Automatic Generation of High-Resolution Thermal 3D Building Envelope Models exploiting UAV Imagery

Elia Ferrari, Jonas Meyer, Andreas Koch, Stephan Nebiker

Institute of Geomatics, FHNW University of Applied Sciences and Arts Northwestern Switzerland, 4132 Muttenz, Switzerland – (elia.ferrari, jonas.meyer, stephan.nebiker)@fhnw.ch , andreas.koch@students.fhnw.ch

Keywords: TIR, Thermal inspection, 3D model, 3D reconstruction, Camera alignment, UAV.

Abstract

Buildings are major contributors to global energy consumption, with the thermal performance of their envelopes playing a crucial role. Detecting thermal bridges, which compromise insulation, is essential for energy efficiency. To efficiently detect thermal bridges, thermal infrared (TIR) imagery is widely used through visual inspections, more recently by exploiting sensors mounted on unmanned aerial vehicles (UAVs). While RGB images have been extensively used in Structure-from-Motion and Multi-View Stereo processes, applying these techniques to TIR images presents challenges due to lower resolution and inconsistent colour spaces. To overcome the challenges posed by TIR imagery, approaches from different fields investigated the integration of TIR images with other data to support the alignment. Our approach improves upon these methods by using a DJI Mavic 3 Enterprise Thermal UAV to collect RGB and TIR datasets simultaneously. Our guided image alignment and camera rig estimation approach accounts for unknown camera calibration, misalignment, and lever arm parameters, ensuring robust alignment of TIR images with a total error of 5 pixels. With this approach, the geometric accuracy of the resulting point cloud reached an RMSE of 0.13 m. Finally, thermal calibration values collected on site were applied to correct the thermal images, improving temperature value accuracy for 3D model texturing with a temperature deviation of 2.8 °C. The developed method requires no prior camera calibration, TIR image pre-processing, or ground control points, permitting a complete automation of the process.

1. Introduction

1.1 Motivation

Buildings make significant contributions to global energy consumption and greenhouse gas emissions. The thermal performance of the building envelope has a significant impact on overall energy consumption. To efficiently detect thermal bridges (weak points in the thermal insulation), thermal infrared (TIR) imagery is widely used through visual inspections (Zheng et al., 2020). However, often true to scale data, such as point clouds or 3D models of building envelopes, enriched with thermal information, are needed for locating heat leaks or for planning refurbishments of buildings. Compared to RGB images, TIR images show lower geometric and radiometric resolution, lower dynamic range (Luhmann et al., 2013), are affected by unsharp definitions of discontinuities and small details (e.g. blurred edges) (Conte et al., 2018). Hence, structure from motion (SfM) and multi view stereo (MVS) based processing alone is often insufficient in terms of spatial resolution and accuracy (Dlesk and Vach, 2019). Various strategies focusing on the fusion of TIR images with RGB images (Dlesk et al., 2022), point clouds (Elias et al., 2023; Hoegner and Stilla, 2018; Javadnejad et al., 2020; Lin et al., 2019) or 3D models (Alba et al., 2011; Iwaszczuk and Stilla, 2017) from other sources address these geometrical issues. However, such approaches impair their applicability due the complexity of tasks such as TIR camera calibration, operation of high-end data acquisition systems, TIR image pre-processing, and image alignment, as well as the dependence on available and current data when relying on external sources, like 3D city models.

Early research with experimental multi-head sensor systems for Unmanned Aerial Vehicle (UAV) systems date back more than 15 years (Nebiker et al., 2008). With the advent of commercial off-the-shelf UAV systems with multi-sensor heads, RGB and TIR images can be easily acquired simultaneously (Yue, 2024). If the UAV additionally includes an RTK-GNSS module, accurate pose priors can be determined. Such UAV systems allow for the efficient capturing of buildings in just a few minutes. SfM based 3D model generation, especially alignment of TIR images, however, still poses major challenges and requires elaborate workflows (Patrucco et al., 2022).

In this paper we propose a fully automated process, based on the SfM software Agisoft Metashape (Agisoft LLC, 2023) to create thermal 3D building envelope models. We only use simultaneously captured RGB and TIR images from a UAV and accurate pose priors, without the need for GCPs, TIR image pre-processing or known camera calibration. Our main contributions are:

- A guided image alignment and rig estimation process
- A fully automated process from raw RGB and TIR to a 3D thermal building envelope model
- A qualitative and quantitative analysis of first results

1.2 Related Work

The number of applications and studies using UAVs has rapidly increased in recent years due to several reasons. UAVs have become more affordable and dependable, and they can be equipped with different sensors and used for multiple applications. In particular, the use of TIR sensors has been researched in a number of works, ranging from agriculture and forestry (Maes et al., 2017), heritage asset documentation (Patrucco et al., 2022) to buildings and infrastructures thermal analysis (Javadnejad et al., 2020; Ramón et al., 2022).

Standard procedures for RGB imagery processing in mapping and 3D modelling now employ modern SfM-MVS workflows (Deliry and Avdan, 2021). In recent years, these approaches have been applied to thermal infrared images to obtain products such as point clouds or 3D models of buildings with thermal information (Ramón et al., 2022).

In the approach of Zheng et al. (2020), they collected imagery with a UAV equipped with a professional TIR camera, focusing on the 3D reconstruction directly from TIR images. They subsequently applied a thermal correction to the imagery using the temperature deviation of aluminium foil employed for the targets as reference. The analysis with the collected reference measurements showed an absolute temperature accuracy of 5°C. Large scale approaches cannot rely on single targets of aluminium foil. Therefore, Striewski et al. (2024) applied the Thermal Urban Road Normalization algorithm developed by Rahman et al. (2014) to interpolate temperature deviation within a scene and normalize TIR imagery, based on the assumption of roads as pseudo invariant objects. Conversely, Conte et al. (2018) exploited TIR imagery collected with a plane to generate a large scale thermal orthomosaic of a city. In their approach they added a pre-processing step, evaluating different radiometric enhancement methods. This improved the effectiveness of the process with SfM algorithms, generating more tie points in the sparse cloud and a slightly higher density of the dense cloud.

Ramón et al. (2022) highlighted the challenges of 3D reconstructions directly from TIR imagery, which have a limited field of view (FoV) as well as low spatial resolution and therefore generate incomplete point clouds. In addition, the image alignment is made more arduous by the lack of distinctive features.

To face the previously mentioned challenges associated with TIR images, different approaches have been developed. Iwaszczuk and Stilla (2017) exploited an existing 3D building model and its uncertainties for a co-registration of TIR imagery, refining the exterior orientation parameters of the camera. Other studies instead exploited data collected by newly developed low-cost multi-sensor UAVs, capable of simultaneously capturing visible and thermal infrared images. This offers significant advantages for a combined RGB and TIR image alignment with an improved geometric quality of the 3D reconstruction (Patrucco et al., 2022; Zhang et al., 2023). In these studies, dual-sensor datasets were used to investigate the advantages of an integrated reconstruction approach, evaluating them on single buildings or façades. Zhang et al. (2023) implemented a rectification method for TIR images to allow the fusion with RGB images and validated it on a building façade with a fusion error of 5.7 pixels between RGB and TIR images. In contrast, the approach of Patrucco et al. (2022) implemented a three-step workflow to align the images, using the estimated poses of RGB images to improve the TIR image alignment. However, this approach assumes that no misalignment or lever arm between the two cameras persists. The resulting average checkpoint RMSE showed values about 40% higher than the Ground Sampling Distance (GSD) and in one case inferior values than the 3D reconstruction with pure TIR images. Similarly, Maes et al. (2017) adopted the estimated poses of RGB images to enhance the TIR image alignment. Additionally, they also adopted the previously estimated camera parameters from a camera calibration. In contrast, Javadnejad et al. (2020) showed that thermal 3D models benefit from the combination of RGB and TIR images. They performed image alignment (RGB only) and projected thermal information to 3D points by known pre-calibrated lever arm and misalignment values. Despite the added value of a pre-calibration of the thermal camera, low-cost dual-sensors systems are usually unstable, and the internal camera parameters can therefore vary from flight to flight. Thus, for UAV-based data collection a self-calibration is preferable (Elias et al., 2023).

López Ruiz et al. (2021) presented an approach to determine absolute temperature values from a point cloud generated using RGB images. In their investigations, the data collected with a dual-sensor system have been combined, removing the image distortions and projecting the TIR images on the RGB images via transformation matrices. In this way the temperature values can be assigned to the corresponding points of the point cloud. Accordingly, Dlesk et al. (2022) exploit dual-sensor data to create a point cloud augmented with thermal information. They adopted a fixed camera system to remap TIR images onto RGB images exploiting the known relative rotation and translation between the two cameras. The method is computationally expensive and requires an additional visibility test for a correct interpretation. As an alternative, they generated four-channel images adding the TIR information to the RGB images and recomputing the image alignment with the new images. This resulted in faster image processing. However, it resulted in lower accuracies of the checkpoints in comparison to the first method.

A combined evaluation of RGB and TIR images should enable a more robust and accurate generation of thermal 3D building models. Such geometrically precise 3D models would allow a much better assessment of the entire building envelope and planning of renovation measures.

2. Materials and Methods

2.1 Instruments

In this paper, a DJI Mavic 3 Enterprise Thermal (M3T) is used for data acquisition. The UAV is equipped with a multi-sensor camera head, which can capture RGB and TIR images simultaneously. The thermal data is recorded with an accuracy of 2 °C and an image resolution of 0.3 MP, while the wide-angle camera could capture 48 MP RGB images. However, the UAV system limits the wide-angle camera resolution to 12 MP if RGB and TIR data are collected simultaneously. As Bruno and Forlani (2023) showed, another limitation is caused by the electronic shutter of the UAV, which directly influences the flight plan and speed due to the rolling shutter effect. However, in comparison to the UAV used by Bruno and Forlani (2023), a DJI Phantom 4 Pro, the electronic shutter of the M3T performs a full sensor readout two and half times faster and does not need post-processing compensation in low-speed flight mode, as demonstrated by Zmejevskis (2024). In addition, the UAV was equipped with a real-time kinematic (RTK) module, which determines image positions with an accuracy of up to 1 cm + 1 ppm horizontally and 1.5 cm + 1 ppm vertically (DJI, 2023).

To validate absolute temperature values, reference measurements were carried out with a FLIR E40 thermal camera, which enabled independent thermal measurements with an accuracy of 2 °C (FLIR, 2011).

Along with reference temperature measurements, checkpoints have been distributed to enable an accuracy analysis of the results. The targets were measured using a combination of RTK GNSS Rover and total station, allowing the data to be collected within a standard deviation of ±5 cm.

2.2 Study Area

The study area is a detached building in a rural environment. The building is inhabited and heated. It is equipped with a photovoltaic installation on the roof and, according to construction documentation, is poorly insulated. The surrounding terrain has a gentle uniform slope in one direction and just a few elements that obstruct the view, such as trees. As illustrated in Figure 1, seven checkpoints, five on the ground and two on the façade, have been materialised, to provide information about the accuracy of the generated products during analysis.

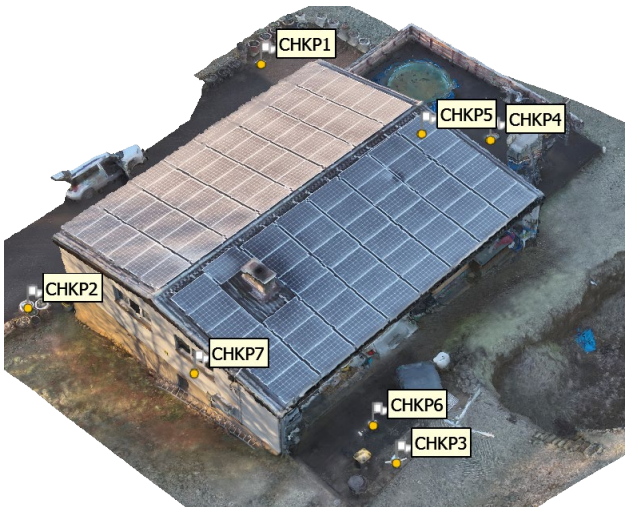

Figure 1. Study object and checkpoint distribution.

No GCPs were placed to maintain fully automated data processing without operator intervention. All checkpoints have been realised using special targets made of aluminium foil, which has a low thermal emissivity. As shown in Figure 2, the targets are clearly visible in the RGB and TIR images. The coordinates of all targets on the ground have been determined by multiple RTK-GNSS measurements, while additional checkpoints on the façades have been measured with a total station. All checkpoints are expected to have a standard deviation of less than 5 cm.

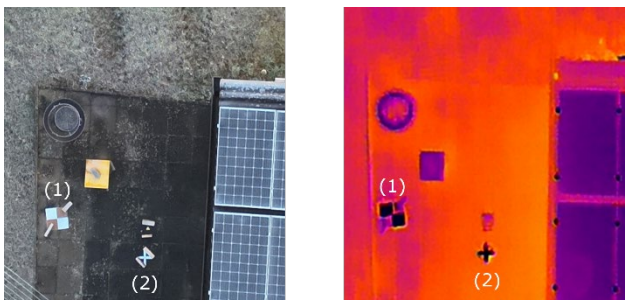

Figure 2. Example of the point materialization. Classical target enhanced with aluminium foil (1) and a cross made of aluminium foil (2) in an RGB (left) and TIR image (right).

2.3 Data Acquisition

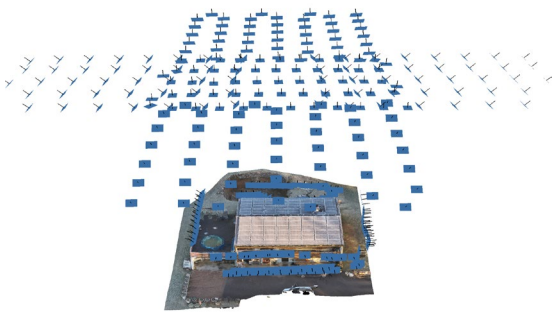
Figure 3. Configuration of different flight missions (nadir, oblique and close range) to completely capture the building of interest.

The study area was captured with a UAV DJI M3T, described in section 2.1. To completely capture the building and its façades, which are partly obscured by eaves and balconies, 362 RGB and TIR images were captured in three configurations: nadir, oblique and close-range (Figure 3). Nadir and oblique images were captured from 30 metres above ground and an angle of 45° was chosen for the oblique configuration. The close-range images were recorded with an average object distance of 5 metres. Table 1 shows the number of images per configuration and the average GSD for RGB and TIR in each configuration.

configuration	GSD [mm]		Total images RGB / TIR	forward / side overlap
	RGB	TIR		
nadir	10.8	39.6	60	80% / 75%
oblique (centre pixel)	15.3	56.0	172	70% / 75%
close range	1.8	6.6	130	-- / 75%

Table 1. Ground sampling distances and number of images per image type and configuration.

The flight mission was conducted in mid-December 2023 in the early morning to ensure minimum solar irradiation on the one hand, and good quality RGB images on the other. At the start of the mission, a thermal sensor calibration was performed to establish the optimal temperature range and emissivity of the measured object. In addition, environmental values, such as distance from the object, humidity, emissivity and reflected temperature, have been registered to automatically adjust the thermal measurements in post processing. The maximum speed of the UAV was set to 3 m/s resulting in a data acquisition time of around 16 minutes. To enable georeferencing, the RTK module of the UAV was used to record precise image poses and to facilitate the image alignment process. At the same time, a FLIR E40 thermal imaging camera was used to collect reference data before and after each flight to analyse the absolute temperature values in order to ensure an average value valid for the entire flight. The measurements were taken at a distance of around 2 meters from the building façade, on window frames or thermal bridges of the construction, which were also visible on UAV images. To this end, three points of interest on the north, west and east façade have been identified and measured.

2.4 Guided Image Alignment and Camera Rig Estimation

To create a combined 3D model with thermal information, both RGB and TIR images need to be aligned precisely. Standard image alignment procedures usually consider each image separately. While the alignment of RGB imagery with sufficient overlap has become a standard procedure, several works show that the alignment of TIR images still poses challenges caused by low geometric and radiometric resolution as well as small fields of views (Akçay, 2021; Dlesk et al., 2022; Javadnejad et al., 2020). Furthermore, the establishment of matches between RGB and TIR images is likely to fail due to radiometric differences and lack of distinctive features on the TIR images (Dlesk et al., 2022). To increase the stability of the image alignment process we define a multi-sensor rig where the RGB camera is the primary sensor and the TIR camera the secondary sensor defined by its relative orientation (lever-arm T_{TIR}^{RGB} and misalignment R_{TIR}^{RGB}) with respect to the primary sensor. Since the calibration parameters of both, the RGB and TIR camera are unknown, a self-calibration of both cameras is performed. However, the translational aspect of the relative orientation correlates strongly with the TIR camera's focal length and principal point, meaning that these camera parameters can also be described by shifting the relative orientation accordingly (Luhmann, 2010).

Initial tests showed that the alignment of the RGB images was successful, but the alignment of the TIR images failed when estimating all unknown parameters during bundle adjustment. According to Remondino and Fraser (2006) we assume that the

bundle adjustment process fails because of the strong correlation of the lever-arm and the intrinsic camera parameters in combination with a weak network geometry due to few key point correspondences in the TIR images.

Under the assumption of accurate image pose priors (0.05 m and 10° standard deviation for position and rotation components respectively) we developed a guided image alignment and camera rig estimation process consisting of three main steps:

a) **Initial image alignment**
Projection centres of RGB and TIR sensors are identical (lever-arm $T_{TIR}^{RGB} = (0,0,0)^T$). Estimation of misalignment and all intrinsic camera parameters of the RGB and TIR sensor except for the focal length of the TIR sensor, which is fixed to the initial value obtained from metadata, during the bundle adjustment.

b) **Camera optimization and lever-arm estimation**
Introduction of the lever-arm $T_{TIR}^{RGB} = (-0.02,0,0)^T$ (manually measured values in metres) with a standard deviation of 0.001 m. Misalignment and camera intrinsics parameters are treated as in the previous step.

c) **Camera optimization and TIR camera calibration**
Estimation of all values of the TIR camera.

For step b) and c) the estimated values of the previous steps are used as approximate values for the current optimization step. The introduced standard deviation for the camera poses and the lever-arm are left unchanged. Additionally, rolling shutter compensation is disabled due to the low-speed flight mode used and the previously described shutter speed of the DJI M3T (section 2.1).

2.5 Temperature Values Correction and Conversion

The TIR images collected with the M3T are saved as radiometric JPEGs, which is a binary format just used to display the coloured TIR images. With the purpose of enabling further processing, the temperature information encoded in the radiometric JPEG files needs to be converted as it cannot be read directly. With the help of the DJI Thermal SDK, the data have been corrected using the collected object distance, humidity, emissivity and reflected temperature and then saved as standard raw files with just one channel containing the corrected absolute temperature values for each pixel (DJI, 2022).

2.6 Process Automation

In the last part of our approach, we focused on the automation of the whole workflow depicted in Figure 4, integrating all the processing steps within a script implemented in Python. To this end, we exploited the Application Programming Interface (API) of the Agisoft Metashape (Agisoft LLC, 2023) software to automate the image alignment and camera rig estimation, as well as for dense point cloud and 3D model generation. The integration of the DJI Thermal SDK in the script allowed the automation of thermal image conversion and correction. Finally, the converted images were replaced and automatically processed for the texture generation, again with the Agisoft Metashape API. This enabled the fully automated generation of a realistic and high-resolution thermal 3D building envelope model with absolute temperature values.

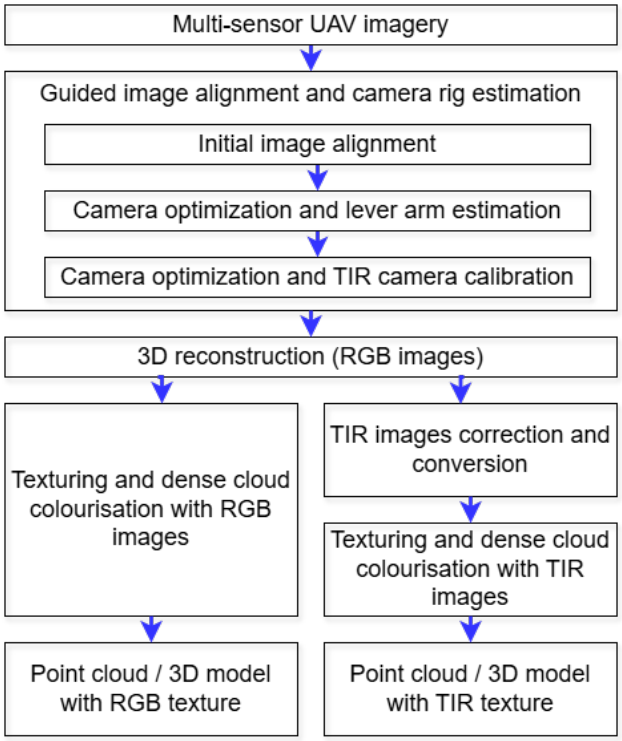

Figure 4. Workflow of our automated photogrammetric process.

2.7 Point Cloud and Texture Evaluation

After defining the camera rig and image alignment strategy, both geometrical accuracy and absolute temperature accuracy of the results were evaluated. For the geometrical analysis, the dense point cloud resulting from the aligned RGB images has been used as reference for a comparison with both, the point cloud from the aligned TIR images and the point cloud resulting from the combined alignment of TIR and RGB images. To this end, the same point cloud section has been extracted from all three point clouds and the deviation from the reference has been computed.

Due to the lower geometric resolution, the image alignment of TIR images is significantly less accurate than the alignment of RGB images (Conte et al., 2018). In order to investigate the impact on the final product, a 3D mesh was generated using in turn the combined point cloud, RGB and TIR images, and the point cloud resulting from RGB images only. The resulting meshes were textured with TIR images aligned with the combined approach and with RGB images from standard alignment, respectively. The measured checkpoints attached to the façades were used to calculate the deviation of the TIR texture from the RGB texture (reference).

Finally, the evaluation of the absolute temperature values aims to compare the generated thermal texture generated from the corrected TIR images with the reference values of the FLIR 40 thermal camera. For this analysis, the measurements of the three reference points on the building's façades were compared with the measurements on the UAV's TIR imagery. In this context, the thermal values have been calculated from the average of six UAV's images, whereby the average temperature of all values within a radius of six pixels was used for each image.

3. Experiments and Results

3.1 Guided Image Alignment and Rig Estimation

Our proposed image alignment and camera rig estimation process is evaluated by measuring all seven checkpoints (Figure 1) within the aligned images. This is performed for both TIR and RGB images independently. The residuals in checkpoint coordinates are calculated after each of the three steps of the proposed approach: a) initial camera alignment (with identical projection centres for TIR an RGB images, and fixed focal length of TIR camera); b) camera optimization and lever-arm estimation; c) camera optimization and TIR camera calibration. Table 2 shows the average residuals for all seven checkpoints for TIR and RGB images respectively. The residuals are provided as 2D and 3D coordinate differences in object space and as pixel differences in image space.

Imagery	TIR		RGB	
Residuals	object space 2D / 3D [m]	image space [pix]	object space 2D / 3D [m]	image space [pix]
(a)	0.101 / 0.245	5.29	0.030 / 0.055	0.95
(b)	0.101 / 0.251	5.34	0.030 / 0.055	0.95
(c)	0.095 / 0.232	4.92	0.030 / 0.055	0.95

Table 2. Residuals of checkpoint observations in TIR and RGB images after each step of the proposed image alignment process.

Table 2 shows that the residuals of the checkpoints measured in TIR images are lower after our proposed guided image alignment and camera-rig estimation process. However, the estimation of the lever-arm without optimizing the focal length of the TIR camera results in slightly higher residuals than assuming that both projection centres are identical. In contrast, the residuals of checkpoints measured in RGB images do not change after the initial alignment process, meaning that the estimation of the lever-arm, misalignment and intrinsic camera parameters have no influence on the RGB image poses.

3.2 Point Cloud and Texture Evaluation

3.2.1 Point Cloud Evaluation: As introduced in section 2.7, the three resulting point clouds, from RGB images only, from TIR images only and from the combination of RGB and TIR images, were compared. In this comparison the first one served as reference to examine the deviations of the latter two.

The geometric accuracy analysis of the combined point cloud showed that around 40% of the points show less than 2 cm deviation from the reference. 90% of the points show differences of up to 5 cm from the reference, resulting in a total RMSE of 0.13 m. Similarly, the comparison of the point cloud generated with aligned TIR images only with the reference point cloud resulted in a deviation of 2 cm for 45% of the points. However, only 77% of the points showed a deviation from the reference point cloud of less than 5 cm, resulting in a higher total RMSE of 0.19 m.

The point cloud resulting from the combined method showed a remarkably high point density of 2102 pts/m^2, comparable to the point cloud resulting from pure RGB image alignment with 2154 pts/m^2. In contrast, the point cloud resulting from TIR images only has significantly higher noise and about one third of the density of the reference point cloud (642 pts/m^2).

3.2.2 Texture Evaluation: The investigation showed deviations of the TIR texture from the RGB texture averaging 5.5 cm in position (2D) and averaging 3 cm in height. The worst value has been encountered at the check point CHKP5, placed on the west façade, with a difference from the reference (RGB texture) of 10 cm in position. As shown in section 3.1 the accuracy of the TIR image alignment with residuals of 4.9 pixels is significantly lower than that of the RGB images with 0.95 pixels (Table 2). Considering the GSD of the close-range images with 6.6 mm for TIR images and 1.8mm for RGB images (Table 1), the obtained differences can be explained by the uncertainty of the image alignment and the resulting deviations in the object space.

In addition, as shown in Figure 5, a visual comparison of the two textures was carried out. It showed that the structure of the photovoltaic system and the different insulation layers between the ground floor and the first floor are easily recognisable and can be correctly located in the 3D model. Despite a maximum deviation of 10 cm, the overall models showed satisfactory results for the application of a thermal 3D model.

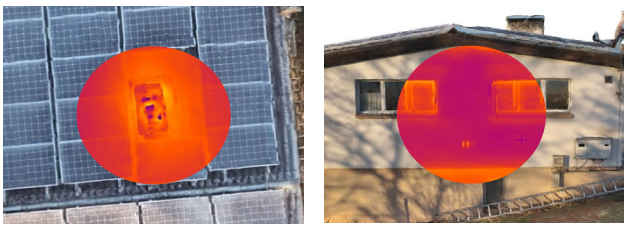

Figure 5. Sections of RGB texture (in the background) and thermal texture (circular, in the centre): roof view (left), east façade (right).

3.3 Absolute Temperature Values

In the analysis of the accuracy of absolute temperature values, the reference temperature values of the FLIR E40 have been compared with those calculated from the M3T imagery, as described in section 2.5. The differences summarized in Table 3, show no systematic deviation and lie within the simple standard deviation of the instrument accuracy of 2.8 °C.

Reference point	FLIR E40 (reference)	UAV M3T (actual)	ΔT (reference - actual)
	Average / Std. Dev.	Average / Std. Dev.	
East façade	2.1 °C ± 0.8	0.8 °C ± 0.1	1.3 °C
West façade	-3.4 °C ± 0.7	-0.6 °C ± 0.3	-2.8 °C
North façade	-0.5 °C ± 0.3	-3.2 °C ± 0.2	2.7 °C

Table 3. Absolute temperature differences between average reference value (FLIR E40 thermal camera) and average value of UAV (M3T) TIR imagery.

4. Discussion

The combined processing of RGB and TIR images is beneficial for image alignment as it can address the challenges posed by TIR imagery. This can avoid using radiometric enhancement methods to improve TIR image alignment and prevent applying two transformations to obtain the original temperature values (Conte et al., 2018). An additional challenge arises from the differing ideal conditions required for capturing RGB and TIR data simultaneously. RGB image quality depends on adequate natural light, requiring daylight conditions to capture high-

contrast, sharp visuals. In contrast, TIR imaging benefits from minimal solar radiation, as thermal readings become more reliable and less influenced by external heat sources, such as direct sunlight. This discrepancy between real and ideal conditions presents a trade-off when collecting integrated RGB-TIR data, as optimisation for one sensor may compromise the other. This reduces the possibilities for data collection around twilight periods when ambient light is low enough to avoid substantial heat from solar radiation, yet sufficient for capturing usable RGB images.

A further consideration is the choice of sensor used for flight planning, which directly affects either the TIR or RGB imagery. In this study, the RGB sensor was selected for mission planning. While this ensured adequate overlap and coverage in RGB images, it led to lower overlap in TIR images due to the narrower FoV and consequently impacting image alignment accuracy and point cloud density of TIR images. On the other hand, using the TIR sensor for mission planning results in capturing more RGB images, leading to increased data volume and processing demands.

Integrating all processing steps within a single script automates the entire workflow, including guided image alignment and camera rig estimation, absolute temperature value correction and conversion, 3D modelling and texturing, thus eliminating time-consuming manual interactions. The automated pipeline leverages RTK-GNSS positioning instead of GCPs, simplifying and speeding up data acquisition and processing. However, the reliance on RTK-GNSS technology brings with it a dependency for the automation process, which could limit the application to data coming from UAVs equipped with RTK-GNSS only.

The proposed guided image alignment and camera rig estimation steps provided an alignment accuracy of approximately 5 pixels, slightly higher than the 3-4 pixels accuracy obtained in GCP-supported workflows (Conte et al., 2018). While adequate for generating visually accurate thermal 3D models, this variance suggests that the use of GCPs could further benefit the alignment accuracy. However, special coded targets should be employed, with the intention of guaranteeing the complete automation of the process.

The data fusion allows for generating a thermal point cloud with the higher accuracy and density of a RGB point cloud. The inclusion of both RGB and TIR imagery enhances the quality of the 3D point cloud, not just generating a point cloud density like the standard RGB point cloud, but also reaching a maximal deviation from the reference point cloud of 5 cm for 90% of the points. The total RMSE of 0.13 m can be considered as good in comparison with results obtained with other methods, which yielded a RMSE between 0.2 and 0.22 m (Javadnejad et al., 2020). However, using the RGB point cloud as reference a dependency between TIR and RGB data is introduced, since both data have been collected simultaneously with the same platform. A feasible alternative for the geometric analysis consists in an independent measurement system, such as a terrestrial laser scanner, which can provide additional and unrelated information about the geometry accuracy of the approach.

The deviation between RGB and TIR textures presented in section 3.3 most likely result from a combination of inaccurate parameters of the external orientation of the TIR sensor and stitching errors due to the low geometric resolution and dynamic range of TIR imagery. Possibilities to further increase the accuracy of the thermal textures could lie in a more precise approach for estimating the relative orientation (misalignment and lever arm) and a compromise in flight planning to ensure greater overlap of the TIR images.

The evaluation of the absolute temperature values in section 3.4 supports the method's applicability for thermal assessment, as it allows for effective representation of temperature variation across building surfaces with an accuracy higher than other approaches (Dlesk et al., 2022; Zheng et al., 2020). It also demonstrates that, the temperature correction with calibration values is a crucial component for obtaining accurate thermal values.

5. Conclusions and Outlook

In this study, we presented an automated workflow for creating 3D thermal models using data from a dual-camera UAV system equipped with an RTK-GNSS module. By leveraging simultaneously captured RGB and TIR images with multi-camera head and eliminating the need for GCPs, we achieved a fully automated process. The implemented three-step method for a guided image alignment and camera rig estimation enabled a combined processing of TIR and RGB images, resulting in alignment residuals of approximately 5 pixels at measured checkpoints. Furthermore, from these aligned images, a TIR point cloud with an enhanced density was generated, with a deviation from the reference under 5 cm for 90% of the points and a total RMSE of 0.13 m. Finally, the temperature corrections applied to TIR images produced thermal textures with a standard deviation of 2.8 °C. While differences in optimal conditions for capturing RGB and TIR imagery pose limitations for simultaneous data collection, the integration of TIR and RGB datasets enhances the visualization and analysis of building thermal performance.

Future research could focus on refining alignment accuracy by developing automated GCPs measurement methods with coded targets and improving the estimation of camera calibration, lever-arm and misalignment of multi-sensor heads. Additionally, validating this approach with an independent system, such as terrestrial laser scanning, could provide further insights into its geometric accuracy. Incorporating façade-mounted reference sensors during data acquisition could also enhance the reliability of absolute temperature values. These advancements would further improve the method's accuracy.

References

Agisoft LLC, 2023. Agisoft Metashape Professional, Version 2.1.2.

Akçay, Ö., 2021. Photogrammetric analysis of multispectral and thermal close-range images. Mersin Photogrammetry Journal, 3, 29–36. doi.org/10.53093/mephoj.919916.

Alba, M.I., Barazzetti, L., Scaioni, M., Rosina, E., Previtali, M., 2011. Mapping Infrared Data on Terrestrial Laser Scanning 3D Models of Buildings. Remote Sensing, 3(9), 1847–1870. doi.org/10.3390/rs3091847.

Bruno, N., Forlani, G., 2023. Experimental Tests and Simulations on Correction Models for the Rolling Shutter Effect in UAV Photogrammetry. Remote Sensing, 15(9), 2391. doi.org/10.3390/rs15092391.

Conte, P., Girelli, V., Mandanici, E., 2018. Structure from Motion for aerial thermal imagery at city scale: Pre-processing, camera calibration, accuracy assessment. ISPRS Journal of Photogrammetry and Remote Sensing, 146, 320–333. doi.org/10.1016/j.isprsjprs.2018.10.002.

Deliry, S.I., Avdan, U., 2021. Accuracy of Unmanned Aerial Systems Photogrammetry and Structure from Motion in Surveying and Mapping: A Review. J Indian Soc Remote Sens, 49(8), 1997–2017. doi.org/10.1007/s12524-021-01366-x.

DJI, 2023. Specs - DJI Mavic 3 Enterprise - DJI Enterprise. https://enterprise.dji.com/mavic-3-enterprise/photo (10 May 2024).

DJI, 2022. DJI Thermal SDK, Version 1.4.

Dlesk, A., Vach, K., 2019. POINT CLOUD GENERATION OF A BUILDING FROM CLOSE RANGE THERMAL IMAGES. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-5/W2, 29–33. doi.org/10.5194/isprs-archives-XLII-5-W2-29-2019.

Dlesk, A., Vach, K., Pavelka, K., 2022. Photogrammetric Co-Processing of Thermal Infrared Images and RGB Images. Sensors, 22(4), 1655. doi.org/10.3390/s22041655.

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. doi.org/10.1016/j.ophoto.2023.100041.

FLIR, 2011. FLIR E-Serie. https://www.flir-infrarotkameras.de/WebRoot/Store12/Shops/61587589/4DFB/9E3E/2D90/6279/C31D/C0A8/29BA/0647/FLIR_E_series.pdf (16 October 2024).

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, 1–19. doi.org/10.1080/17686733.2018.1455129.

Iwaszczuk, D., Stilla, U., 2017. Camera pose refinement by matching uncertain 3D building models with thermal infrared image sequences for high quality texture extraction. ISPRS Journal of Photogrammetry and Remote Sensing, 132, 33–47. doi.org/10.1016/j.isprsjprs.2017.08.006.

Javadnejad, F., Gillins, D.T., Parrish, C.E., Slocum, R.K., 2020. A photogrammetric approach to fusing natural colour and thermal infrared UAS imagery in 3D point cloud generation. International Journal of Remote Sensing, 41(1), 211–237. doi.org/10.1080/01431161.2019.1641241.

Lin, D., Jarzabek-Rychard, M., Tong, X., Maas, H.-G., 2019. Fusion of thermal imagery with point clouds for building façade thermal attribute mapping. ISPRS Journal of Photogrammetry and Remote Sensing, 151, 162–175. doi.org/10.1016/j.isprsjprs.2019.03.010.

López Ruiz, A., Jurado, J.M., Ogayar, C., Feito, F., 2021. An optimized approach for generating dense thermal point clouds from UAV-imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 182, 78–95. doi.org/10.1016/j.isprsjprs.2021.09.022.

Luhmann, T., 2010. Erweiterte Verfahren zur geometrischen Kamerakalibrierung in der Nahbereichsphotogrammetrie (Habilitation Thesis). Deutsche Geodätische Kommission, Reihe C, Nr. 645.

Luhmann, T., Piechel, J., Roelfs, T., 2013. Geometric Calibration of Thermographic Cameras, in: Kuenzer, C., Dech, S. (Eds.), Thermal Infrared Remote Sensing, Remote Sensing and Digital Image Processing. Springer Netherlands, Dordrecht, pp. 27–42. doi.org/10.1007/978-94-007-6639-6_2.

Maes, W., Huete, A., Steppe, K., 2017. Optimizing the Processing of UAV-Based Thermal Imagery. Remote Sensing, 9(5), 476. doi.org/10.3390/rs9050476.

Nebiker, S., Annen, A., Scherrer, M., Oesch, D., 2008. Light-Weight Multispectral Sensor for Micro UAV - Opportunities for Very High Resolution Airborne Remote Sensing. Int Arch Photogram Rem Sens Spatial Inform Sci, 37.

Patrucco, G., Giulio Tonolo, F., Sammartano, G., Spanò, A., 2022. SFM-BASED 3D RECONSTRUCTION OF HERITAGE ASSETS USING UAV THERMAL IMAGES. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B1-2022, 399–406. doi.org/10.5194/isprs-archives-XLIII-B1-2022-399-2022.

Rahman, M., Hay, G., Couloigner, I., Hemachandran, B., 2014. Transforming Image-Objects into Multiscale Fields: A GEOBIA Approach to Mitigate Urban Microclimatic Variability within H-Res Thermal Infrared Airborne Flight-Lines. Remote Sensing, 6(10), 9435–9457. doi.org/10.3390/rs6109435.

Ramón, A., Adán, A., Javier Castilla, F., 2022. Thermal point clouds of buildings: A review. Energy and Buildings, 274, 112425. doi.org/10.1016/j.enbuild.2022.112425.

Remondino, F., Fraser, C., 2006. Digital camera calibration methods: Considerations and comparisons. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 36(5), 266–272. doi.org/10.3929/ETHZ-B-000158067.

Striewski, F., Comi, E.L., Tiefenbacher, F., Lack, N., Battaglia, M., Bleisch, S., 2024. Application, Adaption and Validation of the Thermal Urban Road Normalization Algorithm in a European City. Workshop on Visualisation in Environmental Sciences (EnvirVis). doi.org/10.2312/ENVIRVIS.20241135.

Yue, K., 2024. Multi-sensor data fusion for autonomous flight of unmanned aerial vehicles in complex flight environments. Drone Syst. Appl., 12, 1–12. doi.org/10.1139/dsa-2024-0005.

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, 113710. doi.org/10.1016/j.enbuild.2023.113710.

Zheng, H., Zhong, X., Yan, J., Zhao, L., Wang, X., 2020. A Thermal Performance Detection Method for Building Envelope Based on 3D Model Generated by UAV Thermal Imagery. Energies 13, 6677. doi.org/10.3390/en13246677.

Zmejevskis, L., 2024. DJI Mavic 3 has no mechanical shutter. Sensor readout speed explained. https://www.pix-pro.com/blog/dji-mavic-3-rolling-shutter (11 May 2024).