Zhu and Jutzi *Urban Informatics* (2026) 5:4
https://doi.org/10.1007/s44212-025-00094-8

Urban Informatics
The International Society for Urban Informatics

RESEARCH Open Access

Automatic generation of thermal point clouds from TIR images and LiDAR point clouds

Jingwei Zhu[1] and Boris Jutzi[1,2*]

Abstract

3D thermal models are associated with building inspection and energy efficiency evaluation. Fusing Thermal infrared (TIR) images with MLS (Mobile Laser scanning) point clouds enables the generation of thermal point clouds, which combine detailed geometric data with thermal attributes at each 3D point. RGB images are typically used to reconstruct a 3D point cloud and apply thermal textures to the model. Therefore, the generated thermal point cloud heavily relies on accurate RGB reconstruction and scale estimation. In this contribution, we introduce a novel image-feature alignment method to directly co-register TIR images with MLS point clouds. The intensity images are generated from the point clouds, and corresponding feature points are matched with the TIR images. With the estimated corresponding points, the pose can be calculated, and the thermal textures are projected onto the MLS point clouds for thermal point cloud generation. Our method achieves results comparable to manual labeling with a projection error of RMSE 3.4 pixels, offering an efficient and reliable solution to generate 3D thermal models for building energy evaluations.

Keywords Point clouds, Co-registration, Image matching, TIR images, Thermal point cloud

1 Introduction

Monitoring building energy efficiency is essential for promoting urban sustainability, reducing energy consumption, and enhancing indoor thermal comfort. Among the various energy demands of a building, heating and cooling systems typically dominate overall usage. Consequently, evaluating thermal insulation performance, detecting anomalies, and identifying potential energy leakages have emerged as key priorities for governments, industry stakeholders, and researchers (Renganayagalu et al., 2024).

Infrared Thermography (IRT) imaging, a non-invasive and cost-effective sensing technology, has been widely used to capture surface temperature distributions. Unlike point-based thermometers, a thermal camera records thermal radiation across entire surfaces, enabling rapid detection of thermal anomalies such as heat loss, insulation failure, or water intrusion. However, Thermal Infra-Red (TIR) images are inherently 2D and have limited spatial resolution, narrow Field of View (FoV), and poor geometric fidelity, making it difficult to interpret thermal data in complex urban environments.

To overcome these limitations, recent studies (Macher & Landes, 2022) have explored the integration of TIR images with Light Detection and Ranging (LiDAR) point clouds, which provide accurate 3D geometric information through dense spatial sampling. The fusion of thermal texture and geometric data enables the construction of 3D thermal point clouds, where each point in space is associated with a temperature value. These enriched point clouds facilitate a more intuitive and comprehensive understanding of building thermal behavior as the high-resolution three-dimensional thermal information allows precise localization of heat loss on building envelopes.

*Correspondence:
Boris Jutzi
boris.jutzi@kit.edu
[1] Chair of Photogrammetry and Remote Sensing, Technical University of Munich, Arcisstraße 21, 80333 München, Germany
[2] Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology, Kaiserstraße 12, 76131 Karlsruhe, Germany

Despite its potential, generating accurate thermal point clouds from TIR and mobile laser scanning (MLS) data remains a non-trivial challenge. The two modalities differ in resolution, dimensionality, and sensing mechanisms: TIR captures radiance in the 7–15 μm range with blurry textures and structural detail, while MLS point cloud provides precise, basically 3D geometry. Aligning these heterogeneous datasets requires robust cross-modal co-registration, which is still an open problem. Registering images to point clouds is typically performed between RGB images and LiDAR point clouds, by estimating the corresponding 2D-3D features or the image pose with respect to the point cloud using deep-learning methods (Wang et al., 2025; Kang et al., 2024), which require large annotated datasets. However, direct registration of TIR images with point clouds for thermal point cloud generation is rarely reported.

In this paper, we address this challenge by proposing an automatic image-based pipeline for generating thermal point clouds by fusion of TIR image sequences and MLS point clouds. The proposed method utilizes projected intensity images derived from the point cloud to establish reliable 2D correspondences with the thermal images, enabling direct association of thermal textures with the original 3D points. The resulting thermal point clouds provide enhanced spatial context and valuable temperature insights for building energy diagnostics.

The main contributions of this research are:

- We propose an image-based algorithm for corresponding point detection of TIR images and MLS point clouds.
- We propose a feasible workflow to automatically co-register the TIR images to the MLS point clouds.
- Our experiments validate the results of co-registration and compare them with the manual results.

The structure of this contribution is organized as follows: In Sect. 2, we summarize the related work, and our proposed method is presented in detail in Sect. 3. The data and experiments are described in Sect. 4, and then the results are shown and discussed in Sects. 5 and 6. Finally, some conclusions are drawn in Sect. 7.

2 Related work

Generating 3D thermal models typically involves mapping 2D thermal textures from TIR images onto 3D geometric models. Model-based methods use preexisting or reconstructed 3D models and project thermal texture with known camera orientations (Hoegner & Stilla, 2018; Weinmann et al., 2012; Marie et al., 2024). Such models are often limited in representing complex surfaces and structures, and precise camera orientations are required for mapping. A widely adopted approach to generate thermal point clouds involves combining photogrammetric techniques with thermal imaging using structure from motion (SfM) (Schonberger & Frahm, 2016). Due to the blurry features and low texture in thermal images, RGB images are often used to support 3D reconstruction and improve image registration (Lin et al., 2025; López et al., 2021). Since thermal and RGB cameras typically operate at different resolutions and fields of view, scale alignment often requires manual estimation or external geo-referencing to ensure consistency with real-world coordinates. Despite point cloud co-registration, integrated systems are widely used, where only the relative pose between sensors needs to be estimated (Brea et al., 2024; Schichler et al., 2025; Qiu et al., 2025). In these approaches, the RGB imagery serves as a link between the TIR image and the high-quality 3D geometry, facilitating indirect alignment of the TIR to the point point cloud. However, the overall quality of the thermal point cloud is still highly dependent on proper illumination in the RGB images, the fidelity of the 3D reconstruction, and accurate scale alignment between modalities.

Beyond RGB-involved methods, some studies have explored direct registration between TIR images and LiDAR point clouds (Zhu et al., 2021), attempting to extract and match cross-modal features without relying on RGB intermediates. However, the robustness of these methods remains limited due to significant modality differences and weak feature correspondence. Furthermore, their scalability to large-scale datasets requires further investigation (Elias et al., 2023).

While recent research has increasingly applied deep learning techniques to cross-modal registration, particularly in RGB-LiDAR scenarios (Li & Lee, 2021; Yew & Lee, 2022; Hao et al., 2024), their adaptation to TIR-LiDAR alignment remains underdeveloped. This is largely due to the scarcity of large-scale annotated datasets and the inherent modality gap between thermal and geometric information.

To address these limitations, we propose a fully image-based method that automatically estimates the relative pose between TIR images and LiDAR point clouds, without relying on RGB assistance or pre-calibrated setups. Our approach leverages geometric priors and structure-aware constraints to enable robust thermal point-cloud generation, bridging the gap between 2D thermal observations and 3D spatial understanding.

3 Method

To bridge the dimensional gap between 2D thermal images and 3D LiDAR point clouds, we propose a image-based framework that maps 3D spatial data into a 2D domain. Instead of reconstructing 3D geometry from 2D images, which is computationally intensive and underconstrained, we project the 3D point cloud onto the image plane with a virtual camera for a virtual image simulation. This dimensional reduction preserves relevant spatial information while facilitating efficient feature matching and transformation estimation.

Figure 1 shows the overall workflow. The MLS point cloud is projected to (A) first generate the virtual image by intensity of the 3D points with the coarse vehicle pose, together with a 3D point matrix embedding the point coordinates. Multimodal feature correspondences are (B) matched between the TIR image and the intensity image. These correspondences are used to estimate the precise image pose (C) via a EPnP (Lepetit et al., 2009). The thermal texture from the TIR image is mapped onto the MLS point cloud to (D) generate a thermal point cloud.

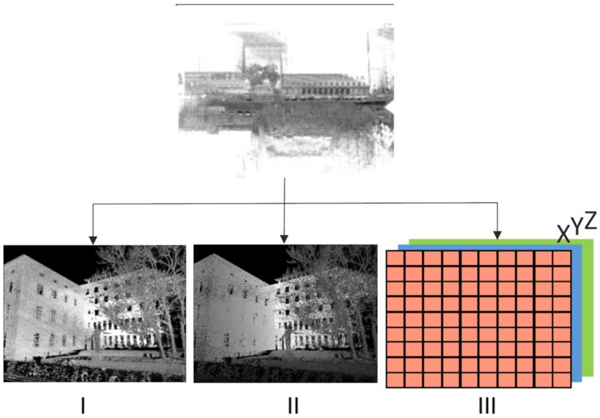

Fig. 2 Images generated from MLS point clouds: (I) Intensity image, (II) Range image, (III) Coordinate map

3.1 Intensity images generation from MLS point clouds

To enable robust 2D-3D feature matching, the MLS point cloud is first converted into an intensity image. The intensity represents the surface reflectance measured by the laser scanner. Depending on the sensor's wavelength, different materials and textures exhibit distinct reflectance, so the intensity encodes textural and material contrasts that are not present in pure geometry, thereby enabling clearer object boundaries and more reliable feature detection. As shown in Fig. 2, while both the intensity and range modalities capture geometric structure, the intensity image provides enhanced representation for reflective or fine-structured surfaces, such as nearby bicycles and building windows, where material reflectivity highlights features that are less pronounced in the range image. Since we cannot directly produce such a geometry-based intensity map, we employ a virtual pinhole camera model similar to our thermal camera to simulate the imaging geometry. The MLS point cloud is projected onto an image plane using the collinearity (projection) equation (Eq. 1). Given an initial virtual camera pose close to the vehicle, each 3D point X_i is projected onto a 2D image coordinate u_i using the intrinsic matrix

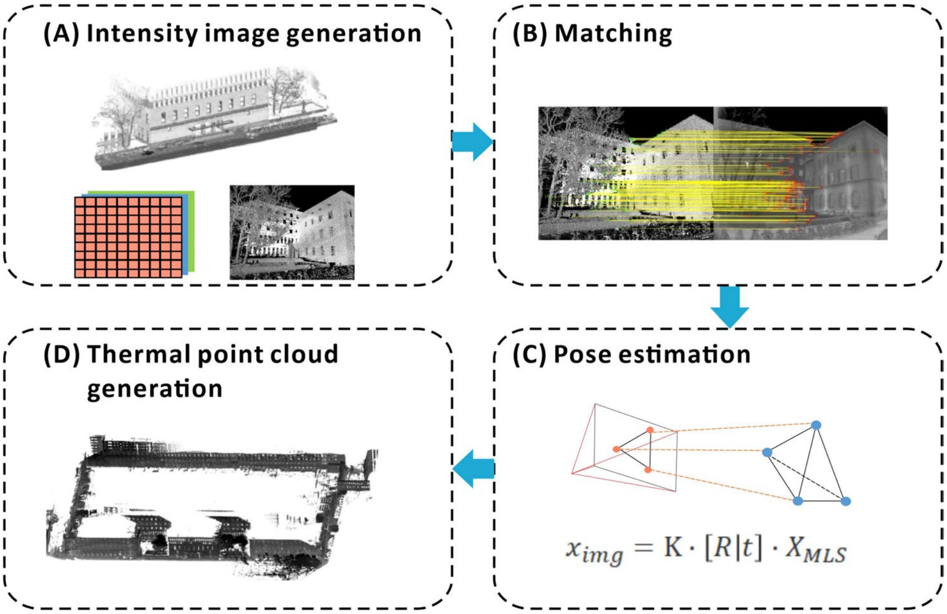
(A) Intensity image generation
(B) Matching
(D) Thermal point cloud generation
(C) Pose estimation
$x_{img} = \mathrm{K} \cdot [R|t] \cdot X_{MLS}$

Fig. 1 Overall workflow for thermal point cloud generation (**A**-**D**)

of the camera K, rotation matrix R, and translation vector t:

$$u_i = K[R|t]X_i \tag{1}$$

$$K = s \cdot \begin{bmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix}, \quad R \in SO(3), \quad t = [X_0, Y_0, Z_0]^T \tag{2}$$

To handle occlusions and ensure physically consistent rendering, a buffer with $20cm$ is applied so that only the nearest point is retained for each pixel. In addition to the intensity image, the range image encodes, for each pixel, the shortest 3D distance from the camera center to the visible surface point, as well as a 3D coordinate map that records the corresponding 3D position of that pixel are automatically generated.

3.2 Multimodal feature matching

Due to the spectral and radiometric differences between TIR and MLS-derived images, classical descriptors (e.g., SIFT (Lowe, 2004a)) are ineffective. We adopt the HAPCG descriptor (Yao et al., 2021), which leverages anisotropic diffusion and phase congruency (Kovesi, 1999) to extract robust features invariant to intensity and orientation changes. Feature detection is performed on both images using phase-based anisotropic filtering, and Harris corner detection (Harris & Stephens, 1988) is then constructed in polar coordinates using gradient orientation histograms. Descriptor similarity is measured via Euclidean distance, and false matches are filtered with Fast Sample Consensus (Wu et al., 2014), which requires that inlier correspondences satisfy an affine geometric consistency within local neighborhoods.

3.3 Pose estimation via PnP

Given 2D-3D corresponding point pairs, the precise camera pose is estimated by solving the Perspective-n-Point (PnP) problem. The objective is to minimize the reprojection error.

$$\text{Reprojection Error} = \sum_{i=1}^{N} \|u_i - \Pi(K, R, T, X_i)\|^2 \tag{3}$$

We employ the efficient PnP (EPnP) algorithm (Lepetit et al., 2009) to estimate the 6-DoF pose between the thermal camera and the 3D LiDAR point cloud. EPnP formulates the 3D-to-2D correspondence problem by expressing all 3D points as a weighted combination of four virtual control points using barycentric coordinates. This formulation transforms the problem into a linear system, which allows for an efficient initial pose estimation in O(n) time complexity, where n is the number of 2D-3D correspondences. Gauss-Newton iteration is applied to minimize reprojection errors for pose refinement.

3.4 Thermal point cloud rendering

Once the TIR camera pose for each image is known, the thermal texture is mapped onto the point cloud. We adopt an indirect rendering strategy Fig. 3b: The point cloud is reprojected into the TIR image frame, and the thermal value of each point is obtained via bilinear interpolation (Gonzales & Wintz, 1987) from the surrounding pixel intensities. This allows for sub-pixel sampling when a projected 3D point does not fall exactly on a pixel center, which is a standard and effective technique in image-to-point cloud fusion. Compared with a ray-tracing strategy (Fig. 3a) that traces each image pixel into 3D and often yields sparse thermal coverage, the indirect approach assigns radiance to every visible 3D point and therefore produces a denser and more visually complete thermal point cloud, which is essential for our application.

To ensure occlusion-aware projection, we use the range image to filter out points beyond visible thresholds. Due to the limited field-of-view of 2D images, a sub thermal cloud can be rendered from the corresponding TIR image. All the clouds are merged by averaging overlapping values, resulting in a consistent and dense thermal point cloud for the whole area.

4 Data and experiments

The test site is selected from TUM2TWIN (Wysocki et al., 2025) dataset (Fig. 4), measured using a mobile platform, which includes two laser scanners and an uncooled bolometer (Zhu et al., 2020). This area is around $140m \times 200m$, with 69,855,517 points. The uncooled thermal imaging camera is cross-mounted and looks backward toward the drive direction. TIR images

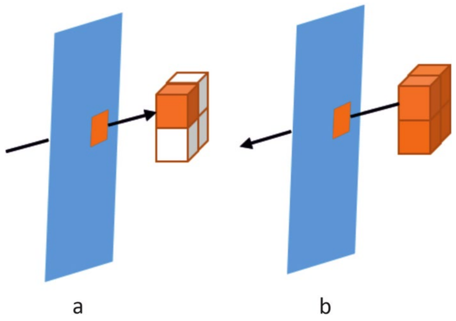

Fig. 3 Thermal rendering strategies: **a** Ray tracing, **b** Indirect projection

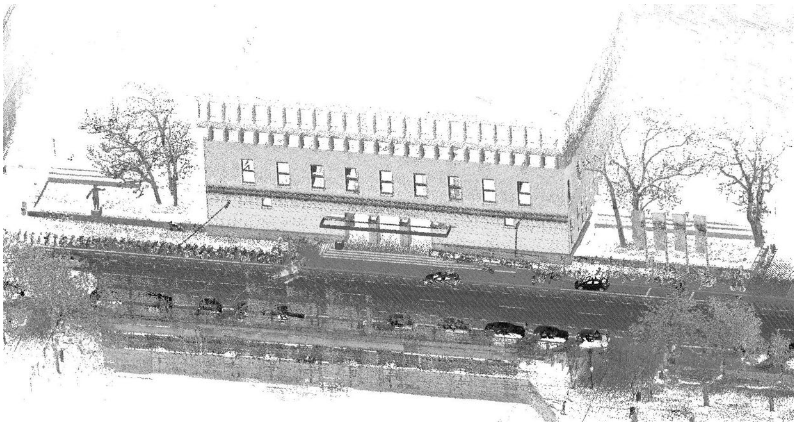

Fig. 4 Testing site selected from TUM2TWIN (Wysocki et al., 2025)

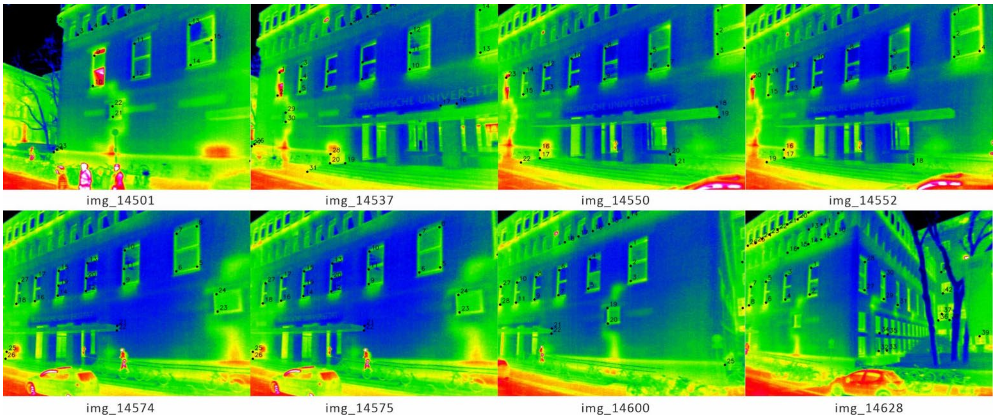

Fig. 5 Manually labeled corresponding points in TIR images with pseudo-color

are provided as 16-bit-TIFFs with lossless compression (LZW) in the size of 640 pixel × 480 pixel. Additional information about the car's position is provided by Global Positiioning System (GPS) and inertial measurement units (IMU) mounted on the vehicle.

In order to estimate the pose calculated with the proposed method, we select eight TIR images and manually label the corresponding image points and 3D points in the point clouds for evaluation. The selected images are presented in Fig. 5. Though only four points are required for pose estimation, we select more than 10 points each to increase the redundancy.

The processing and generation of thermal point clouds were done using c++ and pcl library(1.81) (Rusu & Cousins, 2011). The computer is with 32G RAM, and an i7-6000 @3.4 GHz CPU. The implementation is available in a public repository.[1]

5 Results

This section presents the experimental results and evaluates the performance of the proposed method. The TIR images were calibrated for the intrinsic parameters of the camera. MLS point clouds from two scanners were merged and processed to eliminate noise, redundancy, and outliers. As a result, the point count was reduced by over 50%, leaving 23.2 million points for processing.

Figure 6 shows the generated intensity image alongside its corresponding TIR image. The initial pose of the virtual camera is initialized from the vehicle's GPS position to ensure sufficient overlap with the TIR image. Despite the modality differences, objects such as buildings, vehicles, and trees are recognizable in both representations. However, due to the varying radiance characteristics of different objects (e.g., cars appear bright in TIR but dark in intensity images, whereas trees exhibit the opposite pattern), the generated intensity image suffers from contrast variations and noise artifacts, such as salt-and-pepper noise. To mitigate these projection-induced artifacts,

[1] https://github.com/JingweiFZ/RegImg-tir-pcd

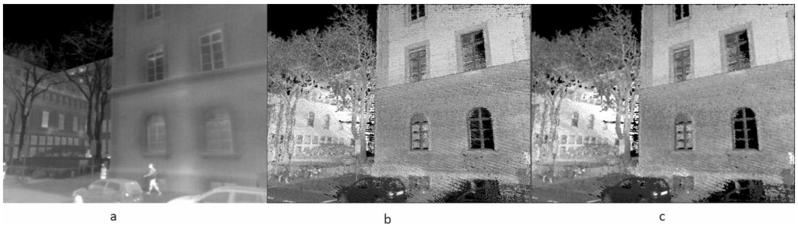

Fig. 6 Example of image 14729: **a** TIR image, **b** generated intensity image, **c** intensity image after median filter

a median filter with 3 × 3 was applied as shown in Fig. 6c. This operation effectively replaces each pixel with the median of its neighborhood, removing these isolated spikes while preserving true edges and structural details for matching.

Figure 7 illustrates matched corresponding points across modalities. The red points are detected feature points in TIR images, and the green ones are those in the corresponding intensity images. Matches are spatially distributed, with clusters along windows and facades. Some outliers are present, but are mitigated through descriptor redundancy and consensus filtering.

Using the HAPCG descriptor, corresponding point pairs are effectively detected through multimodal feature matching between TIR and intensity images. As shown in Fig. 8, an average of 254.00 corresponding points are identified per image, which significantly surpasses the number obtained through manual annotation (average: 28.13 points). This substantial improvement highlights the capability of HAPCG to capture rich and robust feature representations across modalities, even under challenging conditions such as contrast inversion and radiometric inconsistencies. The dense and consistent detection of matched points not only reduces reliance on labor-intensive manual labeling but also ensures a more accurate and repeatable image registration process, which is critical for downstream tasks of 2D-3D fusion and scene reconstruction.

To evaluate camera pose accuracy, we projected labeled 3D points into the 2D image plane using the estimated parameters and compared the results with ground truth (manually) and GPS pose of the vehicle for intial intensity image generation. L1 (Eq. 4) measures the average distance from the projected points to the corresponding image points, while RMSE is the quadratic mean of the differences between the observed values and the predicted ones (Eq. 5). Table 1 shows the L1 and RMSE for each image. According to the result, our method achieves an average RMSE of 3.40 pixels, close to manual annotation (2.53 pixels) and significantly better than GPS (24.13 pixels). Notably, for image 14628, a higher RMSE is observed due to limited scene contrast.

$$\mathrm{L1}(p,\hat{p}) = \frac{1}{N}\sum_{i=0}^{N-1}|p_i - \hat{p}_i| \tag{4}$$

$$\mathrm{RMSE}(p,\hat{p}) = \sqrt{\frac{1}{N}\sum_{i=0}^{N-1}(p_i - \hat{p}_i)^2} \tag{5}$$

To visualize and compare the differences, sub thermal point clouds were generated for each image by projecting

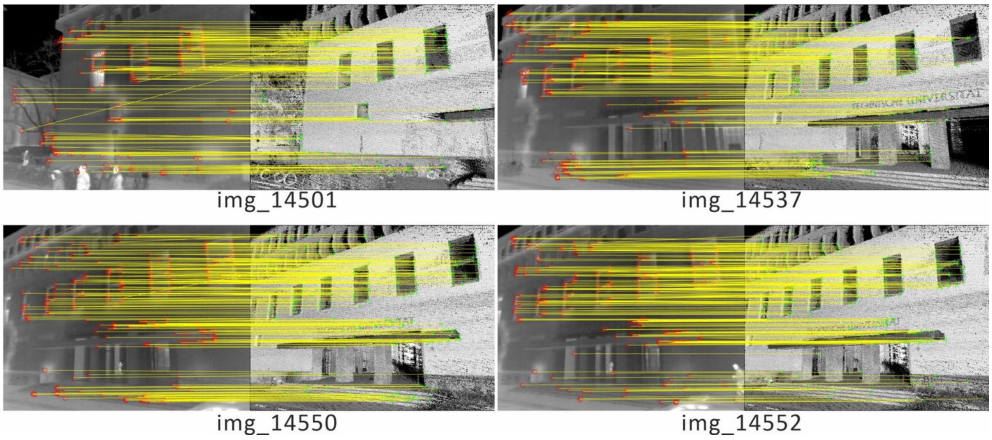

Fig. 7 Matched corresponding points between TIR (red) and intensity (green) images

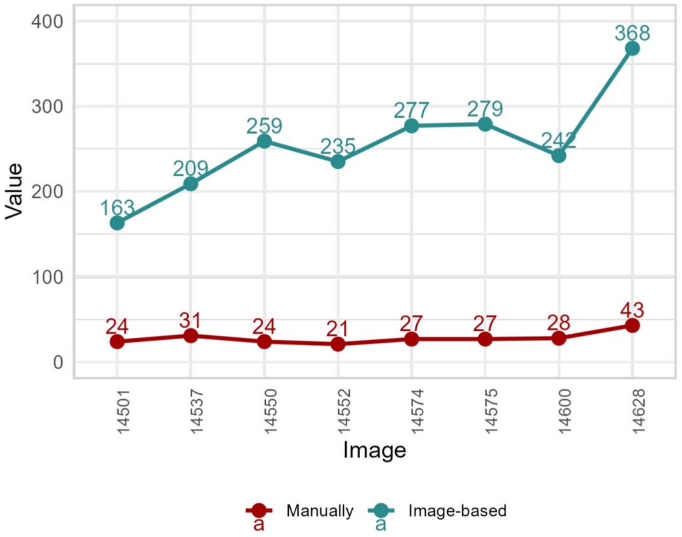

Fig. 8 Number of detected feature points: manual vs. image-based method

Table 1 Pose estimation accuracy comparison (L1 and RMSE in pixels)

	GPS[a]		Proposed		Manual	
Image	**L1**	***RMSE***	**L1**	***RMSE***	**L1**	***RMSE***
14501	37.71	21.52	5.11	3.23	3.48	2.26
14537	42.76	24.82	5.67	3.47	4.16	2.74
14550	39.52	23.75	5.24	3.23	4.70	2.93
14575	43.85	25.96	3.98	2.56	3.18	1.98
14600	38.48	22.81	6.70	3.97	4.43	2.65
14628	37.62	23.50	7.24	4.40	4.30	2.85
14552	41.66	25.15	5.29	3.23	4.10	2.50
14574	43.12	25.55	4.94	3.14	3.66	2.31
Average	40.59	24.13	5.52	3.40	4.00	2.53

[a] Result using initial vehicle GPS position data

the 2D thermal texture onto the corresponding 3D point clouds (Fig. 9). While window and building boundaries align accurately with the geometric point cloud by manually results, distinct shifts attributable to the initial GPS solution are evident in the thermal point clouds, as highlighted by the red rectangular annotations. The image-based method highly improved the initial pose from the GPS data, and the features are visually accessible. Compared to GPS-based projections, which show visible misalignment, our results closely match manually labeled outputs without requiring human annotation.

Finally, the fused thermal point cloud for the test site is presented in Fig. 10, providing a comprehensive 3D representation of surface temperature distribution. Warmer regions are indicated by colors closer to red, while cooler areas shift towards yellow. Notably, the roads exhibit significantly higher temperatures than surrounding buildings, likely due to their heat-retaining asphalt materials and direct solar exposure. In addition, several linear vertical patterns of elevated temperature can be observed on the façades, particularly near window areas and building entrances. These thermal anomalies are aligned with the structural layout of the windows and doors, suggesting the presence of heat leakage paths. Such patterns may correspond to internal heating elements, such as radiators or pipelines, transferring thermal energy from the building interior to its external surfaces.

6 Discussion

The proposed workflow is training-free and data-efficient, offering strong cross-modal matching without the overhead of model training. By leveraging the proposed HAPCG descriptor, our method effectively captures modality-invariant features, enabling accurate and dense

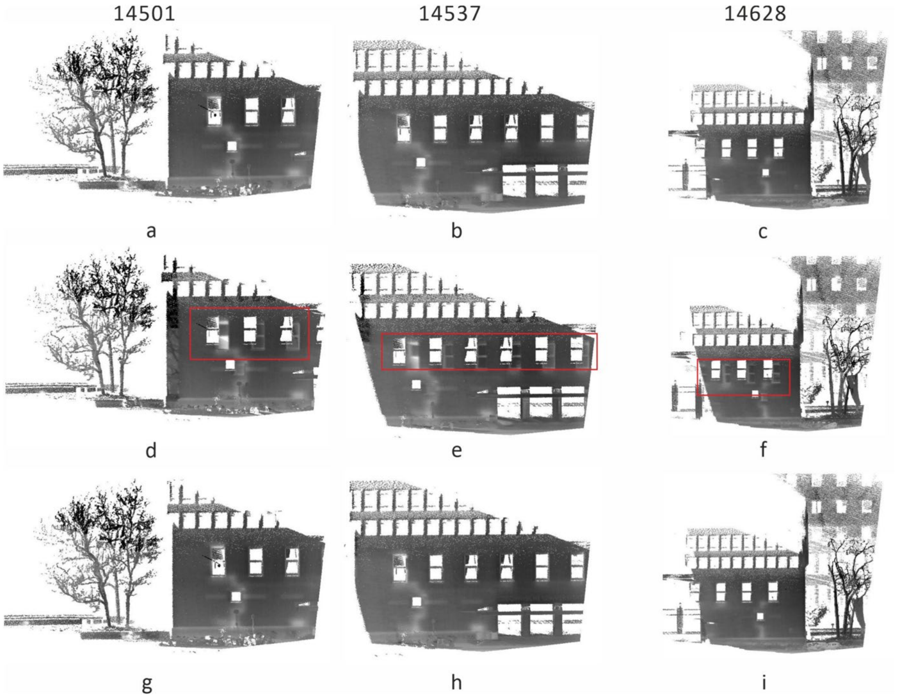

Fig. 9 Thermal point clouds by rendering thermal texture to the point cloud: **a**-**c** manually labeled result, **d**-**f** initial GPS pose, **g**-**i** image-based method

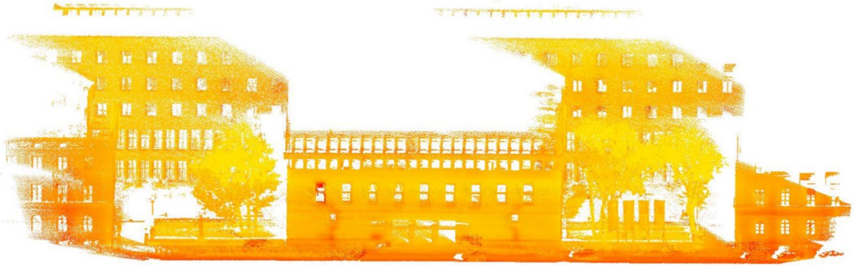

Fig. 10 Fused thermal point cloud of the test site

correspondences between TIR and intensity images. This facilitates reliable camera-pose estimation and multimodal registration. Compared to manual annotation, the automatic feature-matching approach yields a significantly greater number of correspondences, thereby improving registration completeness and reducing manual effort.

To further position our approach with respect to common feature descriptors, we evaluated representative baselines including the classical hand-crafted SIFT (Lowe, 2004b), the learning-based matcher SuperGlue (Sarlin et al., 2020), and the cross-modality descriptor MIRRIFT (Geng et al., 2025) on eight representative image pairs. SIFT and SuperGlue rarely produce more than a handful of geometrically verified matches and often fail to meet the minimum requirement for stable pose estimation, with visual inspection confirming many false matches. MIRRIFT achieves higher inlier counts than SIFT and SuperGlue but remains well below HAPCG. In contrast, HAPCG consistently detects one to two orders of magnitude more accurate correspondences in all pairs tested. This superior performance is mainly attributed to HAPCG's ability to integrate cross-modal texture modeling with geometric constraints, allowing it to remain robust under strong spectral and structural discrepancies.

These results demonstrate that while classical or purely learning-based descriptors can be effective in single-modality RGB scenarios, they are inadequate for the challenging thermal intensity cross-modality matching required here. The superior and robust performance of HAPCG underscores the importance of incorporating domain-specific geometric constraints and cross-modal texture modeling for accurate and dense 2D correspondence, which is critical for downstream camera pose estimation and thermal point-cloud generation.

The resulting thermal point cloud captures fine-scale thermal variations across the urban scene, demonstrating both geometric precision and radiometric consistency. Distinct thermal patterns associated with different objects such as roads, façades, and heating structures can be clearly visualized and analyzed. These results highlight the model's capacity not only to reconstruct 3D geometry but also to provide interpretable thermal information linked to the underlying structural and functional characteristics of the environment.

Despite these advantages, some limitations remain. First, feature matching performance may deteriorate in areas with low texture or homogeneous thermal responses, such as glass surfaces or occluded regions. Second, generating intensity images from point clouds inherently compresses the spatial richness of the 3D data, potentially leading to a loss of geometric detail. Future work could also explore hybrid strategies that combine the adopted HAPCG descriptor with lightweight learning-based methods to reduce annotation requirements and further improve robustness in texture-poor or occluded areas.

7 Conclusion

We presented an automatic image-based method for generating thermal point clouds by fusing TIR images with MLS point clouds. The key innovation lies in converting 3D point cloud into 2D intensity images, enabling robust multimodal feature matching in the image domain. This dimensional alignment allows for accurate pose estimation using HAPCG descriptors and EPnP optimization, followed by thermal texture projection onto 3D geometry.

The proposed framework significantly improves pose accuracy compared to GPS-based initialization, achieving an average RMSE of 3.4 pixels, comparable to manually labeled ground truth but without labor-intensive annotation. The generated thermal point clouds effectively visualize building-scale heat distributions and structural features such as pipelines and façade radiation, demonstrating their potential for energy diagnostics and infrastructure monitoring.

Acknowledgements

We would like to express our gratitude to the Leonhard Obermeyer Center (LOC) at the Technical University of Munich (TUM) and Fraunhofer IOSB for their support in facilitating this work.

Authors' contributions

J.Z. wrote the main manuscript text, and B.J. revised it. All authors reviewed the manuscript.

Funding

No funds, grants, or other support were received for conducting this study.

Data availability

Data cannot be shared openly but are available upon request from the authors.

Declarations

Competing interests

The authors declare no competing interests.

Received: 8 August 2025 Revised: 11 November 2025 Accepted: 15 November 2025
Published online: 01 February 2026

References

Brea, A., García-Corbeira, F. J., Tsiranidou, E., Peláez, G. C., Díaz-Vilariño, L., & Martínez, J. (2024). Low-cost thermal point clouds of indoor environments. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 48*, 99–105.

Elias, M., Weitkamp, A., & Eltner, A. (2023). Multi-modal image matching to colorize a slam based point cloud with arbitrary data from a thermal camera. *ISPRS Open Journal of Photogrammetry and Remote Sensing, 9*, 100041.

Geng, Z., Yang, B., Pi, Y., Fan, Z., Dong, Y., Huang, K., & Wang, M. (2025). MIRRIFT: Multimodal Image Rotation and Resolution Invariant Feature Transformation. *IEEE Transactions on Geoscience and Remote Sensing, 63*, 1–16.

Gonzales, R. C., & Wintz, P. (1987). *Digital image processing*. Addison-Wesley Longman Publishing Co. Inc.

Hao, M., Zhang, Z., Li, L., Dong, K., Cheng, L., Tiwari, P., & Ning, X. (2024). Coarse to fine-based image-point cloud fusion network for 3d object detection. *Information Fusion, 112*, 102551.

Harris, C., & Stephens, M. (1988). A combined corner and edge detector. In *Alvey vision conference*, volume 15 (pp. 10–5244). Citeseer.

Hoegner, L., & Stilla, U. (2018). Mobile thermal mapping for matching of infrared images with 3d building models and 3d point clouds. *Quantitative Infrared Thermography Journal, 15*(2), 252–270.

Kang, S., Liao, Y., Li, J., Liang, F., Li, Y., Zou, X., Li, F., Chen, X., Dong, Z., & Yang, B. (2024). CoFiI2P: Coarse-to-Fine Correspondences-Based Image to Point Cloud Registration. *IEEE Robotics and Automation Letters, 9*(11), 10264–10271.

Kovesi, P. (1999). Image features from phase congruency. *Videre Journal of Computer Vision Research, 1*(3), 1–26.

Lepetit, V., Moreno-Noguer, F., & Fua, P. (2009). Epnp: An accurate o(n) solution to the pnp problem. *International Journal of Computer Vision, 81*, 155–166.

Li, J., & Lee, G. H. (2021). Deepi2p: Image-to-point cloud registration via deep classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 15960–15969). IEEE.

Lin, D., Yang, N., Miao, Q., Cui, X., & Xu, D. (2025). True 3d thermal inspection of buildings using multimodal uav images. *Journal of Building Engineering, 100*, 111806.

López, A., Jurado, J. M., Ogayar, C. J., & Feito, F. R. (2021). An optimized approach for generating dense thermal point clouds from uav-imagery. *ISPRS Journal of Photogrammetry and Remote Sensing, 182*, 78–95.

Lowe, D. G. (2004a). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision, 60*, 91–110.

Lowe, G. (2004b). Sift - the scale invariant feature transform. *International Journal of Computer Vision, 2*(91–110), 2.

Macher, H., & Landes, T. (2022). Combining tir images and point clouds for urban scenes modelling. In *XXIV ISPRS Congress "Imaging today, foreseeing tomorrow", Commission II 2022 edition, 6–11 June 2022, Nice, France*, volume 43 (pp. 425–431). ISPRS.

Marie, E., Lecomte, V., Landes, T., Macher, H., & Delasse, C. (2024). Temporal and thermal visualization by fusion of thermal images and 3d mesh. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII–2/W8–2024*, 319–326.

Qiu, Z., Martínez-Sánchez, J., & Arias, P. (2025). Fusion of thermal images and point clouds for enhanced wall temperature uniformity analysis in building environments. *Energy and Buildings, 339*, 115781.

Renganayagalu, S. K., Bodal, T., Bryntesen, T.-R., & Kvalvik, P. (2024). Optimising energy performance of buildings through digital twins and machine learning: Lessons learnt and future directions. In *2024 4th International Conference on Applied Artificial Intelligence (ICAPAI)* (po. 1–6). IEEE.

Rusu, R. B., & Cousins, S. (2011). 3d is here: Point cloud library (pcl). In *2011 IEEE international conference on robotics and automation* (pp. 1–4). IEEE.

Sarlin, P.-E., DeTone, D., Malisiewicz, T., & Rabinovich, A. (2020). Superglue: Learning feature matching with graph neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 4938–4947). IEEE.

Schichler, L., Festl, K., & Solmaz, S. (2025). Robust multi-sensor fusion for localization in hazardous environments using thermal, lidar, and gnss data. *Sensors, 25*(7), Article 2032.

Schonberger, J. L., & Frahm, J.-M. (2016). Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4104–4113). IEEE.

Wang, G., Zheng, Y., Wu, Y., Guo, Y., Liu, Z., Zhu, Y., Burgard, W., & Wang, H. (2025). End-to-end 2D–3D registration between image and LiDAR point cloud for vehicle localization. *IEEE Transactions on Robotics, 41*, 4643–4662.

Weinmann, M., Hoegner, L., Leitloff, J., Stilla, U., Hinz, S., & Jutzi, B. (2012). Fusing passive and active sensed images to gain infrared-textured 3d models. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXIX–B1*, 71–76.

Wu, Y., Ma, W., Gong, M., Su, L., & Jiao, L. (2014). A novel point-matching algorithm based on fast sample consensus for image registration. *IEEE Geoscience and Remote Sensing Letters, 12*(1), 43–47.

Wysocki, O., Schwab, B., Biswanath, MK., Greza, M., Zhang, Q., Zhu, J., Froech, T., Heeramaglore, M., Hijazi, I., Kanna, K., Pechinger, M. TUM2TWIN: Introducing the Large-Scale Multimodal Urban Digital Twin Benchmark Dataset. arXiv preprint arXiv:2505.07396. 2025. https://arxiv.org/abs/2505.07396

Yao, Y., Zhang, Y., Wan, Y., Liu, X., & Guo, H. (2021). Heterologous images matching considering anisotropic weighted moment and absolute phase orientation. *Geomatics and Information Science of Wuhan University, 46*(11), 1727–1736.

Yew, Z. J., & Lee, G. H. (2022). Regtr: End-to-end point cloud correspondences with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 6677–6686). IEEE

Zhu, J., Gehrung, J., Huang, R., Borgmann, B., Sun, Z., Hoegner, L., Hebel, M., Xu, Y., & Stilla, U. (2020). TUM-MLS-2016: An annotated mobile lidar dataset of the tum city campus for semantic point cloud interpretation in urban areas. *Remote Sensing, 12*(11), Article 1875.

Zhu, J., Xu, Y., Ye, Z., Hoegner, L., & Stilla, U. (2021). Fusion of urban 3d point clouds with thermal attributes using mls data and tir image sequences. *Infrared Physics & Technology, 113*, 103622.

Publisher's Note