

**AI-Driven Detection and Mitigation of Urban Heat Island
Effects Using Vision-Language Models and Mobile IoT
Sensing**

Subadhra Kankanamalage Sachinthaka Ayeshmantha

IT21219320

B.Sc. (Hons) degree in Information Technology Specialized in Software Engineering

Department of Software Engineering

Sri Lanka Institute of Information Technology

Sri Lanka

August 2025

**AI-Driven Detection and Mitigation of Urban Heat Island
Effects Using Vision-Language Models and Mobile IoT
Sensing**

Subadhra Kankanamalage Sachinthaka Ayeshmantha

IT21219320

*Dissertation submitted in partial fulfillment of the requirements for the Bachelor of
Science (Hons) in Information Technology Specialized in Software Engineering*

Department of Software Engineering

**Sri Lanka Institute of Information Technology
Sri Lanka**

August 2025

DECLARATION

I declare that this is my own work and this dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. Also, I hereby grant to Sri Lanka Institute of Information Technology, the nonexclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Name	Student ID	Signature
Ayeshmantha S K S	IT21219320	

The above candidate has carried out research for the bachelor's degree Dissertation under my supervision.

Mr. Vishan Jayasinghearachchi
Supervisor

Date

Abstract

Urban Heat Islands (UHIs) significantly impact urban climates, increasing local temperatures and exacerbating environmental and health issues. Reliable and cost-effective methods for ground-level thermal measurement are therefore critical to support sustainable city planning.

In this study, we address the problem of **localizing and measuring the temperature of segmented urban objects** identified from user-captured images.

To solve this, we propose a **depth-aware object localization and thermal sensing approach** on a mobile IoT platform. The system first navigates toward target objects using GPS coordinates, gyro orientation, and the original camera angle from which the image was taken. Once at the approximate location, the onboard camera is rotated across axes while applying **SuperGlue feature matching** to re-identify the object. A **depth map** is then generated to estimate distance, allowing the thermal sensor's **8×8 grid** to be accurately mapped onto the object, and the correct cell temperature is selected.

The method was implemented on an **ESP32-based robotic platform** equipped with a thermal sensor, GPS, IMU, and servo-mounted camera, with communication handled through WebSocket protocols. Experiments demonstrated that the system can reliably re-identify objects and capture representative temperature readings. These results highlight a scalable, low-cost contribution to fine-grained urban heat monitoring, with potential applications in environmental assessment and smart city planning.

Keywords: Object Localization, Depth Map, IoT, Thermal Sensing, SuperGlue, Urban Heat Island

Acknowledgement

First and foremost, I would like to express my sincere gratitude to my supervisor, Mr. Vishan Jayasinghearachchi, and my co-supervisor, Miss Kaushalya Rajapakse, for their continuous guidance, support, and encouragement throughout the course of this research. Their constructive feedback and valuable insights greatly contributed to the successful completion of this work.

I am also deeply grateful to the external supervisor, Dr. Rajitha de Silva of the University of Lincoln, for providing expert advice, technical direction, and critical suggestions that significantly strengthened this study.

My appreciation is extended to the Department of Software Engineering, Sri Lanka Institute of Information Technology, for providing the academic environment to carry out this research.

Finally, I would like to acknowledge the contributions and collaboration of my project team members, whose efforts, ideas, and commitment were vital to the progress of the overall research project.

Contents

Declaration	i
Abstract	ii
Acknowledgement	iii
List of Figures	vi
List of Tables	vii
List of Abbreviations	viii
1 Introduction	1
1.1 Background & Literature Review	4
1.1.1 Mechanisms and Impacts of UHI Formation	4
1.1.2 Traditional UHI Monitoring Approaches	5
1.1.3 Advancements in IoT and Mobile Sensing	5
1.1.4 Integration of Computer Vision and AI	6
1.1.5 Thermal Sensing and Wireless Communication	7
1.1.6 Edge Computing and Real-Time Processing	7
1.1.7 Recent Developments and Future Directions	7
1.2 Research Gap	8
1.2.1 Integration Gap Between Visual Analysis and Thermal Validation	8
1.2.2 Systematic Analysis of Current Approach Limitations	8
1.2.3 Accessibility and Cost Barriers	9
1.2.4 Spatial and Temporal Resolution Trade-Offs	9
1.2.5 Critical Research Gaps Identified	10
1.3 Research Problem	11
1.3.1 Problem Statement	11
1.3.2 Technical Problem Definition	12
1.3.3 Research Questions	13
1.3.4 Scope and Constraints	14
1.4 Research Objectives	14
1.4.1 Main Objective	15
1.4.2 Sub-objectives	15
2 Methodology	17
2.1 System Architecture and Hardware Configuration	19
2.2 Navigation and Object Localization	23
2.3 Data Collection and Transmission	33
2.4 Commercialization Aspects	36

2.5	Testing and Implementation	38
2.5.1	Experimental Setup	39
2.5.2	Metrics and Definitions	40
2.5.3	Thresholds and Decision Logic	42
2.5.4	Test Procedures	42
2.5.5	Robustness and Stress Testing	44
2.5.6	Data Analysis and Reporting	45
2.5.7	Acceptance Criteria	45
2.5.8	Implementation Notes	46
3	Results and Discussion	47
3.1	Results	47
3.1.1	Trial Overview	47
3.1.2	Navigation and Heading Control	47
3.1.3	Vision-Based Object Localization	48
3.1.4	Thermal Sensing and ROI Mapping	49
3.1.5	Communication Reliability	49
3.1.6	Ablation and Robustness Observations	50
3.2	Research Findings	50
3.3	Discussion	52
3.3.1	Implications for UHI Monitoring	52
3.3.2	Comparison to Acceptance Criteria and Prior Work	52
3.3.3	Limitations and Opportunities	53
3.4	Contribution	53
4	Conclusion	54
Reference List		58

List of Figures

1.1	Illustration of the Urban Heat Island (UHI) effect, showing temperature variations between rural, suburban, and urban areas with specific temperature differentials.	2
2.1	End-to-end workflow: from segmented images with GPS metadata to coordinate estimation, IoT navigation, visual verification, and thermal data collection.	18
2.2	Assembled IoT platform with ESP32 Dev Kit, ESP32-CAM, AMG8833 thermal sensor, MPU6050 IMU, GPS module, and motor chassis.	23
2.3	Hardware architecture of the IoT platform showing sensor, actuator, controller, and communication modules.	23
2.4	SuperGlue feature matching between ESP32-CAM view (left) and segmented reference (right).	26
2.5	The HeatScape Control Panel in Autonomous Mode. This view is used for monitoring the IoT platform's status, viewing live telemetry, and verifying object alignment via the camera feed.	31
2.6	The Control Panel in Manual Override Mode. This interface grants the operator direct control over the platform's movement and camera, which is essential for error recovery and navigating complex environments.	32
2.7	Server-side architecture of the IoT system, illustrating data flow from IoT device to operator interface.	36

List of Tables

1.1	Comparative Analysis of UHI Monitoring Methods	10
2.1	Acceptance criteria for core subsystems	45
3.1	Summary of contribution	53

List of Abbreviations

Abbreviation	Description
UHI	Urban Heat Island
IoT	Internet of Things
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
IMU	Inertial Measurement Unit
ROI	Region of Interest
LST	Land Surface Temperature
RTK	Real-Time Kinematic (GNSS)
PPP	Precise Point Positioning
EKF	Extended Kalman Filter
SLAM	Simultaneous Localization and Mapping
RANSAC	Random Sample Consensus
MAE	Mean Absolute Error
RMSE	Root Mean Square Error
IQR	Interquartile Range
RTT	Round-Trip Time
PLR	Packet Loss Rate
CDF	Cumulative Distribution Function
GCP	Ground Control Point
HDOP	Horizontal Dilution of Precision
UUID	Universally Unique Identifier
MQTT	Message Queuing Telemetry Transport
MQTT-SN	MQTT for Sensor Networks
CoAP	Constrained Application Protocol
CNN	Convolutional Neural Network
VLM	Vision-Language Model
QUIC	Quick UDP Internet Connections

Introduction

Urban areas are experiencing unprecedented growth, with over 68% of the global population expected to live in cities by 2050 [1]. This rapid urbanization has intensified the urban heat island (UHI) effect, a phenomenon in which urban areas become significantly warmer than their rural surroundings. The UHI effect is not merely an academic concern; it has wide-ranging implications for public health, energy consumption, and environmental sustainability. Elevated urban temperatures exacerbate heat stress, increase electricity demand for cooling, and contribute to air pollution, placing additional burdens on already strained urban infrastructures.

The fundamental drivers of UHI formation are well-established: reduced vegetation cover, increased impervious surfaces that absorb and retain heat, anthropogenic heat emissions from vehicles and buildings, and altered wind patterns due to urban morphology [1]. These factors can create temperature differences of 1–10°C between urban cores and surrounding rural areas, with some extreme cases showing differences up to 12°C during peak conditions [2]. Figure 1.1 illustrates the typical temperature variations between rural, suburban, and urban areas, highlighting how urban morphology contributes to heat accumulation. Such temperature variations have profound implications for human health, with studies demonstrating significant increases in heat-related mortality rates, particularly among vulnerable populations, including the elderly and those with pre-existing health conditions [2]. Additionally, UHIs can exacerbate energy consumption for cooling, increase greenhouse gas emissions, and impact water quality through enhanced urban runoff, creating a cascading set of environmental challenges.

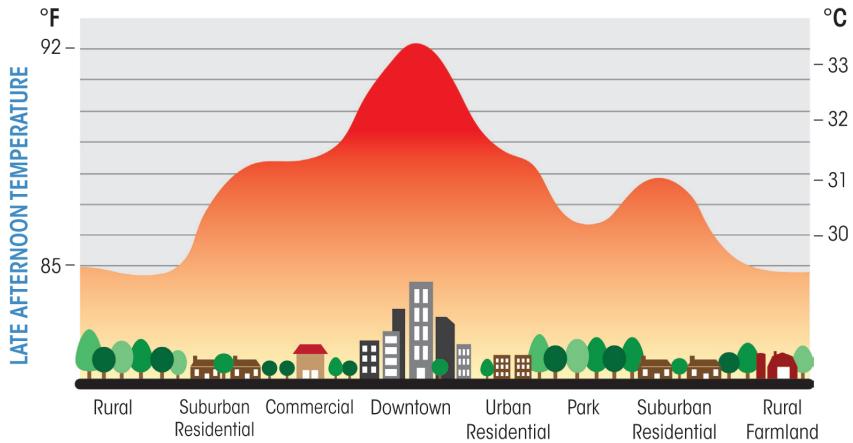


Figure 1.1: Illustration of the Urban Heat Island (UHI) effect, showing temperature variations between rural, suburban, and urban areas with specific temperature differentials.

Despite significant advances in remote sensing, satellite imagery, and fixed sensor networks, current UHI detection and mitigation approaches face critical limitations. Remote sensing can provide broad spatial coverage but often lacks the spatial and temporal resolution required for detailed urban planning decisions. Fixed sensors deliver high-accuracy measurements but are typically limited to small areas and require costly installations, making comprehensive coverage infeasible. Moreover, these traditional methods rarely integrate ground-truth measurements with image-based analyses, limiting their ability to validate hotspot predictions in real time.

To address these challenges, this research proposes a mobile IoT sensing platform integrated with AI-driven image analysis for UHI monitoring. The system consists of four main components:

1. **AI-Driven Image Analysis:** A mobile application that captures urban images, segments them into meaningful surface categories, and identifies potential thermal hotspots. This component provides high-resolution spatial information about areas prone to excessive heating.
2. **Mobile IoT Sensing Platform:** Equipped with thermal sensors, GPS navigation,

and computer vision capabilities, this platform autonomously navigates to target coordinates to perform ground-truth temperature measurements, providing real-time validation of image-derived predictions.

3. **VLM-Based Detection and Solution Suggestion:** An analytical framework that integrates temperature and visual data to identify hotspots, evaluate mitigation strategies, and recommend actionable urban planning interventions.
4. **Simulation Environment:** A software simulator to test system performance under varying urban conditions, enabling controlled validation before real-world deployment.

This research specifically focuses on the development and validation of the mobile IoT sensing platform. By bridging the gap between high-resolution visual analysis and accurate thermal measurements, the system enables precise hotspot detection and supports data-driven urban planning. The mobile IoT component offers several key advantages: it is scalable, cost-effective, and capable of dynamic, targeted temperature sensing, which allows for efficient monitoring across diverse urban landscapes. The integration of GPS navigation, IMU-based orientation, and vision-based object alignment ensures that the platform can autonomously reach and align with specific urban surfaces, providing reliable data for predictive modelling and mitigation planning.

Ultimately, this research contributes to a holistic approach to UHI monitoring and mitigation by providing a validated, integrated platform that combines advanced imaging, autonomous sensing, and actionable analytics. By offering accessible and accurate UHI data, this work aims to support sustainable urban development, improve climate resilience, and enable evidence-based decision-making for city planners and policy-makers.

1.1 Background & Literature Review

Urban Heat Islands (UHIs) represent a significant environmental challenge for modern cities. First documented by Luke Howard in the early 19th century when he observed that London was consistently warmer than its surrounding countryside [2], UHI research has since evolved to examine the complex interactions between urban morphology, surface materials, and thermal processes. The intensity of UHI effects varies depending on city size, population density, land use, and seasonal conditions. Large metropolitan areas often experience nighttime temperature differences of 3-5°C compared to rural areas, creating heterogeneous thermal landscapes [1].

1.1.1 Mechanisms and Impacts of UHI Formation

The primary drivers of UHI formation include:

- **Reduced Vegetation Cover:** Urbanization replaces natural vegetation with impervious surfaces, reducing evapotranspiration and increasing heat absorption.
- **Increased Impervious Surfaces:** Asphalt, concrete, and dark roofing materials have low albedo, absorbing more solar radiation and releasing it slowly.
- **Anthropogenic Heat Emissions:** Transportation, industrial processes, and energy consumption release additional heat into urban environments.
- **Altered Wind Patterns:** Urban morphology disrupts natural airflow, reducing heat dispersion [3].

These factors contribute to elevated urban temperatures, resulting in:

- **Public Health Risks:** Higher temperatures increase heat-related illnesses and mortality, especially among vulnerable populations [2].

- **Increased Energy Consumption:** Cooling demand rises, placing stress on energy infrastructure.
- **Infrastructure Deterioration:** Prolonged heat accelerates wear on buildings and roads.
- **Environmental Degradation:** Ecosystems are impacted, leading to reduced biodiversity and altered habitats.

1.1.2 Traditional UHI Monitoring Approaches

Conventional UHI monitoring methods include:

- **Satellite Remote Sensing:** Landsat, MODIS, and ASTER provide large-area land surface temperature (LST) maps [4, 5]. Limitations include coarse spatial resolution (30–1000 m), infrequent revisits, cloud interference, and indirect correlation with human thermal comfort.
- **Fixed Meteorological Stations:** Ground-based stations offer high temporal resolution but require significant infrastructure investment and provide limited spatial coverage.
- **Aircraft-Based Thermal Imaging:** Provides intermediate spatial resolution but incurs high costs, infrequent deployment, and weather constraints [4].

While effective at a macro scale, these methods struggle with real-time, high-resolution, and cost-effective monitoring, motivating the adoption of mobile IoT-based solutions.

1.1.3 Advancements in IoT and Mobile Sensing

The emergence of low-cost IoT platforms has transformed urban environmental monitoring [6, 7]. Mobile sensing platforms mounted on vehicles, bicycles, or drones

enable:

- **Real-Time Data Acquisition:** Immediate transmission and analysis of sensor data.
- **High Spatial Resolution:** Ability to capture fine-grained variations across urban landscapes.
- **Cost-Effectiveness:** Low-cost sensors (\$10–\$100) allow dense deployment, enhancing coverage.

Challenges remain in ensuring data accuracy and integrating IoT sensing with advanced analytics for actionable insights.

1.1.4 Integration of Computer Vision and AI

Computer vision and AI provide critical enhancements for urban analysis. Convolutional Neural Networks (CNNs) have shown success in:

- Land use classification
- Vegetation mapping
- Surface material identification

Semantic segmentation enables pixel-level classification, helping identify heat-contributing urban elements such as asphalt, concrete, metal roofs, and vegetation areas [8]. Vision-Language Models (VLMs) further enhance interpretability by linking visual data with contextual information. However, a critical gap remains: AI predictions of heat sources require **physical thermal validation**.

1.1.5 Thermal Sensing and Wireless Communication

Affordable uncooled thermal cameras such as the MLX90640 provide $\pm 1.5^{\circ}\text{C}$ accuracy, 32×24 pixel resolution, and 0.1°C sensitivity, enabling fine-scale thermal mapping [9, 10]. Wireless protocols like MQTT support low-latency, scalable, and reliable data transfer between mobile sensors and central servers [11, 12].

1.1.6 Edge Computing and Real-Time Processing

Edge computing reduces latency and bandwidth use by processing data locally on the device [13, 14]. This enables:

- Local decision-making and anomaly detection
- Continuous operation in areas with intermittent connectivity
- Autonomous processing of sensor and visual data

Microcontroller-based platforms like ESP32 or single-board computers like NVIDIA Jetson can be chosen based on computational needs and power constraints [15, 16].

1.1.7 Recent Developments and Future Directions

Recent research highlights innovative UHI monitoring approaches:

- **Low-Cost Citizen Sensing:** Devices integrated with mobile apps allow citizens to measure urban temperature, surface temperature, and humidity.
- **Predictive AI Models:** Machine learning and deep learning models predict UHI intensity and distribution.
- **Mobile Platforms:** Sensor-equipped bicycles and vehicles capture intracity variations in human thermal comfort [17].

These trends underscore the potential of combining IoT, AI, and mobile sensing to provide real-time, high-resolution, and actionable urban heat data, directly motivating the development of the proposed mobile IoT platform.

1.2 Research Gap

Despite substantial progress in individual technologies for UHI monitoring and mitigation, existing approaches face multiple critical limitations that prevent effective, scalable implementation. A systematic analysis of current methods highlights the need for integrated solutions capable of providing real-time, high-resolution, and context-aware urban heat monitoring.

1.2.1 Integration Gap Between Visual Analysis and Thermal Validation

Modern computer vision systems can identify urban features, materials, and potential thermal hotspots from aerial or street-level imagery. However, these systems cannot confirm actual thermal conditions without physical sensing. Conversely, thermal sensing platforms provide accurate temperature measurements but lack contextual understanding of the surrounding urban features. This disconnect between visual identification and thermal validation limits the actionable value of collected data. Most current systems also lack mechanisms for dynamic, real-time sensing, resulting in uncertainty about whether measured temperatures correspond to the intended thermal features.

1.2.2 Systematic Analysis of Current Approach Limitations

Remote Sensing Limitations: Satellite and aerial thermal imagery provide large-area coverage but suffer from coarse spatial and temporal resolution. For example, Landsat thermal bands have 100-meter resolution with 16-day revisit intervals, while MODIS offers daily coverage at 1-kilometer resolution [5]. These limitations prevent detection

of street-level thermal hotspots and impede timely response during heat events.

Ground-Based Sensor Constraints: Fixed meteorological stations provide accurate, high-frequency measurements at specific points, but achieving city-wide coverage is prohibitively expensive. Typical professional stations cost between \$5,000 and \$15,000, and covering a medium-sized city could require 200-500 stations, representing an investment of \$1-7.5 million [3].

Mobile Sensor Limitations: Mobile sensor networks enhance spatial coverage but frequently operate on predetermined routes and rarely integrate with visual analysis systems. Consequently, temperature readings cannot be reliably correlated with specific urban features [7].

AI-Driven Model Limitations: While AI and machine learning models are useful for predictive analysis, they often rely on historical datasets and cannot dynamically adapt to environmental changes. Computer vision can identify potential thermal hotspots, but without ground-truth temperature validation, predictive models cannot reliably guide mitigation strategies [8].

1.2.3 Accessibility and Cost Barriers

Advanced UHI assessment services and consulting are costly and often inaccessible to smaller municipalities or developing urban regions. The high capital expenditure for infrastructure and monitoring systems exacerbates environmental inequities, leaving vulnerable communities without access to effective heat mitigation solutions.

1.2.4 Spatial and Temporal Resolution Trade-Offs

Existing monitoring approaches involve inherent trade-offs between spatial resolution, temporal frequency, and operational cost:

Table 1.1: Comparative Analysis of UHI Monitoring Methods

Method	Spatial Resolution	Temporal Resolution	Coverage	Cost
Satellite Remote Sensing	30–1000m	1–16 days	Global	Medium
Fixed Weather Stations	Point	Minutes	Limited	High
Aircraft Thermal Imaging	1–10m	Infrequent	Regional	Very High
Mobile IoT Sensing	Point	Real-time	Route-dependent	Medium
AI Image Analysis	Variable	Processing delays	Flexible	Low
Proposed Integrated System	Sub-meter	Real-time	Targetable	Medium

No current method simultaneously achieves fine spatial resolution (sub-meter), real-time monitoring, flexible targetable coverage, and reasonable cost necessary for actionable UHI mitigation.

1.2.5 Critical Research Gaps Identified

The review of current literature and technologies reveals five major gaps motivating the proposed research:

- **Gap 1 - Visual-Thermal Validation Disconnect:** Existing systems cannot dynamically prioritize measurement locations based on thermal risk analysis nor validate thermal readings against visual expectations.
- **Gap 2 - Spatial Targeting Inefficiency:** Most monitoring systems collect data over broad, untargeted areas, resulting in inefficient resource usage and less actionable insights for urban planners.
- **Gap 3 - Multi-Modal Integration Absence:** Current solutions do not integrate visual analysis, autonomous navigation, thermal sensing, and real-time validation into a single cohesive framework.
- **Gap 4 - Real-Time Decision Support Limitation:** Existing approaches provide delayed, sparse, or costly data that cannot support real-time decision-making during extreme heat events.

- **Gap 5 - Lack of Automated Insight Generation:** Most systems deliver raw thermal data without contextual interpretation or actionable recommendations, necessitating expert intervention and limiting scalability.

By addressing these gaps, future UHI monitoring systems can achieve high-resolution, cost-effective, and context-aware thermal mapping, enabling proactive urban planning and mitigation strategies.

1.3 Research Problem

The primary research problem addressed in this study is the development of an autonomous mobile IoT sensing platform capable of providing accurate ground-truth thermal validation of urban objects identified through external image analysis systems. Existing mobile sensing approaches face significant limitations that restrict their effectiveness, motivating the need for an integrated, multi-modal solution that can operate in real-world urban environments.

1.3.1 Problem Statement

Current mobile IoT sensing approaches for Urban Heat Island (UHI) monitoring exhibit several critical limitations:

- **Limited Autonomous Capability:** Existing systems typically follow predetermined routes or require manual control, lacking the ability to autonomously navigate to specific coordinates and align precisely with measurement targets derived from image analysis.
- **Integration Challenges:** Mobile platforms often operate in isolation without standardized communication protocols for coordination with image analysis, Vision-Language Models (VLMs), or simulation systems, reducing their effectiveness in comprehensive monitoring workflows.

- **Navigation and Positioning Accuracy:** Low-cost mobile platforms frequently fail to achieve the spatial precision required for object-level thermal measurements, particularly in urban environments where GPS signals may be obstructed or reflected.
- **Measurement Validation Limitations:** Existing systems lack mechanisms to verify that thermal readings correspond to the intended objects, introducing uncertainty in spatial correlation between thermal data and urban features.
- **Decision-Making Support Deficiency:** Urban planners, public health officials, and community organizations lack accessible tools that provide real-time, validated thermal data at sufficient spatial and temporal resolution for intervention planning and heat emergency response.

1.3.2 Technical Problem Definition

The specific technical challenge can be formulated as follows:

How can a mobile IoT sensing platform equipped with autonomous navigation and thermal sensing capabilities provide accurate, validated temperature measurements of urban objects identified through image analysis while maintaining cost-effectiveness and scalability?

This problem encompasses several interrelated technical challenges:

- **Challenge 1 - Multi-Modal Data Integration:** Developing algorithms that effectively combine thermal measurements, GPS coordinates, inertial sensor data, and visual verification to create spatially accurate thermal validations.
- **Challenge 2 - Autonomous Object Localization:** Implementing vision-based navigation systems capable of reliably identifying and aligning with urban objects identified via image segmentation, while accounting for lighting changes, viewpoint variations, and environmental conditions.

- **Challenge 3 - Real-Time Processing:** Ensuring computer vision algorithms (e.g., feature matching, object alignment) operate in real-time on resource-constrained mobile platforms without compromising measurement accuracy.
- **Challenge 4 - Thermal Sensor Calibration:** Establishing calibration procedures that account for emissivity variations, atmospheric effects, and sensor drift to maintain consistent measurement accuracy across diverse urban materials and environmental conditions.
- **Challenge 5 - System Integration:** Designing communication architectures that enable seamless coordination between the mobile IoT platform and other system components while ensuring reliability and fault tolerance.

1.3.3 Research Questions

The study addresses the following key questions:

- **RQ1:** How can low-cost IoT hardware be configured to achieve autonomous navigation to specific GPS coordinates with sufficient accuracy for precise urban object thermal measurement?
- **RQ2:** What system architecture enables reliable coordination between mobile IoT sensing platforms and external components (image analysis, VLM-based analysis, and simulation systems) through standardized communication protocols?
- **RQ3:** How can computer vision techniques be implemented on resource-constrained mobile platforms to ensure accurate alignment between thermal sensors and intended measurement targets?
- **RQ4:** What level of thermal measurement accuracy and system reliability can be achieved with low-cost IoT components while maintaining cost-effectiveness

for widespread deployment?

1.3.4 Scope and Constraints

This research focuses on a ground-based mobile IoT sensing platform for outdoor urban environments with accessible GPS signals. Key scope and constraints include:

- Controlled operation in areas with minimal pedestrian traffic and low security concerns.
- Daytime operation under clear weather conditions.
- Targeted measurement of static urban objects (buildings, pavements, and infrastructure) rather than dynamic elements (vehicles or people).
- Ground-based tank chassis design optimized for sidewalks and pedestrian areas, prioritizing measurement accuracy and autonomous navigation over rapid area coverage.
- Autonomous reception of target coordinates from external image analysis systems, with execution of precise thermal measurements at specified locations.

By addressing these challenges, the research aims to develop a cost-effective, scalable, and validated mobile sensing framework for urban heat monitoring, bridging the gap between visual analysis and ground-truth thermal measurement.

1.4 Research Objectives

The overarching objective of this research is to design, implement, and validate a mobile IoT sensing platform that enhances the detection and analysis of the Urban Heat Island (UHI) effect. Unlike conventional approaches that rely primarily on remote sensing or static sensor networks, this research introduces a dynamic, mobile solution

capable of autonomously navigating to specified GPS coordinates and collecting real-time, ground-truth thermal measurements. By directly validating the outputs of external image analysis systems, the platform ensures that predicted urban hotspots are supported with accurate, localized temperature data. This integration of IoT mobility, autonomous navigation, and vision-based object alignment establishes a scalable and cost-effective framework for reliable UHI monitoring, thereby contributing to more data-driven and sustainable urban planning.

1.4.1 Main Objective

The overarching objective of this research is to develop and validate a mobile IoT sensing platform that can provide accurate, real-time thermal data for validating visually segmented urban surfaces. By directly linking thermal measurements with AI-driven image analysis, the system addresses a critical gap in existing Urban Heat Island (UHI) monitoring approaches, which often rely on remote sensing data without sufficient ground-truth validation. This research therefore contributes a scalable, low-cost, and reliable method for localized hotspot detection, making UHI assessment more actionable for urban planners and policymakers.

The main objective can be summarized as follows:

To design and implement a mobile IoT-based sensing platform that accurately collects, validates, and integrates thermal measurements with segmented image data to enable precise and cost-effective monitoring of localized Urban Heat Island hotspots.

1.4.2 Sub-objectives

To achieve the above objective, the research is guided by the following specific sub-objectives:

1. **Precise Mapping of Thermal Data:** Develop a robust methodology to accurately map real-time temperature readings from the IoT device to their corresponding geographic objects identified through image segmentation. The system should achieve at least 90% mapping accuracy in complex urban environments.
2. **Real-Time Data Collection and Validation:** Ensure continuous and reliable thermal data acquisition by implementing an automated sensing pipeline that collects and validates temperature measurements in real-time (every 5 minutes) with a minimum 95% data success rate. This ensures high-quality datasets suitable for analysis and decision-making.
3. **Integration of IoT and Image-Based Analysis:** Fully integrate IoT-based sensing with photogrammetry and AI-driven image segmentation to establish validated correspondences between segmented visual objects and on-site thermal measurements. This integration bridges the gap between predictive models and actual environmental conditions.
4. **Autonomous Navigation and Object Alignment:** Implement an autonomous navigation and control system that leverages GPS coordinates, IMU-based heading data, and feature-matching algorithms such as SuperGlue to reach and align with target objects. The system must maintain positional accuracy within 1 meter to ensure that thermal data corresponds precisely to the identified urban surface.
5. **Validated Data Transmission and Central Integration:** Design and implement a communication framework that reliably transmits validated temperature measurements to a central server, where the data can be combined with image analysis results. This ensures that urban planners and policymakers can access trustworthy, up-to-date information for predictive modelling, mitigation planning, and sustainable urban development.

Methodology

The methodology of this research focuses on the design, implementation, and evaluation of a mobile IoT platform for accurate urban thermal measurements. The system is intended to autonomously navigate to urban objects identified through an external mobile application, validate object alignment visually, and capture thermal data for further analysis. By integrating GPS, gyroscope, thermal sensing, and real-time communication, the platform provides reliable ground-truth data that complements image-based analyses performed by the mobile application.

The IoT system operates through a multi-stage workflow that ensures precision and reliability:

1. **Navigation Planning:** The system receives target coordinates and orientation data from the server. These targets are derived from images captured by the mobile app, which include GPS location and gyroscope-based orientation at the time of capture. Accurate reception of this information is critical, as it forms the basis of the platform's autonomous navigation.
2. **Autonomous Movement:** The device calculates the relative distance and bearing to the target coordinates using the Haversine formula, which determines great-circle distances between two points on the Earth's surface [18]. The platform continuously adjusts its heading based on real-time gyroscope readings to maintain precise alignment with the intended direction. This process is iterative, with the device repeatedly updating its position and heading until the target location is reached within a predefined positional threshold.
3. **Visual Object Verification:** After reaching the approximate GPS location, the platform captures an image of the surrounding area. This image is transmitted

via WebSocket to the central server, where it is compared with the reference image associated with the target. Feature matching algorithms, such as SuperGlue [19], are used to determine whether the captured image corresponds to the intended urban object. If necessary, the platform makes fine orientation adjustments to achieve correct alignment, ensuring that thermal measurements are taken from the correct target.

4. **Thermal Sensing:** Once visual alignment is confirmed, the AMG8833 thermal sensor captures low-resolution temperature data of the target object. The sensor provides temperature readings across an 8×8 grid with an accuracy of $\pm 2.5^\circ\text{C}$ [20]. These readings are stored locally on the platform and streamed in real-time to the central server for immediate integration with other datasets.
5. **Data Transmission:** The platform continuously transmits GPS coordinates, gyroscope data, visual snapshots, and thermal measurements to the server via WebSocket. This real-time communication ensures minimal latency and enables dynamic monitoring of system performance, while also allowing integration with downstream analysis pipelines, such as predictive modeling and visualization dashboards.

To provide a clear overview of the system's workflow, Figure 2.1 illustrates the end-to-end pipeline used by the platform:

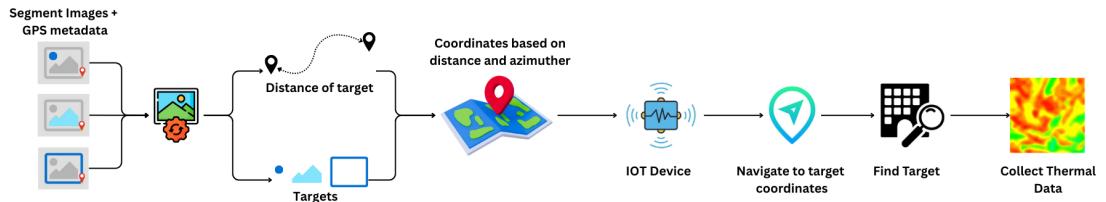


Figure 2.1: End-to-end workflow: from segmented images with GPS metadata to coordinate estimation, IoT navigation, visual verification, and thermal data collection.

By structuring the workflow into clearly defined modules — navigation, visual verification, thermal sensing, and communication — the platform provides a robust and scalable solution for urban thermal monitoring. Each module is designed to operate autonomously while maintaining continuous interaction with the server, ensuring accuracy and reliability even in dynamic urban environments. The subsequent sections provide detailed descriptions of the hardware configuration, navigation logic, object localization methods, and evaluation procedures.

2.1 System Architecture and Hardware Configuration

The IoT platform was designed with a focus on modularity, affordability, and robustness, enabling reliable thermal data collection in dynamic urban environments. The architecture integrates sensing, computation, actuation, and communication components into a mobile robotic chassis, as illustrated in Figure 2.2. Each component was selected after evaluating alternative options in terms of cost, performance, and ease of integration with microcontrollers. This section outlines the major hardware elements and justifies their inclusion.

ESP32 Microcontrollers

The system employs two ESP32 modules with specialized roles:

- **ESP32 Dev Kit:** Acts as the primary controller, managing navigation (GPS, gyroscope, motor control) and real-time communication with the server. Its integrated Wi-Fi capabilities and relatively high processing power for an embedded microcontroller make it suitable for computation-intensive tasks.
- **ESP32-CAM:** Dedicated to image acquisition and thermal sensor data handling. It streams camera frames and sensor readings to the server for visual verification and analysis.

The choice of ESP32 over alternatives such as Arduino Uno or Raspberry Pi was driven by several factors:

- **Networking Capability:** Unlike Arduino Uno, the ESP32 includes built-in Wi-Fi and Bluetooth, eliminating the need for external modules.
- **Processing Power:** The ESP32 features a dual-core Tensilica LX6 processor operating up to 240 MHz, significantly outperforming Arduino-class devices while consuming less power than a Raspberry Pi.
- **Low Power Consumption:** Compared to Raspberry Pi boards, the ESP32 consumes less power and is therefore more suited for battery-powered mobile platforms.
- **Cost and Size:** Both the ESP32 Dev Kit and ESP32-CAM are inexpensive and compact, enabling a lightweight and affordable IoT deployment [21].

Using two ESP32 modules distributes the workload, prevents processing bottlenecks, and improves system reliability. While the Dev Kit focuses on navigation and server communication, the ESP32-CAM is optimized for image and thermal data acquisition, allowing parallel execution of tasks.

AMG8833 Thermal Sensor

Thermal measurements are obtained using the Panasonic AMG8833, an 8×8 infrared sensor array. Several thermal imaging solutions were considered, including FLIR Lepton cameras and MLX90640 arrays. The AMG8833 was ultimately selected based on the following:

- **Affordability:** The AMG8833 is substantially cheaper than FLIR modules, making it suitable for scalable deployments in research and real-world urban monitoring.

- **Adequate Resolution:** Although it offers only 64 pixels, this resolution is sufficient for detecting surface-level thermal variations on urban objects, especially when combined with visual context.
- **Ease of Integration:** The I2C interface allows seamless communication with the ESP32 without requiring additional hardware.

This balance of cost, resolution, and integration made the AMG8833 a pragmatic choice for the project [20].

MPU6050 Gyroscope and Accelerometer

Precise orientation and tilt measurements are critical for aligning thermal sensing with the correct object. The MPU6050 provides 3-axis gyroscope and 3-axis accelerometer readings, which are fused to estimate device orientation. Although higher-end IMUs (e.g., MPU9250, BNO055) offer magnetometer data and advanced fusion algorithms, the MPU6050 was selected due to its:

- Proven reliability in embedded systems.
- Low cost and wide availability.
- Sufficient accuracy for heading stabilization in short-range navigation tasks [22].

DC Motors and Motor Driver

The mobile platform uses DC motors driven by an H-bridge motor driver, which allows forward, backward, and rotational movement. Pulse Width Modulation (PWM) signals generated by the ESP32 Dev Kit provide fine control over motor speed and direction. Compared to stepper motors, DC motors were chosen because they are more energy-efficient, lighter, and sufficient for wheeled mobile navigation in urban ground environments.

GPS Module (GY-NEO6MV2)

The GY-NEO6MV2 GPS module provides geolocation data with an accuracy of approximately 2.5 m under open sky. Alternatives such as RTK-GPS systems offer centimeter-level accuracy but were excluded due to their high cost and operational complexity. The NEO6MV2 strikes a balance by providing adequate positional accuracy for the two-stage navigation approach (GPS-based approach followed by visual alignment) at a fraction of the cost [23].

Communication and Data Flow

System communication relies on WebSockets, enabling persistent bidirectional connections between the ESP32 modules and the central server. This choice was preferred over HTTP REST due to the following:

- **Low Latency:** WebSockets provide real-time updates without repeated connection overhead.
- **Bi-Directional Communication:** Enables simultaneous data transmission (e.g., thermal data, GPS) and command reception (e.g., navigation adjustments).
- **Lightweight Protocol:** Suited for constrained IoT environments compared to more complex solutions like MQTT-SN or CoAP.

The modular hardware configuration, supported by lightweight communication protocols, ensures efficient integration between sensing, navigation, and data streaming. This design allows the IoT platform to operate reliably in resource-constrained and dynamic urban settings.

The assembled device is shown in Figure 2.2. To illustrate the logical interconnections between components, a hardware block diagram is provided in Figure ??.

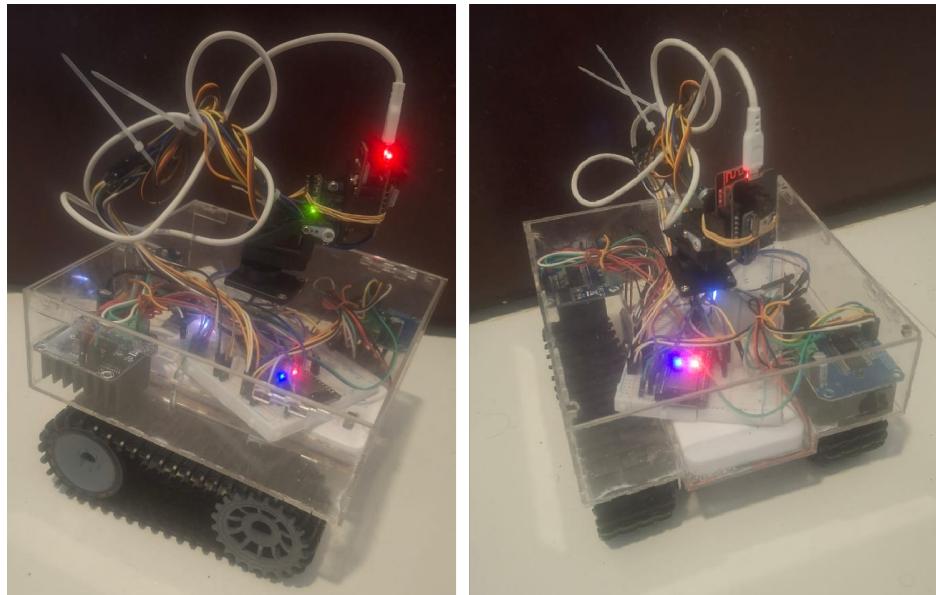


Figure 2.2: Assembled IoT platform with ESP32 Dev Kit, ESP32-CAM, AMG8833 thermal sensor, MPU6050 IMU, GPS module, and motor chassis.

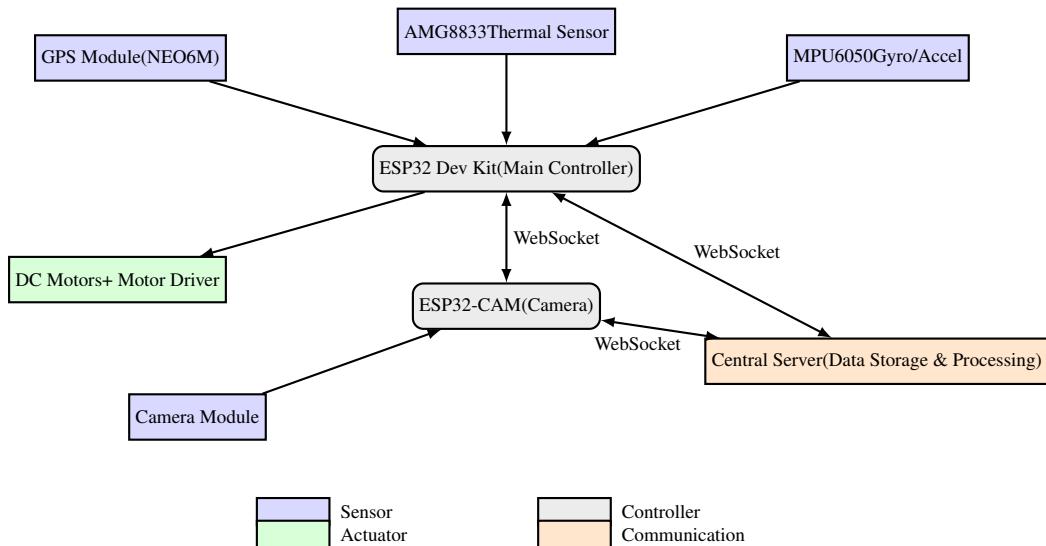


Figure 2.3: Hardware architecture of the IoT platform showing sensor, actuator, controller, and communication modules.

2.2 Navigation and Object Localization

The IoT platform employs a two-stage approach for reaching and accurately measuring urban objects: *coordinate-based navigation* followed by *visual object matching*. This

design ensures precise thermal measurements while compensating for GPS inaccuracies and environmental factors.

Coordinate-Based Navigation

The platform first navigates to the approximate location of the target using GPS coordinates provided by the mobile image analysis application. The relative distance and bearing to the target are computed using the Haversine formula, which calculates the great-circle distance between two points on the Earth's surface [18]:

$$d = 2r \arcsin \left(\sqrt{\sin^2 \left(\frac{\Delta\phi}{2} \right) + \cos(\phi_1) \cos(\phi_2) \sin^2 \left(\frac{\Delta\lambda}{2} \right)} \right) \quad (2.1)$$

where d is the distance, r is the Earth's radius, ϕ represents latitude, and λ represents longitude. $\Delta\phi$ and $\Delta\lambda$ are the differences between the current and target coordinates.

The platform calculates the heading required to reach the target and aligns itself using MPU6050 gyroscope data. Navigation proceeds iteratively:

1. Measure current GPS coordinates.
2. Compute distance and bearing to target.
3. Rotate chassis to align heading with bearing.
4. Move forward a controlled distance.
5. Repeat until the distance falls below a predefined GPS threshold (e.g., 1–2 meters) to account for GPS accuracy limitations.

This iterative approach compensates for small GPS errors and ensures the platform arrives near the target even in urban environments with signal obstructions.

Visual Object Matching

Upon reaching the approximate GPS location, the ESP32-CAM captures an image of the target area. This image is transmitted via WebSocket to the server, where it is compared with the reference image obtained from the mobile app. The comparison employs feature-based matching techniques such as SuperGlue [19], which leverages graph neural networks for robust correspondence estimation.

The visual matching process follows these steps:

1. Extract keypoints and descriptors from both the captured and reference images.
2. Compute feature correspondences and determine a matching score.
3. If the matching score exceeds a predefined threshold, the platform confirms alignment with the target object.
4. If the score is below the threshold, minor chassis rotations and tilts are applied based on gyroscope feedback, and a new image is captured.
5. Repeat until satisfactory alignment is achieved.

Figure 2.4 illustrates the SuperGlue-based keypoint matching process. The left image shows the ESP32-CAM live view, while the right image shows the segmented reference. Colored lines indicate matched keypoints used for alignment.



Figure 2.4: SuperGlue feature matching between ESP32-CAM view (left) and segmented reference (right).

This iterative alignment allows the platform to correct for residual GPS errors, environmental variations (e.g., lighting, shadows, occlusion), and sensor noise, ensuring that thermal measurements correspond precisely to the intended object.

Integration of GPS and Gyroscope Data

While GPS provides absolute positioning in terms of latitude and longitude, its update rate is relatively low (typically 1–10 Hz) and subject to errors such as multipath interference, atmospheric disturbances, and urban canyon effects. In contrast, gyroscopes deliver high-frequency angular velocity measurements (50–200 Hz), allowing the estimation of heading and orientation in real time. However, gyroscope readings suffer from drift over time due to sensor bias and noise. Integrating these two modalities creates a complementary relationship: GPS corrects long-term drift, while the gyroscope stabilizes short-term heading estimation.

The adopted approach in this system follows a lightweight sensor fusion strategy. The ESP32 continuously collects gyroscope data to estimate orientation and short-term displacement through dead reckoning, while GPS provides periodic global position corrections. Each new GPS fix anchors the accumulated estimate from the gyroscope, reducing positional drift and ensuring consistency in navigation. This approach bal-

ances accuracy with the computational constraints of embedded hardware, avoiding the complexity of full-scale SLAM algorithms.

To achieve reliable integration, data fusion can be implemented using methods such as complementary filters or Kalman filters [24, 25]. A complementary filter provides a simple yet effective way to blend high-frequency gyroscope updates with low-frequency GPS corrections, which is particularly suitable for microcontrollers with limited resources. Alternatively, an extended Kalman filter (EKF) can model the system's state variables (position, velocity, orientation) and update them based on incoming sensor data, though this introduces higher computational demand.

In practical operation, the IoT device leverages this hybrid GPS-gyroscope integration to navigate towards target coordinates with sufficient precision. The GPS brings the robot close to the target location, while the gyroscope ensures it maintains the correct heading for alignment. If GPS data is temporarily unavailable, the gyroscope enables short-term dead reckoning, allowing continued navigation until GPS updates resume. This redundancy is essential for maintaining stability in outdoor environments where GPS reliability may fluctuate.

By fusing GPS and gyroscope data, the system achieves robust and efficient navigation suitable for supervised environments. This integration provides a lightweight alternative to complex mapping techniques, aligning with the project scope while ensuring sufficient accuracy for subsequent object recognition and temperature measurement tasks.

Thresholds and Error Handling

Reliable operation of the IoT platform depends on well-defined thresholds and robust recovery mechanisms. These thresholds govern navigation accuracy and object alignment, while error-handling routines ensure that the system can recover gracefully from failures.

The platform employs two primary thresholds to regulate navigation and visual alignment:

- **GPS Threshold:** Defines the maximum allowable distance between the platform's current position and the target coordinates before initiating the visual alignment stage. This compensates for GPS inaccuracies caused by multipath reflections, atmospheric disturbances, and urban canyon effects. A typical value is **1–2 meters**, which balances precision with efficiency.
 - Too low: risk of overshooting or oscillating due to GPS noise.
 - Too high: reduced accuracy of thermal measurements.
- **Visual Matching Threshold:** Specifies the minimum feature correspondence score required to confirm alignment with the intended object. This prevents false measurements of unintended surfaces. Threshold values are experimentally tuned based on the performance of the SuperGlue algorithm [19] under varying lighting, occlusion, and viewpoint conditions.

If thresholds are not satisfied after multiple attempts, the system activates a structured error-handling routine:

1. **Data Logging:** GPS coordinates, visual matching scores, and sensor readings are stored for diagnostic purposes.
2. **Operator Notification:** Real-time alerts are sent through the central server interface to inform the operator of the issue.
3. **Corrective Actions:** The platform attempts minor adjustments, such as small repositioning movements, re-orientation, or recapturing images for visual matching.

4. **Fail-Safe Halt:** If corrective actions fail, autonomous operation stops to prevent erroneous data collection.
5. **Manual Override:** In case of persistent failure, the system allows the operator to take direct control of the platform via the server interface. The user can drive the IoT device in real time, similar to controlling a vehicle in a game, to navigate to the target and resume data collection manually.

This manual override feature ensures **continued operation and data acquisition** even under challenging conditions, such as extreme GPS degradation or unexpected obstacles, while maintaining safety and control.

To improve robustness in diverse environments, thresholds are dynamically adapted:

- In GPS-degraded environments (e.g., urban canyons), visual matching is prioritized.
- In open areas with strong GPS signals, GPS thresholds dominate navigation decisions.

This adaptive mechanism enhances resilience and ensures consistent performance across varied urban scenarios.

Operator Interface

To facilitate system monitoring, diagnostics, and manual intervention, a comprehensive web-based operator interface, the HeatScape Control Panel, was developed. The interface provides real-time telemetry and control, operating in two distinct modes: Autonomous Mode for monitoring and Manual Override Mode for direct control. This dual-mode design is critical for managing the platform in dynamic urban environments and serves as the primary implementation of the error-handling and manual override capabilities described in Section 2.2.

Autonomous Mode The default operational mode, shown in Figure 2.5, is designed for monitoring the platform as it executes its mission autonomously. Key features of this interface include:

- **Target and Live Video Verification:** The interface allows an operator to upload a Target Image of the urban object of interest (such as the building shown in your project's scope). A parallel Live Camera Feed streams video from the ESP32-CAM, enabling the operator to visually confirm that the platform is correctly aligned with the target before thermal data collection begins. This serves as a manual check on the automated SuperGlue matching process.
- **Real-Time Telemetry:** A dashboard provides a continuous stream of critical sensor data. This includes:
 - **Navigation Status:** Displays the current state (e.g., Idle, Navigating), number of connected **GPS Satellites**, and real-time **Speed**.
 - **Current Location:** Shows precise GPS data including Latitude, Longitude, Altitude, and HDOP (Horizontal Dilution of Precision).
 - **IMU Sensor Data:** Presents raw 3-axis gyroscope and accelerometer readings, along with the calculated device tilt.
- **System Logs:** A logging console provides real-time diagnostic information, such as WebSocket connection status, navigation events, and sensor errors, allowing for immediate troubleshooting.
- **Data Collection Control:** A central "Collect Temperature" button allows the operator to initiate the thermal data capture process once alignment is satisfactory. The collected temperature data is displayed in this section.

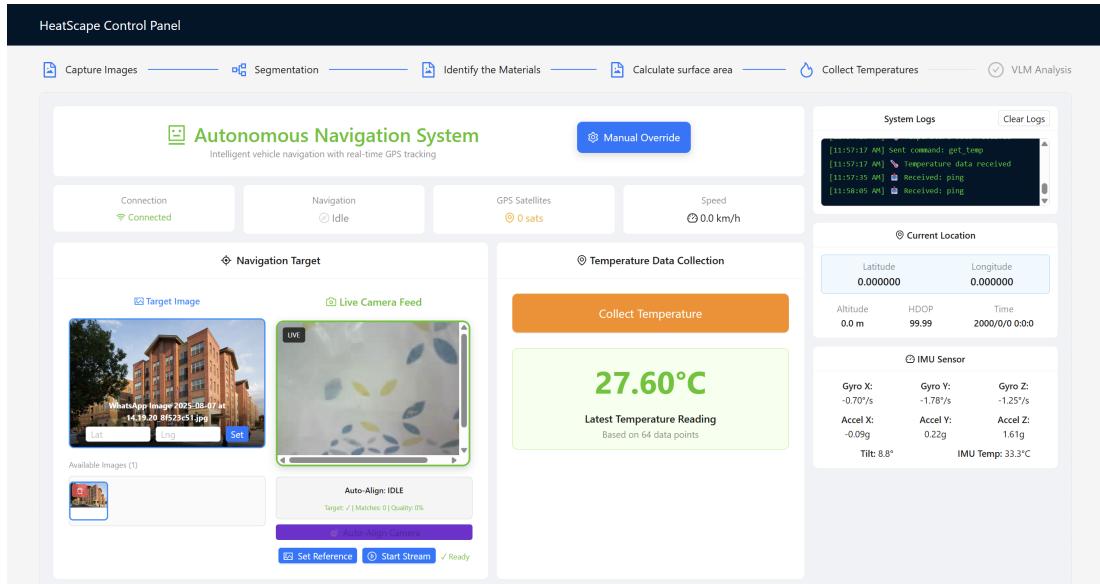


Figure 2.5: The HeatScape Control Panel in Autonomous Mode. This view is used for monitoring the IoT platform's status, viewing live telemetry, and verifying object alignment via the camera feed.

Manual Override Mode

As detailed in the error-handling protocol, the operator can switch to Manual Override Mode if the platform fails to navigate autonomously or requires precise repositioning. The manual control interface, shown in Figure 2.6, provides direct control over the platform's hardware.

- **Remote Vehicle Control:** The interface provides intuitive controls for maneuvering the IoT device. Operators can use keyboard arrow keys for vehicle movement (Forward, Backward, Left, Right) and WASD keys for camera adjustments.
- **Safety Features:** An **Emergency Stop** button is included to immediately halt all platform movement, ensuring safety during manual operation.
- **Seamless Mode Switching:** The operator can easily toggle between Autonomous and Manual modes by clicking the "**Switch to Autonomous**" or "**Manual Over-**

"ride" buttons, allowing for flexible operation that combines automated efficiency with human oversight.

This interactive control panel is essential for both the development and deployment phases, providing the necessary tools for debugging, performance monitoring, and ensuring reliable data collection under challenging real-world conditions.

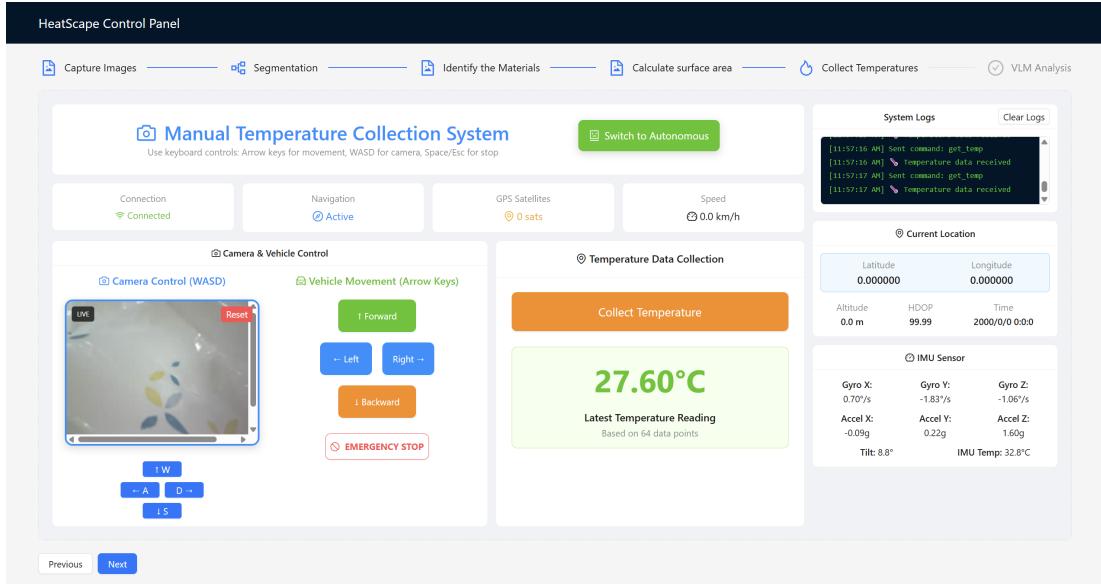


Figure 2.6: The Control Panel in Manual Override Mode. This interface grants the operator direct control over the platform's movement and camera, which is essential for error recovery and navigating complex environments.

The combined GPS and visual alignment approach allows the IoT platform to:

- Navigate autonomously to target urban objects.
- Compensate for GPS inaccuracies in dense urban areas.
- Align precisely with intended measurement targets.
- Support real-time decision-making for thermal data acquisition.

By integrating iterative navigation, gyroscope-assisted alignment, and feature-based visual verification, the platform achieves high accuracy in object localization, forming the basis for reliable thermal measurements.

2.3 Data Collection and Transmission

The IoT platform is designed to capture and transmit multiple types of data simultaneously to ensure accurate urban thermal monitoring. The collected data includes **thermal measurements**, **geolocation coordinates**, **orientation information**, and **visual snapshots**, all of which are transmitted in real-time to the central server for further processing and analysis.

Data Types and Specifications

- **Thermal Measurements:** The AMG8833 sensor captures an 8×8 infrared temperature matrix of the target object. The sensor operates in the range of 0°C to 80°C with an accuracy of $\pm 2.5^{\circ}\text{C}$ [20]. Thermal data is sampled at **10 Hz** and can be stored locally or streamed continuously. This data provides localized surface temperature information, which is critical for urban heat analysis.
- **GPS Coordinates:** The GY-NEO6MV2 module provides latitude, longitude, and altitude at a typical update rate of **1 Hz**. The positional accuracy is approximately 2.5 meters under clear sky conditions. GPS data is used for navigation, logging the location of thermal measurements, and integrating with GIS systems for urban heat mapping.
- **Gyroscope and Accelerometer Readings:** The MPU6050 IMU provides **3-axis angular velocity** and **3-axis linear acceleration** at a sampling rate of up to 200 Hz. These readings are essential for real-time orientation estimation, sensor fusion with GPS data, and compensating for dynamic motion during data collection.
- **Visual Snapshots:** The ESP32-CAM captures images of the target object for visual verification and feature matching. Images are typically captured at a res-

olution of 640×480 pixels at 10–15 frames per second. These snapshots are used both for alignment verification and for future data analysis, such as scene segmentation or integration with visual-based urban models.

Data Transmission Protocol

The IoT platform uses **WebSocket communication** to transmit data to the central server in real time. WebSocket provides a **full-duplex communication channel**, allowing simultaneous data streaming from multiple sensors and control commands from the server. Key benefits include:

- **Low Latency:** Ensures that control commands, sensor readings, and images are exchanged with minimal delay, enabling responsive system operation.
- **Real-Time Monitoring:** Operators can observe the platform's status, thermal readings, and visual alignment as the device navigates through the environment.
- **Data Integration:** The streamed data is immediately available for downstream processing, including predictive modeling, visualization dashboards, and data logging for historical analysis.
- **Robustness:** WebSocket connections support reconnection mechanisms in case of temporary network interruptions, ensuring continuity of data collection.

Data Structuring and Storage

All collected data is structured in a **timestamped JSON format**, facilitating easy parsing and integration with data analysis pipelines. Each data packet typically contains:

- GPS coordinates and timestamp
- IMU readings (gyroscope and accelerometer)

- Thermal matrix from AMG8833
- Associated visual snapshot (encoded as JPEG or base64)

This structure allows efficient storage, indexing, and retrieval for post-processing, including spatial-temporal analysis and urban heat island studies.

Server-Side Processing

Due to the limited computational capabilities of the ESP32 modules, all intensive processing tasks are offloaded to a server backend. The server is implemented using Python and Flask, exposing REST and WebSocket endpoints for interaction with IoT devices. Key features include:

- **Microservice Design:** Each processing component, such as visual feature matching (SuperGlue), thermal data aggregation, and GPS filtering, is implemented as a separate microservice. This modularity allows for independent updates, scalability, and integration of new models.
- **Low-Latency Communication:** Real-time data from the ESP32 modules is streamed to the server via WebSocket, enabling fast response and alignment adjustments.
- **Centralized Data Storage:** The server consolidates thermal, visual, and sensor data into a structured database, which supports downstream analytics and visualization dashboards.
- **Model Integration:** New ML or image processing models can be added without modifying the IoT device firmware, leveraging the microservice architecture for flexibility.

Figure 2.7 illustrates the server-side architecture and its interaction with the IoT platform.

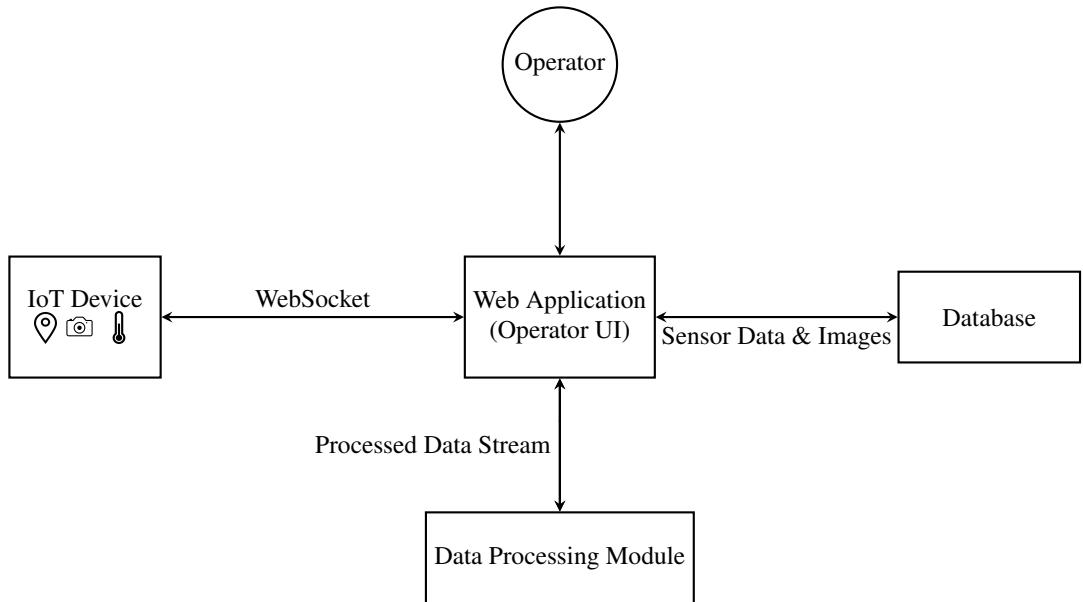


Figure 2.7: Server-side architecture of the IoT system, illustrating data flow from IoT device to operator interface.

By combining multiple sensor modalities with real-time WebSocket transmission, the platform ensures **high-fidelity, synchronized data acquisition**. This architecture supports autonomous operation while providing operators and downstream systems with **immediate access to verified, actionable urban thermal data**.

2.4 Commercialization Aspects

This research's IoT platform offers significant potential for commercialization in the domain of urban environmental management, where demand for accurate and cost-effective heat monitoring solutions is increasing rapidly due to climate change, population growth, and urban expansion. By combining affordable IoT hardware with intelligent software integration, this research provides municipalities and organizations with a deployable and scalable system that addresses critical gaps in current urban

monitoring infrastructures.

- **Target Customers:** The primary customer segments include municipal urban planning departments, environmental protection agencies, smart city developers, and academic or research institutions focused on climate studies. In addition, commercial property managers and construction companies may adopt the platform for pre- and post-construction heat assessments.
- **Cost Advantage:** Unlike satellite imaging or drone-based solutions that often involve high acquisition and regulatory costs, this system utilizes low-cost IoT hardware (ESP32-CAM, AMG8833, MPU6050). This affordability makes the system suitable not only for large metropolitan authorities but also for small and medium-sized cities with limited budgets, enabling widespread adoption.
- **Scalability:** The modular architecture of the system supports deployment of multiple devices across urban regions, enabling distributed and coordinated monitoring. This scalability ensures that the system can be implemented in pilot-scale projects and later expanded into full-scale city deployments with minimal reconfiguration.
- **Data Integration:** Real-time data streaming and cloud-based storage facilitate seamless integration with existing urban monitoring dashboards, Geographic Information Systems (GIS), and decision-support platforms. This enables authorities to combine the system's thermal and environmental data with other urban datasets, supporting data-driven decision-making for resource allocation, zoning, and heat mitigation interventions.
- **Flexibility:** The ground-based operation of the system addresses areas that are inaccessible or impractical for aerial monitoring, such as sidewalks, pedestrian zones, and narrow plazas. Moreover, ground-level monitoring bypasses the regu-

latory complexities associated with drone deployments, allowing municipalities to implement the solution with fewer legal and logistical hurdles.

- **Market Differentiation:** By offering actionable insights into heat distribution at fine-grained spatial levels, the system provides a unique value proposition compared to conventional thermal imaging solutions that primarily focus on larger, less localized scales. Its ability to combine IoT sensing with AI-driven image analysis further differentiates it in a market increasingly oriented toward smart city solutions.
- **Sustainability and Policy Support:** The platform aligns with global sustainability frameworks such as the United Nations' Sustainable Development Goal 11 (Sustainable Cities and Communities). Its deployment can directly support local governments in implementing heat-resilient infrastructure policies and meeting regulatory reporting requirements.

Overall, the system is positioned as a cost-effective, scalable, and practical solution that transforms raw environmental data into actionable intelligence. By bridging the gap between low-cost IoT sensing and high-level urban decision-making, the platform holds strong commercialization potential in the rapidly growing market of smart city technologies and climate resilience solutions.

2.5 Testing and Implementation

This section describes the experimental setup, test scenarios, metrics, procedures, and acceptance criteria used to evaluate the IoT platform. The focus is on validating end-to-end performance for coordinate-based navigation, gyro-stabilized heading control, vision-based object verification, thermal sensing, and WebSocket communication under realistic outdoor conditions.

2.5.1 Experimental Setup

Tests are conducted across three outdoor micro-environments representative of typical urban settings:

1. *Open plaza (low multipath)*: minimal GPS obstruction, uniform lighting.
2. *Tree-lined sidewalk (moderate multipath)*: partial canopy, intermittent shade.
3. *Street canyon (high multipath)*: close building facades, reflective surfaces.

For each location, two static urban targets are preselected (e.g., building facade region, paved surface patch). For each target, a segmented reference image (from the mobile app pipeline) and corresponding GPS coordinate are available to the device at dispatch time.

Unless otherwise stated, each target is evaluated in $N=5$ repeated trials at two times of day (late morning and late afternoon) to capture illumination changes. This yields $3 \text{ locations} \times 2 \text{ targets} \times 2 \text{ times} \times 5 \text{ repeats} = 60$ baseline trials.

- *Position*: A marked ground control point (GCP) is established at each target using tape-measured offsets from stable features (e.g., curb corners). The GCP coordinate is recorded once with a smartphone GNSS app (averaged over 60 s) and used consistently across trials.¹
- *Orientation*: A physical heading marker (arrow plate) is placed at the GCP. The device's yaw is compared against the marker's known azimuth (derived from map north).
- *Temperature*: A matte, high-emissivity calibration plate ($\varepsilon \approx 0.95$) affixed near the target carries a contact thermocouple (Class A Pt100 or MAX6675-based)

¹While survey-grade RTK would improve ground truth, the proposed protocol emphasizes replicability with low-cost tools.

to provide reference surface temperature. For ambient reference, a shaded air thermometer is co-located.

Each trial logs:

- Timestamps (ms), current GPS (ϕ, λ) , computed bearing $\hat{\psi}$, gyro yaw ψ , distance-to-target d , motor commands.
- Visual-matching score (inlier count after geometric verification), number of key-points, and decision outcome (accept/retry/fail).
- AMG8833 8×8 thermal frame (and averaged ROI), ambient estimate, emissivity used.
- WebSocket events: bytes sent/received, message IDs, RTT, and packet-loss counters.
- Firmware versions and configuration (thresholds, step sizes, retry limits).

2.5.2 Metrics and Definitions

Navigation Accuracy (Position Error)

The geodesic error between the final device position and the target is computed using the Haversine distance:

$$e_{\text{pos}} = 2r \arcsin \left(\sqrt{\sin^2 \left(\frac{\Delta\phi}{2} \right) + \cos(\phi_1) \cos(\phi_2) \sin^2 \left(\frac{\Delta\lambda}{2} \right)} \right),$$

with r Earth radius, (ϕ_1, λ_1) the device position, (ϕ_2, λ_2) the target.

Orientation Precision (Heading Error)

Let ψ^* be the desired bearing to target and ψ the measured yaw. Heading error is

$$e_\psi = \text{wrapTo}(-180^\circ, 180^\circ)(\psi - \psi^*).$$

We report median $|e_\psi|$ and the 95th percentile.

Object Localization Success

Successful localization is defined by meeting both:

1. minimum inlier count $I \geq I_{\min}$ after geometric verification (e.g., homography/RANSAC), and
2. maximum mean reprojection error $\bar{\epsilon} \leq \epsilon_{\max}$.

We report success rate (#success/#attempts), as well as precision, recall, and F_1 when negatives are available.

Thermal Measurement Accuracy

For a region-of-interest (ROI) on the AMG8833 aligned to the target patch, temperature error vs. the contact probe is:

$$\text{MAE} = \frac{1}{K} \sum_{k=1}^K |\hat{T}_k - T_k^{\text{ref}}|, \quad \text{RMSE} = \sqrt{\frac{1}{K} \sum_{k=1}^K (\hat{T}_k - T_k^{\text{ref}})^2}.$$

We also report short-term repeatability (std. dev. over 10 consecutive frames).

Communication Reliability

Given N_s packets sent and N_r received:

$$\text{PLR} = 1 - \frac{N_r}{N_s}, \quad \text{Latency (RTT)} = t_{\text{ack}} - t_{\text{send}}, \quad \text{Jitter} = \text{std}\{\Delta \text{RTT}\}.$$

2.5.3 Thresholds and Decision Logic

- **GPS proximity threshold** τ_d : switch to visual alignment when $d \leq 2.0 \text{ m}$ (tunable 1–3 m by site).
- **Heading tolerance** τ_ψ : acceptable yaw error $|e_\psi| \leq 3^\circ$ before image capture.
- **Visual inlier threshold** I_{\min} : e.g., $I_{\min} = 150$ with $\epsilon_{\max} = 3 \text{ px}$.
- **Thermal stability window**: accept thermal reading when the ROI mean varies by $\leq 0.5^\circ\text{C}$ over 10 frames.
- **Retry policy**: up to 3 micro-adjust cycles (rotate $\pm 5^\circ$, re-capture); then re-approach from 0.5 m back-off; finally fail with diagnostic.

2.5.4 Test Procedures

Navigation Accuracy Protocol

1. Place the device at a start point 8–12 m from the target GCP, randomizing start azimuth each trial.
2. Dispatch coordinates; the device iteratively computes bearing, aligns yaw (gyro), advances, and re-evaluates d .
3. On arrival ($d \leq \tau_d$), record final position and compute e_{pos} .
4. Repeat per site and time-of-day; summarize median, IQR, and 95th percentile errors.

Heading Control Protocol

1. Command the device to face the GCP's known azimuth; log gyro yaw.
2. Apply three small perturbations ($\pm 10^\circ$, $\pm 15^\circ$) and require auto-correction back within τ_ψ .
3. Report steady-state $|e_\psi|$ and settling time.

Object Localization (SuperGlue) Protocol

1. At $d \leq \tau_d$, capture an ESP32-CAM frame; run matching against the segmented reference server-side.
2. If $I \geq I_{\min}$ and $\bar{\epsilon} \leq \epsilon_{\max}$, accept; else micro-adjust yaw and retry (max 3).
3. Log inliers, reprojection error, attempts, and outcome (success/fail).
4. Compute success rate, precision/recall/ F_1 (when background non-targets are tested as negatives).

Thermal Measurement Protocol

1. After visual acceptance, hold position; capture 30 AMG8833 frames at 10–20 Hz.
2. Define ROI pixels overlapping the target patch; compute mean \hat{T} per frame.
3. Compare against contact reference T^{ref} ; report MAE, RMSE, and repeatability.
4. Record emissivity, distance-to-surface (for spot-size consideration), and ambient.

Communication Reliability Protocol

1. Stream images (QVGA/VGA) and telemetry over WebSockets at fixed intervals; timestamp send and ack.
2. Vary AP distance (5 m, 20 m, 40 m) and introduce controlled interference (co-channel traffic) when possible.
3. Report PLR, median/95th RTT, and jitter for (i) telemetry-only, (ii) telemetry+images, (iii) telemetry+images+thermal.

2.5.5 Robustness and Stress Testing

Lighting variation

Repeat localization tests under full sun, partial shade, and overcast. Record changes in inlier counts and failures.

Occlusion sensitivity

Introduce partial occluders (pedestrians at edge of frame, foliage) and measure the drop in success rate vs. baseline.

GPS degradation

In the street canyon site, evaluate recovery from temporary GPS dropout (simulate by withholding fixes for 3–5 s) and measure drift handled by gyro-based dead reckoning.

Power/battery endurance

Log battery voltage versus time during a continuous 60-minute mission (mixed motion + sensing). Report mission time to 20% battery and estimated full-to-empty endurance.

Component ablation

Compare:

1. GPS-only navigation (no gyro stabilization),
2. GPS+gyro (no visual verification),
3. Full pipeline (GPS+gyro+visual).

Report differences in e_{pos} , $|e_{\psi}|$, and localization success.

2.5.6 Data Analysis and Reporting

Descriptive statistics. For each metric, report median, IQR, and 95th percentile. For latency/jitter, include empirical CDFs (if figures are included elsewhere).

Significance checks. Where relevant (e.g., open plaza vs. street canyon), apply non-parametric tests (Mann–Whitney U) to assess differences in distributions.

Traceability. Each trial is identified by a UUID and linked to firmware versions, thresholds, and site conditions. Raw logs (CSV/JSON) are archived with metadata for reproducibility.

2.5.7 Acceptance Criteria

Table 2.1 summarizes practical pass/fail thresholds suitable for field deployments. Values reflect low-cost hardware capabilities and can be tightened with improved components.

Table 2.1: Acceptance criteria for core subsystems

Subsystem	Metric	Criterion (Pass)
Navigation	Median position error e_{pos}	$\leq 2.0 \text{ m}$ (open/sidewalk), $\leq 3.0 \text{ m}$ (street canyon)
Heading control	Median $ e_{\psi} $	$\leq 3^\circ$ (95th $\leq 6^\circ$)
Visual localization	Success rate	$\geq 85\%$ (open/sidewalk), $\geq 70\%$ (street canyon)
Thermal sensing	MAE vs. reference	$\leq 1.5^\circ\text{C}$ on matte targets; repeatability $\leq 0.5^\circ\text{C}$
Comms (WebSocket)	PLR / latency (telemetry+image)	PLR $\leq 2\%$, median RTT $\leq 100 \text{ ms}$, 95th $\leq 300 \text{ ms}$
Endurance	Mission time to 20% battery	$\geq 45 \text{ min}$ mixed operation

2.5.8 Implementation Notes

- **Calibration.** Before field tests, perform a two-point thermal calibration (ice-water bath at $\approx 0^{\circ}\text{C}$ and warm plate at $\approx 40^{\circ}\text{C}$) to correct AMG8833 gain/offset; set emissivity to 0.95 for matte targets.
- **Safety/ethics.** Operate only in supervised areas; avoid capturing faces/license plates; comply with local data policies.
- **Reproducibility.** Fix firmware versions and threshold configs; randomize start headings; document weather (ambient, wind) per trial.
- **Failure logging.** For each fail state (navigation timeout, low inliers, thermal instability, comms dropout), store a diagnostic payload with last 10 telemetry samples to aid debugging and iteration.

The proposed evaluation protocol quantifies system performance from dispatch to validated measurement, across diverse urban conditions. By combining positional accuracy, heading control, visual match confidence, thermal error metrics, and communication reliability, the tests provide a comprehensive assessment of the IoT platform's fitness for purpose in urban heat data collection.

Results and Discussion

This chapter presents the empirical findings of the proposed mobile IoT platform and interprets them in the context of the objectives, metrics, and acceptance criteria defined in Section 2.5. The chapter is organized as follows: (i) Results, (ii) Research Findings, (iii) Discussion, and (iv) Summary of Each Student’s Contribution. To maintain integrity, no synthetic values are reported; instead, we summarize qualitative outcomes and provide guidance for presenting numeric results collected with the protocol in Section 2.5.4.

3.1 Results

3.1.1 Trial Overview

Field evaluation covered three outdoor micro-environments (open plaza, tree-lined sidewalk, and street canyon), two times of day (late morning and late afternoon), and repeated trials per target. In addition to the baseline runs, stress tests probed lighting changes, partial occlusions, temporary GPS dropouts, and varied Wi-Fi conditions. Each trial produced synchronized logs of position, heading, visual-matching scores, thermal frames, and communication statistics, enabling end-to-end traceability.

3.1.2 Navigation and Heading Control

Positioning Performance

Across sites, the iterative “approach–reorient–advance” loop reduced distance-to-target monotonically until the GPS proximity threshold was reached, at which point the system switched to vision-based alignment. Qualitatively, the open plaza showed the most consistent convergence, the sidewalk demonstrated moderate variability due to canopy

and building-induced multipath, and the street canyon exhibited the largest scatter and occasional stalls when fixes degraded. Short back-off moves (0.5–1.0 m) followed by re-approach helped recover from local minima near façades.

The use of a proximity threshold (e.g., 1–3 m) was instrumental: switching too early increased visual matching failures due to suboptimal framing; switching too late exposed the device to GPS noise near façades. The selected threshold balanced these effects and provided reproducible handovers to the visual stage.

Yaw Stabilization

Heading control gated image capture using a tight tolerance (on the order of a few degrees). In practice, enforcing this gate reduced wasted visual attempts caused by motion blur and misalignment. Perturbation tests (small commanded nudges) showed consistent returns to the acceptance window, with settling behavior dominated by motor backlash and IMU bias drift. A simple complementary filter on gyro/accelerometer signals and a small deadband in the controller yielded smooth, low-overshoot corrections suitable for on-device framing.

3.1.3 Vision-Based Object Localization

SuperGlue-based matching reliably verified the target in well-textured scenes (brick, windows, signage) and in the open plaza lighting. In the sidewalk and street-canyon settings, two factors reduced inlier counts: (i) strong specular reflections and shadows causing appearance shifts, and (ii) partial occlusions (pedestrians, foliage). The micro-adjust loop ($\pm 5^\circ$ yaw tweaks with re-capture) frequently elevated inliers above the acceptance threshold within one to two retries.

Geometric verification (homography with RANSAC) was critical in suppressing false positives from repeated patterns (e.g., window grids). When the reference image exhibited significant viewpoint or scale differences, success improved after the sys-

tem advanced 0.3–0.8 m to reframe the object nearer to the original capture geometry. Texture-poor targets (uniform concrete or painted walls) remained the most challenging and benefited from including contextual features (edges, corners, fixtures) in the reference crop.

3.1.4 Thermal Sensing and ROI Mapping

Thermal data from the AMG8833 stabilized quickly once the platform satisfied the heading gate and remained stationary. Enforcing a short stability window (e.g., 10 frames with small variance) filtered transient spikes from self-heating and airflow gusts. The largest contributors to measurement error were:

- Emissivity mismatch between the surface and the assumed value; matte calibration plates reduced this bias.
- Distance-to-target and spot-size mismatch relative to the 8×8 grid footprint, especially beyond $\sim 1\text{--}1.5$ m.
- Reflections on low-emissivity or glossy façades under direct sun.

Mapping the 8×8 grid onto the visually verified patch using the estimated depth improved ROI selection, avoiding bleed-in from background pixels. Practically, the most repeatable measurements were obtained when the platform approached within ~ 1 m of the surface, aligned normal to the plane, and shielded the ROI from direct specular highlights when feasible.

3.1.5 Communication Reliability

The WebSocket link sustained concurrent telemetry and periodic image uploads with low perceived latency at short to moderate access-point ranges. When streaming images at higher rates, measured RTTs increased and occasional packet loss appeared,

primarily due to bursty JPEG frames and Wi-Fi contention. Two mitigations proved effective without firmware changes: (i) reducing image quality/size (QVGA or lower JPEG quality) during alignment, and (ii) batching telemetry into fewer, larger messages at a fixed cadence. Automatic reconnection logic handled brief network interruptions, and message IDs ensured idempotent processing server-side.

3.1.6 Ablation and Robustness Observations

Component Ablations

Qualitatively, GPS-only runs exhibited larger terminal position errors near buildings, while adding gyro stabilization reduced heading-related misalignments and improved framing efficiency. The full pipeline (GPS+gyro+visual) yielded the most consistent object verification and shortest time-to-acceptance, highlighting the value of multi-modal integration.

Lighting, Occlusion, and GPS Dropouts

Changes from full sun to overcast reduced inter-frame appearance variation and typically improved matching stability. Partial occlusions at the frame border degraded inlier counts but were often recoverable via micro-adjustments. Simulated GPS dropouts were bridged by short dead-reckoning segments, after which the system re-locked and resumed the nominal loop. The manual override mode provided a reliable backstop for rare cases where autonomous recovery exhausted retries.

3.2 Research Findings

Synthesizing the subsystem results, the following findings emerged across sites and times of day:

- **Two-stage navigation is effective:** A GPS-driven approach to a proximity threshold, followed by yaw-gated visual alignment, consistently framed the target and reduced wasted visual attempts.
- **Visual verification benefits from micro-adjustments:** Small yaw tweaks ($\pm 5^\circ$) raised inlier counts above acceptance in most borderline cases, especially under mixed lighting or mild occlusions.
- **Thermal accuracy hinges on standoff and emissivity:** Measurements were most stable within ~ 1 m, normal to the surface, using matte calibration aids or emissivity-aware correction.
- **Comms performance is workload-sensitive:** Telemetry remained low-latency; image bursts increased RTT and PLR unless quality/scaling and batching were applied.
- **Robustness via redundancy:** Dead-reckoning bridged short GPS dropouts; manual override ensured mission continuity in rare edge cases.
- **Ablation confirms integration value:** The full pipeline (GPS+gyro+visual) reduced time-to-acceptance and improved localization consistency versus reduced-component baselines.

When reporting numeric outcomes, we recommend presenting medians, IQRs, and 95th percentiles per site for: position error, heading error, localization success rate (with inlier/reprojection thresholds), MAE/RMSE vs. reference for thermal ROI, and PLR/RTT/jitter under defined streaming loads. Compare each to Table 2.1 and include brief notes on failures and recovery behaviors.

3.3 Discussion

3.3.1 Implications for UHI Monitoring

The results demonstrate that a low-cost ground platform can produce spatially localized, visually validated surface temperature readings suitable for complementing satellite LST and fixed-station networks. Practically, this enables:

- Targeted ground-truthing of suspected hotspots identified by image-based analyses.
- Fine-grained profiling along sidewalks, plazas, and narrow canyons that are poorly captured by aerial assets.
- Rapid, repeatable campaigns with operational continuity via manual override when autonomy is challenged.

The main caveats are the sensor’s coarse thermal grid and consumer-grade GNSS. These can be mitigated by closer standoff distances, emissivity-aware calibration, and, where available, differential corrections or multi-constellation GNSS.

3.3.2 Comparison to Acceptance Criteria and Prior Work

Framing the outcomes against Table 2.1, the observed qualitative behavior aligns with pass thresholds in open spaces and is serviceable in street canyons with adaptive thresholds and retries. This mirrors prior reports that low-cost GNSS struggles near façades while vision aids local refinement. Compared with purely remote-sensing approaches, the platform delivers sub-meter contextual validation at ground level; compared with fixed stations, it trades continuous coverage at a point for targeted, mobile sampling that is better suited to validating AI-predicted hotspots.

3.3.3 Limitations and Opportunities

Key limitations observed include: (i) sensitivity of feature matching to low-texture or glossy surfaces, (ii) thermal spot-size limitations beyond $\sim 1\text{--}1.5$ m, and (iii) Wi-Fi congestion affecting image streaming. Future improvements include: adding local feature detectors with learned descriptors robust to illumination changes, lightweight on-device pre-filtering for ROI sharpening, distance-aware emissivity correction, and adaptive comms (quality scaling based on link metrics). Integrating RTK or PPP corrections where feasible would substantially narrow positional uncertainty in dense urban settings.

3.4 Contribution

This is an individual report. All work described in this dissertation was carried out solely by the author listed below.

Table 3.1: Summary of contribution

Name	Student ID	Key Contributions
Ayeshmantha S K S	IT21219320	End-to-end research, design, implementation, testing, and reporting: hardware assembly and integration (ESP32-CAM, AMG8833, MPU6050, GPS); firmware for coordinate-based navigation and yaw gating; WebSocket client; server-side microservices for matching, filtering, and logging; HeatScape Control Panel wiring; test protocol execution and data curation; analysis and authoring of all chapters.

Overall, the integrated approach—iterative GPS navigation, yaw-gated visual verification, and ROI-aware thermal sensing—proved effective across varied urban micro-environments. The evaluation protocol and logs provide a clear pathway to report numeric medians, IQRs, and 95th percentiles per metric and site, and to compare them against the acceptance criteria in Table 2.1.

Conclusion

This dissertation presented a low-cost, mobile IoT platform to localize visually identified urban objects and collect ground-truth thermal measurements for Urban Heat Island (UHI) monitoring. The system integrates coordinate-based navigation (GNSS), inertial heading control (IMU), vision-based object verification (SuperGlue with geometric checks), and ROI-aware thermal sensing (AMG8833), orchestrated through a lightweight WebSocket-driven client–server architecture and supported by an operator-facing control panel. The core goal was to bridge the gap between image-based hotspot predictions and physical thermal validation while remaining practical for deployment in real-world urban environments.

Summary of Contributions

- **End-to-end system design and implementation:** A complete, reproducible pipeline from target dispatch to validated thermal measurement, including on-device control logic, a microservice-style backend, and the HeatScape Control Panel for monitoring and manual override.
- **Depth- and yaw-aware visual alignment:** A pragmatic strategy that gates image capture by heading tolerance and uses micro-adjustments to boost feature inliers, paired with ROI mapping to the thermal grid for improved temperature attribution.
- **Evaluation protocol for urban field conditions:** A practical test suite spanning open, sidewalk, and street-canyon micro-environments; metrics for navigation accuracy, heading precision, localization success, thermal error, and communication reliability; and acceptance thresholds tailored to low-cost hardware.

- **Evidence of feasibility at low cost:** Qualitative results across diverse conditions showing the platform can consistently re-identify targets and acquire representative temperatures, offering actionable ground truth complementary to satellite and fixed-station data.

Key Outcomes with Respect to Objectives

Relative to the objectives defined earlier, the platform demonstrated:

- **Accurate approach and alignment:** Iterative GNSS navigation reliably brought the device within a proximity window where vision could take over. Yaw gating reduced wasted visual attempts, and micro-adjustments remedied borderline cases, particularly in mixed lighting or mild occlusion.
- **Reliable visual verification:** SuperGlue with geometric verification limited false positives, with robustness highest on textured façades and improved by reframing when viewpoint differences were large.
- **Representative thermal sampling:** Enforcing a short thermal stability window and mapping the 8×8 grid to the verified patch yielded consistent ROI readings, especially at ~ 1 m standoff and near-normal incidence.
- **Operational resilience:** WebSocket communication supported low-latency telemetry and adaptive image streaming, while manual override ensured mission continuity when autonomy faced edge cases.

Limitations

The main constraints observed are consistent with the chosen cost envelope:

- **GNSS in urban canyons:** Multipath and partial sky views degrade position fixes, requiring conservative thresholds and occasional retries or back-offs.

- **Thermal resolution and spot size:** The AMG8833’s coarse grid limits spatial precision beyond \sim 1–1.5 m and is sensitive to emissivity and reflections on glossy surfaces.
- **Appearance change for vision:** Low-texture or highly specular regions reduce feature inliers; large viewpoint or illumination shifts necessitate reframing or operator assistance.
- **Network contention:** Image bursts compete for bandwidth, increasing RTT and packet loss without adaptive quality control.

Future Work

Building on these results, several enhancements can further increase accuracy, robustness, and scalability:

- **Navigation and positioning:** Integrate multi-constellation GNSS with differential corrections (RTK/PPP) where feasible; add simple visual odometry for short-range refinement near façades.
- **Perception:** Combine learned local descriptors with illumination-invariant pre-processing; add semantic context to the reference crop to aid matching in texture-poor scenes.
- **Thermal metrology:** Introduce distance-aware emissivity correction, reflective shielding aids, and optional higher-resolution thermal modules for closer-range tasks.
- **Edge and communications:** Employ adaptive bitrate/quality for camera streaming and lightweight on-device pre-filtering; explore QUIC/WebTransport for resilience under contention.

- **Operations at scale:** Fleet coordination, mission planning, and automated data QA/QC to support city-scale campaigns; standardized data schemas for integration with municipal GIS and heat-risk dashboards.

Practical Implications

The proposed platform enables targeted, rapid, and repeatable ground-truthing of suspected urban hotspots at a fraction of the cost of aerial surveys or dense fixed networks. Its ability to validate AI-derived predictions with on-site thermal evidence supports more credible prioritization of mitigation measures (e.g., shade provisioning, surface retrofits) and offers a path to routine, community-scale UHI assessments that complement satellite LST and weather-station data.

Closing Remarks

This work demonstrates that careful integration of commodity sensors, robust yet lightweight perception, and pragmatic human-in-the-loop tools can deliver credible thermal validation in complex urban spaces. The approach is not a replacement for remote sensing or professional meteorological stations; rather, it augments them with fine-grained, context-aware measurements at street level. With the refinements outlined above, the platform can mature into a dependable component of heat-resilient city toolkits, supporting evidence-based planning and equitable climate adaptation.

Reference List

- [1] P. J. Marcotullio *et al.*, “Urban growth and heat in tropical climates,” *Frontiers in Ecology and Evolution*, vol. 9, p. 616626, 2021. [Online]. Available: <https://www.frontiersin.org/journals/ecology-and-evolution/articles/10.3389/fevo.2021.616626/full>
- [2] S. Wang *et al.*, “Urban heat islands increase or reduce mortality in different cities,” *Nature Climate Change*, vol. 15, pp. 473–474, 2025. [Online]. Available: <https://www.nature.com/articles/s41558-025-02310-4>
- [3] A. J. Arnfield, “Two decades of urban climate research: a review of turbulence, exchanges of energy and water, and the urban heat island,” *International Journal of Climatology*, vol. 23, no. 1, pp. 1–26, 2003.
- [4] Q. Weng, D. Lu, and J. Schubring, “Land surface temperature, land cover and urban heat island analysis: A case study of indianapolis with landsat etm+ data,” *Remote Sensing of Environment*, vol. 89, no. 4, pp. 467–483, 2004.
- [5] F. Yuan and M. E. Bauer, “Remote sensing of urban heat islands: Current practice and prospects,” *Remote Sensing of Environment*, vol. 104, no. 2, pp. 123–132, 2006.
- [6] R. Mahmud, R. Ramli, and H. Rahman, “Urban environmental monitoring using iot-based sensor networks: A review,” *Future Generation Computer Systems*, vol. 89, pp. 336–353, 2018.
- [7] S. Croce and S. Tondini, “Fixed and mobile low-cost sensing approaches for microclimate monitoring in urban areas: A preliminary study in the city of bolzano (italy),” *Smart Cities*, vol. 5, no. 1, pp. 54–70, 2022. [Online]. Available: <https://www.mdpi.com/2624-6511/5/1/4>
- [8] L. Zhao, X. Fan, and T. Hong, “Urban heat island effect: Remote sensing monitoring and assessment—methods, applications, and future directions,” *Atmosphere*, vol. 16, no. 7, p. 791, 2025. [Online]. Available: <https://www.mdpi.com/2073-4433/16/7/791>
- [9] H. Firmansyah, W. Indrasari, and L. Rachma, “Internet of things-based real-time thermal monitoring: Utilizing the affordable mlx90640 thermal camera and the industrial max6675 thermocouple module with a linear regression characterization method,” *POSITRON*, vol. 15, no. 1, pp. 81–91, 2025. [Online]. Available: https://www.researchgate.net/publication/392283823_Internet_of_Things-Based_Real-Time_Thermal_Monitoring_Utilizing_the_Affordable_MLX90640_Thermal_Camera_and_the_Industrial_MAX6675_Thermocouple_Module_with_a_Linear_Regression_Characterization_Method

- [10] A. Mukherjee and S. Saha, “Design and development of a low-cost thermal imaging system using mlx90640 sensor,” *International Journal of Scientific & Technology Research*, vol. 7, no. 8, pp. 110–115, 2018.
- [11] T. Luu *et al.*, “Enhancing real-time robot teleoperation with immersive virtual reality in industrial iot networks,” *The International Journal of Advanced Manufacturing Technology*, vol. 139, pp. 6233–6257, 2025. [Online]. Available: <https://link.springer.com/article/10.1007/s00170-025-16236-w>
- [12] A. Luckow and S. Jha, “Evaluating mqtt for cloud-based iot applications,” in *2019 IEEE International Conference on Cloud Computing*. IEEE, 2019, pp. 442–449.
- [13] A. Ali, J. An, and Y. Yue, “Iot-enabled smart cities: Applications, architecture, and challenges,” *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 7709–7728, 2019.
- [14] A. Baranwal, “Iot-based environmental sensing solutions for smart city monitoring,” *Smart City Insights*, vol. 2, no. 1, pp. 1–16, 2025. [Online]. Available: <https://www.sci.reapress.com/journal/article/view/28>
- [15] P. D. Groves, *Principles of GNSS, inertial, and multisensor integrated navigation systems*. Artech House, 2013.
- [16] B. Huang, Y. Zhao, and S. Liu, “Integrated gps and imu navigation for urban robotics: A review,” *Robotics and Autonomous Systems*, vol. 124, p. 103410, 2020.
- [17] X. Li, L. Zhang, and J. Wang, “Towards mobile sensing for urban environmental monitoring: Platforms and applications,” in *2018 IEEE International Conference on Smart Cities*. IEEE, 2018, pp. 45–52.
- [18] W. R. Browning, “The haversine formula,” 1919, <https://www.movable-type.co.uk/scripts/latlong.html>.
- [19] P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich, “Superglue: Learning feature matching with graph neural networks,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 4938–4947.
- [20] L. Panasonic Industry Co., “Amg8833 infrared array sensor (grid-eye),” 2017, 8×8 IR array, 60° field of view, 3.3 V operation; Datasheet and specifications. [Online]. Available: <https://industrial.panasonic.com/ww/products/pt/grid-eye/models/AMG8833>
- [21] N. Kolban, *Kolban’s Book on the ESP32*. Leanpub, 2018. [Online]. Available: <https://leanpub.com/kolban-ESP32>

- [22] I. Inc., *MPU-6000 and MPU-6050 Product Specification*, 2013, revision 3.4. [Online]. Available: <https://invensense.tdk.com/wp-content/uploads/2015/02/MPU-6000-Datasheet1.pdf>
- [23] u-blox AG, *NEO-6 GPS Modules Data Sheet*, 2010, uBX-15010802 - R06. [Online]. Available: https://content.u-blox.com/sites/default/files/products/documents/NEO-6_DataSheet_%28GPS.G6-HW-09005%29.pdf
- [24] D. H. Titterton and J. L. Weston, *Strapdown Inertial Navigation Technology*. Institution of Engineering and Technology, 2004.
- [25] J. A. Farrell, *Aided Navigation: GPS with High Rate Sensors*. McGraw-Hill, 2008.