ASM Tango A Visual-Inertial Man-Out-of-the-Loop Guidance Architecture for GPS-Denied Autonomous Aerial Platforms

Arjun Kaushal[1], Rabinder Henry[2]
[1]*B.Tech Student (Robotics Engineering), uGDX School of Technology, Atlas Skilltech University, Mumbai, India*
[2]*Course Director and Project Supervisor, uGDX School of Technology, Atlas Skilltech University, Mumbai, India*

***Abstract*—ASM Tango is a subsonic, modular autonomous aerial demonstrator designed to validate a Man-Out-of-the-Loop (MOOTL) guidance architecture under GPS-denied, communications-denied operational conditions. The platform integrates a 3D-printed aerodynamic airframe with two independent perception layers—Visual and Orientational—to be arbitrated by a centralized root sensor fusion engine. The Visual layer employs a Parent-Child configuration: the Parent layer performs real-time object detection and target classification via onboard neural inference; the Child layer provides pixel-level thermal-optical cross-validation to detect adversarial infrared manipulation at the detection level, upstream of any guidance decision. The Orientational layer maintains continuous six-degree-of-freedom inertial and attitude state. Individual perception components have been validated in controlled environments. Root fusion integration via Software-in-the-Loop simulation is the active development frontier. The graceful-degradation architecture—sustaining guidance integrity under single-layer denial—presents direct applicability to autonomous space-grade GNC systems where GPS is permanently absent.**

***Index Terms*—MOOTL guidance, GPS-denied navigation, visual- inertial fusion, IR cross-validation, autonomous aerial systems**

I. INTRODUCTION

Autonomous guidance under GPS denial is an operationally confirmed requirement. Adversarial jamming and spoofing of GPS/INS-coupled navigation is documented in contested environments [1, 2], and platforms relying on a single primary sensing modality fail catastrophically when that modality is denied. Contemporary fire-and-forget systems—including the SCALP-EG, JASSM, and Taurus KEPD 350—employ GPS/INS-coupled mid-course navigation combined with Imaging Infrared (IIR) or Semi-Active Laser (SAL) terminal seekers. Each architecture retains independently exploitable failure modes: GPS jamming and spoofing, SAL designator dependency, IIR countermeasure deployment, and datalink emission detectability.

ASM Tango addresses the guidance challenge through two independent, cross-validated perception layers [3]. The Visual layer provides scene understanding and target classification via onboard neural inference. The Orientational layer maintains continuous inertial state and translates arbitrated guidance decisions into actuator commands. A centralized root fusion engine arbitrates both layers and is designed to sustain guidance integrity if either is degraded.

The platform's primary architectural contribution is pixel- level thermal-optical cross-validation within the Visual layer: a mechanism that detects adversarial infrared manipulation at the detection level rather than at the seeker level. This is structurally distinct from conventional IIR seeker hardening approaches.

II. RELATED WORK

A. Autonomous Aerial Guidance

Contemporary fire-and-forget systems employ Digital Scene-Matching Area Correlator (DSMAC) techniques against pre-stored target imagery, paired with INS/GPS mid-course navigation [1]. IIR terminal seekers demonstrate substantially greater resistance to laser jamming than reticle-based seekers, limiting the effectiveness of Directed Infrared Countermeasures (DIRCM) against imaging focal plane arrays [4]. However, IIR seekers remain susceptible to scene alteration and spatially aware thermal countermeasures. Countering imaging seekers is an active and unresolved problem in countermeasure research [4].

SAL-dependent systems introduce an operational dependency on a forward-positioned designating asset and a detectable laser emission signature. Platforms retaining datalink capability for mid-course updates introduce RF emission signatures detectable even where Low Probability of Intercept wave- forms are employed.

DoD Directive 3000.09 formalizes the fan-out ratio constraint: human operator-in-loop architectures impose a structural ceiling on concurrent autonomous agent deployment that MOOTL design resolves by construction [5].

B. Visual-Inertial Odometry

Visual-Inertial Odometry (VIO) fuses camera and IMU measurements to sustain localization in GPS-denied environments without external positioning infrastructure [6]. Tight-coupling formulations—where raw feature measurements and IMU pre-integrations are jointly estimated—outperform loose-coupling approaches in dynamic or low-texture environments [7]. Feature-based SLAM approaches extend these formulations to closed-loop map reuse [8]. ASM Tango's root fusion engine targets a state estimation framework consistent with these for- mutations.

C. Adversarial IR Countermeasures

DIRCM systems are effective against point-source reticle- based seekers but substantially less effective against imaging IIR systems employing spatial discrimination [4]. Detection- level cross-validation—comparing simultaneous with-IR and without-IR feeds of the same scene at the pixel level—has not been documented in existing low-cost demonstrator architectures.

III. SYSTEM ARCHITECTURE

ASM Tango comprises three integrated subsystems: the aerodynamic airframe, a dual-layer perception stack, and a root sensor fusion engine. The architecture targets graceful degradation: guidance integrity is sustained when either perception layer is degraded or denied.

A. Visual Perception Layer

The Visual layer operates in a Parent-Child configuration. Both units image the same scene simultaneously, providing in- dependent but spatially co-registered feeds that enable pixel- level cross-validation.

Parent layer. An embedded computing platform running an edge AI accelerator performs real-time object detection, target classification, and cluster identification via onboard neural inference [9]. A wide-FOV dual-channel optical sensor captures simultaneous visible and infrared imagery. Feature recognition enables GPS-denied environmental navigation without external positioning infrastructure. The system autonomously distinguishes target profiles from non-threat subjects, with concur- rent skeletal posture estimation providing additional discriminative context [10]. Fig. 1 illustrates a representative Parent- layer inference output.

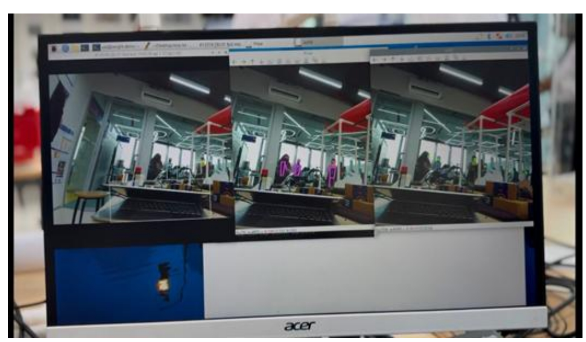


Figure 1. Parent layer neural inference output. Simultaneous YOLO- based object detection, skeletal posture estimation, and facial feature identification on a single frame. Onboard execution via edge AI accelerator.

Child layer. A secondary embedded platform hosts a thermal imaging module and an IR-filtered optical sensor restricted to visible wavelengths. The IR-filtered channel operates in parallel with the Parent's wide-FOV sensor, providing a simultaneous without-IR feed of the same scene. Cross-referencing the Parent's with-IR output against the Child's without-IR output at the pixel level enables detection of adversarial IR manipulation. Fig. 2 shows a representative Child-layer thermal acquisition.

The thermal imaging module and IR-filtered sensor share a common orientational calibration, establishing spatial co- registration between the visible and thermal domains. This enables pixel-level cross-validation rather than coarse decision- level comparison—a structural requirement for reliable dis- crimination between genuine metallic targets and thermal countermeasures.

Cross-validation constraint. A Parent-layer detection can- not be actioned by the root fusion engine without thermal confirmation from the Child layer. This constraint is independent of detection confidence scores from the inference pipeline.

Figure 2. Child layer thermal acquisition. Infrared feed captured for pixel-level cross-validation against the Parent layer visible channel. Spatial co-registration between thermal and visible domains has been established.

B. Orientational Perception Layer

The Orientational layer employs a flight controller providing six degrees of freedom: three-axis orientation (roll, pitch, yaw) and three-axis inertial state. Guidance commands arbitrated by the root fusion engine are translated here into actuator signals driving the control surfaces. Updated inertial and orientational measurements are fed back to the root system on each cycle for closed-loop state estimation. An external GNSS module is carried as standby redundancy only. It is not a primary navigation input.

C. Root Sensor Fusion Engine

The root fusion engine arbitrates inputs from both perception layers, weights each layer's contribution, resolves inter- layer conflicts, and issues unified guidance commands to the Orientational layer for actuator execution. Inter-layer weighting responds dynamically to degradation events, redistributing authority to the remaining functional layer.

The fusion architecture targets a joint state estimation formulation operating on visual feature measurements and inertial pre-integrations within a single state vector, consistent with tight-coupled VIO literature [7]. The arbitration logic is grounded in optimal state estimation theory [12].

IV. PHYSICAL DEMONSTRATOR

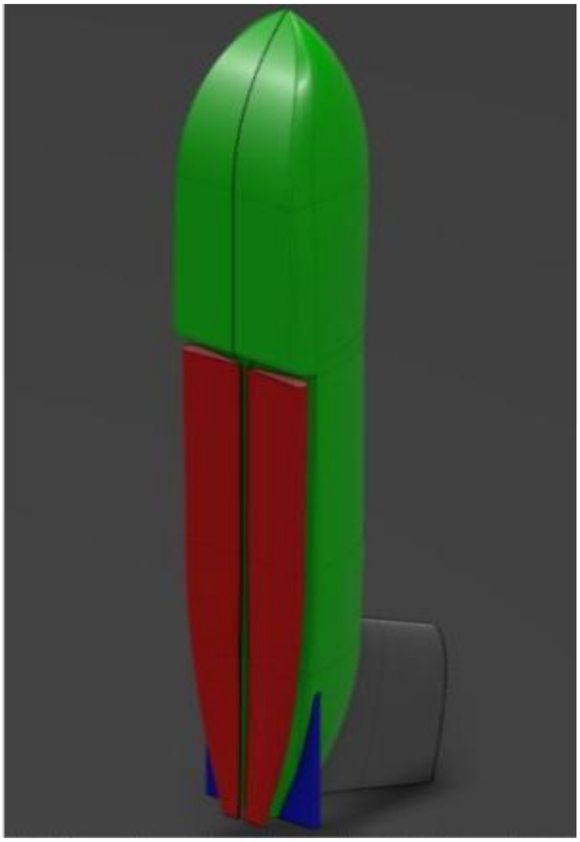
A. Airframe

The physical demonstrator is a 3D-printed aerodynamic shell geometrically inspired by the SCALP-EG cruise missile pro- file: low-observable fuselage and mid-body deployable wing configuration. The body is fabricated in PLA at 0.2 mm layer height and 30–35% gyroid infill. The assembled demonstrator is approximately 1.7 m in length with a deployed wingspan of 900 mm.

The body is segmented for modular integration and disassembly, with each segment joined by positive alignment pins and threaded fastening. The tail employs a cruciform arrangement of four independently actuatable fin surfaces. Fig. 3 presents the CAD isometric and the fabricated article.

B. Validated Functions

The following functions have been demonstrated on the physical article: (a) Wing deployment—servo-actuated mid- body deployment from stowed to cruise configuration, with



(a) CAD isometric render showing the low-observable fuselage profile, mid- body wing bay, and cruciform tail arrangement.

(b) Fabricated PLA article with segmented construction, alignment pins, mid- body wing panels, and cruciform fin surfaces fitted.

Figure 3. ASM Tango aerodynamic demonstrator. Top: CAD isometric. Bottom: fabricated physical article at approximately 1.7 m length with 900 mm deployed wingspan.

mechanical transport lock; (b) Fin deflection—scripted and manual control of four fin surfaces for pitch and yaw authority demonstration; (c) Attitude telemetry—continuous roll, pitch, and yaw output from the Orientational layer to a monitoring terminal; (d) Visual inference—onboard neural inference for object detection and classification on the Parent layer; (e)Thermal-optical feed capture—simultaneous with-IR and without-IR image acquisition on the Child layer; pixel-level cross-validation logic under integration.

V. DEVELOPMENT STATUS

A. Component-Level Validation

Individual perception components have been validated in controlled environments. The Parent layer's edge AI accelerator and dual-channel optical sensor have been confirmed for detection and classification inference. The Child layer's thermal and IR-filtered channels have been validated for simultaneous feed acquisition; spatial co-registration between the two channels has been established. The Orientational layer's flight con- troller has been validated for continuous 6-DOF inertial state estimation and telemetry output.

Passive RF environmental monitoring has been validated using a Software-Defined Radio (SDR) receiver [11]. Fig. 4 shows a representative energy-domain acquisition of the Mum- bai Airport Control Tower VHF channel at 118.1 MHz, con- firming the platform's passive spectrum sensing capability. The SDR subsystem operates in receive-only mode; no transmission capability is present or intended.

Figure 4. Passive RF acquisition via Software-Defined Radio. Energy stream at 118.1 MHz (Mumbai Airport Control Tower VHF). Receive- only operation; no transmission capability.

B. Root Fusion Integration

Root fusion integration is the active development frontier. Software-in-the-Loop (SIL) simulation of the arbitration logic against simulated single-layer denial scenarios is ongoing. Hardware-in-the-Loop integration follows SIL completion.

C. TRL Assessment

ASM Tango is assessed at Technology Readiness Level 3: analytical and experimental critical function proof of concept. Individual perception components have been tested at the com- ponent level. TRL 4—functional verification in a laboratory environment—is the current programme target.

VI. DISCUSSION

A. Detection-Level IR Cross-Validation

Conventional countermeasure hardening targets the seeker: spatial discrimination and kinematic rejection reduce thermal decoy effectiveness at the seeker level [4]. ASM Tango's cross- validation operates upstream of any guidance decision. Be- cause the constraint is architectural—a detection cannot be actioned without thermal confirmation, regardless of inference confidence—the mechanism is not defeatable by modifications to the countermeasure that affect only the seeker's terminal response.

Pixel-level co-registration is the enabling requirement. Coarse decision-level comparison introduces the ambiguity that spatially aware countermeasures exploit. Eliminating that ambiguity at the detection level, rather than at the seeker level, is the architectural contribution.

B. GPS-Denied Navigation

IMU dead-reckoning accumulates drift over time. Visual feature tracking corrects this drift by providing position up- dates relative to observed scene structure [6]. The tight-coupled state estimation formulation targeted by the root fusion engine jointly optimizes both measurement sources, outperforming loose-coupling approaches in dynamic or feature-sparse environments [7].

C. MOOTL Scalability

As concurrent autonomous agent counts scales, per-operator cognitive load grows proportionally under human-in-loop paradigms. MOOTL removes this proportionality. A self- contained arbitration and guidance system directly enables multi-agent coordination architectures where operator over- sight at the individual platform level is operationally untenable [5].

D. Space Domain Applicability

GPS is permanently absent beyond Earth orbit. The visual- inertial fusion architecture validated through ASM Tango is transferable to lunar surface navigation and Martian autonomous platforms without modification to the fusion logic. The graceful-degradation principal maps directly onto deep- space communication latency and sensor anomaly environments where ground intervention during an anomaly is structurally impossible.

VII. CONCLUSION

ASM Tango presents a visual-inertial MOOTL guidance architecture for autonomous aerial platforms operating without GPS or datalink dependency. The platform integrates two in-dependent perception layers—a Visual layer in Parent-Child configuration with pixel-level thermal-optical cross-validation, and an Orientational layer providing continuous 6-DOF inertial state—arbitrated by a centralized root fusion engine designed for graceful degradation under single-layer denial.

The aerodynamic demonstrator has been fabricated and validated for wing deployment, fin actuation, and attitude telemetry. Individual Visual and Orientational layer components have been validated at the component level. Root fusion integration via SIL is the active development frontier.

The detection-level thermal-optical cross-validation mechanism represents an architectural contribution absent from existing low-cost demonstrator platforms. The visual-inertial fusion methodology transfers directly to lunar, Martian, and orbital autonomous GNC applications where GPS denial is the permanent operational baseline.

REFERENCES

[1] P. Zarchan, Tactical and Strategic Missile Guidance, 6th ed. AIAA Progress in Astronautics and Aeronautics, 2012.

[2] T. E. Humphreys, B. M. Ledvina, M. L.

Psiaki, W. O'Hanlon, and P. M. Kintner, "Assessing the spoofing threat: Development of a portable GPS civilian spoofer," in Proc. ION GNSS, Savannah, GA, Sep. 2008, pp. 2314–2325.

[3] R. Henry, "Intelligent Air-to-Surface Missile (ASM)— Student Demonstrator," in Project Overview Specification: B.Tech Robotics, 1st Batch, uGDX School of Technology, Atlas Skilltech University, Mumbai, India, Inter- nal Technical Specification, 2025.

[4] R. Decher and B. Landis, "Countermeasures against imaging infrared seekers," in Proc. SPIE Infrared Tech- nology and Applications, vol. 1689, 1992, pp. 258–270.

[5] U.S. Department of Defense, Directive 3000.09: Auton- omy in Weapon Systems, Nov. 2012, updated 2023.

[6] T. Qin, P. Li, and S. Shen, "VINS-Mono: A robust and versatile monocular visual-inertial state estimator," IEEE Trans. Robotics, vol. 34, no. 4, pp. 1004–1020, 2018.

[7] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, "On-manifold preintegration for real-time visual-inertial odometry," IEEE Trans. Robotics, vol. 33, no. 1, pp. 1– 21, 2017.

[8] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: An open- source SLAM system for monocular, stereo, and RGB-D cameras," IEEE Trans. Robotics, vol. 33, no. 5, pp. 1255– 1262, 2017.

[9] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proc. IEEE Conf. Computer Vision and Pattern Recogni- tion (CVPR), Las Vegas, NV, Jun. 2016, pp. 779–788.

[10] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields," IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 1, pp. 172–186, 2021.

[11] J. Mitola, "Software radio architecture," IEEE Commun. Mag., vol. 33, no. 5, pp. 26–38, May 1995.

[12] D. Simon, Optimal State Estimation: Kalman, H-Infinity, and Nonlinear Approaches. Wiley-Interscience, 2006.