CCC 2025 **Proceedings of the Creative Construction Conference** (2025)
Edited by: Miroslaw J. Skibniewski & Miklós Hajdu & Žiga Turk
https://doi.org/10.22260/CCC2025/0043

HISTORICAL BUILDING ROOF STRUCTURE EXTRACTION BASED ON UAV HYPERSPECTRAL IMAGERY RECONSTRUCTION

Jihan Zhang[1], Duanqin Hong[1,2], Yijun Huang[1], Wenxing Hong[2], **Xi Chen**[1*], Ben M. Chen[1]

1 The Chinese University of Hong Kong, Hong Kong, China
2 Xiamen University, Xiamen, China

Abstract

The sustainable preservation of large-scale, actively inhabited historical villages requires detailed and scalable methods for material-aware structure extraction and health monitoring. Traditional Unmanned Aerial Vehicle (UAV) RGB imagery, while accessible, suffers from poor adaptability in complex environments, particularly when attempting to distinguish between diverse roof types, repair materials, and naturally degraded components under occlusion or varying illumination. To address these limitations, we propose a three-stage framework that reconstructs multispectral imagery from UAV-acquired RGB inputs and leverages prompt-driven semantic segmentation for heritage architecture understanding. First, high-resolution RGB images are captured via UAVs and rigorously filtered to ensure geometric and radiometric quality. These images are then processed using a deep multispectral reconstruction network (MST++), generating spectral cubes that enhance material differentiation. In the final stage, we introduce a semantic extraction pipeline grounded in a custom-built knowledge base of architectural materials and typologies. Semantic prompts (e.g., “damaged clay tile roof”, “metal-repaired roof”) are generated from this knowledge base and used to guide segmentation via the prompt-aware SAM2 model. Experimental results on aerial data from Lai Chi Wo village in Hong Kong demonstrate that the reconstructed multispectral imagery provides superior performance in segmenting complex features such as metal roofs and rooftop vegetation, compared to RGB inputs. The enhanced spectral sensitivity enables improved alignment between semantic prompts and physical features, yielding accurate material-aware maps. This approach supports long-term heritage building monitoring and sets a foundation for proactive, informed conservation strategies for diverse traditional architectural landscapes.

Keywords: Multispectral imagery, Roof segmentation, Unmanned aerial vehicle (UAV).

© 2025 The Authors. Published by the International Association for Automation and Robotics in Construction (IAARC) and Diamond Congress Ltd.

Peer-review under responsibility of the scientific committee of the Creative Construction Conference 2025.

1. Introduction

The preservation and sustainable management of historical architecture, particularly in active ancient villages, present significant challenges due to their expansive scale, intricate structures, and the necessity for ongoing health monitoring. Traditional ground-based surveys are often labor-intensive and time-consuming, making them impractical for large-scale assessments. Unmanned Aerial Vehicles (UAVs) have emerged as a transformative tool in this context, offering efficient, non-invasive means to capture high-resolution imagery over extensive areas. This capability is especially beneficial for surveying complex roof structures, which are critical indicators of a building's structural integrity and historical authenticity.

Conventional UAV-based imaging predominantly relies on Red-Green-Blue (RGB) sensors. While effective for general visualization, RGB imagery often falls short in distinguishing between diverse roofing materials and conditions, particularly under challenging environmental factors such as shadows, vegetation cover, and varying lighting conditions. These limitations hinder the accurate extraction of architectural features necessary for detailed analysis and conservation efforts.

To address these challenges, our research explores the innovative application of hyperspectral reconstruction techniques to UAV-acquired RGB imagery. By reconstructing multispectral data from

**Corresponding author email address: xichen002@cuhk.edu.hk*

standard RGB images, we aim to enhance the spectral resolution of the captured imagery without the need for expensive hyperspectral sensors. This approach leverages the unique spectral signatures of different materials, facilitating more precise differentiation of roofing materials and structural conditions. The enhanced spectral information significantly improves the segmentation and analysis of complex roof structures, enabling more accurate assessments of their condition and composition.

Our methodology involves a two-step framework: initially, we employ advanced segmentation models to isolate rooftop areas from the background and mitigate environmental interferences. Subsequently, we apply optimized edge detection and line fitting algorithms to extract and analyze roof boundary features. This process allows for the accurate measurement of architectural parameters, such as eave lengths, which are essential for monitoring structural changes and planning conservation interventions.

The integration of hyperspectral reconstruction into UAV-based imaging represents a significant advancement in the field of cultural heritage preservation. It offers a cost-effective, scalable, and non-invasive solution for the comprehensive monitoring of historical buildings. By enhancing the spectral and spatial resolution of aerial imagery, our approach provides deeper insights into the structural and material characteristics of ancient architecture, thereby supporting more informed and effective conservation strategies.

1.1. Related works

The application of multispectral reconstruction technology based on UAVs in historical village contexts has gained significant attention in recent years, addressing the limitations inherent in conventional RGB imaging methods. RGB imaging, though easily accessible, often struggles to accurately differentiate building materials and defects due to its limited spectral information, especially in complex environmental conditions such as varied illumination, shadows, and vegetation coverage [1]. In response to these challenges, multispectral imaging has demonstrated considerable potential in enhancing the discrimination and classification of historical architectural materials and conditions. For instance, Sanchez and Quiros applied multispectral imaging involving visible and near-infrared bands to classify 14 categories of historical building materials, achieving a maximum classification accuracy of 81.6%, albeit noting significant dependency on consistent lighting conditions [2]. This highlights the criticality of spectral diversity for accurate material characterization but also the inherent challenges associated with limited spectral bands.

Furthermore, hyperspectral imaging has emerged as a complementary and more precise method, offering significantly greater spectral resolution. Kurz et al. demonstrated the effectiveness of hyperspectral imaging for detailed identification of material characteristics, emphasizing its superior capability for diagnosing and discriminating subtle differences among construction materials, particularly in detecting facade deterioration and structural defects [3]. Similarly, research by Peng et al. successfully employed hyperspectral imagery combined with advanced deep learning methods to detect subtle defects in bridge structures, further underscoring the efficacy of integrating spectral and spatial data for structural health monitoring [4]. Recent studies have also explored the potential synergy achieved by combining multispectral imagery with complementary datasets, such as LiDAR, to enhance classification accuracy [5].

Moreover, Zahiri et al. specifically compared the effectiveness of multispectral and hyperspectral imaging techniques in facade material classification tasks, highlighting that hyperspectral imaging consistently outperformed multispectral imaging in distinguishing building materials due to its broader and finer spectral coverage [1]. However, multispectral imaging still provided substantial practical advantages by offering a cost-effective, more accessible, and quicker alternative for large-scale or routine monitoring applications, particularly when hyperspectral imaging resources were not readily available.

In conclusion, while multispectral reconstruction technology based on UAV imagery has shown promise in historical village monitoring, effectively enhancing material and damage discrimination beyond conventional RGB methods, the highest accuracy and robustness are generally attained through hyperspectral imaging or integration of multispectral imagery with other spectral or spatial datasets. Continued research in data fusion and algorithm development is critical for further leveraging the full

potential of multispectral reconstruction technology in the precise and sustainable management of historical architectural heritage.

2. Methodology

The methodology proposed in this study consists of an integrated three-stage framework designed to leverage UAV-derived multispectral reconstruction imagery for enhanced structure extraction and material analysis of historical village architecture as in Fig. 1. Firstly, UAV data acquisition is systematically performed, involving meticulous flight-path planning and pre-processing of collected RGB images to ensure optimal spatial coverage and image quality. Subsequently, multispectral reconstruction is implemented to transform standard RGB imagery into multispectral data, significantly enriching spectral information critical for differentiating building materials. Lastly, a semantic analysis stage is introduced, incorporating the construction of a domain-specific knowledge base that comprehensively categorizes historical building materials and structural features. This knowledge base supports semantic prompting for precise image segmentation and facilitates the generation of material-aware semantic information. Ultimately, this third stage exploits the spectral sensitivity of multispectral imagery to accurately characterize material distributions across building structures, thus providing valuable statistical insights into architectural conditions and enabling targeted conservation and maintenance efforts.

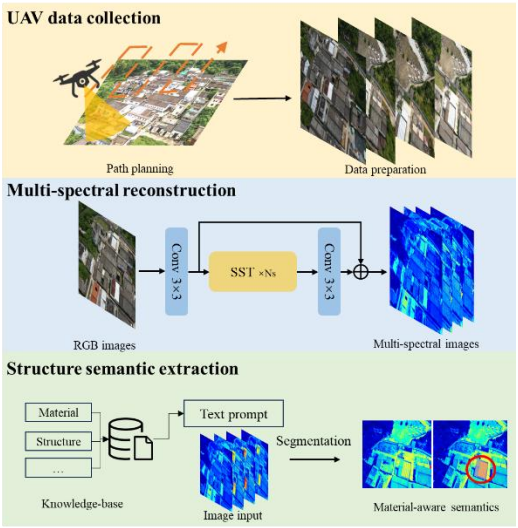


Fig. 1. Overall framework.

2.1. UAV data collection

To ensure robust and scalable analysis of historical village structures, UAV-based aerial imagery acquisition was performed with a focus on systematic path planning, metadata recording, and stringent quality control. The data collection stage aimed to produce a consistent and high-quality dataset, free from redundancy or acquisition-induced artifacts, thus enabling reliable downstream multispectral reconstruction and semantic analysis.

A grid-based flight path was automatically generated using a waypoint planning algorithm, maintaining sufficient overlap between adjacent images—at least 80% forward and 60% lateral overlap—to support photogrammetric consistency. Both nadir and oblique views were captured to address occlusion due to roof inclination and vegetation coverage. The optimal flight altitude H was calculated based on the desired Ground Sampling Distance (GSD):

$$GSD = \frac{H \cdot s}{f} \tag{1}$$

Where GSD denotes the ground sampling distance (cm/pixel), H is the UAV altitude (cm), s is the camera sensor pixel size (cm), and f is the focal length of the lens (cm).

A target GSD of approximately 2 cm/pixel was adopted to preserve fine roof features such as tile patterns and eave boundaries. All images were geotagged using onboard GPS/IMU, and Ground Control Points (GCPs) were deployed to support georeferencing precision.

For the flight metadata, each UAV image I_k was represented as:

$$I_k = (R_k, G_k, B_k, x_k, y_k, z_k, \theta_k) \quad (2)$$

Where R_k, G_k, B_k represent the raw color channels of the image. x_k, y_k, z_k denote the UAV's spatial location in a global reference frame (WGS84). θ_k includes the camera orientation angles (yaw, pitch, roll). This spatiotemporal metadata allowed for subsequent image alignment, geo-registration, and reconstruction operations.

In this study, image pre-processing primarily focused on rigorous quality control and dataset curation, rather than conventional enhancement operations. The objective was to ensure that only high-quality, geometrically consistent images were included for subsequent multispectral reconstruction and semantic analysis. A range of criteria was applied to identify and exclude images likely to introduce errors in spatial or spectral modeling. First, images exhibiting significant geometric distortion—often resulting from extreme UAV pitch or roll angles—were automatically flagged and removed. In parallel, motion blur and defocus were assessed using Laplacian variance-based sharpness metrics, with images below a defined sharpness threshold excluded. Exposure quality was evaluated through histogram analysis, and images showing overexposure (saturation) or underexposure (low signal-to-noise ratio) were discarded. To further reduce redundancy and ensure diversity in the dataset, a feature-based image similarity check was performed. Finally, a semi-automated manual review was conducted to remove images occluded by vegetation, temporary structures, or other noise sources.

2.2. Multi-spectral reconstruction

Multispectral image reconstruction is fundamentally an ill-posed inverse problem where the goal is to recover dense spectral information $S \in \mathbb{R}^{\{H\times W\times 3\}}$ from standard RGB images $I_{RGB} \in \mathbb{R}^{\{H\times W\times 3\}}$, where H, W are the image dimensions and C $\gg$ 3 is the number of target spectral channels (e.g., 31 bands from 400–700 nm). This inverse mapping is modeled as:

$$I_{RGB} = \int_{\lambda} S(x, y, \lambda) \cdot R_{\lambda} d\lambda + \varepsilon \quad (3)$$

where S(x, y, λ) is the spectral reflectance at position (x, y), $R_{\lambda} \in \mathbb{R}^3$ denotes the camera's spectral response function for RGB channels, and ε is measurement noise or model error.

In reconstruction, we aim to approximate the inverse function $F_{\theta}: \mathbb{R}^{\{H\times W\times 3\}} \rightarrow \mathbb{R}^{\{H\times W\times C\}}$ parameterized by a deep neural network with learnable weights θ:

$$\hat{S} = F_{\theta(I_{RGB})} \quad (4)$$

To model this mapping effectively, we employ the MST++ network, which introduces spectral-wise multi-head self-attention (S-MSA) to better exploit spectral continuity and band-wise dependencies—a key property of natural hyperspectral signals. MST++ adopts a coarse-to-fine multi-stage architecture, wherein each stage refines the spectral estimation at increasing resolution. A progressive learning loss is applied at each stage:

$$L_{total} = \sum_{i=1}^{N} \alpha_i \cdot \left|\left|\hat{S}^{(i)} - S_{GT}\right|\right|^1 \quad (5)$$

Where $\hat{S}^{(i)}$ is the reconstructed spectral image at stage i, and α_i are weight coefficients.

In our case, the MST++ model was deployed on UAV-captured RGB imagery over traditional villages, enhancing the material discrimination capability by revealing subtle differences in reflectance between roofing materials such as clay tiles, slate, and aged wood—often indistinguishable in RGB images [6].

2.3. Structure semantic extraction

Semantic understanding of UAV imagery over historical villages poses unique challenges due to the inherent complexity of their architectural layouts, heterogeneous roof structures, and long-term material

transformations. Conventional instance or semantic segmentation models—often trained on urban or modern datasets—fall short when applied directly to such environments. For instance, off-the-shelf rooftop segmentation models fail to distinguish between fine-grained variations such as partially damaged tiled roofs, repaired metal overlays, or historically preserved flat earthen roofs, which are common in living heritage settlements.

To address this, we propose a knowledge-guided semantic analysis framework specifically tailored to the structural and material diversity of traditional villages. At its core, we construct a domain-specific knowledge base, wherein architectural elements are categorized based on their structural typologies (e.g., gable roofs, hipped roofs, flat terraces) and material characteristics (e.g., clay tile, stone slab, metal sheet, wooden beam). This ontology enables more precise and interpretable downstream segmentation.

From this knowledge base, we design semantic prompts that guide vision-language-driven segmentation. Representative prompts include: “partially damaged tile roof”, “modern metal-repaired roof”, “traditional flat mud roof” and “sloped roof with clay tiles”. These prompts are input into a language-conditioned segmentation pipeline based on prompt-aware Segment-Anything (SAM2), which extends the SAM with natural language grounding [7] [8]. The model dynamically generates semantic masks conditioned on the input prompt, allowing flexible and interpretable segmentation of architectural elements that may not be explicitly annotated in training datasets.

Furthermore, we find that applying the prompt-conditioned segmentation on multispectral reconstructed imagery, rather than RGB inputs, leads to significantly improved performance. The enhanced spectral features reveal subtle reflectance differences between materials that may be visually indistinguishable in RGB (e.g., oxidized metal versus aged ceramic tiles, or newer wood inserts versus original beams). This spectral sensitivity allows the model to better align its semantic masks with real-world structural variations.

As a result, we are able to generate material-aware semantic maps of historical village scenes, which serve as valuable assets for downstream conservation workflows—such as condition-based maintenance planning, anomaly detection, and digital documentation of repair interventions.

3. Preliminary results

To validate the proposed framework, we conducted a field experiment over Lai Chi Wo, a well-preserved and actively inhabited historical village located in the northeastern New Territories of Hong Kong. A DJI Mavic 2 Enterprise Advanced UAV was employed for data collection, capturing aerial imagery with both nadir and oblique perspectives to comprehensively cover the architectural surfaces.

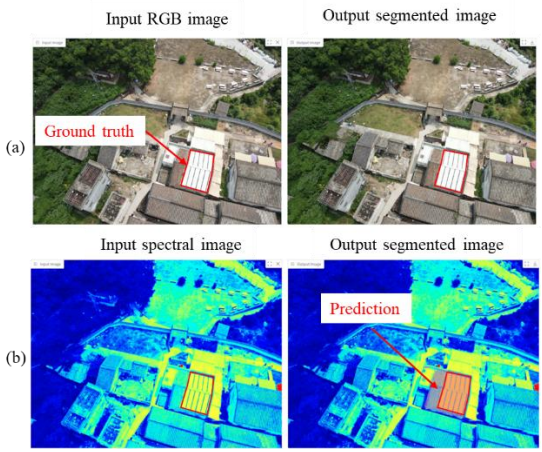
A total of 212 raw RGB images were initially captured. Following the image screening and quality control procedures described in Section 3.1.3—including checks for motion blur, redundancy, occlusion, and under-/overexposure—a final set of 66 high-resolution images (8000×6000 pixels) was retained for further analysis. These images were subsequently processed using the MST++ model to reconstruct corresponding multispectral representations.

Our analysis focused on comparing the effectiveness of semantic segmentation using RGB imagery versus reconstructed multispectral images. Specifically, we applied the prompt-aware SAM2 model to perform text-prompted segmentation tasks guided by material- or structure-specific prompts. To assess segmentation performance across different spectral bands, we employed several standard semantic segmentation metrics, including F1 score, which balances precision and recall.

We observed that different spectral bands offer varying degrees of sensitivity to architectural and environmental features:

-The blue band (450–495 nm) exhibited strong reflectance from metal surfaces, allowing clearer segmentation of metallic repair overlays that were visually ambiguous in RGB due to lighting variation or reflections (see Fig. 2). Quantitatively, the F1 score increased from 0.00% (RGB baseline) to 78.05%, indicating a substantial enhancement in segmentation performance for metal-related architectural features.

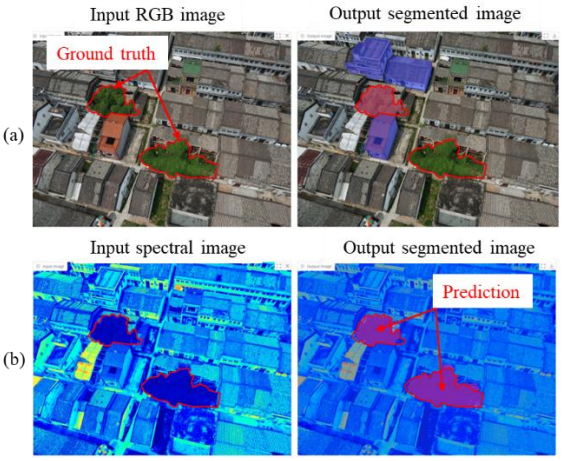
-The near-infrared band (700–750 nm) effectively suppressed background noise and enhanced the contrast of vegetation, particularly when detecting plant overgrowth within roof openings or house interiors (see Fig. 3). The F1 score improved significantly from 24.52% (RGB) to 92.77%, demonstrating the band's effectiveness in vegetation-related segmentation tasks.



(c) Performance metrics

	RGB	Blue
Precision	0	0.6458
Recall	0	0.9863
F1 score	0	0.7805
IoU	0	0.6400

Fig. 2. Comparison result of segmentation on text prompt "metal roof" between RGB and blue-band spectral image.



(c) Performance metrics

	RGB	Near-red
Precision	0.1918	0.9420
Recall	0.3397	0.9139
F1 score	0.2452	0.9277
IoU	0.1397	0.8652

Fig. 3. Comparison result of segmentation on text prompt "vegetation in the house" between RGB and near-red-band spectral image.

These results highlight the intrinsic limitations of RGB-based segmentation in complex historical village environments. Due to shadows, weathering, heterogeneous materials, and non-standard architectural forms, RGB imagery often failed to produce coherent segmentation boundaries. In contrast, multispectral reconstruction significantly improved segmentation accuracy, particularly in edge regions and visually degraded zones. Overall, the preliminary results affirm that our approach—combining spectral reconstruction, prompt-conditioned segmentation, and domain-informed semantic reasoning—provides a robust and generalizable solution for structure extraction and material analysis in heritage management scenarios.

4. Conclusion

This study presents a novel and cost-effective framework for the structure extraction and material analysis of historical village architecture using UAV-acquired RGB imagery and deep learning-based multispectral reconstruction. By leveraging spectral reconstruction via MST++, semantic prompt conditioning, and material-aware segmentation using SAM2, we demonstrate a scalable method for generating detailed semantic maps of complex rural heritage environments. Our experiments over Lai Chi Wo village confirm that the proposed approach significantly outperforms RGB-only baselines in

segmentation accuracy, particularly in cases involving complex material differentiation, shadow interference, and vegetation occlusion.

The findings illustrate the clear advantages of integrating spectral diversity and domain-informed prompting into the architectural monitoring pipeline, especially in heritage settings where conventional methods fail to generalize. Moving forward, we will continue to expand and refine this framework by integrating more robust prompt-learning strategies, validating generalizability across diverse village typologies, and exploring real-time deployment for adaptive conservation planning.

Acknowledgements

This study was funded supported in part by the Research Grants Council of Hong Kong SAR under Grants 14209623 and 14200524, and in part by the InnoHK initiative of the Innovation and Technology Commission of the Hong Kong Special Administrative Region Government via the Hong Kong Centre for Logistics Robotics.

References

[1] Zahiri, Z., Laefer, D. F., Kurz, T., Buckley, S., & Gowen, A. (2022). A comparison of ground-based hyperspectral imaging and red-edge multispectral imaging for façade material classification. Automation in Construction, 136, 104164, https://doi.org/10.1016/j.autcon.2022.104164.

[2] Sanchez, J., & Quiros, E. (2017). Semiautomatic detection and classification of materials in historic buildings with low-cost photogrammetric equipment. Journal of Cultural Heritage, 25, 21–30, https://doi.org/10.1016/j.culher.2016.11.017.

[3] Kurz, T., Buckley, S., & Howell, J. (2022). Close-range hyperspectral imaging for geological field studies: workflow and methods. International Journal of Remote Sensing, 43(15), 5940–5961, https://doi.org/10.1080/01431161.2012.727039.

[4] Peng, X., Wang, P., Zhou, K., Yan, Z., Zhong, X., & Zhao, C. (2025). Bridge defect detection using small sample data with deep learning and hyperspectral imaging. Automation in Construction, 170, 105900, https://doi.org/10.1016/j.autcon.2024.105900.

[5] Kuras, A., Brell, M., Rizzi, J., & Burud, I. (2021). Hyperspectral and Lidar Data Applied to the Urban Land Cover Machine Learning and Neural-Network-Based Classification: A Review. Remote Sensing, 13(17), 3393. https://doi.org/10.3390/rs13173393.

[6] Cai, Y., Lin, J., Lin, Z., Wang, H., Zhang, Y., Pfister, H., Timofte, R. and Van Gool, L., 2022. Mst++: Multi-stage spectral-wise transformer for efficient spectral reconstruction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 745-755), https://doi.org/10.1109/CVPRW56347.2022.00090.

[7] Ravi N, Gabeur V, Hu YT, Hu R, Ryali C, Ma T, Khedr H, Rädle R, Rolland C, Gustafson L, Mintun E. Sam 2: Segment anything in images and videos. arXiv preprint arXiv:2408.00714. 2024 Aug 1, https://doi.org/10.48550/arXiv.2408.00714.

[8] Liu S, Zeng Z, Ren T, Li F, Zhang H, Yang J, Jiang Q, Li C, Yang J, Su H, Zhu J. Grounding dino: Marrying dino with grounded pre-training for open-set object detection. InEuropean Conference on Computer Vision 2024 Sep 29 (pp. 38-55). Cham: Springer Nature Switzerland, https://doi.org/10.1007/978-3-031-72970-6_3.