

# Georeferencing of Photovoltaic Modules from Aerial Infrared Videos using Structure-from-Motion

Lukas Bommes<sup>1</sup>, Claudia Buerhop-Lutz<sup>1</sup>, Tobias Pickel<sup>1</sup>, Jens Hauch<sup>1</sup>, Christoph Brabec<sup>1,2</sup>, and Ian Marius Peters<sup>1</sup>

<sup>1</sup>*Forschungszentrum Jülich GmbH, Helmholtz-Institute Erlangen-Nürnberg for Renewable Energies (HI ERN)*

<sup>2</sup>*Institute Materials for Electronics and Energy Technology, Universität Erlangen-Nürnberg (FAU)*

Correspondence to i.peters@fz-juelich.de

## ABSTRACT

*To identify abnormal photovoltaic (PV) modules in large-scale PV plants economically, drone-mounted infrared (IR) cameras and automated video processing algorithms are frequently used. While most related works focus on the detection of abnormal modules, little has been done to automatically localize those modules within the plant. In this work, we use incremental structure-from-motion to automatically obtain geocoordinates of all PV modules in a plant based on visual cues and the measured GPS trajectory of the drone. In addition, we extract multiple IR images of each PV module. Using our method, we successfully map 99.3 % of the 35084 modules in four large-scale and one rooftop plant and extract over 2.2 million module images. As compared to our previous work, extraction misses 18 times less modules (one in 140 modules as compared to one in eight). Furthermore, two or three plant rows can be processed simultaneously, increasing module throughput and reducing flight duration by a factor of 2.1 and 3.7, respectively. Comparison with an accurate orthophoto of one of the large-scale plants yields a root mean square error of the estimated module geocoordinates of 5.87 m and a relative error within each plant row of 0.22 m to 0.82 m. Finally, we use the module geocoordinates and extracted IR images to visualize distributions of module temperatures and anomaly predictions of a deep learning classifier on a map. While the temperature distribution helps to identify disconnected strings, we also find that its detection accuracy for module anomalies reaches, or even exceeds, that of a deep learning classifier for seven out of ten common anomaly types. The software is published at <https://github.com/LukasBommes/PV-Hawk>.*

## I. INTRODUCTION

The large amount of global installed solar photovoltaics (PV) and expected future growth require automatic image analysis for adequate quality control. As PV modules may develop defects due to environmental influences, aging or incorrect handling, PV plants need to be inspected regularly to ensure safe operation and maximum yield. Due to the large size of most PV plants, inspection is only economic if highly automated [1]. Thus, recent years have seen a surge in automated PV plant inspection systems, such as the ones by Zefri et al. [2], Pierdicca et al. [3], Henry et al. [4], and Carletti et al. [5]. These systems rely on drones equipped with a thermal infrared (IR) camera, that enables detection of abnormal PV modules based on their thermal signature [6]. The large amounts of acquired IR images are automatically processed by computer vision algorithms, which typically detect

PV modules in the images, predict module anomalies, and localize each module in the PV plant.

In this work, we focus on the localization of PV modules in large-scale plants. Localization is a crucial task as it enables targeted repairs of abnormal modules. However, it is also notoriously difficult to identify the correct module among millions of identically looking and densely packed modules from a highly repetitive video with only a limited viewport. Previous works attempted to solve this problem by stitching adjacent video frames of a PV plant row into a panorama image [7]–[9]. This approach was successful, yet only works well for short video sequences. And, as also shown in our previous work [10], panorama stitching requires manual selection of the video frames for each row and provides the module location only relative to other modules. Niccolai et al. [11] also use panorama stitching and additionally match each row panorama to a CAD plan. While this yields absolute module locations, it requires a CAD plan, which is not always available and, even if it is available, is by no means standardized across different PV plants.

Other works explore direct georeferencing of PV modules in each image based on the measured GPS position and altitude of the drone [12], [13]. Georeferencing requires a centimeter-accurate Realtime Kinematics GPS (RTK-GPS) and is prone to GPS measurement errors as no additional visual cues are considered. Further, georeferencing is limited to nadiral images, which may contain sun reflections and exhibit sub-optimal contrast compared to images taken under the optimal viewing angle. Being limited to nadiral images also makes drone operation more difficult.

Another method for module localization is the creation of an orthophoto from a few high-altitude images [14]–[16]. Orthophotos allow visualizing the temperature distribution of the entire PV plant. One issue with this approach is that it is not always possible to take images from high altitudes, e.g. if there are nearby streets. Furthermore, a low spatial resolution and possible visual artefacts impede accurate detection of abnormal modules based on the orthophoto alone.

This work presents a new method for PV plant inspection based on aerial IR videos. As opposed to the related works, our method is fully automated, provides the absolute geocoordinates of each PV module instead of a relative location, works on long video sequences of large-scale plants, requires no CAD plan, works with both standard GPS and RTK-GPS and is not limited to nadiral videos. Furthermore, videos can be acquired from low flight

altitudes and multiple high resolution images of each module are obtained, which are important for downstream analysis.

Our method builds on our previous work [10], but features a more general approach for PV module localization based on structure-from-motion (SfM) [17], [18] to obtain absolute geo-coordinates of the PV modules in a plant. In addition, PV module images are extracted from each video frame and tracked over subsequent frames. Based on the extracted images, module anomalies can be detected with a deep learning classifier [19] and visualized on a map. This enables quick assessment of the health state of the entire PV plant and helps performing targeted repairs. Similarly, module temperatures can be mapped across the PV plant. Temperature mapping allows detecting abnormal modules by comparison with neighbouring modules. This approach can replace more complex deep learning classifiers for detecting abnormal modules, as we will show. As opposed to the related works, our method relies on both visual cues and measured GPS trajectory for georeferencing. This improves robustness to GPS measurement errors and allows to use standard GPS instead of RTK-GPS. We further use videos instead of individual images. Video analysis speeds up data acquisition and works not only with automatic waypoint flights, but also with manual flights performed ad-hoc for small and irregular plants. Videos also yield larger amounts of data as each PV module is captured in multiple video frames, which is beneficial for training machine learning algorithms on the extracted data. Compared to our previous work, requirements on the flight trajectory are less stringent, and, in principle, plants with non-row layouts, such as rooftop plants, can be processed. Having a single tool for different types of PV plants is more cost-effective and requires less maintenance than multiple plant-specific solutions. We also show that for regular plants, multiple rows can be scanned simultaneously, which significantly increases throughput.

## II. METHOD

This section introduces our method for fully automatic extraction and georeferencing of PV modules from aerial IR videos. For an overview see fig. 1. After acquisition with the drone, IR videos of a PV plant are split into individual frames and the GPS trajectory of the drone is extracted and interpolated. Following Bommes et al. [10], PV modules are segmented by Mask R-CNN [20], tracked over subsequent frames, extracted and stored to disk. To georeference PV modules, a subset of keyframes is selected based on travelled GPS distance and visual overlap. Subsequently, a georeferenced 3D reconstruction of the PV plant is obtained by incremental SfM alongside the 6-DOF camera pose of each keyframe. This requires calibrated camera parameters, which are obtained beforehand. The known keyframe poses are then used to triangulate observed PV modules into the 3D reconstruction, yielding the desired module geocoordinates.

### A. Camera Model and Calibration

Several steps of our pipeline use a calibrated pinhole camera model to project 3D scene points into image coordinates and to triangulate image points into a 3D reconstruction of the scene. Lens distortion is modelled by a Brown-Conrady radial distortion model [21] with five distortion coefficients.

Calibration is performed once for each camera using OpenCV's [22] calibration method with around 150 images of a chessboard calibration target (see fig. 2). The target consists of foil patches applied to a polymer panel, providing sufficient contrast in the IR image due to different emissivities. We obtain best results when capturing calibration images outside on a cloudy day.

### B. Drone Flight and Video Acquisition

Our method is intended to be used with IR videos acquired by a drone, which scans one or multiple rows of a PV plant at an altitude of 10 m to 30 m and at a velocity that ensures blur-free images. Acquisition should take place under clearsky conditions and solar irradiance above  $700 \text{ W m}^{-2}$ . Similar to our previous method [10], both nadiral and non-nadiral videos can be processed and the camera angle and flight velocity may be varied during the flight. For accurate georeferencing of the SfM reconstruction the drone needs to travel a sufficient distance in at least two orthogonal directions. Furthermore, the flight altitude should be kept approximately constant in case standard GPS is used and no accurate altitude measurement is available. These requirements are much less restrictive than those of our previous work, resulting in higher flexibility and robustness. It is, for example, no problem, if the drone moves non-monotonically along a plant row, or if the same row is scanned multiple times. Furthermore, situations, in which the scanned row is cropped at the top or bottom of the frame, or in which other rows become visible in the camera viewport, can be handled.

For compatibility with the remaining processing steps, we split the acquired IR videos into individual 16-bit grayscale images, convert each image to Celsius scale, normalize to the interval  $[0, 255]$  using its minimum and maximum temperature value, convert to 8-bit, and finally, perform histogram equalization.

### C. Segmentation, Tracking and Extraction of PV modules

These steps correspond to our previous work [10] and are therefore described only briefly. A Mask R-CNN instance segmentation model, which is trained on a photovoltaic-specific dataset, obtains a binary segmentation mask for each PV module in each video frame. After fitting a quadrilateral to each mask, the underlying image region is extracted, warped to a rectangular region by a homography and stored as a 16-bit radiometric image file. A tracking algorithm associates masks of the same PV module over subsequent frames and assigns a unique tracking ID to each module. The tracking ID is then used to group the extracted image patches of each module.

### D. Preprocessing of the GPS Trajectory

The drone records its latitude and longitude in WGS-84 coordinates at a rate of 1 Hz. As we do not use RTK-GPS, the measured altitude is unreliable and we assume it as unknown in the subsequent steps. To match the rates of GPS measurements to the higher frame rate of the camera, we perform piecewise linear interpolation of the GPS trajectory and sample a GPS position for each frame. Prior to this, we transform the trajectory from WGS-84 coordinates to local tangent plane (LTP) coordinates [23]. LTP coordinates are Cartesian with their origin at or near the inspected site. This enables accurate interpolation and enhances numerical stability in the subsequent SfM procedure.

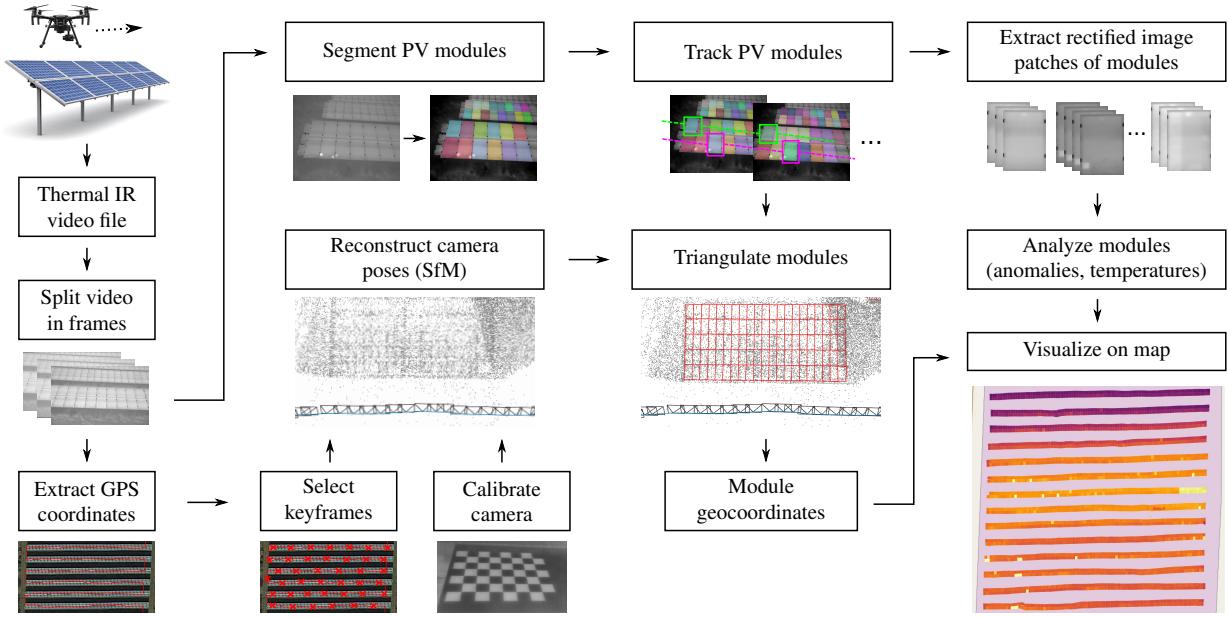


Figure 1: Overview of our method for automatic extraction and georeferencing of PV modules from aerial IR videos.

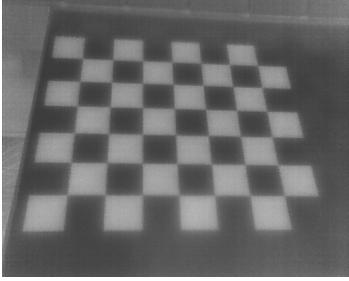


Figure 2: Exemplary IR image of the camera calibration target.

#### E. Selection of Frames for Reconstruction

In this step we select a subset of partially overlapping video frames for the subsequent SfM procedure, which we call keyframes. Subsampling the frames keeps the computational cost of the SfM procedure, which is quadratic in the number of frames, within an acceptable range. It also decouples SfM from the video frame rate, simplifying the use of different cameras. SfM further benefits from the larger parallax between any two keyframes, ensuring more accurate triangulation of scene points.

We select a frame as a keyframe if i) its distance to the previous keyframe along the GPS trajectory exceeds 0.75 m, or if ii) its intersection over union (IoU) with the previous keyframe is smaller than 85 %. To obtain the IoU, ORB features [24] of the frame and the previous keyframe are extracted and matched. A homography is estimated from the matches, which projects the bounding rectangle of the frame onto that of the previous keyframe. The IoU is then the intersection area of both rectangles divided by their total area.

One advantage of capturing videos over individual images, is the ability to adjust the overlap between images after the data is already captured.

#### F. Reconstruction of Camera Poses with SfM

In this step the 6-DOF camera pose of each keyframe is reconstructed using OpenSfM, an incremental SfM library [25]. Inputs are the calibrated camera parameters and the selected keyframes with their GPS positions in LTP coordinates. Due to unavailability of reliable measurements we set the GPS altitude to zero and fix the dilution of precision (DOP) to 0.1 m. Outputs are the rotation and translation of each keyframe in a LTP coordinate system and a 3D point cloud of reconstructed scene points, which is not further needed. An example is shown in fig. 4a. In the following, we explain briefly how the SfM library works.

1) *Feature detection and matching:* The SfM library first finds HAHOG features [26], i.e. characteristic points, in each keyframe. Overlapping frames are then found by matching these features between pairs of frames. To limit the search space matches are computed only for frame pairs which are at most 15 m apart.

2) *Initialization of the reconstruction:* One frame pair with sufficient parallax is selected for initialization of the reconstruction. The pose of the first frame is set as world coordinate origin. The pose of the second frame relative to the first frame is estimated with the five-point algorithm [27] or, in case of a planar scene, by decomposing a homography [28]. An initial set of 3D scene points is triangulated from the matched feature points in both frames.

3) *Iterative reconstruction:* Starting from the initial frame pair the other keyframes are added incrementally to the reconstruction. In each iteration the frame with most matches to any of the reconstructed frames is selected. Its pose is estimated from observed 3D scene points in the reconstruction and their corresponding 2D projections in the frame by solving the perspective-n-point problem [29]. Subsequently, new scene points are triangulated from feature points shared between the newly added frame and other frames in the reconstruction. Afterwards, the entire reconstruction is rigidly transformed, so that camera positions best

align with their measured GPS positions. In regular intervals bundle adjustment optimizes all reconstructed camera poses and scene points simultaneously by minimizing the reprojection error of the scene points in all frames. Here, camera positions are kept close to their measured GPS positions. Additionally, camera parameters are refined. We use these refined parameters in all subsequent steps.

*4) Post-processing of the reconstruction:* Under some circumstances the reconstruction of a long video sequence can fail partially. This results in multiple partial reconstructions each with a different LTP coordinate origin. To register all partial reconstructions in a common LTP coordinate frame, we transform each partial reconstruction to WGS-84 coordinates using the reconstruction-specific LTP origin. We then transform the reconstruction back to LTP coordinates, this time using the common LTP origin. This common origin is arbitrarily set to the origin of the first partial reconstruction.

#### G. Obtaining Geocoordinates of PV Modules

Once the keyframe poses are reconstructed, we triangulate the corner/center points of segmented PV modules into the reconstruction, yielding corresponding LTP geocoordinates. Examples of this are shown in fig. 4b and 4c. Due to inaccuracies in the module segmentation a robust triangulation procedure and subsequent refinement of the obtained LTP coordinates are required.

*1) Triangulation of PV modules:* For each tracked module, we obtain pixel-coordinates of the four corner points and the center point in all keyframes, in which the module is visible. Modules observed in less than two keyframes are skipped as they are likely spurious detections and triangulation is impossible. The five module points are then undistorted with the calibrated Brown-Conrady model, and triangulated from all possible pairs of keyframes, in which they are observed. The so triangulated points are only retained if the following two conditions are met: i) The angle between the two viewing rays is larger than  $1^\circ$  for all five points, and ii) none of the reprojection errors of the five points exceeds a threshold of 5 pixels. As there are typically several pairs of keyframes observing the same module, we get a noisy set of triangulated modules (see fig. 3a). We fuse them robustly by computing the median of corresponding points (see fig. 3b).

*2) Merging of duplicate detections:* A PV module may be lost during tracking and reappear a few frames later with a different tracking ID, resulting in multiple overlapping triangulations of the module in the reconstruction (see fig. 3b). This step identifies and fuses such duplicates. To this end, for each keyframe, all triangulated modules are projected back into the frame. Two or more modules are identified as overlapping if the mean Euclidean distance between the corresponding four corner points and the center point is smaller than 20 pixels. To merge overlapping modules, module points are re-triangulated according to the procedure above, this time using all keyframes of the overlapping modules. The result is shown in fig. 3c.

*3) Refinement of triangulated modules:* The triangulation is further refined by moving nearby PV module corners closer together, yielding a smoother result (see fig. 3d). To this end, we build a graph containing all triangulated module points as vertices  $\mathcal{P}$  and edges between those points that are close to another. Points are considered close if their Euclidean distance is

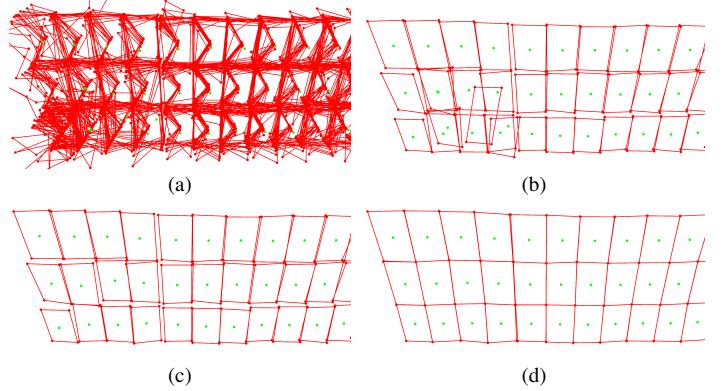


Figure 3: Steps of PV module triangulation: a) initial triangulation from all keyframe pairs, b) after computing median points, c) after merging duplicates, and d) after iterative refinement.

at most 1 m in the reconstruction and 20 pixels in projected image coordinates. Given this graph we use the g2o graph optimization framework [30] to obtain refined module points  $\mathcal{P}^*$  by optimizing the following objective

$$\mathcal{P}^* = \arg \min_{\mathcal{P}} \sum_{\langle i,j \rangle \in \mathcal{C}} \rho_h(\mathbf{e}_{ij}^\top \Omega_{ij} \mathbf{e}_{ij}). \quad (1)$$

Here,  $\mathcal{C}$  is the set of pairs of indices for which an edge exist,  $\mathbf{e}_{ij} = \mathbf{P}_i - \mathbf{P}_j$  is the difference between two points, and  $\Omega_{ij}$  is the information matrix, which we set to the identity matrix. The robust Huber cost function  $\rho_h$  reduces the impact of outliers.

We do not apply any further refinements, such as aligning surface normals of modules, or enforcing a rectangular shape, to retain maximum flexibility of our method with respect to the layout of PV modules.

#### H. Final Dataset Structure

After triangulation, LTP coordinates of PV module corners and center points are transformed back to WGS-84 coordinates and stored in a GeoJSON file together with the module tracking ID. Similarly, the extracted IR patches of each module are stored as image files in a directory named after the tracking ID. This dataset structure allows for analysis of the extracted image patches and visualization of results on a map.

### III. EXPERIMENTS & RESULTS

In this section, we apply our method to five different PV plants. We quantify the module extraction success rate together with the georeferencing error, and validate the tools' ability to process multiple plant rows in parallel. We further map predicted module anomalies and module temperatures and investigate, to what extent the temperature distribution can replace a deep learning-based classifier for the detection of abnormal modules.

#### A. Video Dataset

To validate our method, we acquire IR videos of five PV plants in Germany with a combined 35084 PV modules using a drone of type DJI Matrice 210. Tab. 1 contains details of the PV plants, drone flights and weather conditions during data acquisition. Plants A to D are large-scale open-space plants with regular row-based layouts. Plant E is a less regular arrangement

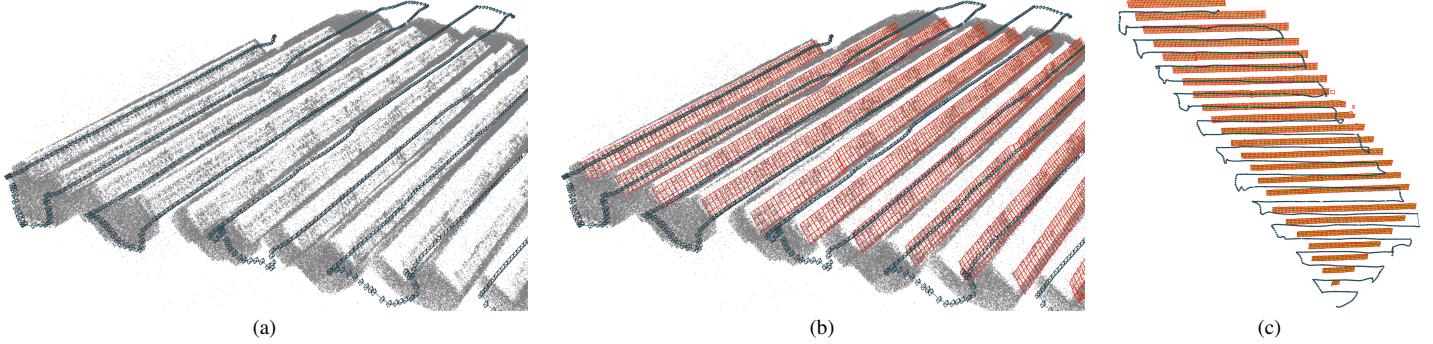


Figure 4: Reconstruction results of the SfM procedure: a) reconstructed feature points (grey) and camera poses (blue line and black camera frustums), b) with triangulated PV modules, c) top-down view on the triangulated modules of an entire PV plant.

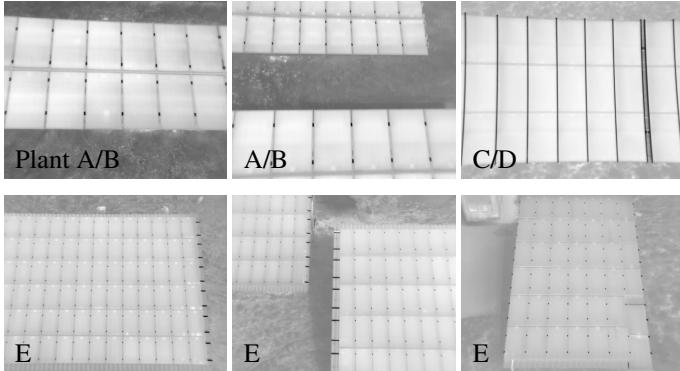


Figure 5: Exemplary IR video frames of the PV plants in our study.

of PV arrays mounted on several rooftops. All plants consist of 60-cell crystalline silicon modules. Videos of plants A, B and E are recorded by a DJI Zenmuse XT2 thermal camera with  $640 \times 512$  pixels resolution, 8 Hz frame rate and 13 mm focal length. For plants C and D we use another variant of the DJI Zenmuse XT2 with 30 Hz frame rate and 19 mm focal length. Fig. 5 shows exemplary video frames from our dataset. Note, that we could not use the dataset from our previous work due to an exchange of our camera and unavailability of calibrated camera parameters.

#### B. Module Extraction Success Rate

We apply our method on the five PV plants in our video dataset and extract over 2.2 million IR images of the 35084 PV modules of all plants (on average 64.1 images per module). Additionally, geocoordinates are obtained for each module as exemplary shown for plant B in fig. 6 (for the other plants see appendix A-A). As detailed in tab. 2, 99.3 % of all modules are successfully extracted and georeferenced. As compared to the 87.8 % success rate of our previous work, we now miss only one in 140 modules instead of one in eight. This 18-fold improvement of the extraction success rate is mostly due to the higher robustness of our new method to errors in the data acquisition process, such as cropping of the scanned row, double acquisition of the same row, or small loops in the drone trajectory. Such an error caused our previous method to loose all modules in an entire plant row. Opposed to that, our new method can handle many of those acquisition errors, and fails at most locally for a few modules. The almost perfect success rate of our new method is important in practice, as every missed module

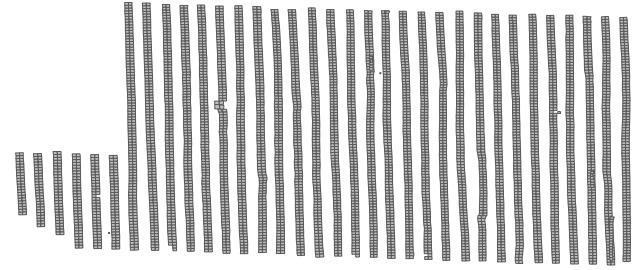


Figure 6: Map with estimated geocoordinates of PV modules in plant B.

is a missed opportunity to increase yield and profitability of the plant. Furthermore, safety critical anomalies (e.g. fire hazards) could be overlooked.

Tab. 2 also contains a detailed breakdown of the failure modes of the 234 modules missed by our method. In total, 13 modules exhibit substantial distortions, and 40 modules are missing in the reconstructions, because they are not covered by sufficiently many video frames to be accurately triangulated. Another 181 modules appear multiple times in the reconstruction because the merging procedure (sec. II-G.2) failed. This happens for modules appearing in video frames, which are temporally far apart. As these frames have a large relative pose error (due to the use of standard GPS) the triangulated modules do not align well and can not be merged correctly. This can most likely be mitigated by using RTK-GPS. Finally, there are 24 false positive modules corresponding to other objects, which are mistaken as PV modules by the Mask R-CNN segmentation model.

#### C. Georeferencing Accuracy

In this section we quantify the accuracy of the georeferenced PV module locations in terms of the root mean square error (RMSE) between estimated LTP geocoordinates  $(\hat{e}, \hat{n})$  and ground truth geocoordinates  $(e, n)$  of  $N$  selected PV modules

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{e}_i - e_i)^2 + \frac{1}{N} \sum_{i=1}^N (\hat{n}_i - n_i)^2}. \quad (2)$$

Here,  $e$  and  $n$  are the east and north positions in the LTP coordinate system. The altitude coordinate is omitted as for mapping only the horizontal error is of interest. Because point

Table 1: Details of PV plants, drone flights and weather conditions in our study. Start and end time are in UTC+2:00. Peak velocity is the 99.9 % quantile of all velocities estimated from position and time delta of subsequent video frames. Weather data is from Deutscher Wetterdienst [31]. We report mean and standard deviation of measurements taken at the nearest weather station every 10 minutes during the flight.

Plant Details			Flight Details					Weather Conditions			
ID	# Modules	Type	Start time	End time	# Frames	Distance	Peak velocity	Air temp.	Global radiation	Wind speed	Wind dir.
A	13 640	open-space	10:28:48	12:40:14	42 272	7612 m	4.1 m s <sup>-1</sup>	25.9 ± 0.5 °C	39.7 ± 1.8 J cm <sup>-2</sup>	2.8 ± 0.4 m s <sup>-1</sup>	WSW
B	5280	open-space	13:37:59	14:14:30	13 715	2929 m	4.1 m s <sup>-1</sup>	26.8 ± 0.4 °C	30.3 ± 6.1 J cm <sup>-2</sup>	3.6 ± 0.4 m s <sup>-1</sup>	SW
C	6210	open-space	12:16:21	12:39:05	34 593	2468 m	6.6 m s <sup>-1</sup>	22.3 ± 0.3 °C	46.2 ± 10.6 J cm <sup>-2</sup>	5.7 ± 0.6 m s <sup>-1</sup>	WNW
D	8460	open-space	11:01:00	11:33:34	50 348	3479 m	7.2 m s <sup>-1</sup>	23.4 ± 0.2 °C	57.4 ± 1.5 J cm <sup>-2</sup>	2.0 ± 0.8 m s <sup>-1</sup>	W
E	1494	rooftops	11:30:24	11:54:02	4527	485 m	4.2 m s <sup>-1</sup>	19.0 ± 0.3 °C	42.8 ± 3.0 J cm <sup>-2</sup>	2.6 ± 0.3 m s <sup>-1</sup>	SE

Table 2: Numbers of PV modules and module image patches extracted from the plants in our dataset. Failures are missing (MM), duplicate (DP), distorted (DS) and false positive (FP) modules.

Plant	# Modules		# Patches		# Failures			
	Total	Extracted	Extracted	∅/Module	MM	DP	DS	FP
A	13 640	13 463	398 221	29.2	18	152	7	13
B	5280	5246	140 120	26.6	6	28	0	3
C	6210	6200	635 437	102.3	4	1	5	2
D	8460	8453	936 867	110.8	7	0	0	4
E	1494	1488	138 008	90.8	5	0	1	2
Total	35 084	34 850	2 248 653	64.1	40	181	13	24

correspondences have to be found manually, we select only every 11th module in every second row of the plant and consider only the top-left module corners. We further limit the accuracy analysis to plant A, as it is the largest plant in our dataset and the only one, for which a ground truth is available. Ground truth positions are obtained from an orthophoto of the plant. This is possible, as this orthophoto exhibits a small RMSE of less than 2 cm, facilitated by the use of RTK-GPS, ground control points, high-resolution visual imagery and a higher flight altitude.

The RMSE for the entire plant is 5.87 m. This is close to the expected 4.9 m accuracy of GPS under open sky conditions [32]. However, the RMSE is not constant for the entire plant, but instead smoothly increases from 0.42 m in the east to 9.39 m in the west. This also becomes evident in fig. 7, which shows the spatial interpolation of the RMSE over the entire plant. This error drift in the SfM reconstruction is most likely caused by the low accuracy and unknown DOP of the measured GPS trajectory of the drone. As the SfM reconstruction consist of seven partial reconstructions and we do not use ground control points, another possible cause is misalignment of the partial reconstructions.

To analyze the distortion of each individual row, we remove the trend in the RMSE distribution. To this end, we align each row with the respective ground truth positions prior to computing the RMSE for the row. The resulting per-row RSME values range from 0.22 m to 0.82 m, indicating low distortion of individual rows. Due to this, accurate localization of PV modules within the plant is possible, despite the large absolute RMSE of 5.87 m.

#### D. Simultaneous Processing of Multiple Rows

One important advantage of our new method is the ability to process multiple PV plant rows simultaneously. We validate this experimentally by acquiring IR videos of the first 12 rows of plant A. We perform three flights, scanning one, two and three rows at a

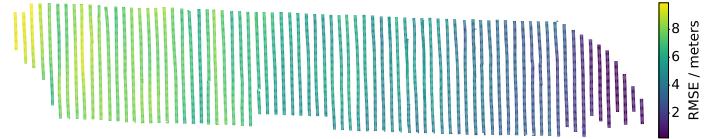


Figure 7: RMSE between ground truth and estimated horizontal geo-coordinates of PV modules in plant A. The RMSE is computed for selected points (black dots) and linearly interpolated on a 4000 × 800 grid (heatmap).

Table 3: Results for simultaneous scanning of one, two and three rows.

	One Row	Two Rows	Three Rows
Flight distance	1307 m	681 m	461 m
Flight duration	707 s	338 s	189 s
Average module resolution	141 px × 99 px	73 px × 50 px	46 px × 33 px
Module throughput	3.36 s <sup>-1</sup>	7.03 s <sup>-1</sup>	12.57 s <sup>-1</sup>

time. Fig. 8 shows exemplary video frames of each flight as well as the reconstructions of modules and flight trajectories produced by our method.

As reported in tab. 3, scanning two and three rows simultaneously speeds up the flight duration by a factor of 2.1 and 3.7, respectively. Module throughput increases accordingly from 3.36 s<sup>-1</sup> to 7.03 s<sup>-1</sup> and 12.57 s<sup>-1</sup>. This means, scanning all 2376 modules of the 12 selected rows takes only 338 or 189 seconds when scanning two or three rows at a time. Additionally, flight distance decreases by a factor of 1.9 and 2.8. This has the benefit of increasing the range of the drone before a battery change is needed. The cost for the improvement in throughput is a two- or threefold reduction in the resolution of extracted module images. Furthermore, we found the manual flight is slightly more complicated when scanning three rows at a time instead of one or two, because it is easier to miscount the rows when shifting over to the next row triplet. However, this is not a limitation when flying autonomously.

This experiment confirms the ability of our method to significantly increase throughput simply by scanning more than one plant row at a time. This is highly relevant in practice, as it significantly reduces duration and cost of the inspection. It is also an improvement over our previous method [10], which could process only one row at a time.

#### E. Mapping Module Anomalies

In this section we apply a deep learning-based binary classifier to the extracted IR image patches of each PV module in plant A,

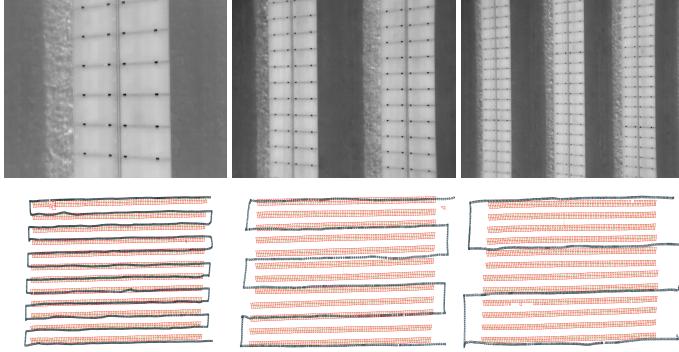


Figure 8: Top row: Exemplary video frames for scanning one, two and three PV plant rows simultaneously. Bottom row: Resulting reconstructions of modules and flight trajectory.

which predicts whether the module is abnormal or not.<sup>1</sup> We then use the estimated module geocoordinates to visualize the distribution of abnormal modules on a map (see fig. 9). Since there are multiple images for each module, we can plot the fraction of images, in which a module is predicted as abnormal. We call this the *anomaly ratio*. As opposed to a simple binary prediction, the anomaly ratio is an approximate indicator for the severity of a module anomaly. This is, because for severe, i.e. clearly visually expressed, anomalies the classifier is more confident, reaching a larger consensus of its predictions over all images of a module.

The so obtained anomaly map enables not only targeted repairs of severely abnormal modules, but also facilitates the identification of fundamental problems of the plant. For the analyzed plant, we find for example, that anomalies occur much more frequently in the bottom row, where modules are closer to the ground, rather than in the top row. A possible explanation for this is the intrusion of moisture into the PV modules near the ground. Being aware of such an issue allows the operator to monitor affected modules more thoroughly and to take action to prevent further damage to the plant.

#### F. Mapping Module Temperatures

Apart from module anomalies, we visualize the spatial distribution of module temperatures in plant A (see fig. 10). Temperatures are obtained from the extracted IR image patches of each module and plotted on a map using the module geocoordinates. For each module the maximum, minimum, mean or median temperature over the module area can be computed. Prior to this, we cut away a few pixels (5 % of the image width) from the image borders to ignore module frames and mounting brackets. To obtain a final temperature value for each module, we take the mean over the values estimated for each of the image patches of the module. As opposed to using a single representative image patch or the maximum over all patches, the mean is more robust to artifacts, which may be present in some of the module images. Of both mean and maximum temperature distributions, we find the maximum temperatures (see fig. 10b) more informative as they are sensitive to the local hot spots typically occurring in abnormal modules.

<sup>1</sup>We use the ResNet-34 convolutional neural network (CNN) classifier from Bommes et al. [19], which is trained with a supervised cross-entropy loss on labelled IR module patches of plant B in the dataset of the original work.

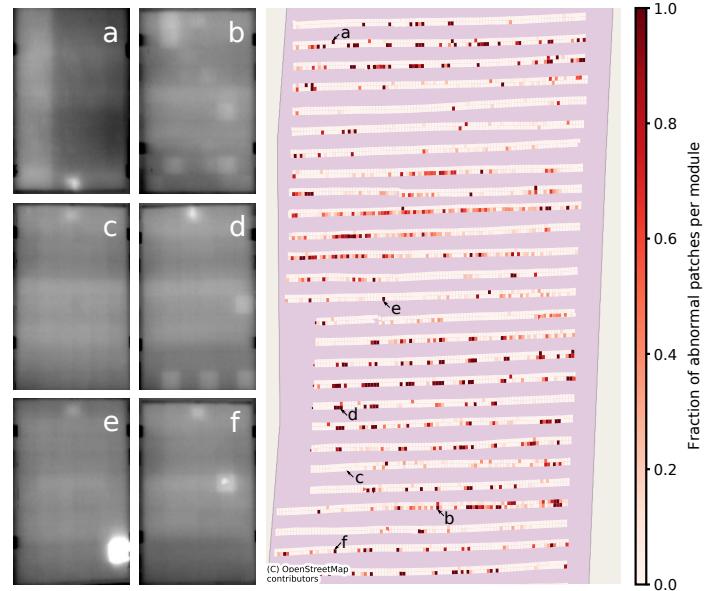


Figure 9: Map of predicted module anomalies in a section of plant A. On the left, IR images of the modules highlighted on the map are shown. Their temperature range is 30 °C (black) to 50 °C (white).

However, both mean and maximum module temperatures reveal the global temperature distribution of the plant, which is not constant, but exhibits a low-frequency pattern with temperature differences of up to 15 K. As module images are acquired over a duration of 132 minutes, possible explanations for this pattern are slow changes in the solar irradiance [33], cloud cover [34], [35], air temperature, wind speed, and camera temperature [36]. The temperature distribution is also affected by local differences in the radiative and convective heat transfer, and by the number of neighbouring modules, leading to cooler modules at the edges of each plant row [37]. Direct use of this temperature distribution for anomaly detection is not possible, as there is no common threshold value, which separates normal from abnormal modules. To account for this, we compute local temperature differences between neighbouring modules. Specifically, we subtract the median of the maximum module temperatures of all neighbouring modules within a radius of 7 m from each module. Fig. 10c shows the resulting relative maximum module temperatures. These relative temperatures are independent of the changes in environmental conditions during the flight, and consequently facilitate detection of abnormal modules by selecting a suitable temperature threshold.

Apart from locally overheated modules, the temperature distribution allows to identify string anomalies. For example in plant A, there is an inactive string (in the middle of the 19th row counted from the bottom), which is clearly visible in the temperature map (see fig. 10). Being able to identify such anomalies is important, as an entire inactive string causes large yield and power losses.

#### G. Anomaly Detection with Module Temperatures

In this section we analyze, whether the relative maximum module temperatures (see fig. 10c) alone are sufficient to accurately identify abnormal modules, and whether they can replace the more complex deep learning-based anomaly classifier from



Figure 10: Map of plant A showing the distribution of mean (a) and maximum (b) module temperatures. Each value is the average over all images showing a module. For higher contrast temperatures are clipped below 30 °C and above 50 °C. In (c) local differences of the maximum module temperature are emphasized by subtracting the median temperatures of neighbouring modules within a radius of 7 m.

sec. III-E. To this end, we manually label all 13463 modules of plant A as healthy or as abnormal with one out of the ten anomaly classes shown in fig. 11. For each module a binary anomaly prediction is obtained by comparing its relative maximum module temperature to a specified threshold value. Similarly, for the deep learning classifier we compare the anomaly ratio (see fig. 9) of the module to a threshold value.

As common in the anomaly detection literature [38]–[40], we quantify the anomaly detection performance as the area under the receiver operating characteristic (AUROC). This metric is independent of a specific threshold value, and therefore, enables a fair comparison of both classifiers, which depend differently on their threshold values. AUROC is defined as the area under the true positive rate  $TPR = TP/(TP + FN)$  plotted against the false positive rate  $FPR = FP/(FP + TN)$  at different threshold values. Here, TP and TN are the numbers of correctly classified abnormal and healthy modules, and FP is the number of healthy modules falsely classified as abnormal and FN the number of abnormal modules falsely classified as healthy.

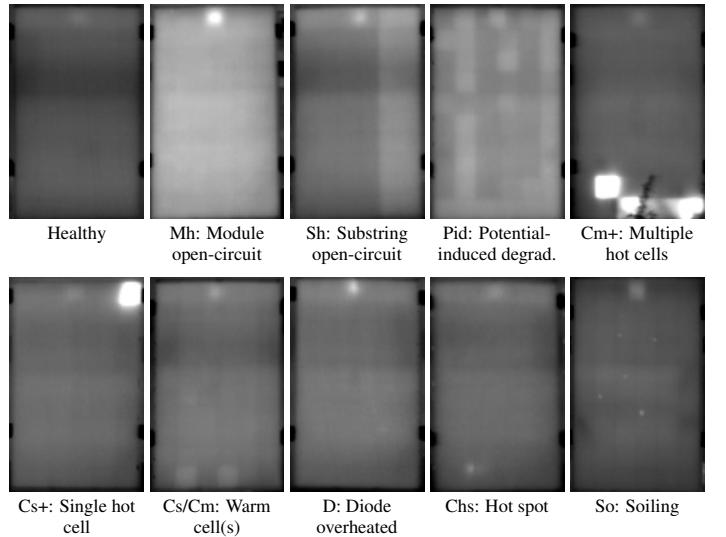


Figure 11: Exemplary IR images of the module anomaly classes in our analysis. Temperature ranges from 30 °C (black) to 60 °C (white). The figure is adopted from our previous work [10].

Tab. 4 reports the resulting AUROC scores for each anomaly class and an overall AUROC score, which considers all anomaly classes. The results indicate that both module temperature distribution and deep learning classifier perform equally and nearly perfect on severe anomalies (Sh, Cs+, Cm+), while they complement each other on the less severe anomaly classes. The deep learning classifier performs better for Pid, Cs and Cm anomalies, which are characterized by low temperature gradients, and are therefore not as accurately identifiable by the temperature distribution. On the contrary, the temperature distribution performs better for D, So, Chs anomalies, which have large temperature gradients and a small spatial extent. The small spatial extent makes detection of these anomalies difficult for a convolutional neural network. Both classifiers perform poorly on homogeneously overheated modules (Mh) because their predictions are based on temperature differences within the image (deep learning classifier) or within the local neighbourhood of modules (temperature distribution). However, using absolute instead of the relative maximum module temperatures allows to accurately identify Mh anomalies (see sec. III-F).

Summing up, the module temperature distribution can supersede a complex deep learning-based anomaly classifier for the detection of seven out of ten common module anomalies in a PV plant. This is beneficial for practical applications because of the simplicity, higher speed and better interpretability of the temperature distribution. Furthermore, no training is required, which saves the effort of creating a labelled training dataset and circumvents the issue of having to generalize from the training to the test dataset.

#### IV. CONCLUSION

In this work, we developed a method for the automatic extraction and georeferencing of PV modules from aerial IR videos, which can be used for fully automatic PV plant inspection. One possible future improvement of our method is the use of centimeter-accurate RTK-GPS instead of standard GPS, which

Table 4: AUROC scores for the detection of module anomalies in plant A by the module temperature distribution versus a deep learning classifier. Scores of the better classifier and scores above 99 % are in bold.

Anomaly	# Modules	AUROC / %	
		Temp. Distribution	Deep Learning Clf.
Mh	22	<b>59.04</b>	52.73
Sh	32	<b>99.78</b>	<b>99.95</b>
Pid	149	76.01	<b>95.92</b>
Cm+	11	<b>100.00</b>	<b>99.61</b>
Cs+	30	<b>99.81</b>	<b>99.90</b>
Cm	420	60.71	<b>86.16</b>
Cs	294	61.65	<b>78.84</b>
D	294	<b>99.31</b>	62.06
Chs	23	<b>89.64</b>	81.53
So	136	<b>79.71</b>	61.29
Overall	1411	74.31	<b>78.04</b>

could reduce the RMSE of module geocoordinates and stabilize the SfM procedure. Similarly, accuracy and stability of the SfM procedure could be improved by using visual videos instead of IR videos, as visual videos provide a higher resolution, wider viewing angle, color information and exhibit lower variation of image intensities [41]. However, this requires accurate temporal synchronization and spatial registration of the visual and IR stream, which is a challenging task. Another future direction is the correlation of the obtained temperature distribution with electrical data, such as power and yield, which could provide additional insights into the health state of a PV plant. Finally, our method could be extended for augmented reality applications by rendering a more immersive 3D model of the plant with overlaid textures, module images and interactive reports for each module.

## V. ACKNOWLEDGEMENTS

This work was financially supported by the State of Bavaria via the project PV-Tera (No. 446521a/20/5) and by BMWi via the project COSIMA (FKZ: 032429A). We sincerely thank the N-Ergie Nürnberg and PV Service Pro Kollnburg for supporting the project. The authors have declared no conflict of interest.

## REFERENCES

- [1] F. Bizzarri, S. Nitti, and G. Malgaroli, “The use of drones in the maintenance of photovoltaic fields,” *E3S Web of Conferences*, vol. 119, 2019. DOI: 10.1051/e3sconf/201911900021.
- [2] Y. Zefri, I. Sebari, H. Hajji, and G. Aniba, “Developing a deep learning-based layer-3 solution for thermal infrared large-scale photovoltaic module inspection from orthorectified big UAV imagery data,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 106, no. 5, p. 102652, 2022. DOI: 10.1016/j.jag.2021.102652.
- [3] R. Pierdicca, M. Paolanti, A. Felicetti, F. Piccinini, and P. Zingaretti, “Automatic faults detection of photovoltaic farms: solAir, a deep learning-based system for thermal images,” *Energies*, vol. 13, p. 6496, 2020. DOI: 10.3390/en13246496.
- [4] C. Henry, S. Poudel, S.-W. Lee, and H. Jeong, “Automatic detection system of deteriorated PV modules using drone with thermal camera,” *Applied Sciences*, vol. 10, p. 3802, 2020. DOI: 10.3390/app10113802.
- [5] V. Carletti, A. Greco, A. Saggese, and M. Vento, “An intelligent flying system for automatic detection of faults in photovoltaic plants,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, pp. 2027–2040, 2019. DOI: 10.1007/s12652-019-01212-6.
- [6] S. Gallardo-Saavedra, L. Hernández-Callejo, and O. Duque-Perez, “Technological review of the instrumentation used in aerial thermographic inspection of photovoltaic plants,” *Renewable and Sustainable Energy Reviews*, vol. 93, pp. 566–579, 2018. DOI: 10.1016/j.rser.2018.05.027.
- [7] F. Grimaccia, S. Leva, and A. Niccolai, “PV plant digital mapping for modules’ defects detection by unmanned aerial vehicles,” *IET Renewable Power Generation*, vol. 11, no. 10, pp. 1221–1228, 2017. DOI: 10.1049/iet-rpg.2016.1041.
- [8] G. Francesco, L. Sonia, and N. Alessandro, “A semi-automated method for defect identification in large photovoltaic power plants using unmanned aerial vehicles,” in *IEEE Power Energy Society General Meeting*, Portland, OR, USA, 2018, pp. 1–5.
- [9] M. Aghaei, S. Leva, and F. Grimaccia, “PV power plant inspection by image mosaicing techniques for IR real-time images,” in *IEEE 43rd Photovoltaic Specialists Conference*, Portland, OR, USA, 2016, pp. 3100–3105.
- [10] L. Bommes, T. Pickel, C. Buerhop-Lutz, J. Hauch, C. Brabec, and I. M. Peters, “Computer vision tool for detection, mapping, and fault classification of photovoltaics modules in aerial IR videos,” *Progress in Photovoltaics: Research and Applications*, vol. 29, no. 12, pp. 1236–1251, 2021. DOI: 10.1002/pip.3448.
- [11] A. Niccolai, F. Grimaccia, and S. Leva, “Advanced asset management tools in photovoltaic plant monitoring: UAV-based digital mapping,” *Energies*, vol. 12, no. 24, p. 4736, 2019. DOI: 10.3390/en12244736.
- [12] P. Addabbo, A. Angrisano, M. L. Bernardi, G. Gagliarde, A. Menella, M. Nisi, and S. L. Ullo, “UAV system for photovoltaic plant inspection,” *IEEE Aerospace and Electronic Systems Magazine*, vol. 33, no. 8, pp. 58–67, 2018. DOI: 10.1109/MAES.2018.170145.
- [13] M. Nisi, F. Menichetti, V. Bramante, T. Tr, B. Muhammad, and R. Prasad, “EGNSS high accuracy system improving photovoltaic plant maintenance using RPAS integrated with low-cost RTK receiver,” in *Global Wireless Summit*, Aarhus, Denmark, 2016.
- [14] D. H. Lee and J. H. Park, “Developing inspection methodology of solar energy plants by thermal infrared sensor on board unmanned aerial vehicles,” *Energies*, vol. 12, no. 15, p. 2928, 2019. DOI: 10.3390/en12152928.
- [15] Y. Zefri, A. ElKettani, I. Sebari, and S. A. Lamallam, “Thermal infrared and visual inspection of photovoltaic installations by UAV photogrammetry—Application case: Morocco,” *Drones*, vol. 2, no. 4, p. 41, 2018. DOI: 10.3390/drones2040041.
- [16] I. Tsanakas, L. Ha, and F. Al Shakarchi, “Advanced inspection of photovoltaic installations by aerial triangulation and terrestrial georeferencing of thermal/visual imagery,” *Renewable Energy*, vol. 102 (Part A), pp. 224–233, 2016. DOI: 10.1016/j.renene.2016.10.046.
- [17] J. L. Schönberger and J.-M. Frahm, “Structure-from-motion revisited,” in *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016, pp. 4104–4113.
- [18] S. Agarwal, N. Snavely, I. Simon, S. M. Seitz, and R. Szeliski, “Building rome in a day,” in *IEEE International Conference on Computer Vision*, Kyoto, Japan, 2009, pp. 72–79.
- [19] L. Bommes, M. Hoffmann, C. Buerhop-Lutz, T. Pickel, J. Hauch, C. Brabec, A. Maier, and I. M. Peters, “Anomaly detection in infrared images of photovoltaic modules using supervised contrastive learning,” *Progress in Photovoltaics: Research and Applications*, 2022. DOI: 10.1002/pip.3518.
- [20] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” in *IEEE International Conference on Computer Vision*, Venice, Italy, 2017, pp. 2980–2988.
- [21] D. C. Brown, “Decentering distortion of lenses,” *Photogrammetric Engineering*, vol. 32, no. 3, pp. 444–462, 1966.
- [22] G. Bradski, “The OpenCV library,” *Dr. Dobb’s Journal of Software Tools*, vol. 120, pp. 122–125, 2000.
- [23] W. Torge and J. Müller, *Geodesy*. De Gruyter, 2012. DOI: 10.1515/9783110250008.
- [24] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “ORB: An efficient alternative to SIFT or SURF,” in *International Conference on Computer Vision*, Barcelona, Spain, 2011, pp. 2564–2571.
- [25] Mapillary, *Opensfm*, <https://github.com/mapillary/OpenSfM>, 2021.
- [26] K. Mikolajczyk and C. Schmid, “An affine invariant interest point detector,” in *European Conference on Computer Vision*, Copenhagen, Denmark, 2002, pp. 128–142.

- [27] D. Nister, “An efficient solution to the five-point relative pose problem,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 6, pp. 756–770, 2004. DOI: 10.1109/TPAMI.2004.17.
- [28] O. D. Faugeras and F. Lustman, “Motion and structure from motion in a piecewise planar environment,” *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 02, no. 03, pp. 485–508, 1988. DOI: 10.1142/S0218001488000285.
- [29] V. Lepetit, F. Moreno-Noguer, and P. Fua, “EPnP: An accurate O(n) solution to the PnP problem,” *International Journal of Computer Vision*, vol. 81, no. 2, pp. 1573–1405, 2008. DOI: 10.1007/s11263-008-0152-6.
- [30] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, “G2o: A general framework for graph optimization,” in *IEEE International Conference on Robotics and Automation*, Shanghai, China, 2011, pp. 3607–3613.
- [31] DWD Climate Data Center (CDC), *Recent 10-minute station observations of air temperature, global radiation, wind speed and wind direction for germany, quality control not completed yet, version recent, last accessed: 08/10/2021*.
- [32] F. van Diggelen and P. K. Enge, “The world’s first GPS MOOC and worldwide laboratory using smartphones,” in *28th International Technical Meeting of the Satellite Division of The Institute of Navigation*, Tampa, FL, USA, 2015, pp. 361–369.
- [33] P. Rajendran and H. Smith, “Modelling of solar irradiance and daylight duration for solar-powered UAV sizing,” *Energy Exploration & Exploitation*, vol. 34, no. 2, pp. 235–243, 2016. DOI: 10.1177/0144598716629874.
- [34] R. Nomura, T. Harigai, Y. Suda, and H. Takikawa, “Second by second prediction of solar power generation based on cloud shadow behavior estimation near a power station,” *AIP Conference Proceedings*, vol. 1807, no. 1, p. 020024, 2017. DOI: 10.1063/1.4974806.
- [35] K. Lappalainen and S. Valkealahti, “Analysis of shading periods caused by moving clouds,” *Solar Energy*, vol. 135, pp. 188–196, 2016. DOI: 10.1016/j.solener.2016.05.050.
- [36] Q. Wan, B. Brede, M. Smigaj, and L. Kooistra, “Factors influencing temperature measurements from miniaturized thermal infrared (TIR) cameras: A laboratory-based approach,” *Sensors*, vol. 21, no. 24, 2021. DOI: 10.3390/s21248466.
- [37] J. Denz, C. Buerhop, C. Camus, I. Kruse, T. Pickel, B. Doll, J. Hauch, and C. Brabec, “Quantitative assessment of the power loss of silicon PV modules by IR thermography and its practical application in the field,” in *37th European Photovoltaic Solar Energy Conference and Exhibition*, Lisbon, Spain, 2020, pp. 1542–1547.
- [38] I. Golan and R. El-Yaniv, “Deep anomaly detection using geometric transformations,” in *Advances in Neural Information Processing Systems (NIPS)*, Montréal, Canada, 2018, pp. 9781–9791.
- [39] L. Bergman, N. Cohen, and Y. Hoshen, “Deep nearest neighbor anomaly detection,” *arXiv preprint arXiv:2002.10445*, 2020. arXiv: 2002.10445. [Online]. Available: <http://arxiv.org/abs/2002.10445>.
- [40] D. Hendrycks, M. Mazeika, and T. Dietterich, “Deep anomaly detection with outlier exposure,” in *International Conference on Learning Representations (ICLR)*, New Orleans, LA, USA, 2019.
- [41] S.-S. Lin, “Review: Extending visible band computer vision techniques to infrared band images,” *Technical Reports (CIS)*, 2001.

## A. APPENDIX

### A. Additional Georeferencing Results

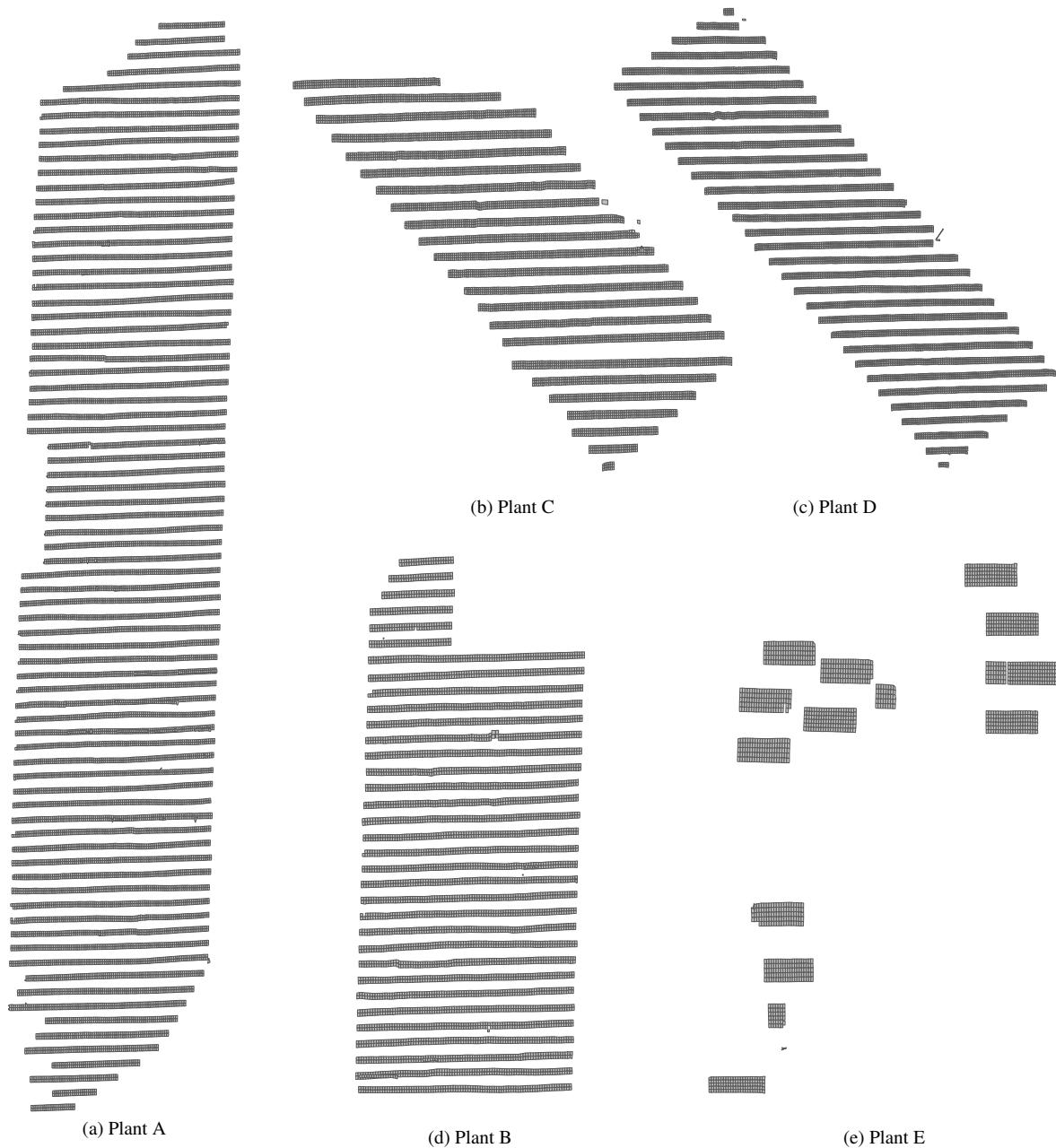


Figure 12: Georeferencing results for all PV plants in our dataset.