusion, a condition frequently encountered in real-world operational settings. Since direct real-world testing of autonomous defense systems involves considerable safety risks, high-fidelity simulation

DOI: 10.17148/IJARCCE.2026.15220

platforms have become essential tools for system development and validation. Sobchyshaka et al. [5] demonstrated that Unreal Engine can support highly immersive and physicsaccurate virtual environments suitable for complex system evaluation. This capability was further applied by Chaudhary et al. [6], whose HEROES platform uses Unreal Engine to train emergency response robots while reducing discrepancies between the simulated and physical deployment. In addition, Nesti et al. [7] introduced SimPRIVE, a framework designed to simulate the interaction of a physical robot within virtual environments, allowing more reliable transfer of simulation results to real-world scenarios. The importance of maintaining human authority within simulated operational environments is emphasized by Karpichev et al. [8], who highlight the critical role of human oversight in collaborative human-robot control frameworks.

Despite these technological advances, the deployment of autonomous weapon systems remains a subject of significant ethical and regulatory debate. Natarajan et al. [9] argue that HITL architectures are essential to ensure accountability and responsible decision-making, particularly when compared to fully autonomous control systems. This viewpoint is consistent with the legal analysis presented by Bhuta et al. [10], who state that meaningful human control is a requirement under international humanitarian law. Further technical concerns are highlighted by Colijn and Podar [11], who discuss the inherent unpredictability and associated risks of removing human supervision from lethal autonomous systems.

From a computational perspective, achieving real-time operation remains a critical design objective. Yun et al. [12] demonstrate that lightweight convolutional neural networks are particularly well suited for embedded target detection tasks where computational resources are limited. Meanwhile, broader multimodal approaches such as the Unified VisionGPT framework proposed by Kelly et al. [13] indicate a shift towards integrating richer contextual information into vision-based AI systems. Nevertheless, for time-sensitive detection tasks, YOLO based architectures remain more practical due to their superior balance between computational efficiency and detection speed compared to earlier region-based models and more complex multimodal frameworks.

III. METHODOLOGY

The system is built using Python, leveraging the OpenCV library for image processing and Ultralytics YOLOv11 for deep learning inference. The system architecture consists of three core modules: the Logic Core, the Localization Module, the HITL Interface, as shown in Figure 1.

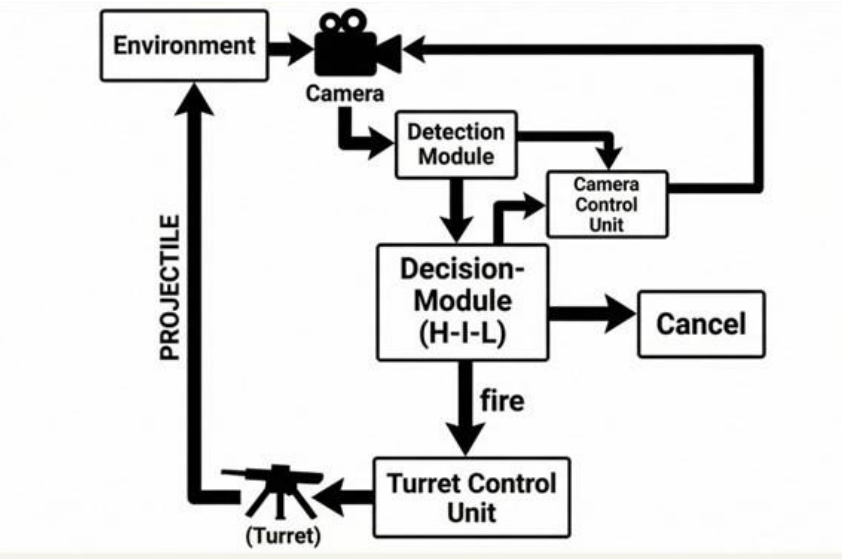

Figure 1: System Design

A. Dual-Model Inference Engine

Unlike single-model systems, AimSense employs a cascading inference pipeline. First, the weapon detection model (YOLOv11) scans each frame for predefined threat classes such as handguns or knives. An early-exit optimization strategy is implemented such that the pipeline terminates immediately if no weapon is detected, thereby reducing computational load and increasing frames per second (FPS) during idle scanning.

DOI: 10.17148/IJARCCE.2026.15220

Second, pose estimation (YOLOv11-Pose) is applied. If a weapon is detected, this secondary model extracts human key points, specifically the wrist and shoulder coordinates. A threat is confirmed through the grabbing logic only when the coordinates of the subject's wrist key point (K_{wrist})fall within the bounding box of the detected weapon $\left(B_{weapon}\right)$with a defined padding threshold (P), expressed as:

$$(B_{x1} - P) < K_{wrist\_x} < (B_{x2} + P)$$

This ensures that holstered weapons or weapons lying on a table do not trigger a false alarm.

B. Target Localization and Grid System

To facilitate real-world response, the 2D video feed is mapped to a tactical grid. In sector mapping, the field of view is divided into a 3 × 3 matrix (sectors A1 through C3). The center point of the threat's bounding box determines the active sector, providing immediate bearing information to the operator.

For distance estimation, the system utilizes the detected shoulder keypoints to estimate range using triangular similarity. Assuming an average human shoulder width ($W_{real} \approx$ 0.45m) and a calibrated focal length (F), the distance (D) is calculated as:

$$D = \frac{W_{real} \times F}{W_{pixel}}$$

C. Human-in-the-Loop Interface

The prototype includes a Tkinter-based GUI that functions as the fire control system. When a threat is validated by the Grabbing Logic, the system transitions from "Scanning" to "Locked" status. The video feed is frozen on the frame containing the evidence. At this stage, the "ENGAGE" and "ABORT" controls are unlocked. This workflow ensures that the AI functions as a detection and tracking mechanism, while final moral and tactical authority remains with the human operator.

IV. IMPLEMENTATION AND RESULTS

The system is currently in the prototype phase and operates on consumer-grade GPU hardware. Regarding detection speed, by utilizing FP16 (Half-Precision floating-point) inference and reducing the input resolution to *640 × 480* pixels during the scanning phase, the system achieves real-time performance.

For localization, the grid overlay successfully identifies the target's sector (e.g., "TARGET LOCKED [B2]"), providing clear and intuitive visual feedback to the operator. The HITL safety protocols successfully interrupt the autonomous loop. In testing, the system effectively ignored neutral subjects and only triggered the "Lock-in" sequence when the specific "wrist-in-box" condition was met.

V. CONCLUSION AND FUTURE SCOPE

The AimSense project demonstrates that integrating pose estimation with object detection significantly reduces false positives in threat scenarios. The current development phase has validated the software logic and the efficacy of the HITL safety protocol.

Future work involves implementing this software in a real-world environment. Future iterations will involve hardware integration, interfacing the Python logic with microcontrollers (e.g., Arduino/ESP32) to control a physical pan-tilt servo turret that aligns with the detected grid sector. Additionally, we aim to implement depth sensing by replacing heuristic distance estimation with LiDAR or stereovision for higher accuracy, and ballistic computation to account for projectile drop. The ultimate goal is to deploy AimSense as a reliable, ethically grounded support tool for security personnel.

REFERNCES

[1] R. Louali, D. Negadi, R. Hamadouche, and A. Nemra, "Design of a Vision-Based Autonomous Turret," Journal of Automation, Mobile Robotics and Intelligent Systems, vol. 16, no. 4, pp. 728–734, 2022, doi: 10.14313/JAMRIS/4-2022/35.

DOI: 10.17148/IJARCCE.2026.15220

[2] K. Keswani, S. Lakhmani, O. Jadhav, S. Zinge, and N. Satpute, "Autonomous AI Sentry Turret with Area Surveillance," International Journal for Research in Applied Science and Engineering Technology (IJRASET), vol. 13, no. 3, pp. 2927–2930, Mar. 2025, doi: 10.22214/ijraset.2025.67970.

[3] Y. Sun et al., "YOLO-E: A Lightweight Object Detection Algorithm for Military Targets," Research Square Preprint, 2024, doi: 10.21203/rs.3.rs-5259808/v1.

[4] Y. Zheng, Y. Jing, J. Zhao, and G. Cui, "LAM-YOLO: Drones-Based Small Object Detection on Lighting-Occlusion Attention Mechanism YOLO," arXiv preprint arXiv:2411.00485v1 [cs.CV], 2024.

[5] O. Sobchyshaka, S. Berrezueta-Guzman, and S. Wagner, "Pushing the Boundaries of Immersion and Storytelling: A Technical Review of Unreal Engine," arXiv preprint arXiv:2507.08142v1 [cs.HC], 2025.

[6] A. Chaudhary, K. Tiwari, and A. Bera, "HEROES: Unreal Engine-Based Human and Emergency Robot Operation Education System," arXiv preprint arXiv:2309.14508v2 [cs.RO], 2025.

[7] F. Nesti, G. D'Amico, M. Marinoni, and G. Buttazzo, "SimPRIVE: A Simulation Framework for Physical Robot Interaction with Virtual Environments," arXiv preprint arXiv:2504.21454v1 [cs.RO], 2025.

[8] Y. Karpichev et al., "Extended Reality for Enhanced Human-Robot Collaboration: A Human-in-the-Loop Approach," arXiv preprint arXiv:2403.14597v3 [cs.RO], 2024, doi: 10.1109/RO-MAN60168.2024.10731170.

[9] S. Natarajan, S. Mathur, S. Sidheekh, W. Stammer, and K. Kersting, "Human-in-the-Loop or AI-in-the-Loop? Automate or Collaborate?," arXiv preprint arXiv:2412.14232v1 [cs.HC], 2024.

[10] N. C. Bhuta, S. Beck, R. Geiß, H.-Y. Liu, and C. Kreß (Eds.), Autonomous Weapons Systems: Law, Ethics, Policy. Cambridge, U.K.: Cambridge University Press, 2016.

[11] A. Colijn and H. Podar, "Technical Risks of (Lethal) Autonomous Weapons Systems," Whitepaper, Encode Justice, n.d.

[12] J. Yun et al., "Real-Time Target Detection Method Based on Lightweight Convolutional Neural Network," Frontiers in Bioengineering and Biotechnology, vol. 10, Art. no. 861286, 2022, doi: 10.3389/fbioe.2022.861286.

[13] C. Kelly et al., "Unified VisionGPT: Streamlining Vision-Oriented AI Through Generalized Multimodal Framework," arXiv preprint arXiv:2311.10125v1 [cs.CV], 2023.