DIRECT CO-REGISTRATION OF TIR IMAGES AND MLS POINT CLOUDS BY CORRESPONDING KEYPOINTS

J. Zhu*, Y. Xu, L. Hoegner, U. Stilla

Photogrammetry and Remote Sensing, Technical University of Munich (TUM), Germany - (jingwei.zhu, yusheng.xu, ludwig.hoegner, stilla)@tum.de

ICWG II/III: Pattern Analysis in Remote Sensing

KEY WORDS: MLS, point cloud, TIR images, co-registration, segmentation, point-based, pose estimation

ABSTRACT:

In this work, we discussed how to directly combine thermal infrared image (TIR) and the point cloud without additional assistance from GCPs or 3D models. Specifically, we propose a point-based co-registration process for combining the TIR image and the point cloud for the buildings. The keypoints are extracted from images and point clouds via primitive segmentation and corner detection, then pairs of corresponding points are identified manually. After that, the estimated camera pose can be computed with EPnP algorithm. Finally, the point cloud with thermal information provided by IR images can be generated as a result, which is helpful in the tasks such as energy inspection, leakage detection, and abnormal condition monitoring. This paper provides us more insight about the probability and ideas about the combining TIR image and point cloud.

1. INTRODUCTION

In the Big Data Era, single sensor alone can hardly provide sufficient information about the target. Therefore, different sensors are used to observe the objects. With the demand of overall understanding of the cities, combining data from multiple sensors and relate information to improve accuracy and specific inferences become a hot topic (Hall, Llinas, 1997). The advantages of geometric features of the 3D model attract researchers working on assigning 2D information to 3D data (Castanedo, 2013, Khaleghi et al., 2013), trying to enrich 3D objects with various properties. (Wang et al., 2017) combining InSAR point clouds and optical image in Urban areas, (Mastin et al., 2009) and (Chen et al., 2004) coregister optical image and Lidar point clouds for visualization, (Weinmann et al., 2014) relate range image to thermal image for object detection. Although researchers presented different methods to combine thermal infrared images with 3D data (Weinmann, 2016, Hoegner, Stilla, 2018), seldom has been done on co-registering thermal infrared image and mobile laser scanner (MLS) point clouds especially in an outdoor situation.

Thermal infrared (TIR) images, acquired by thermographic sensors, depict temperature and emission properties of objects. Different from the optical camera, thermographic sensors detect radiation in the long-infrared range of the electromagnetic spectrum. Since all the objects with a temperature above absolute zeros emit infrared radiation to the environment, thermal infrared images enable us to observe objects, moving process and the thermal properties without visible illumination (Zin et al., 2007, Weinmann et al., 2014, Christiansen et al., 2014). This property helps with thermal inspection of buildings in city areas. The traditional method uses a series of TIR images to inspect the energy usage of the building. The lack of (1) rapid and low-cost data collection and modeling method; (2)measurement for evaluating overall building performance; (3) adequate integrated intelligence for component evaluation; (4)tools for the non-expert decision maker, make it hard to provide reliable

information (Wang et al., 2013). If a thermographic 3D model can be generated by combining TIR images and 3D model, we are able to deal with tasks such as (1) building inspection; (1) energy loss detection; (2) leakage localization, and (4) other tasks such as scene segmentation and classification.

The current strategies for thermographic 3D model generation include: (1) generate 3D thermal model from images, or (2) mapping 2D thermal images to 3D model or point cloud. For the first strategy, the 3D model can be generated by reconstructing 3D scene geometry from thermal images with structure from motion techniques (Westfeld et al., 2015). However, the low resolution of thermal images limits the performance of the generated thermal 3D model. A precise model can be reconstructed by combing point cloud from RGB images and thermal images separately (Hoegner et al., 2016). In the second strategy, co-registering 2D thermal infrared images to 3D models or 3D point clouds can be done with known fixed relative orientation. For automatic method, (Hoegner, Stilla, 2015) proposed an automatic registration method to register the IR images to a given 3D building model. The result gives us an energy or temperature profile of the building. (Iwaszczuk et al., 2012) try to find the best fit between the 3D model and IR images with line segments by RANSAC process. Though using a 3D model for thermal mapping gives us an overall idea of the temperature distribution, such a simple way of expression can hardly represent an as-built design.

To reach a better detailed expression, the point cloud is favored as 3D model data. (Weinmann et al., 2014) proposed the keypoint-based co-registration via the robust matching techniques method to find the corresponding keypoints in intensity image and IR image, and co-register the IR image to range image applying RANSAC-based projection. (Lagüela, Armesto, 2012) extract the line segments from IR images and point cloud separately. With the corresponding lines, the camera position can be estimated by the RANSAC process. Though coregistration of thermal infrared images and point cloud usually reach a good result in indoor situations, seldom work has been

^{*}Corresponding author

done trying to co-register the thermal infrared image to MLS point clouds for a large scene. To fill this gap, we proposed a method to direct co-register the TIR image and the point cloud by corresponding keypoints.

In this paper, we propose a direct point-based method to combine the TIR images and point cloud. The idea is to calculate the camera pose when the photo is taken based on the assumption that the MLS point cloud provides a precise model of the study area. Therefore, the TIR image can be regarded that it was taken from the point cloud with a specific pose. The thermal information for every point in the point cloud can be assigned with the gray value of the corresponding point in the image. This work is the first paper trying to directly co-register the TIR image and point cloud for buildings with corresponding points in the outdoor scene as far as I know. This work demonstrated that the point cloud and TIR image could be combined. Besides, the generated 3D model could explain the thermal information of 3D buildings, which is helpful for thermal interpretation.

This paper is organized in the following way: the first part introduces the background and state-of-art methods for 3D thermographic reconstruction. In the second part, we propose a direct co-registration method based on corresponding keypoints. Then, the results of our experiments based on the method in the second part will be displayed. Finally, we will draw a conclusion based on the results, and discuss the outlook for the topic.

2. METHOD

The proposed process is based on the assumption that the thermal infrared images and the point cloud are generated at the same scene at the same time. This condition ensured that the data contains the information from the same scene, which is the precondition for the matching process. The whole process for the co-registration contains three steps (Figure 1).



Figure 1. Figure placement and numbering

At first, the image and the point cloud are processed separately. Geometric calibration of TIR images is done based on the coordinates of control points and image point. The preprocessing of the point clouds targets to filter the noise, downsampling, filter the irregular objects and crop the scene. Second, keypoints are extracted from TIR images and point clouds, and the corresponding keypoint pairs are selected manually. Based on the selected point pairs, the camera position can be computed by solving PnP (Perspective n point) problem. Finally, a thermographic 3D point cloud can be generated by indirect computing the gray value of corresponding points in the image.

2.1 Proprocessing of TIR images and the point cloud

Camera calibration usually includes geometric calibration and radiometric calibration. Due to the lack of accessibility to the device, we are only able to conduct the geometric calibration by using measured control points and corresponding image points following the photogrammetric calibration method (Luhmann et al., 2013). When the coordinate of control points (X_n, Y_n, Z_n) (in-situ measurements) and their corresponding image coordinates (x_n, y_n) are known, the intrinsic parameters for the camera can be estimated by collinearity equation (1).

$$\begin{aligned} x_i - x_0 &= c_x \cdot \frac{a_{11}(X_i - X_0) + a_{21}(Y_i - Y_0) + a_{31}(Z_i - Z_0)}{a_{13}(X_i - X_0) + a_{23}(Y_i - Y_0) + a_{33}(Z_i - Z_0)} + d_x \\ y_i - y_0 &= c_y \cdot \frac{a_{12}(X_i - X_0) + a_{22}(Y_i - Y_0) + a_{32}(Z_i - Z_0)}{a_{13}(X_i - X_0) + a_{23}(Y_i - Y_0) + a_{33}(Z_i - Z_0)} + d_y(1) \end{aligned}$$

where (X_0, Y_0, Z_0) is the camera position, a_{11} to a_{33} are the coefficients of a 3×3 rotation matrix, and c, x_0, y_0 are the parameters of geometric camera calibration.

The preprocessing of point clouds targets to downsampling the data and filter the noise compared to the given TIR image. Since the raw data contains noise, the wrong points with missing information and isolated points are removed. Then a voxel grid filter is used to reduce the volume of data while keeping the geometric features of the objects. These two steps are common operations in point cloud processing. After that, we are going to filter the unrelated objects and crop the scene regarding the contents in the image.

The precondition for co-registration is that both data contains corresponding components, such as points. Eliminating irrelevant objects, and cropping the scene based on the TIR images information are important in point cloud preprocessing. Irregular objects such as vegetation, road marks, and pedestrians usually cause the problem of nonsense keypoints. A large number of feature points could be extracted around the vegetation, while it is hard to find corresponding feature points in the image. Considering the complexity of the scene and the fact that the buildings are our target, we would like to apply a segmentation process and filter the irrelevant segments based on the assumptions that the buildings can be decomposed to walls abstracted by planes. The short wall which is perpendicular to the front facade was also divided into several segments. The Euclidean cluster extraction (Rusu, 2010) are first applied to filter the small objects in the scene, and then a plane extraction by random sample consensus method (RANSAC) (Fischler, Bolles, 1981) was applied to all the objects. After candidate plane was detected, a Euclidean cluster extraction was applied to filter the remaining small objects.

After removing the irrelevant objects, we would like to crop the scene according to the information in the target TIR image. The point cloud contains all the objects in the scene, while the camera property restricts contents in TIR images. This step is done under the assumption that GPS on the mobile platform records the position when the image was taken, and the thermal camera is mounted on the same platform with a fixed observing orientation. A box with a fixed length of the edge is used as the constraint for point cloud cropping.

2.2 Keypoints extraction and point pairs generation

This step targets to extract keypoints from the point cloud and TIR image, and manually find the corresponding point pairs for camera pose estimation. Co-registration targets to spatial align data such as images or point cloud, and point-based method is the most commonly used strategy. Though this method usually requires higher computational expense, it has the potential to achieve the greatest performance (Liggins II et al., 2017). To enable explicit tracking of features of data to be performed, the extracted keypoints or corners need to be discrete, reliable and meaningful (Charnley, Blissett, 1989). Therefore the corner points of buildings, windows, doors or similar location are ideal keypoints candidates. Keypoints in 2D images and 3D point clouds can be detected separately with current detector, such as Moravec (Moravec, Elfes, 1985), SIFT (Lowe et al., 1999) for 2D images, and Harris3D (Sipiran, Bustos, 2011) for point cloud. Considering the correspondence requirements of 2D and 3D data, detectors with a similar principle would be preferred. Therefore we choose Harris' corner detector (Harris, Stephens, 1988) for the image, and Harris 3D detection (Sipiran, Bustos, 2011) for the point cloud. Note that though SIFT can extract high quality rotation-invariant and transformationinvariant corner points for optical images, SIFT3D detector can hardly detect sufficient keypoints in the point cloud. Besides, among the adapted 3D keypoints detector, Harris 3D proves to be more robust with the evidence that detected points by Harris 3D is less likely to be appearing in other location in the same image (Loog, Lauze, 2010).

After keypoints extraction, corresponding point pairs from TIR images and point clouds are required for the further process. The point pairs are selected manually in our method.

With the corresponding point pairs between thermal IR images and the point clouds, we can compute the camera pose by solving the PnP problem. PnP is the problem of estimating the pose of a calibrated camera given a set of n 3D points and their corresponding 2D projections in the image. The camera pose consists of 6 degrees-of-freedom which are made up of the three rotation variables and 3D translation variables of the camera for the 3D coordinate system. This problem originates from camera calibration and has many applications in computer vision and other areas, including 3D pose estimation, robotics and augmented reality.

Similar to the optical image, thermal IR images are generated by a camera with the perspective projection following Equation 2:

$$u_i = K[R|t]X_i \tag{2}$$

In the equation above, u_i is the coordinate of point in the image, K is the camera intrinsic matrix, R and t describe the rotation and transformation of the virtual camera for local laser scanner coordinate system. X_i is the coordinate of corresponding 3D points in the world coordinate space.

After the image calibration, we got the intrinsic parameters of the thermal IR camera and distortion coefficients. Our task is to get the rotation matrix R and the transformation matrix tfor each IR image. A Efficient PnP algorithm (Lepetit et al., 2009) is adapted. The idea of Efficient PnP method is first to express the n 3D points with four virtual control points, and then compute the camera pose by estimate the coordinates of these control points in the camera referential. At first, each 3D point can be expressed by a linear combination of four spatial points which are not on the same plane. Suppose that we have four control points in the world coordinates namely c_j , j = 1, ..., 4, so any known reference points P_i in the world system can be expressed by:

$$p_i^w = \sum_{j=1}^4 \alpha_{ij} c_j^w, with \sum_{j=1}^4 \alpha_{ij} = 1$$
(3)

Here, α_{ij} are the homogeneous barycentric coordinates which are uniquely defined for different spatial points. Note that the reference point in the camera coordinate system can be expressed as $p_i^c \sum_{j=1}^4 \alpha_{ij} c_j^c$.

In the camera coordinate system, we have:

$$w_i \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = K p_i^c = K \sum_{j=1}^4 \alpha_{i,j} c_j^c \tag{4}$$

where the w_i are scalar projective parameters, $p_{i_i} = 1, ..., n$ are the corresponding image points of reference points. The unknown parameters are the twelve control point coordinates $(X_j^c, Y_j^c, Z_j^c)_{j=1,...,4}$ and n projective parameters $w_{i_{i=1,..,n}}$.

Extending the camera matrix K and control points, a linear system can be formed.

$$Mu = 0 \tag{5}$$

where $u = [(c_1^c)^T, (c_2^c)^T, (c_3^c)^T, (c_4^c)^T]^T$ is a 12-vector of unknowns, and M is a $2n \times 12$ matrix.

Based on the form of Equation5, the solution belongs to the null space of M, and can be expressed as:

$$u = \sum_{i=1}^{N} \beta_i v_i \tag{6}$$

where the set v_i are the columns of the right-singular vector of M corresponding to the N null singular values of M. Considering all the situations, the value of N can range from 1 to 4. Based on the assumption that the Euclidean distance between control points must be constant no matter what coordinate system is used, a small number of quadratic equations can be generated. The strategy to pick the best solution is to compute all the four situations and keep the one that yields the smallest reprojection error (with Equation 7).

$$err = \sum_{i} dist^{2} (K \begin{bmatrix} R|t \end{bmatrix} \begin{bmatrix} p_{i}^{w} \\ 1 \end{bmatrix}, u_{i})$$
(7)

where $dist(\tilde{m}, n)$ is the 2D distance between homogeneous coordinates of point m and point n. With this step, the rotation matrix and transformation will be estimated.

2.3 3D model generation

Thermal 3D model generation is the last step in the whole process, which targets to reconstruct the 3D point cloud with thermal information. Since the camera pose was estimated in the previous step, we can get the thermal information of points in the point cloud by project them to the 2D image plane and find the corresponding image points intensity. Considering the generated image coordinates are not usually located in the pixel grid, a bilinear interpolation was adapted.

3. EXPERIMENTS

The Campus study area is the TUM City $(48.1493^{\circ}N, 11.5685^{\circ}E)$ located in the center of Munich. The MLS laser scanners and the thermal camera are the sensors we need for data acquisition. The sensors are mounted on the MODISSA(Mobile Distributed Situation Awareness (Borgmann et al., 2018)) platform (in Figure 2^1) together with a GPS to provide the location information of the vehicle. The sensor system for MLS point clouds observation composes of two Velodyne HDL-64E. It has 360° Horizontal FOV, 26.9° vertical FOV, and very high data rate. The angular resolution reaches 0.08° in azimuth and -0.4° in the vertical direction. All the objects within 120m range can be observed with up to 2.2 million points per second. The measurements have been georeferenced using post-processed data of an inertial navigation system. The infrared image sequences were acquired with an uncooled microbolometer Jenoptik IR-TCM 640 with a field of view of $65.2^{\circ} \times 51.3^{\circ}$, which was mounted crosswise to the driving direction. The images are provided as 16 bit-TIFFs with lossless compression (LZW) with the size of 640 pixel $\times 480$ pixel. Besides, a file provides additional information about the car position when a specific image is taken.



Figure 2. The MODISSA platform from Fraunhofer IOSB.

3.1 Proprocessing of TIR image and point cloud

The calibration of TIR images required coordinates of control points and corresponding image point coordinates. The control points are measured in-situ with a total station. Then we recorded the coordinates of corresponding image points in the IR image. The principle point is located at (257.78, 246.37). Besides, the radial distortion coefficients are $k_1 = 0.206, k_2 = -0.885$) and tangential distortion coefficients are $p_1 = -0.007, p_2 = -0.006$). The result indicates that Barrel radial distortion dominate in our TIR images.

After computation of intrinsic parameters and distortion parameters, we rectified the images. For example, Figure 3(a) is the corresponding rectified image of Figure 3(b). We found that the quality of images is significantly improved since the curvature of lines, especially the boundary of the roof area is getting more straight. However, the distortions are not fully rectified in some areas such as a downright corner. It is caused by (1) the sparse control points in the corner due to the occlusion by tree leaves, and (2) the difficulty to find the sharp corner points in the TIR image.

Figure 4 shows the original point clouds of target study areas. Geometric information of the scene is precisely described with



Figure 3. Image calibration. (a) Example of orginal image. (b) Corresponding rectified image.

a large amount of 3D points. Besides, many trees are in the scene between buildings in South East direction. The preprocessing of point clouds targets to reduce the volume of the dataset and filter the irrelevant objects. Comparing with the original point clouds (Figure 4), the point clouds after segmentation deleted the crown of trees and most of the building walls and the ground remains. After the first Euclidean cluster segmentation, the vegetation with less dense crowns is removed. Then, by applying a plane segmentation, the ground and building walls are detected. After the final Euclidean cluster segmentation, most of the vegetation is removed. This process reduces the possible wrong keypoints on the tree crowns. Since the TIR image recorded by the camera with certain FOV and the device is fixed on the vehicle with a certain orientation, we crop the point clouds based on the car position information at last. The size of the cropping box was fixed with proper spatial volume to include all the related objects. Figure 5 shows an example of a reduced point cloud. The result clearly shows the geometry of building from the whole scene.



Figure 4. Original Point clouds

3.2 Keypoints extraction

The keypoints extraction targets to detect feature points in the image and point cloud as candidate points for the following coregistration. Figure 6 shows the detected keypoints in TIR images, and most of the corners are detected. However, we found that the keypoints are not isolated but clustered together. The features in thermal infrared images are usually different from those in the optical images. Intensity images in the visual domain typically provide sharp contour with abrupt changes of properties. Thermal infrared images which record thermal radiation of entities in the infrared spectrum rises the problem of low geometric resolution and blurry features especially in lines or contours. The temperature of observed objects and materials with different emission properties are the cause for different looks. Note that, even two objects with the same temperature may appear distinctive in thermal infrared images due to materials with various emissivity coefficients (Weinmann, 2016).

¹© Image Copyright 2019, Fraunhofer IOSB



Figure 5. Point clouds after segmentation. (a) Result after the first Euclidean cluster extraction. (b) Segmentation after plane segmentation by Sample Consensus method. (c) Final result after segmentation. (d) An example of scene cropping.

Temperature, the signal of energy distribution, usually distributes continuously. The phenomena gives a blurred boundary of objects in the images. Besides, the tracks of objects with high temperature can be visible in the form of energy in thermal images, which could be noise. The differences in thermal infrared images and optical images make it difficult to extract keypoint with image operators.

In order to find the keypoints which are more close to the corner of the building, we tried different parameters settings. Figure 6 and Table 1 are the results of 2D Harris' corner detection, which includes window size, operator size, and free parameter k. We found that when operator size is smaller than 9, seldom

keypoints will be extracted. We fix the operator size to 9 and focus on the window size. In Table 1, we set the window size as 3 pixels, 5 pixels, and 7 pixels. When the window size is getting larger, the detected keypoints increased. Considering The blurred boundaries of the objects in TIR images, a larger window size is required to find the gradient difference.



Figure 6. Detected harris keypoints in TIR image. (a) Original image (b) keypoints for test 1. (c) keypoints in test 2. (d) keypoints in test 3.

Nr.	window size	operator size	k	point number
1	3	9	0.05	2911
2	5	9	0.05	8090
3	7	9	0.05	17421

Table 1. Parameter settings for keypoint detection in IR images.

Figure 7 shows the detected keypoints in the point cloud. Most of the keypoints are located on the corner of the window frame or along the edge. However, some keypoints are detected on the wall. The cause of keypoints along the window edge is the uneven distribution of points on the boundary. The noise on the wall is due to the missing points or uneven point density.

Table 2 shows different parameter settings for 3D Harris point detector and Figure 7 are figures of the corresponding points distribution. The radius is the spatial constraint for the neighboring points and threshold is the value to filter the corner from non-corner points by the response. The radius constraint the considered local area for each point. Considering the voxel size is 0.1 for the point clouds, the radius should be larger than this value. In Table 2, we found that when radius decrease from 0.3 to 0.15, fewer points are detected from Figure 7(a) to 7(c). However, when the radius r is set to be too small, as in Figure 7(c), the detected points usually include sort of noise(lots of wrongly detected keypoint on the wall) due to lack of enough neighbor points. Comparing test 2 and test 4, when threshold θ is getting larger, fewer keypoints are detected. In order to have enough keypoints along the boundary or in the corner (best to be in the corner), we set the parameters as $r = 0.2, \theta = 0.01$.

Comparing the keypoints detected in the TIR image (Figure 6) and the point cloud (Figure 7), we found that the detected keypoints which seems related are not on the same location: (1) The image keypoints are located on both side of the window.



Figure 7. Harris keypoints in the point cloud. (a) keypoints in test 1. (b) keypoints in test 2. (c) keypoints in test 3. (d) keypoints in test 4.

Nr.	radius(r)	threshold(θ)	point number
1	0.3	0.01	86
2	0.2	0.01	204
3	0.15	0.01	666
4	0.2	0.02	60

Table 2. Parameter setting for Harris 3D keypoints.

In the point cloud, most of them appear on the left side of the window. (2) The keypoints in the image are in the inner frame of the window surrounding the glasses, while the keypoints on the point clouds close to the wall corner, the outer frame of the window. Therefore, we need to be careful when we select the corresponding keypoints.

3.3 Camera pose estimation and 3D model reconstruction

After keypoints extraction, we could select corresponding point pairs for pose estimation. Based on the requirements, at least four point pairs are needed. The distribution of points is important. It's important not to have all the keypoints on the same plane. Figure 8 shows a range image generated by the point cloud with the estimated pose. Comparing with the corresponding rectified images, the range image record the building from similar observing orientation. The difference is due to the missing data and inaccurate distortion coefficients.

When the camera pose was computed, we can generate a 3D point cloud with thermal information. Figure 9 shows the 3D thermographic point cloud in false color and the referenced corresponding TIR image. The constructed model shows a similar thermal distribution of building facade, and represent it in a 3D model. By checking the window frame and the geometric features of the point cloud, the data matches. From the constructed point cloud, we can easily tell the temperature/emission distribution of the building and find some clues for the possible thermal phenomenon. For example, there is a yellow area under the small window on the right down side of the wall. Besides, all the windows on the left sides are relative lighter green, which indicates a higher temperature compare to the wall. If a sequence of TIR images is applied, we will be able to generate a 3D thermographic point cloud recording the thermal information of the building in all aspects, which could be helpful in thermal distribution representation, energy inspection, leakage detection, and even monitoring the temperature changes.





3.4 Verification

Since we have no as-built model or precise pose information to check the accuracy of the result, we re-project the convex hull points extracted from the wall plane segments (depict the contour of building boundary) to the corresponding image. The result is given in Figure 10. The locations for walls and windows are almost the same in general (with some shift) by comparing the outline of windows and doors, but not perfect match in all the areas especially on the edge of the image. We find that the closer to the image center, the better it matches. The possible reason mainly lies in the estimation of distortion parameters, and probability form the error in pose estimation.

4. DISCUSSION AND OUTLOOK

In this paper, we proposed a process that directly co-registrate TIR image and the point cloud with corresponding keypoints. A 3D point cloud with thermal information is generated, which could help with energy inspection, thermal information interpretation, and leakage localization. Our work demonstrates that it is possible to direct combining TIR image and the point cloud without additional information such as 3D model or optical images. Considering the whole process, several aspects may influence the co-registration result. The first is the difficulty in selecting isolated and reliable key-points in TIR images. Due to the blurry property of the TIR image, the detected key-points



(b)

Figure 9. 3D thermographic model. (a) Generated model with false color. (b) Corresponding rectified image with false color (blue-low temperature, red-high temperature).



Figure 10. Varification

are clustered around the corner which is hard to manually located the precise location. The inaccurate coordinate of keypoints would introduce errors in corresponding point pair selection, which cause impact on the accuracy of the estimated camera pose. The second aspect is the precision of camera calibration. The intrinsic parameters of the camera and the distortion coefficients for image rectification are calculated by camera calibration, which requires equally distributed image points and corresponding control points. Due to the limited field of view by the thermal camera, no sufficient control points could be observed in the image especially in the corner. The estimated camera pose will be influences by imprecise point coordinates from rectified images. Another aspect is the density of point clouds. Since point clouds are acquired by the Mobile laser scanner, which can be regarded as a discrete sampling of the scene, it is not able to record the coordinates of all the corners in the scene. Therefore, we could only find the approximated location of the corner points. The inaccurate coordinates of corresponding key-points will lead to lousy pose and influence the final result.

This work can be regarded as a test of combining TIR images and the MLS point cloud, and several problems could be studied in future work, such as automatic combination, improvement of the model accuracy. This paper not only demonstrates the possibility of direct combining TIR images and the point cloud but also provide us some ideas for further researches in the following aspects: (1) Precisely locate the keypoints in TIR images;(2) Other ways to extract keypoints in the point cloud and TIR image; (3) Automatically find the corresponding keypoints in the point cloud and TIR images; (4) Considering the point-based method is limited by the selection of keypoints in the data, is there any other method for data combination, such as object-based method; (5) The application of thermographic 3D point cloud in energy inspection and object tracking.

REFERENCES

Borgmann, B., Schatz, V., Kieritz, H., Scherer-Klöckling, C., Hebel, M., Arens, M., 2018. DATA PROCESSING AND RECORDING USING A VERSATILE MULTI-SENSOR VEHICLE. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-1, 21–28.

Castanedo, F., 2013. A review of data fusion techniques. *The Scientific World Journal*, 2013.

Charnley, D., Blissett, R., 1989. Surface reconstruction from outdoor image sequences. *Image and Vision Computing*, 7(1), 10–16.

Chen, L.-C., Teo, T.-A., Shao, Y.-C., Lai, Y.-C., Rau, J.-Y., 2004. Fusion of LIDAR data and optical imagery for building modeling. *International Archives of Photogrammetry and Remote Sensing*, 35(B4), 732–737.

Christiansen, P., Steen, K. A., Jørgensen, R. N., Karstoft, H., 2014. Automated detection and recognition of wildlife using thermal cameras. *Sensors (Switzerland)*, 14(8), 13778–13793.

Fischler, M. A., Bolles, R. C., 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Commun. ACM*, 24(6), 381–395.

Hall, D. L., Llinas, J., 1997. An introduction to multisensor data fusion. *Proceedings of the IEEE*, 85(1), 6–23.

Harris, C., Stephens, M., 1988. A combined corner and edge detector. *Alvey vision conference*, 15, Citeseer, 10–5244.

Hoegner, L., Stilla, U., 2015. Building facade object detection from terrestrial thermal infrared image sequences combining different views. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2(3), 55. Hoegner, L., Stilla, U., 2018. Mobile thermal mapping for matching of infrared images with 3D building models and 3D point clouds. *Quantitative InfraRed Thermography Journal*, 0(0), 1–19.

Hoegner, L., Tuttas, S., Stilla, U., 2016. 3D building reconstruction and construction site monitoring from RGB and TIR image sets. 2016 12th IEEE International Symposium on Electronics and Telecommunications (ISETC), 305–308.

Iwaszczuk, D., Hoegner, L., Schmitt, M., Stilla, U., 2012. Line based Matching of Uncertain 3D Building Models with IR Image Sequences for Precise Texture Extraction. *Photogrammetrie - Fernerkundung - Geoinformation*, 2012(5), 511–521.

Khaleghi, B., Khamis, A., Karray, F. O., Razavi, S. N., 2013. Multisensor data fusion: A review of the state-of-the-art. *Information Fusion*, 14(1), 28–44.

Lagüela, S., Armesto, J., 2012. Automatic procedure for the registration of thermographic images with point clouds. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (IAPRS)*, XXXIX(B5), 211–216.

Lepetit, V., Moreno-Noguer, F., Fua, P., 2009. Epnp: An accurate o (n) solution to the pnp problem. *International journal of computer vision*, 81(2), 155.

Liggins II, M., Hall, D., Llinas, J., 2017. Handbook of multisensor data fusion: theory and practice. CRC press.

Loog, M., Lauze, F., 2010. The Improbability of Harris Interest Points. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(6), 1141-1147.

Lowe, D. G. et al., 1999. Object recognition from local scaleinvariant features. *iccv*, 99, 1150–1157.

Luhmann, T., Piechel, J., Roelfs, T., 2013. Geometric calibration of thermographic cameras. *Thermal Infrared Remote Sensing*, Springer, 27–42.

Mastin, A., Kepner, J., Fisher, J., 2009. Automatic registration of LIDAR and optical images of urban scenes. 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009, 2009 IEEE, 2639–2646.

Moravec, H., Elfes, A., 1985. High resolution maps from wide angle sonar. *Proceedings. 1985 IEEE international conference on robotics and automation*, 2, IEEE, 116–121.

Rusu, R. B., 2010. Semantic 3d object maps for everyday manipulation in human living environments. *KI-Künstliche Intelligenz*, 24(4), 345–348.

Sipiran, I., Bustos, B., 2011. Harris 3D: a robust extension of the Harris operator for interest point detection on 3D meshes. *The Visual Computer*, 27(11), 963.

Wang, C., Cho, Y. K., Gai, M., 2013. As-Is 3D Thermal Modeling for Existing Building Envelopes Using a Hybrid LIDAR System. *Journal of Computing in Civil Engineering*, 27(6), 645–656.

Wang, Y., Zhu, X. X., Zeisl, B., Pollefeys, M., 2017. Fusing meter-resolution 4-D InSAR point clouds and optical images for semantic urban infrastructure monitoring. *IEEE Transactions on Geoscience and Remote Sensing*, 55(1), 14–26.

Weinmann, M., 2016. Co-Registration of 2D Imagery and 3D Point Cloud Data. *Reconstruction and Analysis of 3D Scenes*, Springer International Publishing, 111–140.

Weinmann, M., Leitloff, J., Hoegner, L., Jutzi, B., Stilla, U., Hinz, S., 2014. Thermal 3D mapping for object detection in dynamic scenes. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-1(November), 53–60.

Westfeld, P., Mader, D., Maas, H.-G., 2015. Generation of tir-attributed 3d point clouds from uav-based thermal imagery. *Photogrammetrie-Fernerkundung-Geoinformation*, 2015(5), 381–393.

Zin, T. T., Takahashi, H., Toriu, T., Hama, H., 2007. Fusion of Infrared and Visible Images for Robust Person Detection. *Innovative Computing, Information and Control, 2007. ICICIC '07. Second International Conference on,* 1, 310–310.