The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-1/W6-2025
3rd International Workshop on Evaluation and BENCHmarking of Sensors, Systems and
GEOspatial Data in Photogrammetry and Remote Sensing (GEOBENCH), 20–21 November 2025, Wroclaw, Poland

Multi-Modal and Multi-Sensor Photogrammetric Data Fusion Exploiting a New Repository for Infrared Thermography Datasets

Neil Sutherland[1], Luca Morelli[2], Jon Mills[3], Paul Bryan, Stuart Marsh[1], Fabio Remondino[2]

[1] Nottingham Geospatial Institute, University of Nottingham - Nottingham, United Kingdom
- <neil.sutherland><stuart.marsh>@nottingham.ac.uk
[2] 3D Optical Metrology Unit, Bruno Kessler Foundation – Trento, Italy – <lmorelli><remondino>@fbk.eu
[3] School of Engineering, Newcastle University – Newcastle upon Tyne, United Kingdom – jon.mills@newcastle.ac.uk

Keywords: InfraRed Thermography (IRT), data fusion, multi-modal, architectural heritage, SuperPoint, LightGlue.

Abstract

InfraRed Thermography 3D-Data Fusion (IRT-3DDF), an emerging field of research combining 2D thermal images with 3D models, has demonstrated its competency visualising the performance of historic buildings and the behaviour of materials under varying environmental conditions. However, for 3D thermal models to become viable tools in the assessment of architectural heritage, fully-automatic IRT-3DDF methods capable of producing both geometrically- and radiometrically-accurate models require greater investigation. Therefore, using a new repository of multi-modal, multi-sensor and multi-platform datasets, this paper presents a fully-automatic IRT-3DDF method using deep learning-based multi-modal image matching, fusing multiple aerial and terrestrial sensors using combined bundle block adjustments. Results demonstrate the successful orientation several multi-sensor datasets using pre-trained neural networks, achieving sub-centimetre geometric accuracy and radiometrically consistent thermal models suitable for building diagnostics. Future work will assess the generalisability of the proposed pipeline across additional datasets, expanding its application to broader conservation, repair and maintenance (CRM) practices.

1. Introduction

InfraRed Thermography (IRT), an non-invasive, non-contact and non-destructive testing (NDT) technique, represents an established tool for heritage conservators, capable of producing real-time, highly-accurate and easily-interpretable measurements of a building's temperature (Historic England, 2025). Whilst IRT has demonstrated its competency as a stand-alone NDT technique, a growing body of research has emerged fusing complementary 2D and 3D datasets for both contemporary (Ramón et al., 2022) and historic (Adán et al., 2021) buildings. InfraRed Thermography 3D-Data Fusion (IRT-3DDF), a field combining thermal infrared (TIR) images with detailed 'base' geometries, allows for the generation of accurate 3D models supplemented with temperature information. Importantly, the rise of IRT-3DDF from the early-2000s has been driven by the growing availability, affordability and functionality of TIR cameras, often featured within three-dimensional thermal imaging systems (TTISs), manufactured or custom-built units comprising multiple sensors fixed on stationary, mobile or aerial platforms. These critical developments have laid the foundation for the continued development of IRT-3DDF and its subsequent applications for architectural heritage (Sutherland et al., 2025).

However, whilst IRT-3DDF research has grown over the past decade, critical needs remain under-researched. Firstly, IRT-3DDF methods have relied on several broad approaches: the co-registration of independently-oriented TIR and RGB image blocks using ground control points (GCPs) (Wakeford et al., 2019) or iterative closest point (ICP) (Maset et al., 2017); the use of sensors in fixed relative orientation (RO) (Dlesk et al., 2021); or the reprojection of TIR images onto additional geometric (Adán et al., 2020) or parametric (Hoegner and Stilla, 2016) models. However, the effectiveness of these methods is often restricted by: the manual identification of common features or GCPs; initial manual co-registration between the TIR and RGB blocks before ICP; or the inability for surveys to be undertaken independently to optimise image capture (e.g., RGB acquisition during the day and TIR acquisition during the night). Critically, most methods fall short of implementing fully-automated data fusion pipelines, negating their efficiency and scalability.

Additionally, the transformation of infrared radiation into quantitative temperature values is not only a by-product of the intrinsic (i.e., non-uniformity corrections (NUCs), focal plane-array (FPA) sensor temp.) and extrinsic (i.e., environmental conditions, material emissivity) elements acting upon TIR cameras (Wan et al., 2021), but also the processes involved in the pre- and post-processing of TIR images within IRT-3DDF pipelines (i.e., contrast enhancement, texturing strategies) (Lin et al., 2018). It is evident that the barrier to effective radiometric calibration, correction and validation for TIR cameras is in the access to equipment generating accurate and precise temperature measurements (e.g., calibration blackbodies, FPA thermocouples, field radiometers) (Kelly et al., 2019); which, when not present, has resulted in methods that wholly ignore radiometric accuracy. To date, radiometric calibration, correction and validation for IRT-3DDF has been largely neglected, meaning methods citing high accuracy and precision of derived temperature values remain unsubstantiated (Adán et al., 2023).

Finally, the novelty of IRT-3DDF lies not solely in the ability to *observe* thermal phenomena in three dimensions, but to *quantify* these effects for further interpretation and analysis. Although IRT-3DDF has demonstrated its capability to inform conservation, repair and maintenance (CRM) practices, developments from the AECO industry point to the opportunities for IRT-3DDF applied to novel energy performance, thermal comfort and retrofitting strategies (Natephra et al., 2017). Therefore, for IRT-3DDF to be meaningfully employed, more than superficial modelling and visualisation, 3D thermal models must be generated ready for future analysis, purposefully targeted at novel applications and specific outcomes (Sutherland et al., 2023).

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-1-W6-2025-235-2025 | © Author(s) 2025. CC BY 4.0 License.

1.1 Paper Aims

To address these challenges, this paper presents a fully-automatic IRT-3DDF method using *Thermitage*, a new data repository tailored towards the advancement of photogrammetric data fusion methods using IRT. Thermitage provides a collection of multi-modal, multi-sensor and multi-platform datasets of historic buildings that can be used to develop new methods for IRT-3DDF, featuring both geometric and radiometric validation data enabling the accuracy of derived models to be comprehensively assessed. The objectives of this work are as follows:

1. Present a multi-modal, multi-sensor and multi-platform IRT-3DDF method that does not require camera pre-calibration; manual co-registration; or auxiliary reference data.
2. Assess the radiometric accuracy of derived 3D thermal models by comparing observed- (T_{Obs}), 2D pixel- (T_{2D}) and 3D texel- (T_{3D}) temperature values.
3. Demonstrate the value of quantitative 3D thermal models as a tool for architectural heritage and CRM practices.

2. Thermitage

To achieve the listed objectives, this paper introduces *Thermitage*, a curated repository of multi-modal, multi-temporal and multi-sensor thermal images tailored towards architectural heritage[1]. The ambition for Thermitage, sharing an initial collection of datasets generated specifically for IRT-3DDF, is to provide a collaborative repository to advance photogrammetric data fusion methods focussed on the significance of historic buildings. The applications for Thermitage include: (1) the development of new fusion methods utilising Thermitage's datasets; (2) the use of Thermitage to validate or compare derived methods across different case studies; (3) the generation of training and validation data for learning-based benchmarking; and (4) the re-purposing of derived Thermitage's 3D thermal products for additional investigations.

Thermitage currently provides four datasets for the development of IRT-3DDF methods, all of contrasting scale, subject, setting and sensors (Figure 1). The diversity of Thermitage's data, purposefully curating datasets with varying perspectives (i.e., terrestrial vs. aerial), platforms (i.e., cameras vs. drones) and parameters (i.e., sensors in fixed RO vs. asynchronous), provides the means for methods to be rigorously tested to determine their flexibility and robustness. Importantly, Thermitage's datasets can be used to develop methods where specific methodological constraints are imposed, with each dataset providing specific benefits for popular IRT-3DDF methods. Finally, Thermitage encourages the sharing of thermally-attributed point clouds and thermally-textured mesh models in open file formats, allowing derived 3D products from IRT-3DDF methods to be accessed for further analysis.

To provide the necessary information to address the aforementioned challenges, each Thermitage dataset comes with the following metadata: (1) pre-calibrated IO and RO (where appropriate) values for each featured sensor; (2) geometric validation data in the form of measured points or scale bars using thermal-specific markers; (3) radiometric validation data using an infrared thermometer; (4) recorded environmental conditions for TIR image post-processing (inc., wind speed, atmospheric temperature & relative humidity); (5) material emissivity values for the predominant materials in each scene; and (6) documentation on the building's significance, conservation history and current state. The purpose of this information is to provide all the necessary data for derived IRT-3DDF methods to be comprehensively validated both in terms of geometric and radiometric accuracy, whilst also allowing for methods to be developed targetted to specific thermal anomalies.

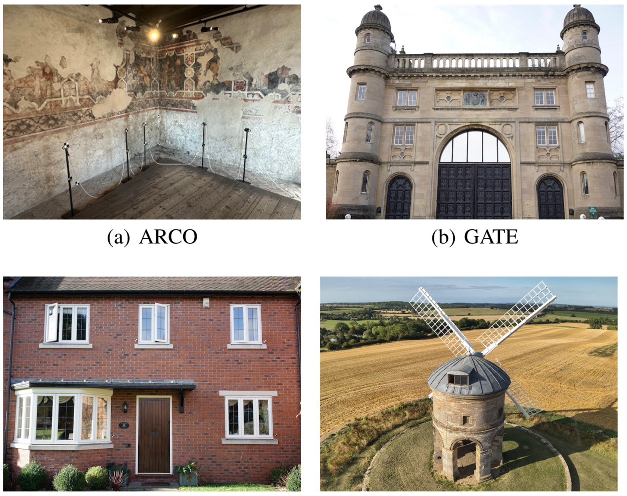

(a) ARCO
(b) GATE
(c) DUNR
(d) WIND

Figure 1. Thermitage's datasets, including: (a) Arco Castle, Italy (ARCO); (b) Lenton Lodge Gatehouse, UK (GATE); (c) Dunrobin House, UK (DUNR); and (d) Chesterton Windmill, UK (WIND).

Datasets	Sensors			Surveys		
	TIR	VIS	RGB	UAV	TLS	SLAM
ARCO	✓	✓	✓	–	–	✓
GATE	✓✓	✓✓	✓✓	–	✓	–
DUNR	✓✓	✓✓	✓✓	✓	–	–
WIND	✓	✓	✓✓	✓✓	–	–

Table 1. Thermitage's datasets, sensors and surveys.

3. Case Study

Utilising one of Thermitage's datasets, Chesterton Windmill (WIND), built between 1632-1633 by Sir Edward Peyto, is a Grade I listed scheduled monument located off the Roman Fosse Way in Warwickshire, UK (Figure 1(d)). Standing 11m tall, Chesterton Windmill features two floors of local limestone and sandstone ashlar with a winchable aluminium cap that can be turned in the direction of prevailing winds. Chesterton Windmill is the earliest tower mill in England; and, whilst the internal wooden staircases and structure no longer remain, still holds many of its original milling components (Pevsner and Pickford, 2016). Chesterton Windmill was restored in April 2024, reintroducing the trellis-style sail frames, stocks (sail blades) and undertaking minor restoration works.

3.1 Cameras

For the implementation of proposed IRT-3DDF method, several datasets were collected of Chesterton Windmill. Firstly, a DJI Mavic 3T (M3T) unmanned aerial vehicle (UAV) was employed to capture aerial imagery around the windmill. The

[1] https://github.com/3DOM-FBK/Thermitage

M3T, with a TIR camera detecting infrared radiation in the 8-14μm spectral range, features an uncooled VO_X microbolometer FPA sensor with a sensitivity of 0.05°C and an accuracy of ±2°C. Prior to data collection, radiometric calibration was undertaken to assess: (1) the accuracy of the derived TIR images against a pseudo-blackbody, (2) the required acclimatisation period of the FPA sensor, and (3) the extent of fixed-pattern noise across the FPA sensor (Wan et al., 2021). Secondly, a Sony α7RII mirrorless camera was employed to collect terrestrial RGB images. Importantly, to differentiate between 'visible spectrum' images derived from the M3T's wide-angle sensor, opposed to images captured from the terrestrial mirrorless camera; the term VIS is adopted for the wide-angle sensor on the M3T, whilst the term RGB is reserved for the Sony α7RII. Specifications for each sensor are listed in Table 2.

Specifications	DJI Mavic 3T		Sony α7RII
	TIR	VIS	RGB
Resolution (pix)	640×512	4000×3000	7952×5304
Sensor Size (mm)	7.7×6.1	6.3×4.8	35.9×24.0
Pixel Pitch (μm)	12.0	1.6	4.5
Nominal Focal Length (mm)	9.1	4.4	35.0

Table 2. Camera specifications for the DJI Mavic 3T (TIR & VIS) and Sony α7RII (RGB).

3.2 Surveys

Aerial and terrestrial surveys of Chesterton Windmill were undertaken on the 9th and 25th August 2025, respectively. For the M3T survey, images were captured between 08:55–09:40AM through passive thermography, with the windmill 'baked' from East-South-East solar radiation throughout the course of the survey. To generate accurate thermal images, field readings for relative humidity and reflected temperature were recorded with an RS PRO RS-91 hygrometer and RS PRO RS-8876 infrared thermometer, respectively. Similarly, an emissivity value (ε) for the limestone ashlar was determined using ASTM E1933-14 (ASTM International, 2022), all included in the post-processing of the derived TIR images.

To acquire images with sufficient overlap for photogrammetric processing, and to mitigate the low resolution of the TIR images, pre-defined 'point-of-interest' flights were orchestrated at different heights around the exterior of the windmill, capturing images at 3sec intervals to achieve $\sim$90% horizontal and $\sim$70% vertical overlap. Additionally, images were acquired of the interior space of the windmill with the drone handheld. Details of the surveys and sample images can be seen in Table 3 and Figure 2, respectively. Finally, a collection of 11 thermal-specific survey markers were temporarily positioned around the structure for referencing and validation. These custom 3D-printed markers, made using milled aluminium (ε=0.91) and rubber (ε=0.37) elements, provided a crosshair identifiable in all image modalities.

3.3 Sensor Configurations

To determine the flexibility of multi-modal image matching for IRT-3DDF, exploiting the diversity of sensors and platforms within Thermitage, several sensor configurations were proposed (Table 4). Firstly, a configuration using TIR+RGB images, captured from contrasting platforms (UAV vs. handheld camera) and positions (aerial vs. terrestrial), were used to determine the performance of multi-modal image matching across wide baselines, perspective changes and time frames. Secondly, TIR+VIS image pairs, captured concurrently during the M3T surveys, were assessed to determine the suitability of a single TTIS for IRT-3DDF. Importantly, this configuration can be undertaken maintaining the assumption of a fixed RO between sensors, a helpful constraint IRT-3DDF sensor fusion (Patrucco et al., 2020). Thirdly, a configuration featuring *all* images (TIR+VIS+RGB) was determined to assess the ability for the fully-automatic pipeline to orient images from three sensors in the same combined bundle block adjustment, the core originality and novelty of the proposed work. Finally, a non-fusion method utilising solely TIR images (TIR-only) is included to assess the suitability of mono-modal image matching using DL-based features and to assess the need for data fusion.

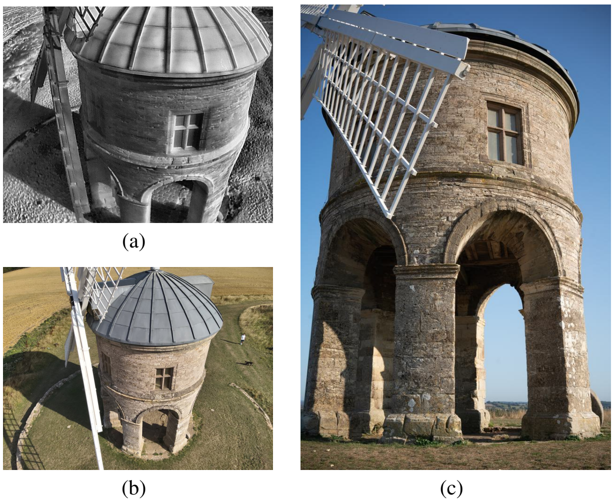

(a)
(b)
(c)

Figure 2. Sample (a) TIR, (b) VIS and (c) RGB images acquired at Chesterton Windmill.

Specifications	DJI Mavic 3T		Sony α7RII
	TIR	VIS	RGB
No. Images (Int./Ext.)	203 (43/160)	203 (43/160)	84 (50/34)
Avg. Distance (m)	4.0/8.0	4.0/8.0	4.0/10.0
Avg. GSD (mm/pix)	5.3/10.6	1.5/3.0	0.5/1.3
Original Res.	640×512	4000×3000	7952×5304
Down. Res.	–	1000×750	994×663

Table 3. Survey specifications for the DJI Mavic 3T (TIR & VIS) and Sony α7RII (RGB).

Sensor Configuration	No. Images (Int./Ext.)	No. Pairs (Exhaustive)
TIR-Only	203 (43/160)	20,503
TIR+RGB	287 (93/194)	41,041
TIR+VIS	406 (86/320)	82,215
TIR+VIS+RGB	490 (136/354)	119,805

Table 4. Proposed sensor configurations and number of image pairs given to DIM.

4. Methodology

The proposed IRT-3DDF pipeline consists of the following steps: (1) the determination of multi-modal correspondence through DL image matching; (2) the use of multi-modal correspondences within combined bundle block adjustments for

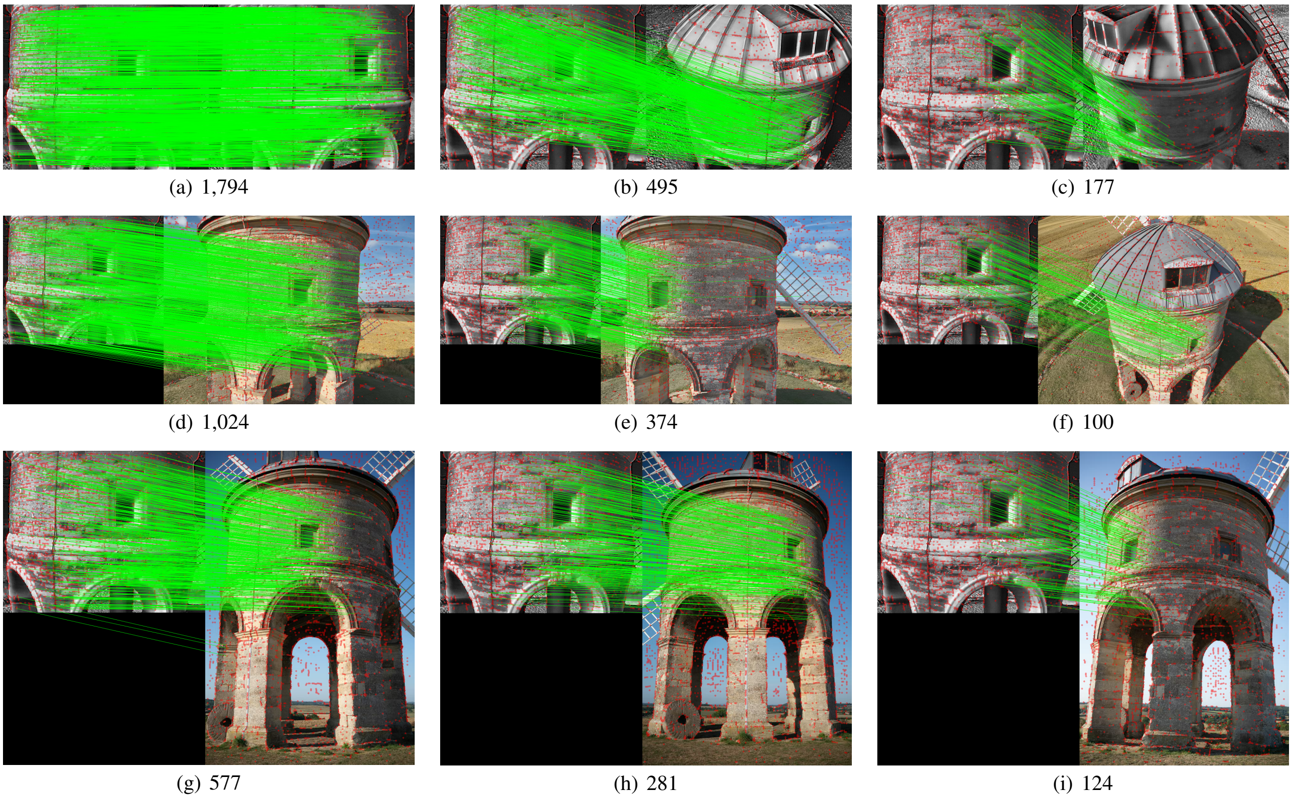

(a) 1,794 (b) 495 (c) 177
(d) 1,024 (e) 374 (f) 100
(g) 577 (h) 281 (i) 124

Figure 3. Example matches from the TIR+VIS+RGB sensor configuration, showing the number of verified matches between an exterior TIR image and additional TIR (a-c), VIS (d-f) and RGB (g-i) images of increasing complexity (left to right).

each listed sensor configuration; (3) a camera pose optimisation to refine derived camera poses; (4) the generation of a 'base' geometry and thermal texture for thermal modelling; and (4) geometric and radiometric validation. Importantly, the proposed IRT-3DDF method is fully-automatic, meaning it does not require pre-calibration for camera parameters, the use of known reference data for orientation, or manual alignment for co-registration.

4.1 Deep-Image-Matching

To determine multi-modal correspondence for IRT-3DDF, Deep-Image-Matching (DIM), a highly-configurable, open-source Python library designed for robust multi-view image matching, was employed (Morelli et al., 2024). DIM gives broad access to numerous state-of-the-art computer vision algorithms, generating local correspondences for bundle adjustments and photogrammetric reconstruction. DIM employs a modular workflow designed for effective image matching, including pipeline selection, image pair selection ('matching strategy') and image pre-processing. Whilst DIM supports various combinations of local features and matching algorithms, SuperPoint (DeTone et al., 2018) and LightGlue (Lindenberger et al., 2023) was chosen as the feature + matcher, respectively.

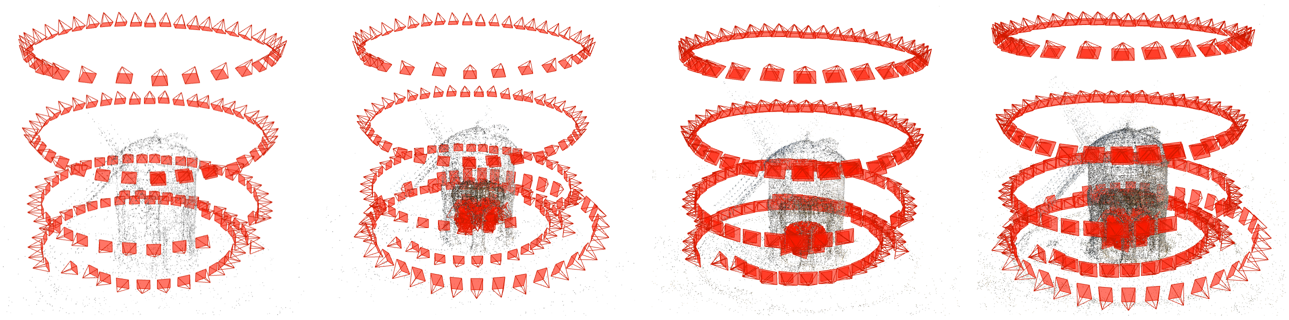
For effective multi-modal matches to be determined, both VIS and RGB images were downsampled from their original resolution to a resolution comparable to that of the TIR images to reduce the scale invariance required for the detector/descriptor (Table 3) (Marelli et al., 2023). A maximum of 8,000 keypoints was extracted from each image, with DIM settings remaining consistent across each configuration, including: no tiling, highest quality image sampling, local features and matchers parameters, and DEGENSAC (Chum et al., 2005) geometric verification error threshold (1pix). DIM's 'bruteforce' matching strategy was used to exhaustively match all images in the sensor configurations, with the number of image pairs shown in Table 4. Furthermore, all TIR images were enhanced with a contrast limited adaptive histogram equalization (CLAHE), increasing detectable features in the TIR images prior to matching (Zhang et al., 2023).

4.2 Combined Bundle Block Adjustments

For each sensor configuration, DIM's generated databases were imported into COLMAP (Schonberger and Frahm, 2016) to incrementally orient all the images, utilising the mono- and multi-modal correspondences. For each sensor (i.e., TIR, VIS or RGB), a distinct camera model was allocated with initial values of the nominal focal length (f) and image centre (u,v) provided along with two radial (k_2,k_3) and two tangential (p_1,p_2) distortion coefficients. During each bundle adjustment, camera parameters, camera poses and sparse 3D tie points were refined through a resection-intersection approach, with the resulting networks exported in Bundler format (Snavely et al., 2006) to Agisoft Metashape for optimisation, thermal modelling and validation.

4.3 Camera Pose Optimisation

As previously stated, due to the limited scale invariance of DL local features, the current pipeline does not match the full resolution images, instead determining multi-modal correspondence from the *downsampled* VIS ($\frac{1}{4}$th res.) and RGB ($\frac{1}{8}^{\text{th}}$ res.) images. Therefore, a camera pose optimisation is proposed, re-introducing the full resolution VIS and RGB images. Firstly, the full resolution VIS and RGB images were processed collectively in a traditional structure-from-motion (SfM)



(a) TIR-Only (15,371) (b) TIR+RGB (33,298) (c) TIR+VIS (56,511) (d) TIR+VIS+RGB (169,129)

Figure 4. Combined bundle block adjustment results showing the 'original' orientation of images from different sensor configurations and the number of derived 3D tie points.

workflow using Agisoft Metashape. Subsequently, the derived camera poses were imported as constrained reference data for each sensor configuration, undertaking an additional 'optimised' bundle adjustment to refine existing camera poses based on the constrained full-resolution data (σ_{xyz}=3.3mm; in line with the GSD of the VIS image block). Importantly, the benefits of this step is that the full-resolution data from all images is used to determine accurate camera poses, with the TIR poses being refined using the highest quality data.

4.4 Thermal Modelling

To generate the final 3D thermal model of Chesterton Windmill, culminating in a thermally-textured 3D mesh model, the full resolution VIS and RGB images were used for a multi-view-stereo (MVS) dense reconstruction, generating the 'base' mesh model upon which thermal images were adhered. This model was imported into the best performing sensor configuration camera network, where the TIR camera poses could be used to texture the solid mesh model. For thermal texturing, DJI Image Processor (Rava, 2025), a post-processing tool where relative humidity, reflected temp. and material emissivity values can be retroactively corrected, was used to incorporate the known environmental conditions. This produced 16-bit TIFF images encoding accurate temperature measurements within each 2D pixel, replacing the CLAHE-enhanced images used during DIM's image matching. In turn, Agisoft Metashape's 'mosaic' texturing approach was used to assign temperature values onto the base geometry, with greater statistical weight given to central image pixel values (deemed 'low frequency' information) and images aligned with façade normals ('high frequency' information) (Agisoft Metashape, 2024).

4.5 Validation

To assess the geometric accuracy of each sensor configuration, a series of scale bars were measured between the thermal-specific- and additional natural markers visible in all modalities. This resulted in 5 ground control scale bars (GCSBs) and 4 check scale bars (CSBs). Notably, the results of each sensor configuration are separated into each modality, allowing quality assessment of each image block (i.e., TIR, VIS or RGB) to be undertaken independently. In addition, the Mean Reprojection Error (MRE) on the thermal-specific markers across all images was used as additional metric for reconstruction accuracy. Furthermore, to assess the camera pose optimisation, and the benefit gained from re-introducing full resolution images, comparison between the sensor configurations before ('Original') and after ('Optimised') camera pose optimisation were assessed using the same metrics.

To demonstrate the radiometric accuracy of the proposed IRT-3DDF pipeline (a largely overlooked component of IRT-3DDF) (Adán et al., 2023), a comparison between observed infrared thermometer- (T_{obs}), 2D pixel- (T_{2D}) and 3D texel (T_{3D}) temperature values was undertaken. During data capture, T_{obs} readings were taken perpendicular to the thermal-specific markers on both crosshair rubber elements. In addition, T_{2D} values were taken from the most perpendicular instance of each marker in the corrected TIFF images, extracted in FIJI (Fiji Is Just ImageJ) (Schindelin et al., 2012). This allows for the holistic assessment of derived temperature values throughout an entire IRT-3DDF pipeline.

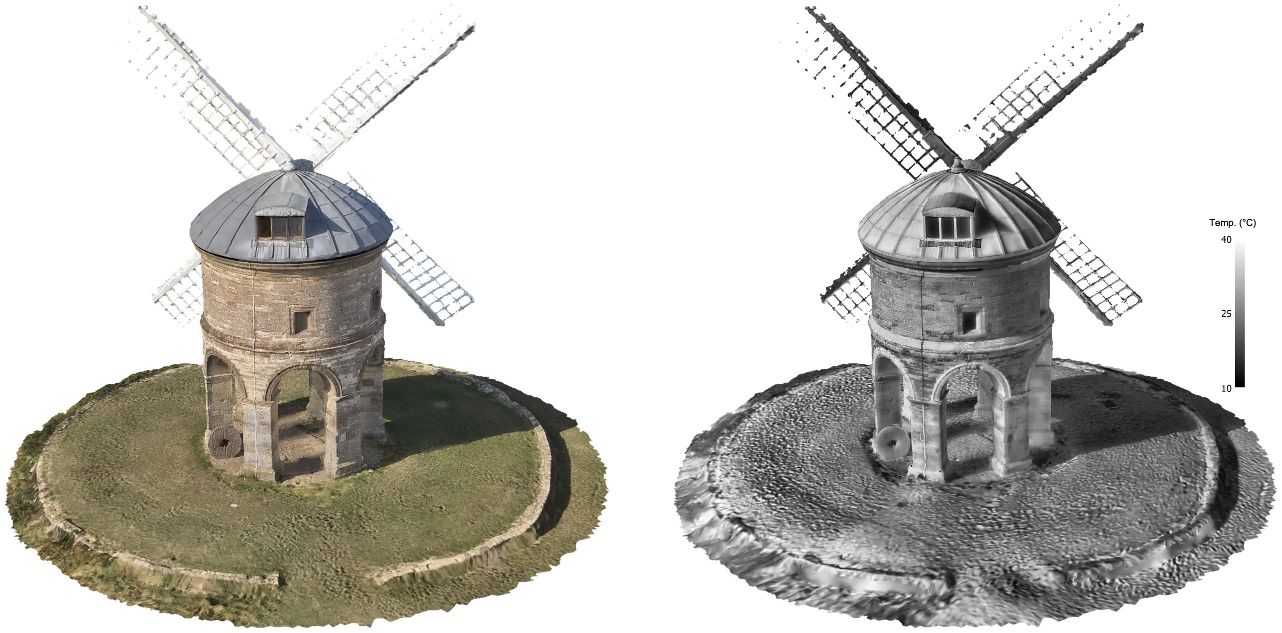
5. Results

5.1 Multi-Modal Image Matching

The results of the proposed IRT-3DDF pipeline show the success of using DL-based neural networks outside of their original training domain, with Superpoint + LightGlue able to determine local correspondences *between* (mono-modal) and *across* (multi-modal) image modalities. Figure 3, showing example matches derived from the multi-modal image matching pipeline, demonstrate the robustness of multi-modal correspondences across challenging baselines, perspectives and building components. Table 5 shows the results of the combined bundle block adjustments for the different sensor configurations, including the number of 3D tie points generated in COLMAP and the number of oriented images separated by region. As expected, data fusion methods outperform the TIR-only block, with the inclusion of multi-modal matches central to successful image orientation. Notably, the TIR-only block fails to co-register any of the interior images with those of the exterior, likely due to insufficient correspondences linking both perspectives. Interestingly, strong results from the inclusion of the tandem 'visible' sensor are apparent, with both blocks containing the VIS block successfully orienting all images. This is likely a product of the similar perspective between tandem TIR and VIS images producing a good number of multi-modal matches. These successful results justify the continued exploration of multi-modal image matching for IRT-3DDF, with similar DL-based approaches requiring future investigation.

5.2 Image Orientation

Assessing the geometric accuracy of the proposed IRT-3DDF pipeline, Table 6 shows the results of the sensor configurations when assessed against the CSBs in each image modal-



(a) RGB

(b) TIR

Figure 5. Chesterton Windmill represented as (a) RGB and (b) TIR textured 3D mesh models.

Sensor Configuration	**No. 3D Points**	**Oriented Images (Tot./Int./Ext.)**		
		TIR	VIS	RGB
TIR-Only	15,371	160/203 0/43 160/160	–	–
TIR+RGB	33,298	190/203 30/43 160/160	–	84/84 50/50 34/34
TIR+VIS	56,511	203/203 43/43 160/160	203/203 43/43 160/160	–
TIR+VIS +RGB	169,129	203/203 43/43 160/160	203/203 43/43 160/160	84/84 50/50 34/34

Table 5. Combined bundle block adjustment results separated by the number of images oriented in total (black), of the building interior (red) and building exterior (blue).

ity (i.e., $\mathrm{RMSE}_{\mathrm{TIR}}$, $\mathrm{RMSE}_{\mathrm{VIS}}$, $\mathrm{RMSE}_{\mathrm{RGB}}$), with Figure 4 visualising the camera networks of the COLMAP bundle adjustments. Firstly, all sensor configurations are able to achieve sub-centimetre accuracy on the CSBs, in line with the listed GSDs of each image block. Secondly, whilst the TIR-only block is not able to orient all images, it still achieves agreeable results in line with the given GSD ($\mathrm{RMSE}_{\mathrm{TIR}}$=9.0mm at 10.6 mm/pix). Thirdly, the TIR+VIS block achieves a superior $\mathrm{RMSE}_{\mathrm{TIR}}$ than the TIR+RGB block, an expected result due to the similar poses between the fixed tandem M3T sensors. This points to a possible benefit for IRT-3DDF use cases when a singular platform is preferred for data collection and when the orientation of TIR images is a priority (notably for IRT-3DDF methods where the reprojection of TIR images onto additional geometric or parametric models is utilised) (Hoegner et al., 2016). Finally, it is evident that results improve with an increasing numbers of images in the combined bundle block adjustment, with the TIR+VIS+RGB achieving the best results across *each* image modality. This represents a significant milestone for IRT-3DDF, co-register three different sensors successfully in a single automated process without fixed RO, included GCPs or auxiliary reference data.

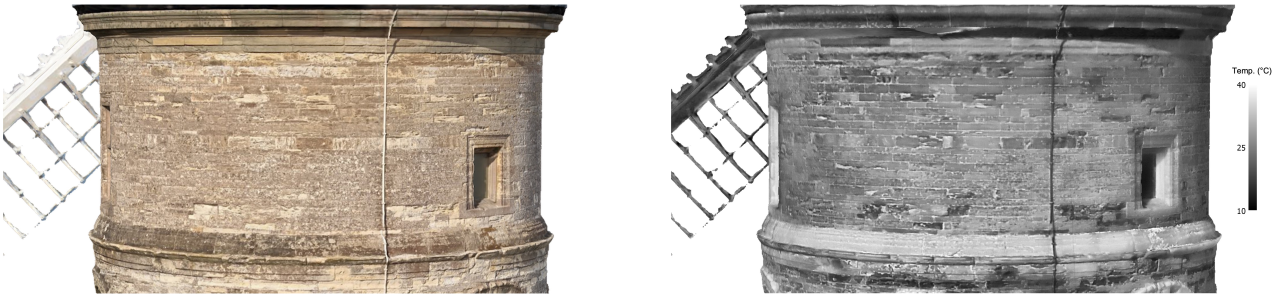
	Sensor Configuration	**MRE (pix)**	**RMSE (CSBs)**		
			TIR (mm)	VIS (mm)	RGB (mm)
Orig.	TIR-Only	0.64	9.04	–	–
	TIR+RGB	0.69	8.17	–	1.89
	TIR+VIS	0.67	6.69	3.00	–
	TIR+VIS+RGB	0.69	6.12	2.01	1.09
Opt.	TIR+RGB	0.69	7.53	–	0.45
	TIR+VIS	0.65	7.05	3.84	–
	TIR+VIS+RGB	0.65	6.59	3.51	0.15

Table 6. Image orientation results from the 'Original' (Orig.) combined bundle block adjustments and 'Optimised' (Opt.) reference-constrained bundle adjustments.

5.3 Camera Pose Optimisation

The results of the camera pose optimisation, introducing constrained camera poses generated from the full resolution VIS and RGB images, can also be seen in Table 6, comparing the RMSE on the CSBs between the 'Original' and 'Optimised' image blocks. Firstly, it is clear that the camera pose optimisation brings $\mathrm{RMSE}_{\mathrm{VIS}}$ and $\mathrm{RMSE}_{\mathrm{RGB}}$ values in line with those achieved on the full resolution images (where $\mathrm{RMSE}_{\mathrm{VIS}}$=3.89mm and $\mathrm{RMSE}_{\mathrm{RGB}}$=0.68mm on the full resolution CSBs), demonstrating the re-introduction of the full resolution camera poses is capable of refining the poses of *all* images in each sensor configuration. Notably, the re-introduction of the full resolution RGB poses for the TIR+RGB sensor configuration improves both the $\mathrm{RMSE}_{\mathrm{TIR}}$ and $\mathrm{RMSE}_{\mathrm{RGB}}$, showing the success of the proposed optimisation to accommodate sensors of significantly different resolution (640×512 vs. 7952×5304). However, for both the TIR+VIS and TIR+VIS+RGB sensor configurations, slight increases in $\mathrm{RMSE}_{\mathrm{TIR}}$ values are observed, likely due to the number and weight of constrained poses within the optimised bundle adjustment. Whilst the results of the camera pose optimisation show the ability to re-introduce information from the full resolution VIS and RGB

https://doi.org/10.5194/isprs-archives-XLVIII-1-W6-2025-235-2025 | © Author(s) 2025. CC BY 4.0 License.



(a) RGB

(b) TIR

Figure 6. Chesterton Windmill's East-South-East sunlit façade represented as (a) RGB and (b) TIR textured 3D mesh models.

images, greater investigation is required to determine whether the camera pose optimisation refines the poses generated from the combined bundle block adjustments within expected limits, or whether the weight of the constrained full resolution poses (here, using an arbitrary value based on the VIS sensor's GSD) hampers the ability for the TIR block to be improved.

5.4 Thermal Modelling

The thermally-textured 3D mesh models from the proposed IRT-3DDF pipeline can be seen in Figure 5, showing Chesterton Windmill modelled in both image modalities. Utilising the 'Optimised' TIR+VIS+RGB camera poses, the success of the proposed IRT-3DDF method can be seen, with the derived TIR poses generating a seamless texture of both the exterior and interior of Chesterton Windmill. Upon inspection, the temperature discrepancy between the high-contrast sunlit and low-contrast shadowed sides can be observed, representing a ∼17°C difference between opposing façades. Furthermore, Figure 6, presenting a close-up of the East-South-East-facing cornices, shows visible signs of stone spalling, biological colonisation and discolouration, with apparent temperature differences between areas of deteriorating and recently restored stone. Interestingly, the lower cornice, perpendicular to the solar radiation, presents a 5°C increase over the main façade. The generated 3D thermal models not only provide meaningful visualisations of Chesterton Windmill for interpretation, but provides the ability for temperature measurements to be extracted in future segmentation and classification approaches. The continued exploration of these models, in the form of temporal analyses, model segmentation and simulation, represent viable avenues of future exploration, notably for their ability to provide actionable insights for CRM practices.

5.5 Validation

Finally, the radiometric validation results at Chesterton Windmill can be seen in Figure 7, scatter-plotting the relationships between in-situ 'observed' infrared thermometer values (T_{Obs}), 2D pixel values from closest image to each measured markers normal ray (T_{2D}), and the 3D texel values of each marker within the resulting 3D thermal model (T_{3D}). Here, each wall of the 3D scatter plot depicts the relationship between two temperature metrics, with the 3D scatter plot representing the relationship (coefficient of determination) between all three metrics. This achieves a R^2 of 0.68, a satisfactory result that points to several areas of further investigation. Upon inspection of each coupled metrics, T_{Obs} vs. T_{3D} achieves the highest R^2 value (0.67), suggesting the suitability of Agisoft's mosaicking approach to combine temperature values from several images. Interestingly, the lowest R^2 comes from the T_{2D} vs. T_{3D} metrics (0.38), likely due to the singular 2D pixel value not being representative of the combined texel value. As suggested by landmark IRT-3DDF studies, the design of a strategic texturing approach is a necessary future development if accurate 2D pixel values are to integrated through a known texturing formula (Hoegner and Stilla, 2016).

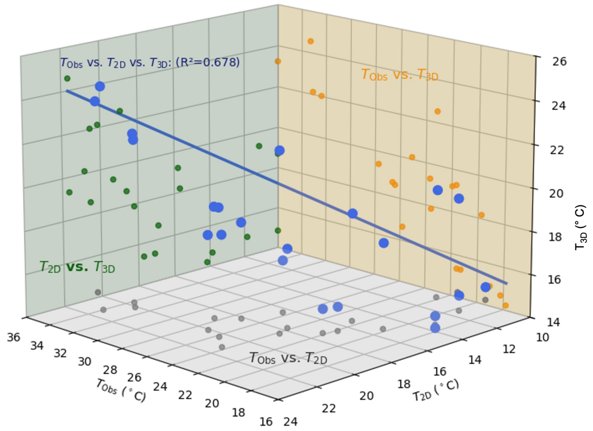


Figure 7. Scatter plots comparing 'Observed' (T_{Obs}), 2D pixel- (T_{2D}) and 3D texel- (T_{3D}) values.

6. Conclusions and Future Work

In summary, this paper outlines a novel IRT-3DDF method using learning-based multi-modal image matching for the co-registration of different sensor configurations. Results demonstrate that the proposed IRT-3DDF pipeline, operating without pre-calibration, manual co-registration or positioning data, successfully orients multiple sensors within the same combined bundle block adjustment. This confirms that DL matching models, trained to be invariant to significant radiometric and geometric distortions not associated with multi-modal sensors, can be applied effectively outside of their intended training domain. In addition, the resulting 3D thermal models not only provide comprehensive visualisations for the temperature of Chesterton Windmill under dynamic environmental conditions, but allows for the extraction of reliable temperature measurements for building diagnostics and further analysis. Future work will look to determine the generalisability of the proposed IRT-3DDF pipeline, applying multi-modal image matching and constrained sensor configurations to additional datasets from Thermitage.

Acknowledgments

This research was funded by the EPSRC Geospatial Systems CDT, grant number: EP/S023577/1. We would like to thank Dr. Maria Peppa for the use of the DJI M3T, funded by the collaborative ALGAE-Watch project. We would also like to thank Jon Bedford and Historic England's Geospatial Survey Team for lending the Sony α7RII.

References

Adán, A., Pérez, V., Ramón, A., Castilla, F. J., 2023. Correction of Temperature from Infrared Cameras for More Precise As-Is 3D Thermal Models of Buildings. *Applied Sciences*, 13(6779).

Adán, A., Pérez, V., Vivancos, J. L., Aparicio-Fernández, C., Prieto, S. A., 2021. Proposing 3D Thermal Technology for Heritage Building Energy Monitoring. *Remote Sensing*, 13(1537).

Adán, A., Quintana, B., García Aguilar, J., Pérez, V., Castilla, F. J., 2020. Towards the Use of 3D Thermal Models in Constructions. *Sustainability*, 12(8521).

Agisoft Metashape, 2024. Agisoft Metashape Professional.

ASTM International, 2022. Practice for Measuring and Compensating for Emissivity Using Infrared Imaging Radiometers.

Chum, O., Werner, T., Matas, J., 2005. Two-View Geometry Estimation Unaffected by a Dominant Plane. *2005 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'05)*, 772–779.

DeTone, D., Malisiewicz, T., Rabinovich, A., 2018. SuperPoint: Self-Supervised Interest Point Detection and Description. *IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 337–349.

Dlesk, A., Vach, K., Pavelka, K., 2021. Transformations in the Photogrammetric Co-Processing of Thermal Infrared Images and RGB Images. *Sensors*, 21(5061).

Historic England, 2025. Understanding the Environmental Performance of Historic Buildings for Conservation. Guidance HEAG329, Historic England, Swindon, United Kingdom.

Hoegner, L., Stilla, U., 2016. Automatic 3D Reconstruction and Texture Extraction for 3D Building Models from Thermal Infrared Image Sequences. *Proc. of the 2016 Int. Conf. on QIRT*, 322–331.

Hoegner, L., Tuttas, S., Xu, Y., Eder, K., Stilla, U., 2016. Evaluation of Methods for Coregistration and Fusion of RPAS-Based 3D Point Clouds and Thermal Infrared Images. *Int. Arch. of the Photogramm. Remote Sens. Spatial Inf. Sci.*, XLI-B3, 241–246.

Kelly, J., Kljun, N., Olsson, P.-O., Mihai, L., Liljeblad, B., Weslien, P., Klemedtsson, L., Eklundh, L., 2019. Challenges and Best Practices for Deriving Temperature Data from an Uncalibrated UAV Thermal Infrared Camera. *Remote Sens.*, 11(567).

Lin, D., Jarzabek-Rychard, M., Schneider, D., Maas, H.-G., 2018. Thermal Texture Selection and Correction for Building Facade Inspection Based on Thermal Radiant Characteristics. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2, 585–591.

Lindenberger, P., Sarlin, P.-E., Pollefeys, M., 2023. LightGlue: Local Feature Matching at Light Speed. *IEEE/CVF Int. Conf. on Computer Vision (ICCV)*, 17581–17592.

Marelli, D., Morelli, L., Farella, E. M., Bianco, S., Ciocca, G., Remondino, F., 2023. ENRICH: Multi-purposE dataset for beNchmaRking In Computer vision and pHotogrammetry. *ISPRS Journal of Photogramm. and Remote Sens.*, 198, 84–98.

Maset, E., Fusiello, A., Crosilla, F., Toldo, R., Zorzetto, D., 2017. Photogrammetric 3D Building Reconstruction from Thermal Images. *ISPRS Annals of the Photogramm. Remote Sens. Spatial Inf. Sci.*, IV-2/W3, 25–32.

Morelli, L., Ioli, F., Maiwald, F., Mazzacca, G., Menna, F., Remondino, F., 2024. Deep-Image-Matching: A Toolbox for Multiview Image Matching of Complex Scenarios. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-2/W4-2024, 309–316.

Natephra, W., Motamedi, A., Yabuki, N., Fukuda, T., 2017. Integrating 4D Thermal Information with BIM for Building Envelope Thermal Performance Analysis and Thermal Comfort Evaluation in Naturally Ventilated Environments. *Building and Environment*, 124, 194–208.

Patrucco, G., Cortese, G., Tonolo, G. F., Spanò, A., 2020. Thermal and Optical Data Fusion Supporting Built Heritage Analyses. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLIII-B3-2020, 619–626.

Pevsner, N., Pickford, C., 2016. *Warwickshire*. The Buildings of England, revised edition edn, Yale University Press, London.

Ramón, A., Adán, A., Javier Castilla, F., 2022. Thermal Point Clouds of Buildings: A Review. *Energy and Buildings*, 274.

Rava, M., 2025. DJI Image Processor.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., Cardona, A., 2012. Fiji: An Open-Source Platform for Biological-Image Analysis. *Nature Methods*, 9(7), 676–682.

Schonberger, J. L., Frahm, J.-M., 2016. Structure-from-Motion Revisited. *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 4104–4113.

Snavely, N., Seitz, S. M., Szeliski, R., 2006. Photo Tourism: Exploring Photo Collections in 3D. *ACM Transactions on Graphics*, 25(3), 835–846.

Sutherland, N., Marsh, S., Priestnall, G., Bryan, P., Mills, J., 2023. InfraRed Thermography and 3D-Data Fusion for Architectural Heritage: A Scoping Review. *Remote Sensing*, 15(2422).

Sutherland, N., Marsh, S., Remondino, F., Perda, G., Bryan, P., Mills, J., 2025. Geometric Calibration of Thermal Infrared Cameras: A Comparative Analysis for Photogrammetric Data Fusion. *Metrology*, 5(43).

Wakeford, Z. E., Chmielewska, M., Hole, M. J., Howell, J. A., Jerram, D. A., 2019. Combining Thermal Imaging with Photogrammetry of an Active Volcano Using UAV: An Example from Stromboli, Italy. *The Photogramm. Record*, 34(168), 445–466.

Wan, Q., Brede, B., Smigaj, M., Kooistra, L., 2021. Factors Influencing Temperature Measurements from Miniaturized Thermal Infrared (TIR) Cameras: A Laboratory-Based Approach. *Sensors*, 21(8466).

Zhang, D., Liu, Y., Zhao, Y., Liang, J., Sun, B., Chu, S., 2023. Algorithm Research on Detail and Contrast Enhancement of High Dynamic Infrared Images. *Applied Sciences*, 13(12649).