SERBIAN JOURNAL OF ELECTRICAL ENGINEERING
Vol. 23, No. 2, June 2026, 321-334

UDC: 681.527.5:629.014.9 DOI: https://doi.org/10.2298/SJEE2602321V

Original scientific paper

Comparison of Theoretical and Experimental Pixel-on-Target Metrics for Early UAV Detection Using Multi-Sensor EO/IR System

Saša Vujić[1], Miloš Radisavljević[1], Miloš Pavlović[1], Dragana Perić[1], Mladen Veinović[2]

Abstract: This paper presents an analysis comparing theoretical model predictions and experimental measurements of the number of pixels on target (PoT) for unmanned aerial vehicles (UAVs) using multi-sensor electro-optical and infrared (EO/IR) systems. The primary objective is to evaluate the fidelity of theoretical range prediction models that use the number of PoT as a metric, comparing these calculated values against experimental field measurements. The data used in this analysis were acquired using both visible (VIS) and mid-wave infrared (MWIR) channels, with the target UAV positioned at known reference distances up to 900 m. The results quantitatively demonstrate the deviations between the two methodologies, providing critical insights into the practical limitations of performance prediction models for early drone detection systems operating under field conditions.

Keywords: Electro-optical and infrared systems, Unmanned aerial vehicles, Pixels on target, Mid-wave infrared, Range prediction, Sensor resolution.

1 Introduction

The proliferation of unmanned aerial vehicles (UAVs) has necessitated the development of robust early detection mechanisms. While traditional radar systems are highly effective for large aerial targets, they frequently exhibit limitations when tracking small, low-altitude UAVs with minimal radar cross-

[1]Vlatacom Institute of High Technologies, Belgrade, Serbia
sasa.vujic@vlatacom.com, https://orcid.org/0000-0002-5673-2230
milos.radisavljevic@vlatacom.com, https://orcid.org/0000-0003-4642-0126
milos.pavlovic@vlatacom.com, https://orcid.org/0000-0002-1710-7926
dragana.peric@vlatacom.com, https://orcid.org/0000-0002-2556-8212
[2]Singidunum University, Belgrade, Serbia
mveinovic@singidunum.ac.rs, https://orcid.org/0000-0001-6136-1895

Colour versions of the one or more of the figures in this paper are available online at https://sjee.ftn.kg.ac.rs sections. Consequently, multi-sensor electro-optical and infrared (EO/IR) systems have become indispensable for the reliable detection, classification, recognition, and identification of these asymmetric threats [1, 2]. While UAVs themselves are increasingly utilized as platforms equipped with EO sensors for tracking moving targets [3], the detection of incoming UAVs from stationary ground systems presents a unique set of challenges related to spatial resolution and contrast.

To accurately predict the effective range of these ground-based EO/IR systems, engineers rely heavily on fundamental optical theories and theoretical modeling [4]. The methodology for evaluating target acquisition has evolved significantly, transitioning from the traditional Johnson criteria based on resolvable cycles across a target [5], to the more sophisticated Targeting Task Performance (TTP) metric [6]. This evolution laid the groundwork for the Night Vision Integrated Performance Model (NV-IPM), which is currently the most widely used tool for predicting sensor range performance [7, 8]. The mathematical foundation of NV-IPM assumes that system performance can be described by linear shift invariant (LSI) subsystem components, which has been the standard for assessing thermal imager ranges [9].

However, modern target acquisition algorithms employ a different approach, using raw detector output intensity to evaluate the specific number of pixels on the target (PoT) [10]. Many manufacturers claim that, when combined with artificial intelligence (AI) detectors, their systems can operate at very low PoT values (e.g., around 4×4 pixels). Meanwhile, current algorithm studies suggest that a two-dimensional PoT area of 50 to 70 is adequate for basic detection [11]. Because non-linear target acquisition algorithms do not strictly adhere to the frequency-domain evaluations used by NV-IPM, an inherent discrepancy exists between idealized mathematical range predictions and the actual performance of sensors processing raw detector signals in the field [11]. Consequently, there is a critical need for experimental verification of such limits in practical field conditions.

This study aims to bridge the gap between theoretical modeling and experimental validation. By establishing a rigorous mathematical model for spatial resolution, this paper calculates the theoretical number of PoT for an observed UAV target across a visible (VIS) and a mid-wave infrared (MWIR) channel. These theoretical values are subsequently compared against experimental data gathered through stationary measurements at predefined reference distances up to 900 m. The discussion analyzes the discrepancies between the theoretical and experimental data, identifying the physical, optical, and atmospheric factors that degrade sensor performance at extended ranges.

2 Theoretical Framework and Mathematical Model

The theoretical evaluation of EO/IR systems relies on accurately modeling the projection of a physical target onto a discrete detector array. This section outlines the mathematical framework utilized to calculate the theoretical spatial resolution of the sensor system. By defining the optical parameters of the cameras and the characteristic dimensions of the UAV, a standardized baseline for the expected number of PoT is established. This theoretical baseline is crucial for subsequent comparison with experimental field data, allowing for the precise quantification of real-world performance degradation.

2.1 Sensor resolution and instantaneous field of view

The primary metric for calculating spatial resolution in electro-optical and infrared systems is the instantaneous field of view (IFOV). The IFOV represents the solid angle through which a single detector element perceives the environment, dictating the minimum resolvable footprint at a specific distance. Utilizing the small-angle approximation, the IFOV is a function of the detector pitch and the optical focal length:

$$\mathrm{IFOV} = \frac{d}{f}, \tag{1}$$

where d is the detector pitch (expressed in millimeters for unit consistency), and f is the focal length of the optical lens (in millimeters). The resulting IFOV is expressed in radians.

2.2 Mathematical model for pixels on target

To determine the theoretical number of pixels occupied by a target at a known distance, the physical projection of the target's image onto the discrete detector array must be calculated. For simplicity in range modeling, this is often expressed inversely by projecting the spatial extent of a single pixel (IFOV footprint) onto the target plane. The one-dimensional linear pixel count across the target's characteristic dimension is determined by dividing the target size by the projected spatial extent of a single pixel [11]:

$$\mathrm{POT}_{\mathrm{1D}} = \frac{W}{R \cdot \mathrm{IFOV}}, \tag{2}$$

where W is the characteristic dimension of the target (in meters), and R is the reference range from the sensor to the target (in meters). Assuming that the drone, as a target of interest, can be approximated by a two-dimensional bounding square for mathematical simplicity, the total two-dimensional pixel area on the target is the square of the linear pixel count.

$$\mathrm{POT}_{\mathrm{2D}} = \left(\frac{W}{R \cdot \mathrm{IFOV}}\right)^2. \tag{3}$$

3 Experimental Setup and Methodology

To validate the mathematical predictions and assess the practical limitations of the multi-sensor EO/IR platform, a controlled field experiment was designed and executed. This section describes the experimental methodology, including the technical specifications of the sensing equipment and the standardized procedures utilized for data acquisition. The primary objective of the experimental setup was to quantify the spatial resolution capabilities of the sensors by capturing imagery of a stationary UAV at precise reference distances. This approach ensures an accurate, experimental extraction of the PoT metrics under real-world atmospheric conditions.

3.1 VIS and MWIR sensor specifications

The experimental stationary measurements were conducted utilizing a multi-sensor EO/IR imaging platform. The system integrates a VIS channel and a MWIR channel, allowing for comparative spatial resolution analysis. The characteristics of MWIR and VIS channels are the following:

– The MWIR thermal sensor features a detector pitch of 10 µm, coupled with a continuous zoom lens capable of a maximum focal length of 1000 mm. The IFOV calculated based on (1) is 10 µrad.

– The VIS sensor utilizes a detector pitch of 4.1 µm, equipped with an optical lens providing a maximum focal length of 800 mm. The IFOV calculated based on (1) is 5.125 µrad.

Since the scope of this study is strictly focused on determining the limits of early UAV detection at maximum operational ranges, all subsequent measurements were conducted exclusively at the maximum focal length for both MWIR and VIS channels.

Fig. 1 – *Experimental setup of the multi-sensor EO/IR platform during the stationary field measurements.*

3.2 Stationary measurement procedure

The experimental data acquisition was designed to systematically evaluate the spatial resolution capabilities of the multi-sensor platform. The EO/IR multisensory imaging system was rigidly mounted on a stationary ground platform, leveled, and precisely aligned to minimize mechanical jitter.

A commercial UAV (DJI Mini 4 Pro) was utilized as a target. To ensure consistency with the theoretical characteristic dimension ($W = 0.18$ m) established in the mathematical model, the UAV was oriented to present a direct frontal aspect relative to the sensor payload. The target was commanded to maintain a stationary hover at four predefined reference waypoints: 100 m, 300 m, 600 m, and 900 m. The exact radial distance to each waypoint was verified using the UAV's internal telemetry and confirmed via a laser rangefinder to ensure positional accuracy. The measurements were conducted under clear sky conditions, with an ambient temperature of 18°C, relative humidity of 55%, and mild wind.

Fig. 2 – *Photo of the commercial drone (UAV) used in experiment - DJI Mini 4 Pro.*

At each reference distance, simultaneous imagery was captured utilizing both the VIS and MWIR channels. To prevent the data from being compromised by software-induced smoothing, the digital video streams were recorded directly from the sensor outputs prior to any display-level interpolation. Most standard display software utilizes interpolation to avoid blocky imagery, which artificially alters the true pixel count and blurs target edges. Therefore, analyzing the uncompressed, raw digital array data is desirable for an accurate experimental assessment.

The extraction of the experimental PoT metrics was conducted during post-processing. To capture the instantaneous hardware resolution limit, the calculations were obtained by processing a single frame (snapshot) for each reference distance. The imagery was analyzed frame-by-frame to determine the precise number of pixels occupied by the UAV silhouette. A human-assisted bounding box methodology was applied equally across both modalities to encompass the visual and thermal signatures of the target.

When extracting this data, the spatial sampling inherent to focal plane arrays must be considered. The precise location of a target's physical edge with respect to the detector array lattice is a random event. This spatial sampling creates an edge location ambiguity of approximately one detector pixel.

4 Results

This section presents a comparison of the theoretical and experimental spatial resolution metrics. To accurately evaluate the target acquisition performance of the multi-sensor EO/IR platform, the findings are divided into two distinct phases. First, Section 4.1 establishes the idealized mathematical baseline by calculating the theoretical PoT across the predefined reference distances (100 m, 300 m, 600 m, and 900 m). Subsequently, Section 4.2 describes the experimental PoT extracted from the uncompressed sensor data during the stationary field measurements. Comparing these two datasets allows for the precise quantification of performance degradation caused by real-world optical limitations, spatial sampling artifacts, and atmospheric attenuation.

4.1 Theoretical PoT values

To accurately establish a theoretical baseline for sensor performance, it is imperative to define the optical cross-section, or characteristic dimension (W), of the target UAV.

Rather than utilizing the maximum physical wingspan, which artificially inflates the theoretical pixels on target due to the empty space between the drone chassis and the propellers, the industry standard methodology dictates using the square root of the target's projected cross-sectional area [12]. This relationship is defined as:

$$W = \sqrt{Width \times Height} \,. \tag{4}$$

The DJI Mini 4 Pro, the commercial UAV utilized in this study, exhibits varying frontal profiles. The maximum measured frontal dimensions of the UAV in flight configuration are approximately 0.36 m in width and 0.09 m in height. Applying (4), this yields a characteristic dimension of $W = 0.18$ m. Alternatively, a slightly smaller profile excluding the extreme propeller sweep (0.30 m in width and 0.075 m in height) yields $W = 0.15$ m. To ensure the most rigorous baseline for the theoretical calculations, the maximum effective characteristic dimension of $W = 0.18$ m was established.

Based on the mathematical model defined in (1) through (3), the expected PoT values were calculated for reference distances of 100 m, 300 m, 600 m, and 900 m. The maximum optical zoom was applied to both the VIS (f_{max} = 800 mm) and MWIR (f_{max} = 1000 mm) channels to calculate the theoretical upper limit of the system's spatial resolution. **Table 1** presents the theoretical linear (POT_{1D}) and area (POT_{2D}) pixel counts.

Table 1

Calculated theoretical PoT values for VIS and MWIR channels at maximum focal length f_{max} (*for drone with dimension* W = 0.18 m).

Distance [m]	VIS Channel (f_{max}= 800 mm)		MWIR Channel (f_{max} = 1000 mm)	
	POT_{1D}	POT_{2D}	POT_{1D}	POT_{2D}
100	351.2	123341	180.0	32400
300	117.1	13712	60.0	3600
600	58.5	3422	30.0	900
900	39.0	1521	20.0	400

4.2 Experimental measurements at reference distances

The experimental PoT values derived from field measurements are summarized in **Table 2**. To ensure high fidelity in our data collection, the uncompressed digital video streams were analyzed directly from the sensor outputs.

Table 2

Experimental PoT values extracted for VIS and MWIR channels at maximum focal length f_{max} (*for drone with dimension* W = 0.18 m).

Distance [m]	VIS Channel (f_{max}= 800 mm)		MWIR Channel (f_{max} = 1000 mm)	
	POT_{1D}	POT_{2D}	POT_{1D}	POT_{2D}
100	314	98475	141	19754
300	102	10440	58	3312
600	53	2838	29	828
900	37	1334	18	324

The imagery was analyzed frame-by-frame to determine the precise number of pixels occupied by the UAV silhouette. A bounding box methodology was applied to encompass the visual and thermal signatures of the target. Unlike the theoretical model which assumes a square area, the experimental POT_{2D} values were calculated by multiplying the discrete pixel width and height of this rectangular bounding box, inherently accounting for the discrepancy caused by the drone's actual shape. During the extraction process, the spatial sampling inherent to focal plane arrays was accounted for, recognizing that the true target edge location is a random event. This spatial sampling creates an edge location ambiguity of approximately one detector pixel.

A side-by-side comparison of the UAV captured by both the VIS and MWIR channels is presented in Fig. 3. Each segment includes an enlarged inset of the drone for enhanced visualization.

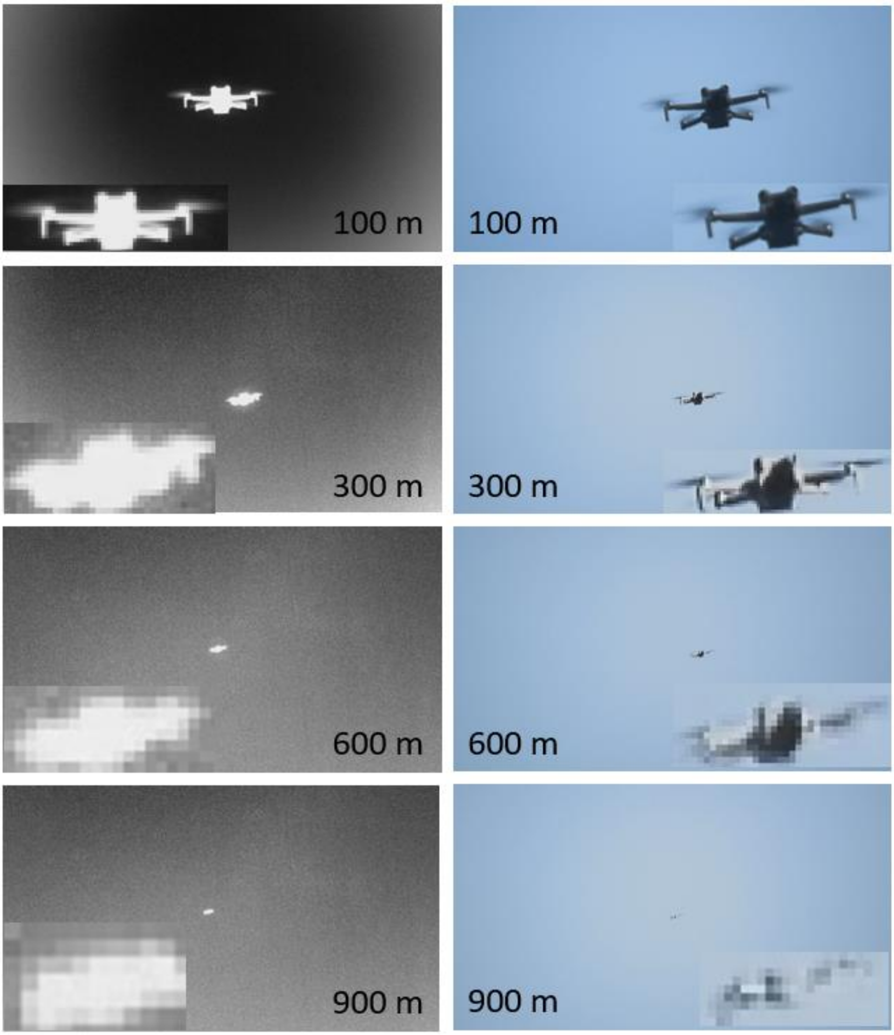

Fig. 3 – *Photos with a side-by-side comparison of the UAV captured by both the VIS and MWIR channels.*

5 Discussion

This section analyzes the discrepancies between the idealized mathematical predictions established in Section 4.1 and the experimental measurements extracted in Section 4.2. A comparison of the datasets reveals that the experimental PoT counts are consistently lower than the theoretical values across all reference distances. For instance, while the theoretical model predicted a linear PoT of 39.0 for the VIS channel at 900 m, the experimental measurement yielded only 37 pixels. Similarly, the MWIR channel dropped from a theoretical 20.0 to an experimental 18 pixels at the same distance. These deviations emphasize that the mathematical model, which assumes perfect point projection based purely on the IFOV, represents an unachievable ideal in field conditions.

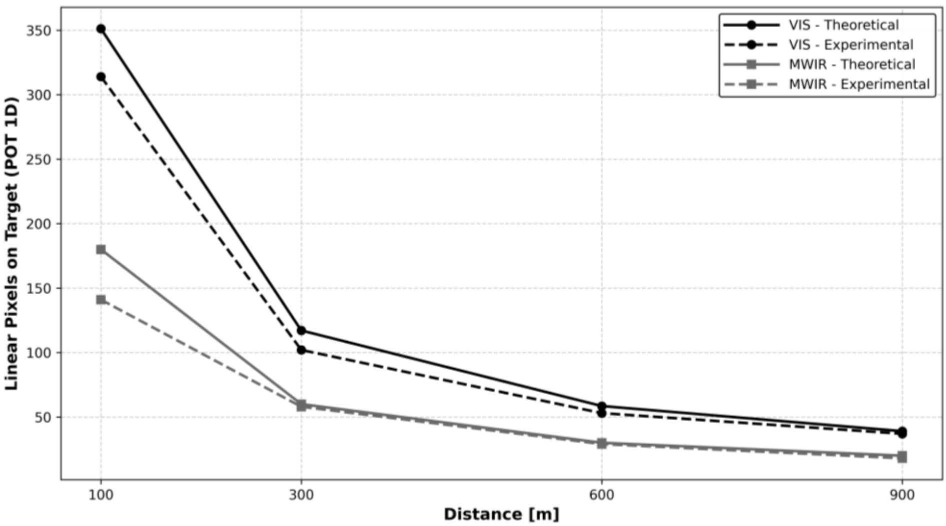

Fig. 4 – *Comparison of theoretical versus experimental linear pixels on target (POT1D) across varying reference distances for the VIS and MWIR imaging channels (Note: Solid lines denote theoretical calculations; dashed lines represent experimental measurements).*

To better understand the severity of spatial resolution degradation, the relative percentage deviation between the theoretical and experimental POT_{1D} was calculated, as illustrated in Fig. 5. While absolute pixel losses appear numerically smaller at extended ranges, the relative error remains significant, providing a clearer context for algorithmic performance.

The maximum relative degradation was observed in the MWIR channel at 100 m, presenting a –21.7% deviation. Furthermore, at 300 m, the MWIR channel exhibited a –3.3% deviation, with the experimental pixel count (58) falling short of the theoretical prediction (60). This confirms that thermal blooming from the rotors does not artificially inflate the target's optical cross-section in this scenario; instead, standard optical and environmental degradation governs the thermal signature.

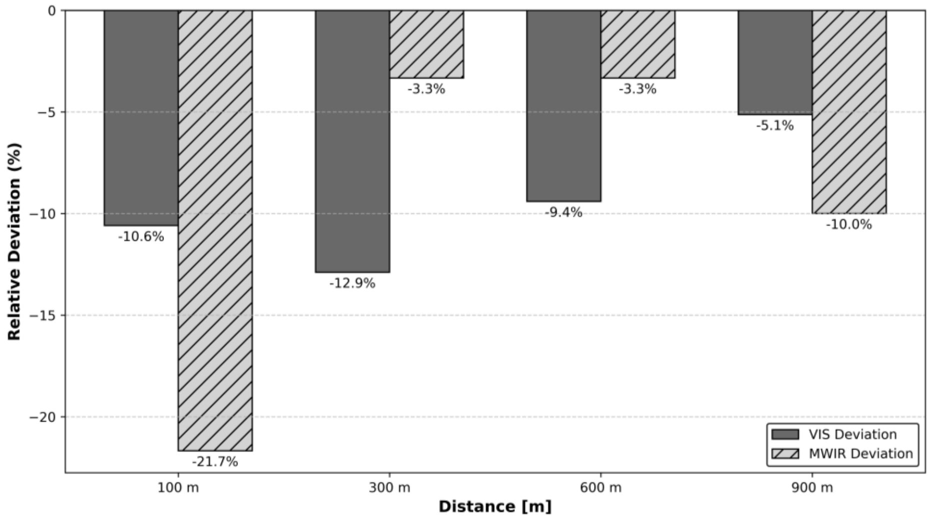

Fig. 5 – *Comparison of theoretical versus experimental linear pixels on target (POT1D) across varying reference distances for the VIS and MWIR imaging channels.*

A comparative analysis of the POT_{2D} results reveals that performance degradation is significantly more severe when evaluated in two dimensions, and the relative error fluctuates inconsistently across distances. This non-linear deviation stems from the inherent imprecision of the bounding box methodology when applied to an asymmetrical target. At closer ranges (100 m and 300 m), the sensor resolves the drone's highly irregular geometry in great detail. Constructing an experimental rectangular bounding box–which inevitably encompasses empty space between the chassis and propellers–and comparing its area to a theoretical idealized square inherently introduces geometric scaling errors. Conversely, at extended ranges (600 m and 900 m), the target subtends very few pixels. In these scenarios, spatial sampling ambiguity (±1 pixel error) and atmospheric blurring make the human-assisted visual thresholding highly subjective. Omitting even a single boundary pixel drastically reduces the total experimental area. Because area is a quadratic function, these methodological and geometric imprecisions compound geometrically, making POT_{2D} a highly sensitive and volatile metric in practical field assessments.

The overall degradation of the experimental spatial resolution can be attributed to a combination of optical, digital, and environmental factors:

– Optical blur and equivalent resolution: In practical sensor designs, the optical blur, often characterized by the Airy disk [4], is frequently comparable to or larger than the detector IFOV. As the focal length and target distance increase, the combination of the optics and detector modulation transfer functions (MTF) dictates that an "equivalent resolution" metric is often more appropriate than relying solely on the nominal detector pitch [4, 12]. Consequently, the target's signature is smeared across adjacent pixels, reducing the sharp edge contrast required to define the bounding box accurately.
– Spatial sampling and edge ambiguity: Digital focal plane arrays spatially sample the scene, and the exact location of the UAV's physical edge relative to the detector array lattice is fundamentally a random event. This spatial sampling creates an inherent edge location ambiguity of approximately one detector pixel. For a small target at 900 m, a ±1 pixel error on each edge represents a highly significant percentage drop in the total area POT_{2D}.
– Methodological and geometric variances: The experimental data extraction is subject to inherent mechanical and methodological tolerances. First, the theoretical model assumes the UAV presents a perfectly orthogonal frontal aspect. In reality, maintaining a stationary hover requires dynamic pitch and roll adjustments to counteract ambient wind, momentarily altering the projected optical cross-section. Future baseline measurements could utilize a tripod-mounted target to completely isolate these geometric variances caused by flight dynamics. Second, the theoretical characteristic dimension (W) approximates the target as an equivalent 2D square. The actual UAV possesses a highly irregular, non-rectangular geometry (a dense fuselage surrounded by thin rotor arms and empty space), complicating the direct correlation between a theoretical area and an experimental bounding box. Finally, the experimental PoT calculation relies on visual thresholding to define the bounding box; in the presence of optical blur and atmospheric scattering, identifying the precise boundary pixel introduces a margin of subjective measurement uncertainty.
– Atmospheric attenuation: While the theoretical model assumes a transparent atmosphere, real-world measurements are subject to atmospheric absorption and scattering. At distances of 600 m and 900 m, ambient humidity and aerosols severely degrade signal transmission. This lowers the apparent contrast of the UAV against the background, causing thinner profile elements (such as propeller arms) to vanish into background noise, thereby reducing the experimental PoT count. To comprehensively account for these complex environmental and thermodynamic imperfections, future theoretical baselines should incorporate advanced atmospheric radiative

transfer codes, such as MODTRAN [13], and detailed thermodynamic target profiling [14], which provide a much more accurate prediction than idealized IFOV models.

– Thermal contrast limitations (MWIR Specific): The MWIR channel exhibits unique discrepancies due to the thermodynamic properties of the target. The theoretical characteristic dimension ($W = 0.18$ m) assumes the entire drone chassis radiates uniformly. In practice, only the motors and battery generate significant thermal contrast (ΔT) against the ambient background. If the UAV body is in thermal equilibrium with the environment, the MWIR sensor primarily detects the localized hot spots. This effectively shrinks the optical cross-section of the target, resulting in a substantially lower experimental PoT compared to the VIS channel.

Ultimately, while the theoretical PoT provides a necessary baseline, the experimental data demonstrates that target acquisition algorithms must account for these real-world degradation factors to ensure reliable UAV detection at maximum operational ranges.

6 Conclusion

This paper presents a comprehensive comparison between theoretical predictions and experimental measurements of spatial resolution for early UAV detection using multi-sensor EO/IR systems. By calculating the theoretical PoT for a specific characteristic dimension ($W = 0.18$ m) and comparing these values against field measurements at distances up to 900 m, the practical limitations of idealized mathematical models were quantified.

The analysis demonstrates that while fundamental optical models, such as those based on the IFOV, are essential for establishing a baseline, they consistently overestimate the effective spatial resolution in field conditions. The experimental data reveal that theoretical models deviated from real-world measurements by up to 12.9% in the visible channel and 21.7% in the thermal channel. This degradation can be attributed to a combination of optical blur, spatial sampling ambiguity, and atmospheric attenuation. Furthermore, the MWIR channel exhibited distinct variations due to the non-uniform thermal contrast of the UAV, where colder chassis components effectively reduced the detectable optical cross-section compared to the theoretical dimensions.

Consequently, to achieve reliable early detection, classification, and tracking of asymmetric aerial threats, it is insufficient to rely solely on theoretical spatial resolution. The development of robust target acquisition algorithms must account for real-world environmental and optical degradation factors. Future work will focus on integrating these experimental correction factors into predictive models, as well as evaluating system performance across variable focal lengths to account for scenarios where maximum magnification is detrimental due to severe atmospheric turbulence, thereby enhancing the operational reliability of multi-sensor EO/IR platforms. Additionally, exploring multi-modal image fusion techniques to combine the high spatial resolution of the VIS channel with the high thermal contrast of the MWIR channel presents a highly relevant topic for further research to maximize the effective PoT.

7 Acknowledgments

The authors would like to express their gratitude to Vlatacom Institute of High Technologies Belgrade for providing the multi-sensor EO/IR platform and the technical support necessary to conduct the stationary field measurements. The work presented here was part of internal project P193.

8 References

[1] F. Svanström, F. Alonso-Fernandez, C. Englund: Drone Detection and Tracking in Real-Time by Fusion of Different Sensing Modalities, Drones, Vol. 6, No. 11, November 2022, p. 317.

[2] S. Liu, G. Li, Y. Zhan, P. Gao: MUSAK: A Multi-Scale Space Kinematic Method for Drone Detection, Remote Sensing, Vol. 14, No. 6, March 2022, p. 1434.

[3] L. Zhang, X. Na, J. Yan, F. Li: Detection, Tracking and Redetection of Moving Target by UAV, Proceedings of the 6th International Conference on Electronic Engineering and Informatics (EEI), Chongqing, China, June 2024, pp. 1349 – 1356.

[4] G. C. Holst: Electro-Optical Imaging System Performance, 6th Edition, JCD Publishing and SPIE, Winter Park, Bellingham, 2017.

[5] T. A. Sjaardema, C. S. Smith, G. C. Birch: History and Evolution of the Johnson Criteria, SANDIA REPORT, Sandia National Laboratories, Albuquerque, 2015.

[6] R. H. Vollmerhausen, E. Jacobs: The Targeting Task Performance (TTP) Metric: A New Model for Predicting Target Acquisition Performance, Technical Report AMSEL-NV-TR-230, Fort Belvoir, 2004.

[7] B. Teaney, J. Reynolds: Next Generation Imager Performance Model, Proceedings of SPIE, Vol. 7662, April 2010, p. 76620F.

[8] R. Vollmerhausen: Night Vision Integrated Performance Model: Impact of a Recent Change on the Model's Predictive Accuracy, Optics Express, Vol. 24, No. 21, October 2016, pp. 23654 – 23666.

[9] D. Perić, B. Livada, M. Perić, S. Vujić: Thermal Imager Range: Predictions, Expectations, and Reality, Sensors, Vol. 19, No. 15, August 2019, p. 3313.

[10] H. Fang, M. Xia, G. Zhou, Y. Chang, L. Yan: Infrared Small UAV Target Detection Based on Residual Image Prediction via Global and Local Dilated Residual Networks, IEEE Geoscience and Remote Sensing Letters, Vol. 19, 2022, p. 7002305.

[11] G. C. Holst, A. Mahalanobis: Comparing Pixels on Target with NV-IPM Range Predictions, Proceedings of SPIE, Vol. 12533, June 2023, p. 1253302.

[12] J. C. Leachtenauer, R. G. Driggers: Target and Environmental Considerations, Ch. 7, Surveillance and Reconnaissance Imaging Systems – Modeling and Performance Prediction, Artech House, Boston, London, 2001.

[13] A. Berk, P. Conforti, R. Kennett, T. Perkins, F. Hawes, J. van den Bosch: MODTRAN6: A Major Upgrade of the MODTRAN Radiative Transfer Code, Proceedings of SPIE, Vol. 9088, June 2014, p. 90880H.

[14] P. A. Jacobs: Thermal Infrared Characterization of Ground Targets and Backgrounds, 2nd Edition, SPIE Press, Bellingham, Washington, 2006.