

# 360-Degree Tri-Modal Scanning: Engineering a Modular Multi-Sensor Platform for Semantic Enrichment of BIM Models

Fiona C. Collins<sup>1</sup>, Florian Noichl<sup>1</sup>, Martin Slepicka<sup>1</sup>, Gerda Cones<sup>1</sup>, André Borrmann<sup>1</sup>

<sup>1</sup> Chair of Computational Modeling and Simulation, School of Engineering and Design, Technical University of Munich, 80333 Munich, Germany - firstname.lastname@tum.de

**Keywords:** point cloud, thermography, sensor fusion, Scan-to-BIM, Scan-vs-BIM.

## Abstract

Point clouds, image data, and corresponding processing algorithms are intensively investigated to create and enrich Building Information Models (BIM) with as-is information and maintain their value across the building lifecycle. Point clouds can be captured using LiDAR and enriched with color information from images. Complementary to such dual-sensor systems, thermography captures the infrared light spectrum, giving insight into the temperature distribution on an object's surface and allowing a diagnosis of the as-is energetic health of buildings beyond what humans can see. Although the three sensor modes are commonly used in pair-wise combinations, only a few systems leveraging the power of tri-modal sensor fusion have been proposed. This paper introduces a sensor system comprising LiDAR, RGB, and a radiometric thermal infrared sensor that can capture a 360-degree range through bi-axial rotation. The resulting tri-modal data is fused to a thermo-color point cloud from which temperature values are derived for a standard indoor building setting. Qualitative data analysis shows the potential for unlocking further object semantics in a state-of-the-art Scan-to-BIM pipeline. Furthermore, an outlook is provided on the cross-modal usage of semantic segmentation for automatic, accurate temperature calculations.

## 1. Introduction

With the growing need to retrofit large numbers of existing buildings, digital processes such as Building Information Modeling (BIM) become relevant for renovation planning. BIM is a method streamlining the work of various stakeholders primarily involved in the planning and construction of buildings by leveraging model-based collaboration across disciplines (Borrmann et al., 2018). The resulting 3D geometric-semantic models can maintain their value throughout the entire lifecycle of the building if they are updated with required as-is information. Such digital models are non-existent or manifest low semantic information depth for buildings constructed before the introduction or during the maturing phase of BIM. As-is data captured using sensing devices such as LiDAR (Light Detection and Ranging) sensors, RGB (Red Green Blue, referring to conventional photography), or TIR (Thermal Infrared) cameras can be used to generate models from scratch or enrich existing ones (Schreyer and Hoque, 2009). Research in the domain of Scan-to-BIM aims to automate these processes (Son et al., 2015). In automatic Scan-to-BIM methods, 3D point clouds are used to reconstruct geometries and semantics of structural elements (e.g., walls, slabs, doors, windows, etc.) with point enrichment methods from Computer Vision (Pan et al., 2022a; Mehranfar et al., 2023). RGB images are often used to add color to the point cloud and allow for the additional recognition and reconstruction of smaller objects such as smoke detectors or signs – useful both for automated and manual processing (Pan et al., 2022b; De Geyter et al., 2023). TIR captures emitted radiation in the infrared range by mapping it onto images as temperature readings. Experts often use this technology to help estimate the energy performance (Pereira et al., 2021) and help decide on the best retrofit strategy (Wang and Cho, 2015). TIR cameras have significantly lower resolution than RGB cameras, and the color contrasts do not necessarily match the physical boundaries of the captured object. Therefore, the interpretation

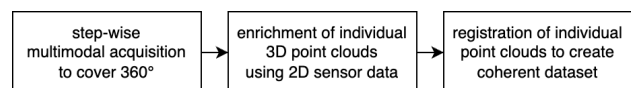


Figure 1. Top-level overview describing the main steps of the presented data acquisition method.

of TIR images is often facilitated by fusing them with RGB images (Zhang et al., 2023) or with 3D point cloud data or digital building models (Hoegner and Stilla, 2018).

This work demonstrates a prototypical tri-modal sensor system hosting LiDAR, RGB, and radiometric TIR sensors with two driven axes (pitch and yaw) for 360° acquisition. The system is engineered in a modular way and is demonstrated in this work mounted on a tripod. Fig. 1 provides an overview of data acquisition and processing steps using the presented system. First, the system collects data from all three sensors step by step in a pre-defined path. The individual point clouds captured by the LiDAR sensor are subsequently enriched with the corresponding RGB and radiometric TIR data. The individual, enriched point clouds are finally registered to form a 360° tri-modal point cloud to create a coherent dataset. The resulting data is rich in information, which is potentially useful for numerous applications provisioning as-is building data to planners. The data is showcased for recognizing building element semantics that are typically overlooked by Scan-to-BIM processes limited on data from single- or dual-sensor systems.

## 2. Related works

Using thermographic data in combination with BIM for energetic health inspection has been the focus of researchers before. On the one hand, thermographic data can be projected onto 3D surface geometry potentially derived from BIM using textures and provided as a basis for detection algorithms or expert in-

spection. Hoegner and Stilla (2015), for instance, use combined facade textures from different acquisition times or viewing directions to document the night cooling effect and identify thermal leakages. On the other hand, researchers seek to enrich the semantic building elements in BIM with as-is thermal properties such as thermal resistances. Ham and Golparvar-Fard (2015), reconstruct a thermal point cloud and consider points in the vicinity of building elements to derive as-is thermal resistances for construction types. Sadhukhan et al. (2020) propose the use of an instance segmentation network for estimating transmittance values of doors, walls, windows, and facades and validate their findings with standards. The enriched BIM models can ultimately be used for energy simulation (Wang and Cho, 2015) and the design of retrofit strategies.

Thermographic images can be acquired in combination with RGB images, and a subsequent photogrammetric reconstruction is used to reconstruct the 3D point cloud (Ham and Golparvar-Fard, 2013; Hoegner and Stilla, 2018). Other methods co-acquire thermographic images with LiDAR data (Zhu et al., 2021; Biswanath et al., 2023). Such dual or tri-modal sensor systems can be mounted on various platforms. For the inspection of roofs, Dahaghin et al. (2019) have attached the TIR and RGB on Unmanned Aerial Vehicles (UAVs) and demonstrate the evident thermal anomalies in the thermal point cloud. For building interiors, combined sensor systems have been proposed as hand-held systems (Ham and Golparvar-Fard, 2015) or as mounts to robotic systems. Borrmann et al. (2014) for instance, fuse the sensor data (LiDAR, RGB, TIR) with a series of calibration steps and register the scans from the robot positions into one global point cloud with simultaneous localization and mapping (SLAM) to help identify heat sources in electrical building systems. A similar robotic system has been proposed more recently by Adán et al. (2019) where the three sensors are additionally placed on a pan-tilt unit, allowing the system to acquire 360° thermal point clouds in horizontal direction at each step the robot takes. Wang and Cho (2015) combine two LiDAR sensors with a TIR camera on a tripod. In summary, existing research lays the basis for multi-sensor systems for building inspection, both indoor and outdoor. However, the exploration of data applications for refining the Scan-to-BIM process remains superficial.

### 3. Method

This Section first describes the hardware setup with its sensor and movable axes. Then, the two cameras' intrinsic and extrinsic calibration steps are documented to find projections between the LiDAR point cloud and the images. After that, suitable poses for acquisition are calculated, and using a series of pitch-yaw angles, a tri-modal 360° point cloud is assembled. Finally, the steps to derive the temperature values from the raw thermal intensity values are described. The reader is referred to Fig. 2 for an overview of the method.

#### 3.1 Hardware setup

The hardware setup comprises three sensing devices acquiring point cloud data, RGB images, and TIR images. The prototype setup alongside the major acquisition coordinate systems and rotating axes is depicted in Fig. 3. The two axes do not intersect and are not coincident with the origin of the world coordinate system. It is thus necessary to apply a series of translations and rotations to register the single scans into a 360° thermal point cloud.

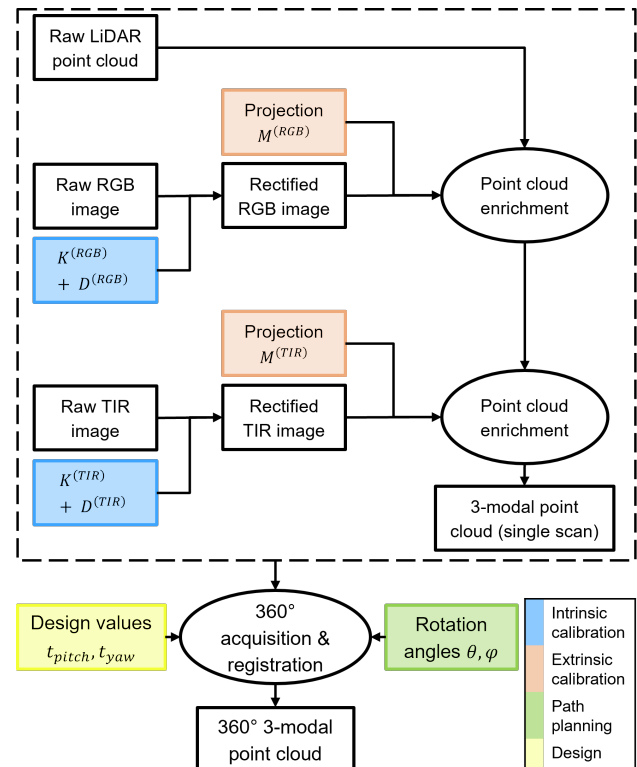


Figure 2. Method overview: A-priori system calibration and stepwise scanning to 360° tri-modal point cloud.

**3.1.1 Sensors** Three different sensing devices are mounted on a rigid 3D-printed bracket, maintaining a consistent relative positioning among them. The bracket itself is mounted on a platform to be rotated, as described in Section 3.1.2. The following sensors are hosted by the proposed system (cf. Fig. 3): A TIR camera (FLIR A50) and a RGB camera (RLC-810A) are mounted on either side of the bracket. The TIR camera senses emitted heat from objects as intensity values. It has an image resolution of 464x348 and a thermal sensitivity of 35mK. The RGB camera is used for RGB data acquisition with an image resolution of 3840x2160. Finally, a LiDAR device (Blickfeld Cube 1) is installed on the top side of the bracket. The device emits a laser beam in the near-infrared spectrum at a wavelength of 905nm. Configuration of acquisition parameters such as Field of View (FoV) is possible via an interface. The FoV of the LiDAR and the scan resolution are configured to capture a maximum number of measurements in the three devices' overlapping FoV. See Section 3.2.3 for more detail.

**3.1.2 Rotation axes** The bracket mentioned in the previous Section is fastened to a platform via two rotating axes (pitch and yaw) while the platform is fixed on a tripod. Each axle is actuated with a stepper motor, which can be controlled via an Arduino microcontroller. The mounted sensors can be aimed in almost any direction around the platform by sending both pitch and yaw rotation angles via a TCP Socket Stream connection to the Arduino. The vertical axle (yaw) can be rotated roughly 360° in either direction, limited only by the connection cables of the sensors and the motors. The horizontal axle (pitch) can be rotated from 0° (sensors facing straight upwards) to roughly 170°. Rotations to larger angles are unnecessary since LiDAR has a minimum acquisition distance. In both directions, the movement is limited by the platform on which the bracket is mounted.

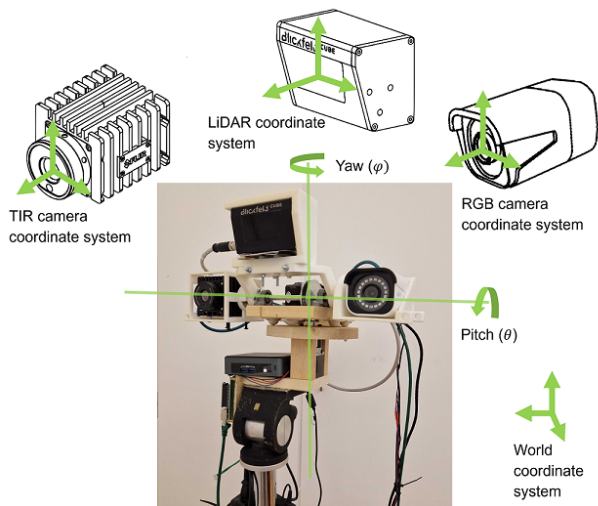


Figure 3. Overview of the hardware setup: Cameras and their coordinate systems, pitch and yaw axes.

Using a Python script, the motion of the mounting system and the data capturing of the sensors can be controlled in a coordinated manner. Taking the overlapping FoV of the sensors (cf. Sec. 3.2.3) into account, a list of scan directions can be defined (cf. Sec. 3.3.1) to achieve up to 360° of data acquisition. This work does not include accurate calibration of the two rotating axes. Nevertheless, the design dimensions of the bracket system and spacings between the axes are used to formalize the translation vectors from the LiDAR coordinate system and the axes. To register a single tri-modal scan in a world coordinate system, the translation vectors and the chosen pitch and yaw rotation matrices are applied for each axis utilizing the rotation formula indicated in Eq. 1. For each axis, the coordinate frame's origin is defined in the center of the physical axle, and its orientation is aligned with the world coordinate frame.

$$P_{rotated}^{(i)} = R_{axis}(\alpha) \cdot (P_{LiDAR}^{(i)} + t_{axis}) - t_{axis} \quad (1)$$

where  $P^{(i)}$  is point  $i$  of point cloud  $P$

$t_{axis} \in \mathbb{R}^3$  notates the translation of the LiDAR origin to the axis origin

$R_{axis}(\alpha) \in \mathbb{R}^{3 \times 3}$  is the axis' rotation matrix

$\alpha \in \mathbb{R}$  is the rotation angle in radians

### 3.2 A-priori geometric calibration

This work focuses on the fusion of data from the three sensing devices. For that purpose, a geometric calibration is necessary to find the projection between the LiDAR point cloud data and the TIR and RGB images, including the respective camera parameters. The perspective projection is formulated in Eqs. 2 and 3.

$$u_i = KM X_i \quad (2)$$

$$M = [R \mid t] \quad (3)$$

where  $i$  =  $i$ -th point  
 $X_i$  = coordinates in LiDAR coordinate system  
 $K$  = camera intrinsic parameter matrix

$R, t$  = rotation and translations between the LiDAR and the camera coordinate system  
 $u_i$  = homogenous image coordinates

The following Sections elaborate on the details of the geometric calibration in two steps: 1. Intrinsic camera calibration is used to find parameters such as focal length, principal point, and distortion coefficients. 2. Extrinsic calibration is performed to find the projection between 3D points and the image planes.

**3.2.1 Intrinsic calibration** The intrinsic camera parameters of the RGB and the TIR camera are calculated with a photogrammetric calibration. The TIR camera is thereby considered a standard camera capturing a different wavelength (Luhmann et al., 2013). For the RLC-810A calibration, an automatic feature detection and matching pipeline using COLMAP (Schonberger and Frahm, 2016) is chosen (cf. Fig 4a). Because this pipeline relies on detecting many distinct feature points, it cannot be used for the FLIR A50 calibration, as the TIR images have low resolution and significantly fewer features. Therefore, using a chessboard plate, a corner-matching algorithm is chosen for the geometric calibration of the TIR camera. An aluminum plate is suspended close to a heat source and overlaid with cardboard with a stamped chessboard pattern. Since the deployed TIR camera has a high thermal sensitivity, the temperature contrasts are high enough for corner detection and matching (cf. Fig 4b).

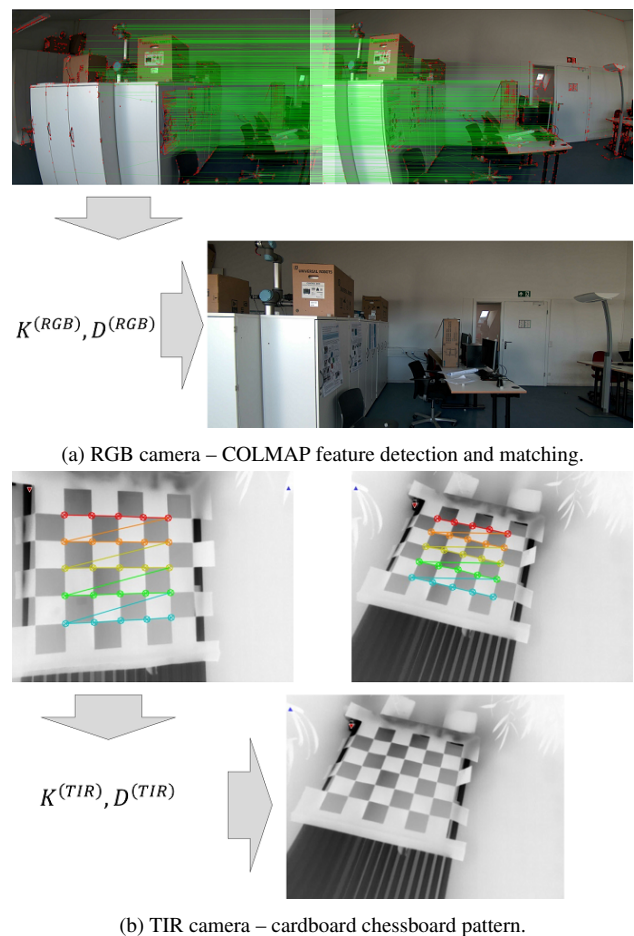


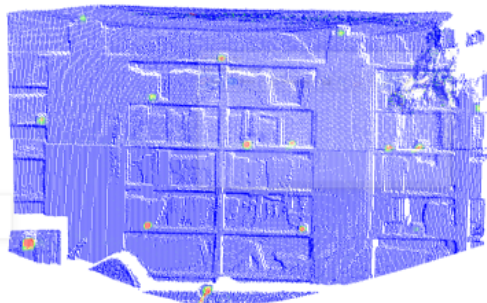
Figure 4. Intrinsic camera calibration and image rectification for two cameras of different resolution and wavelength spectrums.

In the following, the rectified images are meant when speaking of TIR and RGB images. Similarly, from here on,  $K$  is the intrinsic parameter matrix of the virtual cameras, and  $u_i$  denotes coordinates in the rectified images.

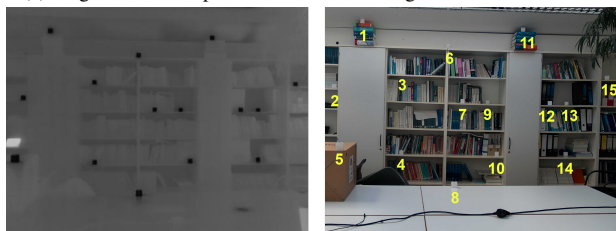
**3.2.2 Extrinsic calibration** Performing the system's extrinsic calibration allows us to find the transformation matrices  $M^{(RGB)}$  and  $M^{(TIR)}$  between the LiDAR coordinates system and the two camera coordinate systems. To easily find the corresponding points  $X_i$  and  $u_i$  in the LiDAR and TIR image, targets, as proposed by Adán et al. (2019), are assembled. Reusable ice cubes are bundled by four and coated with a reflective adhesive film. In total, 15 targets are placed at the calibration scene on a bookshelf at varying distances from the tripod. The cold temperature targets are clearly visible in the TIR image (cf. Fig. 5b). Similarly, the film reflects the laser beam emitted by the LiDAR, and the intensity in the near-infrared light range spikes at the targets (cf. Fig. 5a).

The coordinates of the 15 targets are recorded as corresponding pairs across image frames and point cloud. Given a well-rectified image and the intrinsic parameter matrix  $K$ , Eq. 2 can be solved for the rotation and translation vectors that minimize the re-projection error. The re-projection error is the distance in pixels between the 3D target points re-projected onto the image frame and their corresponding target in the image frame. Solving the pose computation problem with OpenCV's Perspective-n-Point (PnP) with Terzakis and Lourakis (2020)'s method finds  $[R | t]$ .

The resulting Mean Squared Error (MSE) of the re-projections is 1.5 pixels for the LiDAR-TIR calibration and 12.5 pixels for the LiDAR-RGB calibration.



(a) Targets visible in point cloud data at as high reflection intensities.



(b) Targets visible in TIR image (left) and RGB image (right). The RGB image was cropped and the target numbers are added for visualization purposes.

Figure 5. Visible calibration targets across sensors.

**3.2.3 Maximum common FoV** The differences in perspectives between the two lateral cameras lead to a difference in the extent they can cover. This difference occurs due to both the physical gap between the cameras and differences in their individual FoVs. Thanks to the configurable FoV and scan pattern of the LiDAR, the point cloud density in the common FoV

of the two lateral cameras can be optimized (cf. Fig. 6) via the settings in the configuration interface.

A similar effect occurs due to the vertical perspective difference between the LiDAR and the two cameras caused by the bracket design. The cameras, attached below the LiDAR, show slightly different occlusions to the LiDAR. In the camera's perspective view, some areas are occluded by object boundaries in the foreground. This effect is accounted for, by visibility analysis using the method proposed by Pan et al. (2022b).

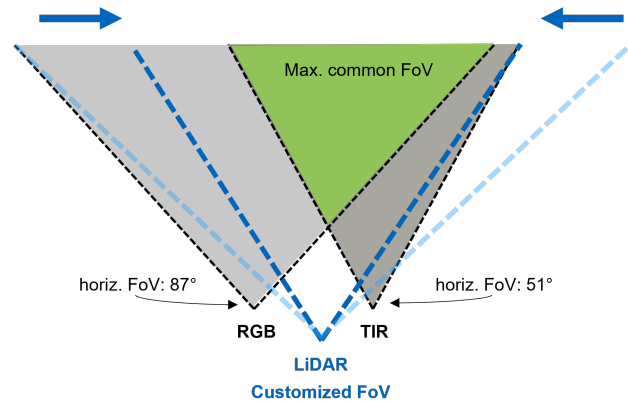


Figure 6. LiDAR customization: FoV and scan resolution to maximize point density in common FoV of the lateral cameras.

### 3.3 360° thermal point cloud

**3.3.1 Pose planning** A scan path should be planned so that the sensor system's full Range of Motion (RoM) is used to collect data. Therefore, the maximum common FoV (cf. Fig. 6) in horizontal ( $\phi_{fov,max}$ ) and vertical direction ( $\theta_{fov,max}$ ) needs to be considered. Within an exemplary RoM of  $\phi_{rom} = 360^\circ$  (yaw) and  $\theta_{rom} = 150^\circ$  (pitch). The first pose of any 360° sequence is defined as vertical up, which is the reference setup with  $\phi = \theta = 0^\circ$ . Exemplary assuming  $\phi_{fov,max} = 50^\circ$  and  $\theta_{fov,max} = 60^\circ$  and desired relative overlap of  $o_{rel} = 5\%$ , 29 poses are required, as depicted in Fig. 7a. The algorithm for this set of poses is described in Algorithm 1. Note that this approach suits 360° horizontal RoM; other approaches are required for selective data acquisition or more constrained systems.

---

#### Algorithm 1: Pose planning for full coverage within RoM

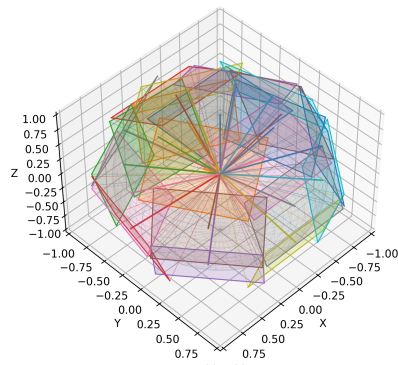
---

```

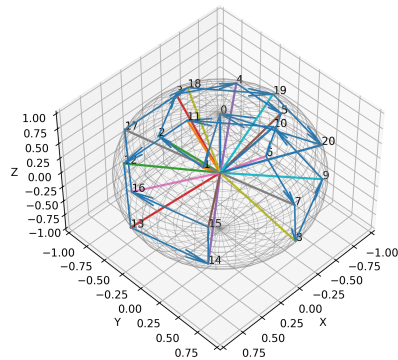
initiate list of poses  $p$  with vertical up
 $p_0 = [\phi_0, \theta_0] = [0, 0]$ 
calculate number of vertical steps and step size
 $n_\theta = \lceil \theta_{rom} / \theta_{fov,max} \cdot (1 + o_{rel}) \rceil$ ,  $\theta_\Delta = \theta_{fov,max} / n_\theta$ 
for  $n \in n_\theta$  do
    calculate length of critical circumference  $l_n$ 
    number of steps, step size
     $m_\phi = \lceil \phi_{rom} / \phi_{fov,max} \cdot (1 + o_{rel}) \rceil$ 
     $\phi_\Delta = \phi_{fov,max} / m_\phi$ 
    for  $m \in m_\phi$  do
        append poses with  $[m \cdot \phi_\Delta, n \cdot \theta_\Delta]$ 
    end
end
return poses
    
```

---

The poses should be visited in a sensible sequence for efficient execution of the data acquisition. To identify a suitable sequence, a fully connected graph  $G = (V, E)$  is created, where the vertices,  $V$ , denote the identified poses and the edges,  $E$ , describe the pairwise angular distance between all



(a) Poses for full RoM coverage, each pose is depicted with a line in 3D according to its orientation and the limits of its FoV as a frustum in the unit sphere.



(b) Pose sequence for efficient execution.

Figure 7. Pose planning for 360° acquisition.

pose pairs. Here, horizontal and vertical angular differences can be weighted with different factors to consider different translations of the drives of the physical axes (e.g., speed). Standard approximation methods can be applied to solve the traveling salesman problem for this graph to obtain an efficient sequence to visit all poses. The exemplary result of this sequence planning is depicted as a numbered path in Fig. 7b.

**3.3.2 Point cloud registration** Given the individual point clouds in the LiDAR coordinate system, a coherent 360° point cloud is assembled. The registration of the individual scans is achieved by a transformation given by system design values and the yaw and pitch angles of a given pose (cf. Section 3.1.2) Since this step has not been preceded by a calibration, some error in the registration is induced. The final 360° point cloud is evaluated with respect to its completeness and registration quality in Section 4.

#### 3.4 From raw infrared intensities to temperature

The result of the previous steps is a 360° tri-modal point cloud including color and raw infrared intensities from the acquired image following the steps in Section 3.2 and 3.3. This Section elaborates on the calculation steps and parameters needed to derive the final temperature value. To derive temperature values from raw signal data, 6 interdependent formulas are required. They involve parameters from 1. the surroundings (e.g. relative humidity, reflective apparent temperature), 2. the observed objects' properties (e.g. emissivity, distance to camera), and 3. the thermographic camera calibration, where the latter are defined in factory calibration and stored as camera metadata. The formulas used are based on IRimage (Pereyra Irujo, 2022), open-

source software for processing infrared images, and were verified against the calculations proposed by the Hardware supplier FLIR. For a detailed insight into the calculations, the reader is referred to Appx. 7.

Since the many parameters involved in the temperature calculation greatly influence the final result, the choice of values must be appropriate for the observed objects. The relevant parameters are often calibrated for a given survey or chosen with care. Here, the parameters for calculating the temperature values are defined based on suggestions from literature. For indoor environments, the ambient atmospheric temperature and reflected apparent temperature were taken as 20°C, as per the camera manufacturer guideline (Teledyne FLIR LLC, 2021) and Dall'O' et al. (2013). An emissivity value of 0.9 was chosen as construction materials have emissivity values between 0.85 and 0.95 (Dall'O' et al., 2013; Teledyne FLIR LLC, 2021). Relative humidity and object distance were taken as 60% and 3m, respectively (STMWI, 2014). The visualization of the TIR images and the thermal point cloud in the following Sections are colorizations of the calculated temperature values.

## 4. Results

The resulting 360° point cloud is visualized in Fig. 8 for a test room of 57m<sup>2</sup>. Chessboard targets are placed on the walls. The quality of the final point cloud is evaluated with respect to completeness and registration accuracy using a benchmark point cloud acquired with a Leica RTC360 terrestrial laser scanner (TLS).

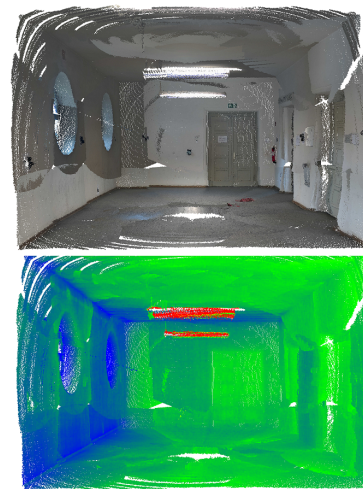


Figure 8. Full 360° trimodal acquisition

Firstly, to assess point cloud completeness, a radius search around each point in the benchmark point cloud is performed with a threshold value  $\lambda[m]$ , yielding an overlap ratio (OR) as defined in Eq. 4.

$$OR = \frac{N_{matched}(\lambda)}{N_{total}}, \quad (4)$$

where  $N_{matched}$  = number of points with match in  $\lambda$   
 $\lambda$  = radius [m]  
 $N_{total}$  = number of points in TLS point cloud

Secondly, the mean relative rotation error (mRRE), and mean relative translation error (mRTE) are computed to quantify the registration error. The true 3D transformations are found by manually aligning the individual scans with the benchmark point cloud using the chessboard targets as support. Table 1, shows the results of the quality assessment. Fig. 9 gives qualitative insights into the point cloud enrichment with RGB and TIR data.

Metric	
mRRE	2.24 [°]
mRTE	0.14 [m]
$OR_{\lambda=0.05}$	0.62
$OR_{\lambda=0.1}$	0.96
$OR_{\lambda=0.15}$	1

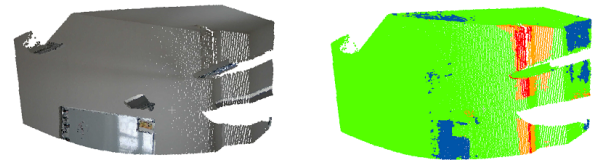
Table 1. Reported results for quality assessment.

### 5. Discussion

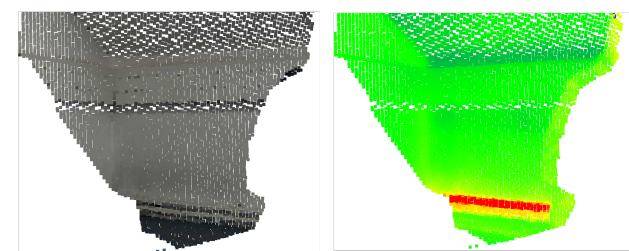
The assembled point cloud is discussed in terms of quality and suitability for contributing to the Scan-to-BIM process. In state-of-the-art Scan-to-BIM pipelines, the spatial arrangements of 3D points are leveraged to recognize shapes (e.g., planes, cylinders, etc.) or semantic clusters (e.g., walls, pipe segments). An accurately registered and complete point cloud is important for such methods. The fusion of RGB data with point cloud data further allows the transfer of findings from a 2D image to a 3D point cluster. The reported camera-LiDAR calibration accuracies, together with the qualitative insights in the previous section, promise valuable transfer of semantics from 2D data to 3D. However, it has to be noted that the point cloud presents some flaws: 1. a registration error due to the lack of calibration, 2. oscillations in the LiDAR detector signal visible in Fig. 8 as a wave pattern in the point cloud. The waves reach 0.06m magnitude, which is reflected in  $OR_{\lambda=0.05}$  and  $OR_{\lambda=0.1}$ . This is a hardware issue already addressed by the vendor in their updated system.

Methods relying solely on LiDAR and RGB data as input face inherent limitations regarding semantic diversity since many building elements are visually barely distinguishable even to the human eye. Adding TIR data as a third mode helps detail the semantics of the BIM model. In Fig. 9a, for instance, it becomes visible that the column-like element in the far back of the room has a significantly higher surface temperature. Based on this additional information, the structure can be reassigned to a semantic class of the HVAC domain. In Fig. 9b, two thin pipes manifesting different temperature values are illustrated. From this difference, a semantic subtype (e.g., supply, return) can be assigned in the Scan-to-BIM pipeline. In future research a method will be developed for automatic detailing of BIM models using tri-modal point clouds and semantic segmentation.

In this work, the temperature is calculated for each point in the tri-modal point cloud using the raw intensity values. The choice of parameters involved in the calculation of the temperature is manual and requires expert knowledge. The most relevant parameters – emissivity and reflected apparent temperature – are found using a sensitivity analysis. The lower the emissivity value, the more influence the reflected apparent temperature has on the calculated object temperature (cf. Fig. 10). It is also found that for indoor environments with typically similar temperatures, and humidity, additional efforts in obtaining accurate distance and humidity values are not required (cf. Fig. 10, right). As illustrated in Fig. 10 (right), maintaining all other



(a) The semantic class from the building element at the back can be changed from e.g., a column to a class from the ventilation/heating domain.



(b) The semantics of the thin pipes can be detailed as supply vs. return.

Figure 9. Tri-modal point clouds enable recognition of detailed semantics for Scan-to-BIM. RGB and TIR image modes (top row), enriched point cloud (bottom row).

values the same, object distance and humidity have errors of  $\pm 0.030$  for a range of 5 to 25m, and  $\pm 0.018$  for a range of 50 – 90%. State-of-the-art Scan-to-BIM methods encompass semantic clustering of 3D points (point cloud semantic segmentation), and the projection of semantic masks from 2D images. Semantics (e.g. element type "pipe") for a given point cluster can thus be assumed. In a thought experiment, a specific material type and thus, its emissivity is assumed given and can be assigned to the previously semantically segmented point clusters. The wall in Fig. 9a is taken as an example. The constant emissivity across the image and the reflection of the environment, cause the magnetic board on the wall to have higher (left side) and lower (right side) temperatures than the wall. By knowing the semantic type by semantic segmentation on RGB images, the board is assigned a lower emissivity value, and the temperature on the left half melts in with the wall in the background. A more accurate and automatic temperature calculation is the result – potentially benefiting industrial applications where precise temperature readings are of importance. Experiments are planned in the future to substantiate this idea.

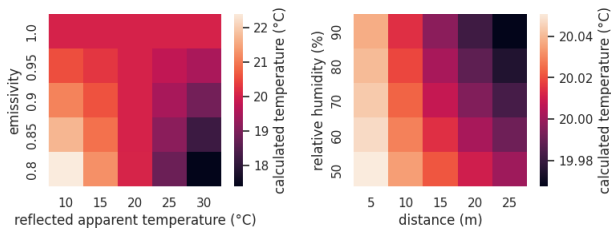


Figure 10. Effects of object emissivity,  $\epsilon$ , reflected apparent temperature,  $T_{refl}$ , relative humidity,  $RH$ , and object distance  $d$ , on calculated object temperature,  $T_{obj}$ .

## 5.1 Limitations

The most apparent error of the system assembly lies in the registration step, where design values were used without calibration. To improve this, the calibration of the movable systems is necessary. Furthermore, the RGB camera currently used has an automatic focus and color balancing, which affects the intrinsic parameters and the RGB values, ultimately resulting in errors in the projection and color differences between single scans (cf. Fig. 8). The modularity of our system allows for an easy change in hardware by reprinting single parts. An update of the LiDAR sensor to avoid the oscillations in the detector signal and a replacement to a more configurable RGB camera are planned. Finally, for the extrinsic sensor calibration part, we suggest researchers use reusable warmth pads instead of ice cubes. Due to the air humidity, the reflective film was quickly covered by condensing water, making the targets invisible in the calibration point cloud.

## 6. Conclusion

In this work, a multi-sensor platform with two rotating axes is introduced to acquire a 360° tri-modal point cloud. The sensing devices are mounted in a modular way, allowing for extensions and modifications in the future. State-of-the-art calibration steps were performed to fuse the sensor data. The step-wise tri-modal scans are registered using design values, and the temperature values are calculated using carefully chosen parameter values from literature. In a series of experiments, the quality and value of the tri-modal point cloud for enriching the Scan-to-BIM pipeline with additional and deeper semantics are discussed. Furthermore, an outlook is provided for leveraging the power of computer vision for a point-wise semantic temperature calculation. In future works, the authors plan to improve the quality of the 360° tri-modal point cloud and gain further insights by using semantic segmentation across modes.

## Acknowledgements

The research presented in this paper is supported by a Georg Nemetschek scholarship from the Technical University of Munich. Furthermore, the research presented is part of the Transregio 277 'Additive Manufacturing in Construction – The Challenge of Large Scale', funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 414265976 – TRR 277.

## References

Adán, A., Prieto, S. A., Quintana, B., Prado, T., García, J., 2019. An Autonomous Thermal Scanning System with Which

to Obtain 3D Thermal Models of Buildings. *Advances in Informatics and Computing in Civil and Construction Engineering*, Springer International Publishing, 489–496.

Biswanath, M. K., Hoegner, L., Stilla, U., 2023. Thermal Mapping from Point Clouds to 3D Building Model Facades. *Remote Sensing*, 15(19).

Borrmann, A., König, M., Koch, C., Beetz, J., 2018. Building information modeling: Why? what? how? A. Borrmann, M. König, C. Koch, J. Beetz (eds), *Building Information Modeling - Technology Foundations and Industry Practice*, 1, Springer, 1–24.

Borrmann, D., Nüchter, A., Dakulović, M., Maurović, I., Petrović, I., Osmanković, D., Velagić, J., 2014. A mobile robot based system for fully automated thermal 3D mapping. *Advanced Engineering Informatics*, 28(4), 425–440.

Dahaghin, M., Samadzadegan, F., Dadras Javan, F., 2019. 3D thermal mapping of building roofs based on fusion of thermal and visible point clouds in uav imagery. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 42number 4/W18, International Society for Photogrammetry and Remote Sensing, 271–277.

Dall'O', G., Sarto, L., Panza, A., 2013. Infrared screening of residential buildings for energy audit purposes: Results of a field test. *Energies*, 6(8), 3859–3878. www.scopus.com. Cited By :89.

De Geyter, S., Bassier, M., De Winter, H., Vergauwen, M., 2023. OBJECT DETECTION AND LOCALISATION FOR BIM ENRICHMENT. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-1/W2-2023, 155–162.

Ham, Y., Golparvar-Fard, M., 2013. An automated vision-based method for rapid 3D energy performance modeling of existing buildings using thermal and digital imagery. *Advanced Engineering Informatics*, 27(3), 395–409.

Ham, Y., Golparvar-Fard, M., 2015. Mapping actual thermal properties to building elements in gbXML-based BIM for reliable building energy performance modeling. *Automation in Construction*, 49, 214–224. <http://dx.doi.org/10.1016/j.autcon.2014.07.009>.

Hoegner, L., Stilla, U., 2015. Building facade object detection from terrestrial thermal infrared image sequences combining different views. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2(3W4), 55–62.

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. <https://doi.org/10.1080/17686733.2018.1455129>.

Luhmann, T., Piechel, J., Roelfs, T., 2013. Geometric calibration of thermographic cameras. *Remote Sensing and Digital Image Processing*, 17(1), 27–42.

Mehranfar, M., Braun, A., Borrmann, A., 2023. Automatic creation of digital building twins with rich semantics from dense RGB point clouds through semantic segmentation and model fitting. *Proc. of the 30th Int. Conference on Intelligent Computing in Engineering (EG-ICE)*.

Pan, Y., Braun, A., Borrmann, A., Brilakis, I., 2022a. 3D deep-learning-enhanced void-growing approach in creating geometric digital twins of buildings. *Proceedings of the Institution of Civil Engineers: Smart Infrastructure and Construction*, 176(1), 24–40.

Pan, Y., Braun, A., Brilakis, I., Borrmann, A., 2022b. Enriching geometric digital twins of buildings with small objects by fusing laser scanning and AI-based image recognition. *Automation in Construction*, 140(0926-5805), 104375.

Pereira, V., Santos, J., Leite, F., Escórcio, P., 2021. Using BIM to improve building energy efficiency – A scientometric and systematic review.

Pereyra Irujo, G., 2022. IRImage: open source software for processing images from infrared thermal cameras. *PeerJ Computer Science* 8:e977. <https://doi.org/10.7717/peerj-cs.977>.

Sadhukhan, D., Peri, S., Sugunaryaj, N., Biswas, A., Selvaraj, D. F., Koiner, K., Rosener, A., Dunlevy, M., Goveas, N., Flynn, D., Ranganathan, P., 2020. Estimating surface temperature from thermal imagery of buildings for accurate thermal transmittance (U-value): A machine learning perspective. *Journal of Building Engineering*, 32.

Schonberger, J. L., Frahm, J. M., 2016. Structure-from-Motion Revisited. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-Decem, 4104–4113.

Schreyer, A. C., Hoque, S., 2009. Interactive 3D Visualization of Building Envelope Systems Using Infrared Thermography and SketchUp. Technical report.

Son, H., Kim, C., Kim, C., 2015. 3D reconstruction of as-built industrial instrumentation models from laser-scan data and a 3D CAD database based on prior knowledge. *Automation in Construction*, 49, 193–200. <http://dx.doi.org/10.1016/j.autcon.2014.08.007>.

STMWI, 2014. Hinweise zum energiesparen, merkbblatt 08 - feuchte wände und schimmelbildung. Technical report, Bayerisches Staatsministerium für Wirtschaft und Medien, Energie und Technologie. Accessed on 12/01/2024.

Teledyne FLIR LLC, 2021. How does emissivity affect thermal imaging? Accessed on 16/11/2023.

Terzakakis, G., Lourakis, M., 2020. A Consistently Fast and Globally Optimal Solution to the Perspective-n-Point Problem. A. Vedaldi, H. Bischof, T. Brox, J.-M. Frahm (eds), *Computer Vision – ECCV 2020*, Springer International Publishing, 478–494.

Wang, C., Cho, Y. K., 2015. Performance Evaluation of Automatically Generated BIM from Laser Scanner Data for Sustainability Analyses. *Procedia Engineering*, 118, Elsevier Ltd, 918–925.

Zhang, C., Zou, Y., Dimyadi, J., Chang, R., 2023. Thermal-textured BIM generation for building energy audit with UAV image fusion and histogram-based enhancement. *Energy and Buildings*, 301.

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 and Technology*, 113.

## 7. Appendix

$$H = RH \cdot \left( e^{1.5587 + 0.06939T_{atm} - 0.00027816T_{atm}^2 + 0.00000068455T_{atm}^3} \right) \quad (5)$$

where  $H$  = water content  
 $RH$  = relative humidity expressed in decimal form  
 $T_{atm}$  = ambient atmospheric temperature in Kelvin

$$\tau_{atm} = X e^{-d^{0.5}(\alpha_1 + \beta_1 H^{0.5})} + (1 - X) e^{-d^{0.5}(\alpha_2 + \beta_2 H^{0.5})} \quad (6)$$

where  $\tau_{atm}$  = atmospheric transmittance  
 $X$  = spectral response parameter X  
 $d$  = object distance in meters  
 $\alpha_1$  = spectral response parameter Alpha 1  
 $\beta_1$  = spectral response parameter Beta 1  
 $\alpha_2$  = spectral response parameter Alpha 2  
 $\beta_2$  = spectral response parameter Beta 2

$$S_{atm} = \frac{R}{J_1 \left( e^{\frac{B}{T_{atm}}} - F \right)} - J_0 \quad (7)$$

where  $S_{atm}$  = the atmosphere camera signal value  
 $R$  = camera calibration parameter Sensor R  
 $B$  = camera calibration parameter Sensor B  
 $F$  = camera calibration parameter Sensor F  
 $J_1$  = camera calibration parameter Global Gain  
 $J_0$  = camera calibration parameter Global Offset

$$S_{refl} = \frac{R}{J_1 \left( e^{\frac{B}{T_{refl}}} - F \right)} - J_0 \quad (8)$$

where  $S_{refl}$  = the reflected camera signal value  
 $T_{refl}$  = reflected apparent temperature in Kelvin

$$S_{obj} = \frac{S_{raw} - S_{atm}(1 - \tau_{atm}) - S_{refl}(1 - \varepsilon)\tau_{atm}}{\varepsilon\tau_{atm}} \quad (9)$$

where  $S_{obj}$  = the object camera signal value  
 $S_{raw}$  = the raw camera signal value  
 $\varepsilon$  = object emissivity

$$T_{obj} = \frac{B}{\log \left( \frac{R}{J_1(S_{obj} + J_0)} + F \right)} \quad (10)$$

where  $T_{obj}$  = calculated object temperature in Kelvin