Vision-based Sensing of Urban Environment using Hardware-Software co-design at Scale

Akshit Gupta, R Venkatesha Prasad
TU Delft, Delft, The Netherlands a.gupta-5@tudelft.nl; r.r.venkateshaprasad@tudelft.nl

***Abstract*—Urban environmental monitoring is crucial for climate-adaptive cities, yet current approaches rely heavily on human inspection, overhead remote sensing, or limited terrestrial RGB imagery, lacking comprehensive multispectral perspectives at street level. To bridge this gap, this paper introduces UrbanScapes, an adaptable and flexible hardware-software co-design platform for large-scale, vision-based sensing of urban environments. Built on a modular architecture, UrbanScapes seamlessly integrates multispectral imaging channels, including Red, Green, Blue (RGB), Near-Infrared, and Thermal, to capture terrestrial imagery. The system is designed for both opportunistic deployment on bicycles, vehicles, or pedestrians and targeted deployments, ensuring high spatio-temporal scalability without requiring hardware redesigns. Crucially, the platform utilizes on-device deep learning models to automatically remove personally identifiable information (PII) at the edge, addressing privacy concerns inherent in street-level data collection. By combining this multispectral terrestrial imagery data with overhead remote sensing imagery, UrbanScapes enables the scalable calculation of physics-based environmental indices, providing a foundational tool for advanced city-wide analysis. (Poster)**

I. Introduction

Urban environmental monitoring is vital for the development of climate-resilient cities. Existing approaches, largely comprising remote sensing imagery and manual human inspections, are limited to overhead views and non-scalable, respectively. In this work, we demonstrate an integrated hardware-software platform, UrbanScapes, for participatory sensing to collect multispectral terrestrial-view data. This data can be combined with overhead remote sensing imagery, enabling comprehensive monitoring of urban environmental parameters at city-wide scales. Specifically, the UrbanScapes hardware comprises visible (red, green, and blue), near-infrared, and thermal imaging sensors that collect terrestrial-view imagery at high spatial and temporal resolutions. Data quality is ensured by using similar bands in overhead remote sensing imagery, which is later processed using computer vision-based deep learning models to understand and quantify various urban environmental parameters. For instance, it is easy to identify different building facade materials using near-infrared channels, to analyze energy leakage from building façades using thermal channels, and to analyze urban greenery using a combination of vision and near-infrared channels. The device is designed for flexible deployment via magnets on bikes, cars, and buses, and can also be carried by humans. Further, all personally identifiable information (PII) is removed on-device using a deep learning model. Hence, allowing for both opportunistic monitoring and targeted monitoring approaches, enabling comprehensive analysis of urban environmental parameters at scale. In this paper, we highlight the design of the system, its architecture, and demonstrate new capabilities after deployment. The dataset collected with an earlier prototype of this device is already available as the Spectrascapes dataset [1].

II. Related Work

Overhead remote sensing imagery (using drones, satellites, and airplanes) generates massive amounts of data across multiple spectral channels, in addition to the standard Red, Green, and Blue (RGB) channels visible to the human eye. These additional spectral channels are useful since they allow the calculation of quantitative, calibrated, physics-based environmental indices. However, most studies in the last decade, such as [2], [3], are often limited by terrestrial RGB imagery from platforms such as Google Street View (GSV), KartaView, and Baidu Maps [4]. Recognizing the value of additional spectral channels in terrestrial imagery, some recent projects have deployed custom hardware to capture them beyond the Red, Green, and Blue bands. For example, in [5], thermal and RGB imaging sensors are utilized to detect energy leakage from building facades, while in [6], thermal and near-infrared imaging sensors are utilized on moving vehicles to detect the health condition of trees. However, these works focus on selecting specific spectral bands based on the application domain, thereby limiting the flexibility of their hardware to capture additional spectral channels. Furthermore, the system designed in these studies does not account for the removal of personally identifiable information (PII). Table I summarizes these works from several aspects.

Work	Custom Platform	Spectral Channels in Imagery			Platform Flexibility	PII removal
		RGB	NIR	Thermal		
[2]	No	✓	✗	✗	Zero	Yes
[3]	No	✓	✗	✗	Zero	Yes
[6]	Yes	✓	✓	✗	Low	No
[5]	Yes	✓	✗	✓	Low	No
This work	Yes	✓	✓	✓	High	Yes

TABLE I: Comparison of recent works in terrestrial view imagery domains in terms of spectral richness, flexibility, PII removal and our contributions.

III. Research contributions

Our platform is shown in Figure 1. Addressing the research gaps (Table I), our approach has the following research contributions – (1) The architecture of Urbanscapes is designed around a central single board computer based on Linux kernel. Using a single-board computer enables support for multiple interfaces and microPython. This enables flexibility in both port (or interface) utilization and software module development for each imaging sensor, with no major redesign required. In our current configuration, a Raspberry Pi 4 was utilized with the three imaging sensors connected and controlled via UART, SPI, and I2C interfaces. (2) The platform allows easy deployment on bicycles, vehicles (such as buses, cars) with magnets, or to be carried by humans. This deployment structure enables scalability in both spatial and temporal coverage. Furthermore, no soldering is required to add new sensors. In the future, dedicated sensing deployments using crowdsourcing platforms such as Premise will enable spatial reconfiguration of sensors to achieve a specific coverage density in an area. (3) Our terrestrial imagery can be validated and calibrated using existing remote sensing datasets since our design uses the same spectral channels as those of satellites, thus ensuring data quality. To ensure PII removal, the YOLO series of models is used after a specific batch of images is captured. This ensures the detection and Gaussian blurring of both human faces and vehicle license plates.

Fig. 1: The UrbanScape device with three imaging sensors and its deployment on a bicycle with magnets.

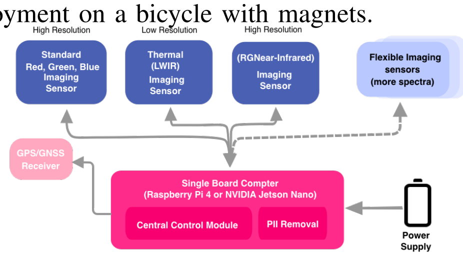

Fig. 2: Imaging data is geo-tagged with a provision for additional imaging sensors with different spectral channels.

IV. System Design and Architecture

The architecture of UrbanScapes is shown in Figure 2. The exploded-view diagram showing the assembly of all components used for different spectral channels is shown in Figure 3. In its current state, the casing design allows a dual-sided view and supports six imaging sensors. The casing is 3D-printed from ASA (Acrylonitrile Styrene Acrylonitrile) filament, providing weather resistance at a low cost. The device measures 26 cm $\times$ 14 cm $\times$ 15 cm and weighs 1.25 kg, including all hardware components and three imaging sensors.

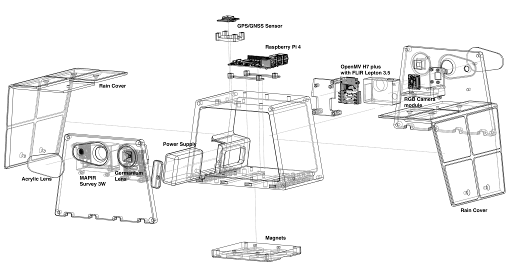

Fig. 3: The assembly of all the components. The design allows imaging sensors to be installed on both sides of the device.

V. Test Deployments and Early Results

An early prototype of the device was deployed in four different urban morphologies in the Netherlands, capturing approximately 17,000 multispectral images over 7 days. The details of deployment, the dataset gathered, and some early use-cases are available at [1]. In addition, using a state-of-the-art deep learning model for 3D reconstruction called ZipMap, we combined terrestrial images captured using this platform with overhead airborne imagery. An example of this reconstruction is shown in Figure 4, for a street in Arnhem, The Netherlands. As seen in the figure, this type of reconstruction enables a comprehensive terrestrial and overhead digital twin of the built environment, with multispectral channels enabling calculation of physics-based environmental indices. A detailed analysis is still in progress.

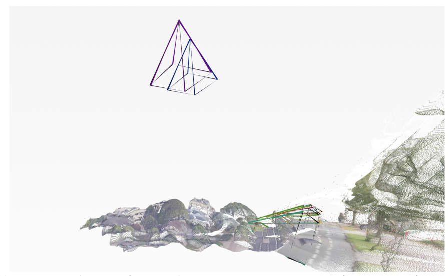

Fig. 4: Comprehensive 3D reconstruction using terrestrial and remote sensing imagery with both overhead and ground-view imagery for a street in Arnhem, The Netherlands. Purple and Blue cones denote airborne imagery data and the remaining cones denote terrestrial imagery data captured using UrbanScapes.

VI. Conclusion and Future

We showcased a hardware-software co-designed platform called UrbanScapes to gather multispectral terrestrial view data with scalability. The architecture enables flexibility of sensors and deployments. This allows integration with remote sensing data to calculate physics-based environmental indices, and on-device deep learning models remove PII. Future work targets temporal data collection using public buses in the Netherlands.

References

[1] A. Gupta, J. Timmermans, F. Biljecki, and R. Uijlenhoet, “The spectrascapes dataset: Street-view imagery beyond the visible captured using a mobile platform,” 2026. [Online]. Available: https://arxiv.org/abs/2604.13315

[2] N. Tarkhan, M. Klimenka, K. Fang, F. Duarte, C. Ratti, and C. Reinhart, “Mapping façade materials utilizing zero-shot segmentation for applications in urban microclimate research,” *Scientific Reports*, vol. 15, no. 1, p. 5492, 2025. [Online]. Available: https://www.nature.com/articles/s41598-025-86307-1

[3] Y. Hou, M. Quintana, M. Khomiakov, W. Yap, J. Ouyang, K. Ito, Z. Wang, T. Zhao, and F. Biljecki, “Global streetscapes – a comprehensive dataset of 10 million street-level images across 688 cities for urban science and analytics,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 215, pp. 216–238, 2024.

[4] F. Biljecki and K. Ito, “Street view imagery in urban analytics and GIS: A review,” *Landscape and Urban Planning*, vol. 215, p. 104217, 11 2021.

[5] M. Klimenka, K. Zhao, R. Hilland, F. Zhang, J. Voogt, and C. Ratti, “Instant infrared: Estimating urban surface temperatures from street view imagery,” *Building and Environment*, vol. 267, p. 112122, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360132324009648

[6] A. Gupta, S. Mora, F. Zhang, M. Rutten, R. Prasad, and C. Ratti, “Greenscan: Toward large-scale terrestrial monitoring the health of urban trees using mobile sensing,” *IEEE Sensors Journal*, vol. PP, pp. 1–1, 07 2024.