Marker-based Extrinsic Calibration for Thermal-RGB Camera Pair with Different Calibration Board Materials

Bilal Sher, Xuchu Xu, Guanbo Chen, and Chen Feng

Tandon School of Engineering, New York University, USA

bas9876@nyu.edu, xx762@nyu.edu, gc2720@nyu.edu, cfeng@nyu.edu

Abstract -
UAVs are used to rapidly and safely collect site data for engineering purposes. Thermal and RGB imagery are often captured for building envelope inspections. On large sites, the data produced can be cumbersome to process, especially if taken in a video format. This paper presents a way to perform image registration, or low-level data fusion, on thermal and RGB images that were simultaneously captured from rigidly fixed cameras - the same type as those found on commercial RGB-Thermal UAVs. The intention is to convert raw captured data into RGB-Thermal tensors that can be used in deep learning applications. Current techniques of geometrically calibrating a thermal camera can include complicated setups that make calibration at extreme angles difficult. Three calibration boards were involved in testing this image registration method: a cardboard and acrylic board, a wood and vinyl board, and a metal and vinyl board. Of the three, only the metal and vinyl calibration board yielded a consistent image, regardless of the time passed between the first and last captured image. This proves that calibration can be conducted cheaply and in any open area without obstructions.

Keywords -

UAVs; Thermal Imagery; Building Envelopes; Data Preparation;

1 Introduction

Thermal cameras are valuable tools for building envelope energy analysis, but they frequently produce distorted images that must be geometrically calibrated before use in computer vision applications like deep learning segmentation networks (See Figure 1).

Image segmentation can be done in multiple ways, including using a deep learning model that accepts two uncalibrated tensors as input. However, this approach requires significant training data to generalize to other drones and camera setups. The workaround is to undistort and align the images into a single tensor with the same field of view. Undistortion is necessary to ensure accurate recognition of shapes across different camera types and different areas of the image. Distortion affects different parts of the image differently; with fisheye lens distortion significantly affects the edges more than the center. Accounting for information from all portions of the image is crucial when automating computer vision tasks.

Geometric calibration determines the intrinsic properties of the camera that can be used to undistort the camera
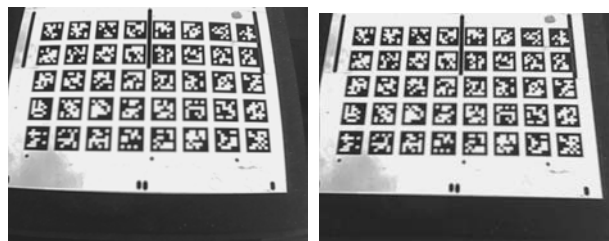

Figure 1. Raw and Processed Thermal Images

image. To learn the camera's calibration matrix and distortion coefficients, a scene with known coordinates is necessary. In RGB photos, the most common method of achieving this is through a checkerboard or a chessboard (See Figure 2). The alternating light-dark pattern identifies a common corner between the squares, and the OpenCV library finds the checkerboard points and performs camera calibration and undistortion based on the corner's screen location.

This paper examines various techniques for geometrically calibrating both thermal and RGB images taken by a drone for use in building envelope energy analysis using deep learning segmentation networks. We explore materials for the easy and simple calibration and image alignment for drones. Additionally, it proposes a method to combine them into a 4D tensor for deep learning network use. Our findings will benefit building energy analysts, engineers, and researchers who employ drones and deep learning segmentation networks for building envelope energy analysis and sustainability.

2 Related Works

2.1 Thermal Camera Calibration

There have been a number of successful attempts to geometrically calibrate a thermal camera. Ursine et al. [1] attempted to geometrically calibrate a thermal camera by creating a checkerboard calibration grid. This grid was constructed with a copper plate that had a painted checkerboard calibration pattern. The problem with such a set up is that multiple calibration rigs would be susceptible to a level of error based on the human painters skills. The
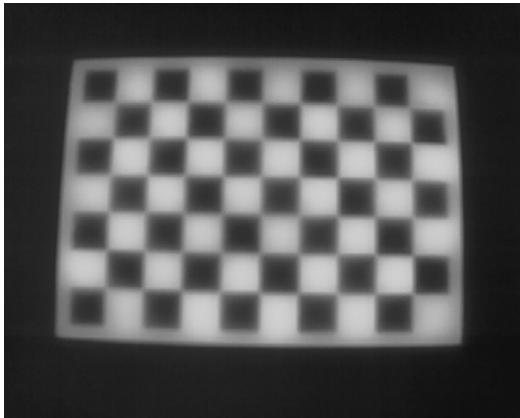
Figure 2. Uncalibrated and distorted thermal image of a checkerboard - corners of the checkerboard are known to lie on straight lines

process of making and manufacturing these calibration grids limits industrial reproducibility.

Shibata et al. [2] are able to ensure industrial reproducibility through the use of an offset grid. While they use a checkboard pattern and the size of the checkers can be ensured to be consistent within a level of machine error, the offset they introduce can result in poor calibration if a camera calibration image is taken at a somewhat large angle. Furthermore, Shibata et al recognize that it is difficult to produce thermal images of checkerboard patterns with a quality and contrast high enough for corner detection and camera calibration. They introduce a novel tone mapping technique to adaptively increase the contrast of their images. The drawback of their method is that they require active heating of their calibration grid this means that camera calibration can only be performed in labs with specific equipment. Producing their set-up is also labour intensive and cannot be easily shipped from one location to another.

Hou et al. [3] were able to achieve results similar to those achieved by Ursine et al. They used a cardboard with square tin foil cut outs to create an emissivity difference that could be create a checkerboard pattern in thermal images. They were able to find the corners in the checkerboard pattern, but they required lab scale conditions to calibrate their thermal camera. Additionally, the calibration kit they have is not robust enough for commercial applications and could easily be damaged in transportation.

Previous papers either use a complicated set up involving active heating [4], resistors [5], or offset plates [2] that make calibration at extreme angles difficult. Other papers use image processing techniques to improve the detectability of points and lines on a thermal image. This paper uses

Apriltags ensuring that it requires no further image processing and can be used to easily calibrate the cameras on a UAV that's flown above and around the board.

2.2 RGB & Thermal Image Alignment

There are a number of successful attempts to combine and register thermal and RGB images. Istenic et al. [6] used a Hough transform to register thermal and RGB images. They required a structured environment with many lines and were not able to accurately account for scale between images in all cases. Knyaz and Moshkantsev [7] developed a technique that created a 3D scene using RGB and thermal information and combined the two 3D scenes to find the transforms between associated RGB and thermal images. Dlesk et al. [8, 9] researched the use of homography to align images with detectable feature points and went on to perform experiments that create RGBT imagery through combining RGT, TBG and RTB imagery.

3 Method

3.1 Cardboard and Acrylic Calibration Board

Initial calibration experiments involved creating a calibration kit with a spatially offset calibration board and heating up protruding parts of the board to create an observable thermal difference between black and white checks in a thermal image. The calibration board was made of black acrylic laser cut squares glued on to a cardboard backing/substrate. Paper with a printed 6 x 8 checkerboard pattern was pasted onto the cardboard sheet and acrylic squares were glued onto the paper. Black sections of the paper were covered by acrylic squares. The cardboard of size 12.5 inches (317.5 mm) x 11.5 inches (292.1 mm). Acrylic squares had a side length of 1.1 inches (27.94 mm) and the printed checkerboard pattern had checks of the same size.

3.2 Wood and Vinyl Calibration Board

A calibration board for thermal imaging was made by spray painting a 300 mm x 420 mm x 3.175 mm thick wood board white and laser cutting it to have 40 mm squares. The squares were connected with 0.1 mm connections to keep them together and prevent falling out. A black vinyl substrate was taped to the back. The board was heated by placing it on a warm surface (40 C to 50 C) to create a temperature difference between the surface of the substrate and the surface of the wood, and images were captured indoors at 23 C to 25 C. The drone flew indoors to hover over the board placed on indoor carpeting, and images were captured over 6 minutes. After the calibration kit had cooled off, it was reheated, and additional images were taken. The connection between the squares was small

enough to not be detected by an RGB corner detection algorithm.

3.3 Metal and Vinyl Calibration Board

A metal Apriltag calibration board was created using a large metal board and vinyl. The metal was 3/8 inches thick and made of unpolished oxidized aluminum, while the vinyl was cut using a Cricut Maker in the NYU Makerspace. The vinyl was applied onto transfer paper, then carefully placed over the metal substrate board, making sure to reduce the chance of bubbles being captured under it. The transfer paper was carefully removed to leave only the vinyl on the metal substrate. Tests were conducted both heated and unheated, indoors and outdoors. In the first qualitative test, the vinyl side of calibration board was heated indoors to produce a thermal differential between the vinyl and the metal substrate, resulting in a strongly visible image. The drone was flown indoors to hover over the calibration board which was put onto carpeted ground and images were captured over a span of 2 minutes and 36 seconds. Noise in the non-covered metal captured in the thermal image became more apparent as the plate lost heat (See Figure 3).
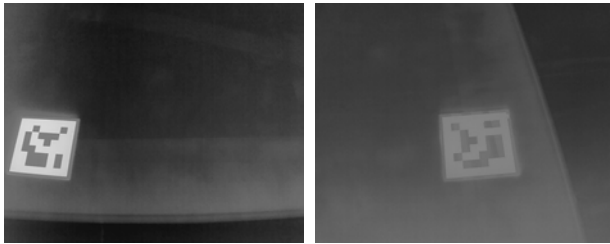

Figure 3. Thermal reflections are more easily visible in the metal portions of the calibration board

In the second test, conducted outdoors on an overcast day, the kit was placed on the concrete ground with an unobstructed view of the sky; the outside temperature on that day was approximately 8 C. The sky emits thermal energy from space which is minimal if the sun doesn't reflect on the calibration board. Overcast clouds helped diffuse the suns energy and reduce its effect on the board. This minimal, passive, diffuse thermal energy reflected differently on the metal than the vinyl and created a sharp clear image that needed no additional heating.In order to calibrate the images, the pyAprilTag library was used to calculate the distortion coefficients and the camera calibration matrix of the two cameras. RGB images that were too large and had to be reduced in size to accurately find the AprilTags. Raw captured thermal images needed to have their color scheme inverted, but no sharpening or filtering was done otherwise.

4 Results

4.1 Thermal Camera Calibration

4.1.1 Cardboard Calibration Kit

Attempts to detect checker corners on thermal images were hindered by uneven heating of the calibration board, resulting in unclear corners. Image processing techniques were employed to improve contrast between acrylic and cardboard, as seen in Figure 4).
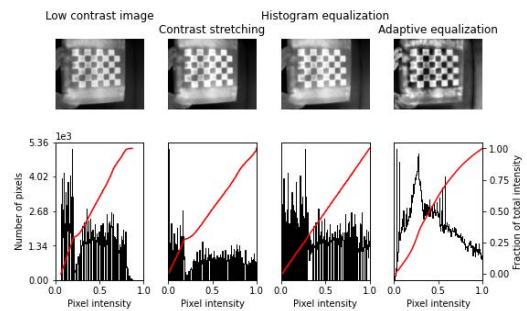

Figure 4. Image enhancement techniques were not able to improve the checker detection on the cardboard and acrylic calibration board due to uneven heating and cooling over the board.

Despite using various image processing techniques, it was not possible to get useful calibration information out of this technique. Although the acrylic squares were able to heat up evenly, the corner detection algorithm failed due to the uneven heating of the cardboard substrate. A better calibration kit was needed with a thinner substrate that would cool off more rapidly than the overlaid checker pattern. 0 corners were detected and 0 images could be used for calibration.

4.1.2 Wood and Vinyl Calibration Board

This calibration board required reheating every 6 minutes. Although some thermal images showed a checkerboard pattern, the raised corners made it difficult for the corner detection algorithm to determine the actual corner, resulting in subpar calibration (See Figure 5).

The only solution was to capture images from a distance, which caused excessive blur and made the algorithm ineffective. This experiment highlighted the need for a planar calibration surface, passive heating for capturing a large number of calibration images, and accurate detection of corners/feature points for precise alignment of RGB and thermal images. 0 corners were consistently detected and 0 images could be used for calibration.
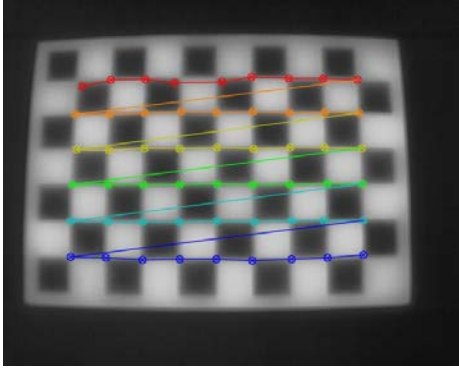
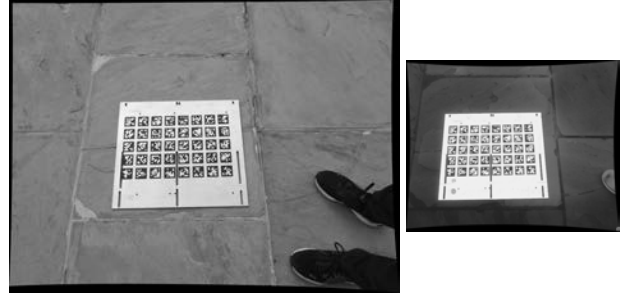Figure 5. Corner detection issues on calibration boards with spatial offsets

4.1.3 Metal and Vinyl Calibration Board

Passively heating the calibration board and orienting it towards the sky negated the need for active heating ensured consistency in the captured images, making the calibration process inexpensive and easily deployable in open areas. The use of vinyl AprilTags on a metal substrate improved the robustness of feature point detection, creating a clear contrast between high-emissivity vinyl and low-emissivity metal. Additionally, the AprilTags helped in finding common points in both RGB and thermal images, although not all AprilTags could be detected in both. Two AprilTags were consistently undetectable due to holes in the metal board, and were thus not used in analysis. Nearly 5500 points were detected using this method.

4.2 RGB & Thermal Image Alignment

4.2.1 AprilTag Detection after Image Reduction

Images were first undistorted using the distortion coefficients and the calibration matrix obtained in the previous step (See Figure 6).

Once images were undistorted, an RGB and thermal image pair was created from the RGB and thermal images taken at the same time of the same scene.

$$\textbf{Image Pair} = (P_i^{RGB}, P_i^{Th}) \quad (1)$$

Each image in the image pair was individually run through the AprilTag detection algorithm to create two sets of detected tag coordinates, X.

$$f(P_i^{RGB}) = \{X_{i,1} \quad \ldots \quad X_{i,n}\} = \mathbb{S}_{RGB} \quad (2)$$

$$f(P_i^{Th}) = \{X_{i,1} \quad \ldots \quad X_{i,n}\} = \mathbb{S}_{Th} \quad (3)$$

The intersection of the two sets of detected tag coordinates creates a new set of tag coordinate pairs.

$$\mathbb{S}_{RGB} \cap \mathbb{S}_{Th} = \mathbb{S}_{TCP} \quad (4)$$

Figure 6. An example of an undistorted image pair

$$\mathbb{S}_{TCP} = \{(X_{i,1}^{RGB}, X_{i,1}^{Th}) \quad \ldots \quad (X_{i,n}^{RGB}, X_{i,n}^{Th})\} \quad (5)$$

Each of the Euclidean coordinate pairs is then homogenized into a vector of homogeneous coordinates. At this point, all the coordinates in an image pair could be normalized and all of the detected coordinates in an image of an image pair could be represented as a coordinate matrix:

$$(P_i^{RGB}, P_i^{Th}) \rightarrow (C_i^{RGB}, C_i^{Th}) \quad (6)$$

The first normalization step is centralization – all coordinates are translated via a translation matrix T such that they are all collectively centered around the origin. The second normalization step is scaling to reduce the average distance from the origin of all centralized points to 1 via a scaling matrix S.

$$\widetilde{C}_i^{RGB_C} = S^{RGB} \cdot T^{RGB} \cdot C_i^{RGB} \quad (7)$$

$$\widetilde{C}_i^{Th_C} = S^{Th} \cdot T^{Th} \cdot C_i^{Th} \quad (8)$$

The normalized points can now be used to assemble an A matrix that will be used to solve a homogeneous linear least squares problem, $Ax = 0$.

$$g(ImagePairs) \rightarrow A \quad (9)$$

After assembling the complete A matrix, singular value decomposition can be used to calculate the homography from normalized RGB image AprilTag coordinates to normalized thermal image AprilTag coordinates.

$$A = U\Sigma V^T \quad (10)$$

The rightmost column of V can be reshaped into a 3 x 3 $\widetilde{H}$ matrix that is the homography matrix between normalized coordinates. In order to establish a baseline level of error for this algorithm, the tag coordinates in the RGB image were transformed into tag coordinates in the thermal image using the following algorithm:

$$C_i^{Th,RGB} = T^{Th^{-1}} \cdot S^{Th^{-1}} \cdot \widetilde{H} \cdot S^{RGB} \cdot T^{RGB} \cdot C_i^{RGB} \quad (11)$$

This could also be represented by the general homography from RGB to thermal images:

$$H = T^{Th^{-1}} \cdot S^{Th^{-1}} \cdot \widetilde{H} \cdot S^{RGB} \cdot T^{RGB} \qquad (12)$$

The X, Y, and overall reprojection error between the thermal image and the transformed RGB image was calculated and compared. The thermal images were taken as the ground truth. RGB images needed to be resized smaller in order for the AprilTags to be detected. This analysis was carried out on RGB images that were resized to be 90% to 10% of the original size in 5% increments. This was done to maximize the number of AprilTags detected and increase the accuracy of the homography between the image pairs. Reprojection error results are summarized in Figure 7.
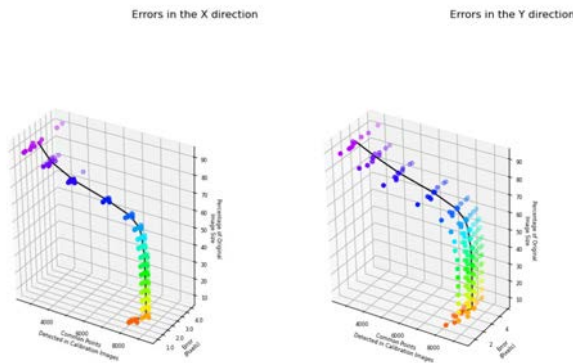

Figure 7. Image Registration Error for homographies between RGB and Thermal images calculated at various RGB image sizes

This general homography could now be applied to overlap an RGB image and a thermal image given a set of corresponding points seen in both images. However, thermal and RGB images often don't share enough features to allow for making correspondences. The differently sized thermal and RGB images meant that a different general homography would be needed to transform RGB images to thermal images for data fusion. An overall flow chart of the process is seen in Figure 8

4.2.2 Homography Calculation

Although a different general homography was required, the normalized homography between all images remained the same. The field of view of the RGB camera was larger than the thermal camera and so there was a portion of the RGB image that would consistently cross over the portion
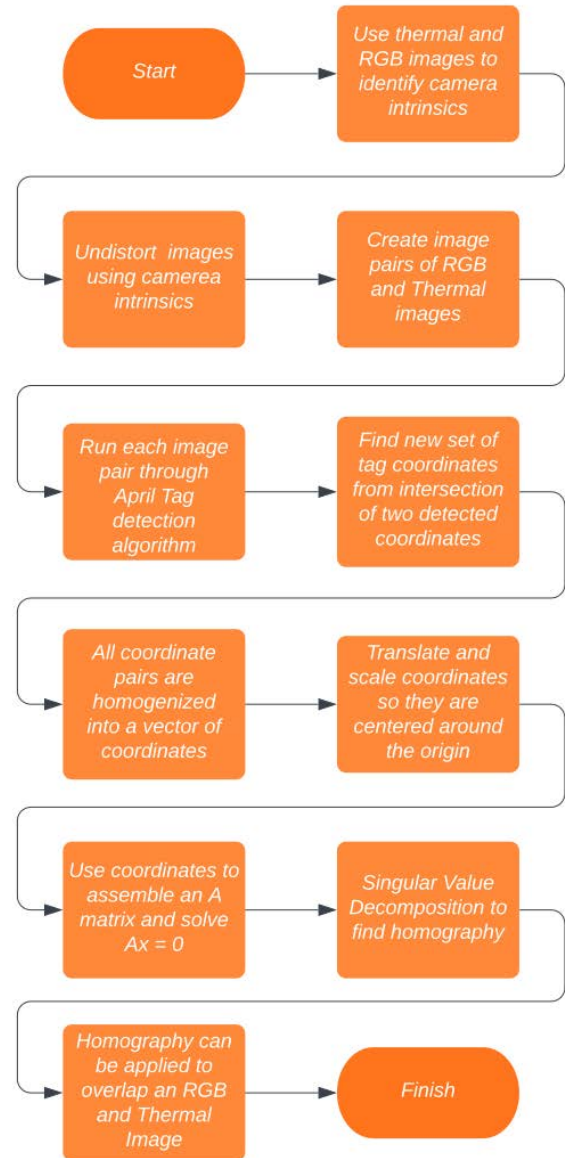

Figure 8. Homography Process Flow Chart

of the thermal image. To calculate the general homography for images between the two cameras, new normalization transforms were calculated. These transforms would not normalize AprilTags corners and centers or some other features observable within both images in an image pair. These would require the following general normalization transforms:

$T_G^{Th} \rightarrow$ General centralizing normalization matrix for thermal images

$T_G^{RGB} \rightarrow$ General centralizing normalization matrix for RGB images

$S_G^{Th} \rightarrow$ General scaling normalization matrix for thermal images

$S_G^{RGB} \rightarrow$ General scaling normalization matrix for RGB images

These thermal image transformations would normalize the thermal image corners and the corresponding points in the RGB image. The normalization points in the thermal image are defined to be based on the image size.

$$P_{N\ G}^{Th} = \begin{bmatrix} 0 & 0 \\ 0 & Height(Th_i) \\ Width(Th_i) & 0 \\ Width(Th_i) & Height(Th_i) \end{bmatrix} \quad (13)$$

However, the normalization points within the RGB image are not precisely known. T_G^{Th}, T_G^{RGB}, and S_G^{Th} can be directly calculated from the size of the RGB and thermal image by using the RGB and thermal image corners. We do not know the precise corresponding normalization points within the RGB image, so we cannot directly calculate the S_G^{RGB}. Instead we must use a Hadamard quotient to relate the S_G^{RGB} and S_G^{Th}. In the previous normalization step, the RGB and thermal images in each image pair were normalized by the scaling normalization matrices S_i^{RGB} and S_i^{Th} that were unique to the images in each image pair. The Hadamard quotient, or the element-wise division, of these two matrices results in a relative scaling matrix that is related to the level of normalization scaling performed for the RGB and thermal images [10].

$$S_i^{RGB} \oslash S_i^{Th} = S_i^{HMD} \quad (14)$$

Given a set of image pairs with corresponding feature points, we could extract a set of scaling normalization matrices for each image pair and this would allow us to obtain a set of Hadamard quotients between the scaling normalization matrices.

$$\{P_i^{RGB}, P_i^{Th}\} \rightarrow \{S_i^{RGB}, S_i^{Th}\} \rightarrow \{S_i^{HMD}\} \quad (15)$$

The set of Hadamard quotients could be averaged to produce a general Hadamard quotient.

$$S_G^{HMD} = \frac{1}{n} \sum_{i=1}^{n} S_i^{HMD} \quad (16)$$

The general Hadamard quotient and the general scaling normalization matrix for thermal images could be used to generate a pseudo general scaling normalization matrix for RGB images.

$$S_G'^{RGB} = S_G^{HMD} \odot S_G^{Th} \quad (17)$$

With $S_G'^{RGB}$, a new general homography could be calculated to transform all RGB images to the thermal image space so that they could be merged into 4 channel RGBT tensors.

$$H_G = T_G^{Th^{-1}} \cdot S_G^{Th^{-1}} \cdot \widetilde{H} \cdot S_G'^{RGB} \cdot T_G^{RGB} \quad (18)$$

This general transformation homography was applied to all the RGB images in the set of RGB and thermal image pairs to produce a new set of warped RGB image and thermal image pairs.

$$\{H_G(P_i^{RGB}, P_i^{Th})\} \quad (19)$$

An AprilTag detection algorithm was used on every image in the new set of image pairs to detect AprilTag corners and center pixel locations. The detected points in the thermal images were used as the ground truth to assess the detected points in the RGB images. Figure 9 shows the pixel errors seen for all detected points in the image pair dataset.
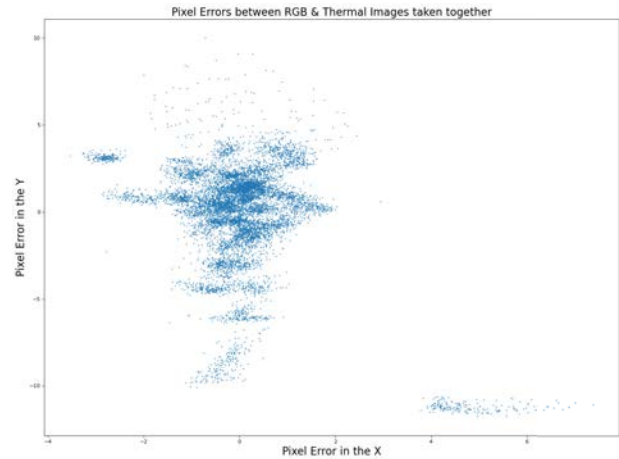

Figure 9. Image Registration Errors between Simultaneously Captured RGB and Thermal Images

5 Discussion

5.1 AprilTag Detection after Image Registration

As the images were downsized from 100% to 80% of their original size, baseline registration accuracy improved until 5500 feature point pairs were available. The experiment revealed that downsizing the RGB image to 50% of its original size was ideal to detect the maximum number of AprilTag points – just under 8000 feature point pairs. Although this didn't improve accuracy, it reduced the impact of outliers on $\widetilde{H}$. This study also established a baseline error rate for the method. Sky-facing plates with

AprilTags and vinyl can be used to create common feature points in RGB and thermal images of the same scene. Metal plates can be placed around a building to improve
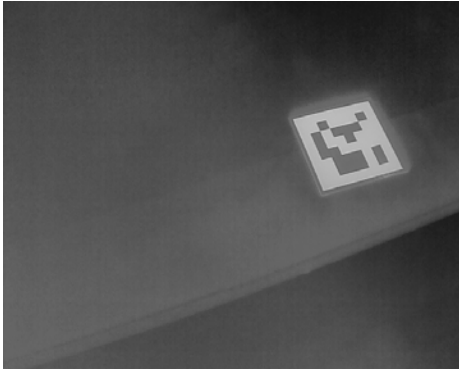

Figure 10. Thermally observed AprilTag marker that could be used to provide common feature points in RGB and thermal images

(a) Poor image registration

(b) Fair image registration

(c) Good image registration

Figure 11. Image registration results showing examples of poor, medium, and fair image registration

the algorithm's image registration if the general homography transform yields subpar results. This is depicted in Figure 10.

This study focuses on calibrating and aligning individual RGB and thermal images from a DJI Mavic 2 Enterprise Advanced drone. Processing and analyzing thermal video footage posed difficulties due to lost frames, loss of information, and blur caused by drone movement. To obtain accurate calibration and alignment, thermal footage needs significant pre-processing, making video analysis impractical for this study. As a result, thermal camera calibration with video clips is not within the scope of this paper.

5.2 Homography Calculation

This experiment showed that thermal and RGB images could be aligned despite differences in distortion and field of view between cameras.

One of the major limitations of how well the RGB and thermal images can be aligned is the relative closeness of the scene to the camera taking the image. Figure 11 shows how the overlap between RGB and thermal images improves as the camera is moved further away from the scene. The nature of building envelope photography is such that drones will always be a large distance away from the scene they are surveying and as a result will not suffer from issues of poor overlap.

This is because as the distance between the 3D location of the AprilTags and the cameras, $\left(X^{Th}, X^{RGB}\right)$, increases, the ratio between the distance between the 3D location of the AprilTags and the cameras and the baseline distance between the two cameras gets progressively
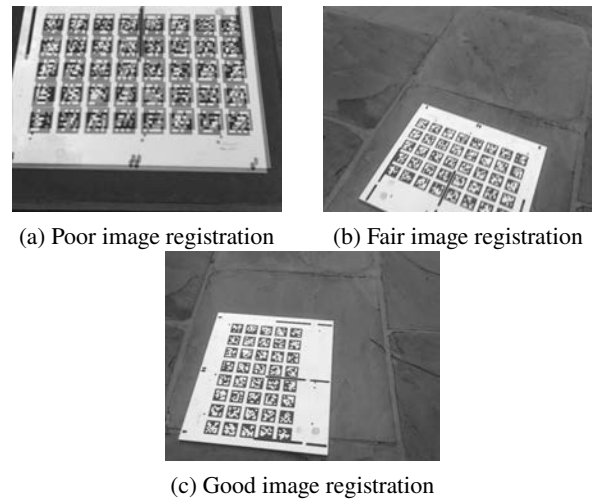
smaller and begins to approach 0. Therefore the angle between the two cameras and the world point changes less as the depth of the world point increases. Figure 12 shows how the 3D to 2D projection of a world point onto the camera image plane changes significantly based on the distance of the world point to the cameras and the baseline distance between the cameras.
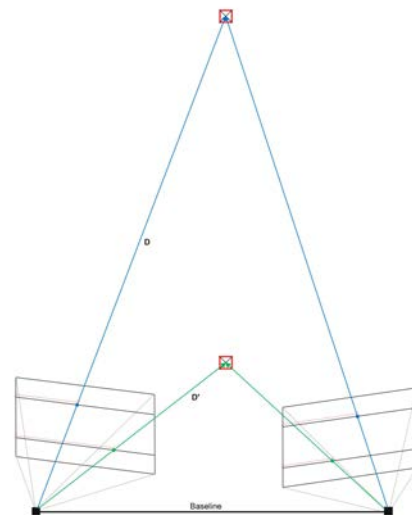

Figure 12. The effect of near-depth objects on image registration via the homography approximation method.

Another limitation is that images taken too closely to-

gether have an influence on the alignment homography and this could result in a less-than-optimal alignment homography. This could be solved through an iterative approach that discards any images from the homography calculation that are far outside the general cluster of errors, this can be visually seen in Figure 9. The image registration algorithm could be further improved by having a better S_G^{HMD} estimation. This estimation can be improved by using RANSAC to filter out images that required an S_G^{HMD} that was significantly different from other images.

6   Conclusion

Using a calibration board made of machine-cut vinyl overlaid onto a metal plate creates a robust calibration rig. Calibration images of this rig aren't affected by spatial offsets and when placed in an open field during the daytime, it does not require active heating. Thermal cameras can be calibrated robustly using vinyl AprilTags and metal plates.

The alignment algorithm is an approximation based on the assumption that the two cameras are positioned relatively close and the image that they are viewing is relatively far away, which is accurate for UAV photography of buildings and building envelopes.

In general, the image registration algorithm had a pixel error of ±5 pixels in the y direction and ±2 pixels in the x direction. These images are fairly well aligned and this algorithm can be used to create custom thermal and RGB image datasets of building envelopes.

References

[1] W. Ursine, F. Calado, G. Teixeira, H. Diniz, S. Silvino, and R. De Andrade. Thermal / Visible Autonomous Stereo Visio System Calibration Methodology for Non-controlled Environments. In *Proceedings of the 2012 International Conference on Quantitative InfraRed Thermography*. QIRT Council. doi:10.21611/qirt.2012.261. URL http://qirt.org/archives/qirt2012doi/papers/QIRT-2012-261.pdf.

[2] Takashi Shibata, Masayuki Tanaka, and Masatoshi Okutomi. Accurate joint geometric camera calibration of visible and far-infrared cameras. 2017(11): 7–13.

[3] Yu Hou, Rebekka Volk, Meida Chen, and Lucio Soibelman. Fusing tie points' RGB and thermal information for mapping large areas based on aerial images: A study of fusion performance under different flight configurations and experimental conditions. 124:103554.

[4] Philip Saponaro, Scott Sorensen, Stephen Rhein, and Chandra Kambhamettu. Improving calibration of thermal stereo cameras using heated calibration board. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 4718–4722, 2015. doi:10.1109/ICIP.2015.7351702.

[5] Michael Gschwandtner, Roland Kwitt, Andreas Uhl, and Wolfgang Pree. Infrared camera calibration for dense depth map construction. In *2011 IEEE Intelligent Vehicles Symposium (IV)*, pages 857–862, 2011. doi:10.1109/IVS.2011.5940515.

[6] R. Istenic, D. Heric, S. Ribaric, and Damjan Zazula. Thermal and visual image registration in hough parameter space. In *2007 14th International Workshop on Systems, Signals and Image Processing and 6th EURASIP Conference Focused on Speech and Image Processing, Multimedia Communications and Services*, pages 106–109. IEEE. ISBN 961-248-036-2.

[7] V. A. Knyaz and P. V. Moshkantsev. JOINT GEOMETRIC CALIBRATION OF COLOR AND THERMAL CAMERAS FOR SYNCHRONIZED MULTIMODAL DATASET CREATING. XLII-2/W18:79–84. ISSN 2194-9034. doi:10.5194/isprs-archives-XLII-2-W18-79-2019.   URL https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLII-2-W18/79/2019/.

[8] Adam Dlesk, Karel Vach, and Karel Pavelka. Transformations in the Photogrammetric Co-Processing of Thermal Infrared Images and RGB Images.   21(15):5061, .   ISSN 1424-8220.   doi:10.3390/s21155061.   URL https://www.mdpi.com/1424-8220/21/15/5061.

[9] Adam Dlesk, Karel Vach, and Karel Pavelka. Photogrammetric Co-Processing of Thermal Infrared Images and RGB Images.   22(4):1655, .   ISSN 1424-8220. doi:10.3390/s22041655. URL https://www.mdpi.com/1424-8220/22/4/1655.

[10] Roger A Horn. The hadamard product. In *Proc. Symp. Appl. Math*, volume 40, pages 87–169, 1990.