Proceedings of the 2022 IEEE
International Conference on Robotics and Biomimetics
December 5-9, 2022, Xishuangbanna, China

Computing Thermal Point Clouds by Fusing RGB-D and Infrared Images: From Dense Object Reconstruction to Environment Mapping

Tanhao Zhang, Luyin Hu, Yuxiang Sun, Lu Li, and David Navarro-Alarcon

Abstract— **Compared with 2D thermal images, visualizing the temperature of objects with their corresponding 3D surfaces provides a more intuitive way to perceive the environment. In this paper, we present an integrated system for large-scale and real-time 3D thermographic reconstruction through fusion of visible, infrared and depth images. The system is composed of an RGB-D and a thermal camera, whose image measurements are aligned with respect to the same coordinate frame. A thermal direct method based on infrared features is proposed and integrated into state-of-art localization algorithms for generating reliable 3D thermal point clouds. The reported experimental results demonstrate that our approach can be used for 3D reconstruction of small and large scale environments based on dual spectrum 3D information.**

I. Introduction

In recent years, thermal imaging has been widely used in a broad range of applications, such as non-contact temperature measurement for medical diagnosis, energy inspection in civil engineering [1], and recently for *thermal servoing* (a new temperature control problem recently introduced in the robotics community [2], [3]). However, most existing applications of thermography are limited to 2D imaging analysis of a scene, e.g., using infrared images for pattern recognition [4]. However, the performance of these systems can be improved by combining thermal images with 3D spatial information, which enables to make quantitative temperature analysis of 3D surfaces of interest [5].

There are various works that follow this depth-thermal fusion approach (either for analysis or simply visualization). However, the majority of current 3D thermal imaging systems are not movable and require considerable post-processing to generate 3D thermograms [6], [7]; Many of these state-of-the-art systems can only use limited frames to generate 3D thermal models [8]. The majority of recent works [9], [8], [6] are based on the KinectFusion algorithm [10], which is one of the most common methods for 3D dense reconstruction. However, one of the shortcomings of this algorithm is that it is only suitable for small-scale reconstruction applications, as it lacks loop closing and global optimization functions; As a consequence, the error of camera pose tracking accumulates as the number of frames increases. Another disadvantage is that the KinectFusion algorithm requires the vision system to move slowly and smoothly around the target; This method utilizes geometric information only, which is obtained by consensus frames, thus, large displacements typically lead to non-optimal solutions [11]. This feature limits the algorithm's applicability in fast moving environments.

In this paper, we present a large-scale robust method for handheld and real-time 3D dual spectrum reconstruction to overcome the limitation mentioned above. Our method is inspired by ORB-SLAM2 [12], which is one popular open source localization algorithms with many available datasets, such [13] and [14]; The comprehensive loop closing capabilities of ORB-SLAM2 makes feasible to conduct large-scale localization tasks. However, ORB-SLAM2 is based on RGB images, which means that the localization result is strongly influenced by illumination conditions [15]. One advantage of the use of thermography is that thermal information in natural environments is largely constant. This consistency feature of thermal imaging fits Grayscale Invariance Assumption (GIA) [16] [17] very well. In this work, we propose a thermal direct method that utilizes thermography and depth information for robust localization results, and integrate this method into the framework of ORB-SLAM2.

The main contribution of this paper are listed as as follows:

- Develop an approach to fuse images from different cameras with different fields of view and different light spectrum sensitivity;
- Propose a new localization method based thermography, and prove the feasibility of the proposed method;
- Combine ORB-SLAM2 with our proposed thermal direct method by using histogram of RGB images, make the above two methods compensate for each other based on texture richness of RGB images.
- Add dual a spectrum mapping node to directly visualize the geometric, color, and thermal information.

The rest of the paper is organized as follows: In Section 2, we reviewed the state-of-art in vision-based location mapping methods, recent research on thermographic technology, and our previous work on the alignment of different spectrum cameras. In Section 3, we present the development of whole system including hardware and software. In Section 4, some results are presented, including the comparison of our proposed thermal direct method and ORB-SLAM2, small-scale and large-scale 3D environment reconstruction with color and thermal information. Finally, we conclude the paper and give future work in Section 5.

This work is supported in part by the Research Grants Council under grant 15212721, and in part by the Jiangsu Industrial Technology Research Institute Collaborative Research Program Scheme under grant ZG9V.

Tanhao Zhang, Luyin Hu, Yuxiang Sun and David Navarro-Alarcon are with The Hong Kong Polytechnic University, Department of Mechanical Engineering, Hung Hom, Kowloon, Hong Kong. (e-mail: dna@ieee.org).

Lu Li is with the Institute of Advanced Manufacturing Technology of the CAS, Changzhou, Jiangsu, China

2022 IEEE International Conference on Robotics and Biomimetics (ROBIO) | 978-1-6654-8109-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ROBIO55434.2022.10011817

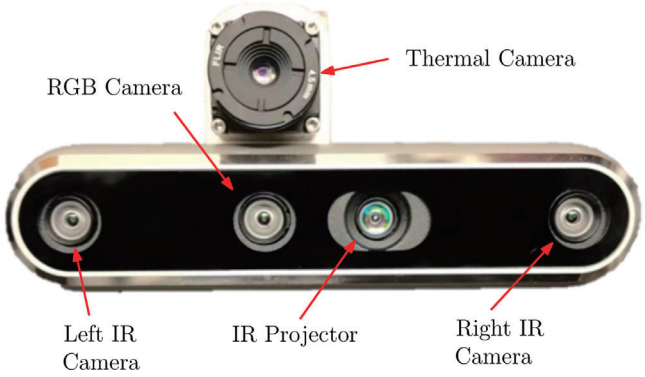
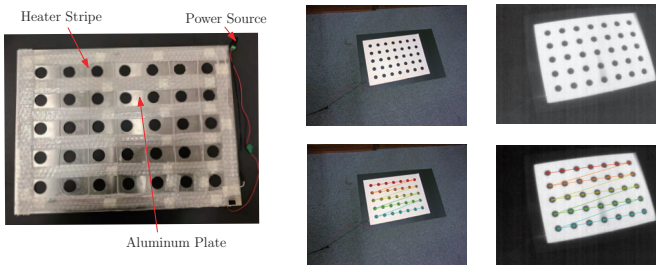


Fig. 1. Visual System: RGB-D camera and Thermal camera attached on 3D printed plate. RGB camera and Thermal camera have the same vertical axis.

TABLE I
SPECIFICATIONS OF CAMERA

Sensor Model	FoV	Resolution	Frame rate	Wavelength
FLIR Boson320	$52deg.$	320×256	60 FPS	8 14 µm
Realsense D455	$87deg.$	640×480	30 FPS	380 800 nm



(a) Back of Calibration Tool

(b) Raw and Labeled RGB Image/ Thermal Image

Fig. 2. Calibration tool and image of RGB camera and thermal camera when doing the calibration. (a) Shows the back of the calibration tool and the main component. (b) Feature point is detected by making high contrast of both RGB and thermal images. The feature point is at the center of the black circle

II. Related Works

A. The application of thermal image

The work in [18] presents a new semantic segmentation network to detect objects in low illumination conditions, that combines RGB images with thermal images in urban scenes; This method is intended to increase the safety of ADAS (Advanced Driver Assistance Systems). The works [19] and [20] show how thermal-color image fusion can be used for semantic segmentation to classify objects. The use of thermography with LiDAR in a SLAM application has been proposed in [21].

B. Alignment of RGB-D and Thermal cameras

The function of alignment is to get the thermal information of each pixel on an RGB image. Some researchers have addressed this issue with methods based on extrinsic calibration on thermal cameras and RGB-D cameras. The mathematical principle behind the calibration of thermal images is virtually the same as in RGB cameras, the only difference is that they typically utilize calibration devices with aluminum foil to reflect long-wave infrared (LWIR) and form high-contrast corners. The main drawback is that it is hard to build a stable heat source for reflection.

Our research team has proposed a new calibration tool that can cross-calibrate multiple vision sensors with different spectral sensibility, including ultraviolet, visible and infrared cameras [22]. The fabrication of this calibration tool if simple and provides comparable accuracy as with traditional visible spectrum calibration devices. In this paper, we continue to use this method to cross-calibrate the intrinsic and extrinsic parameters of a thermal camera and an RGB-D camera.

III. Methodology

A. Multimodal Vision System

The visual system with RGB, depth and thermal information in this paper consists of a FLIR Boson320 thermal camera and a RealSense D455 depth camera. Parameter details of camera are summarized at Table I. This thermal camera is radiometrically calibrated and the sensitivity is less than 60 mK. The thermal and depth cameras are both rigidly fixed on a 3D printed frame to ensure a stable setup for both sensors, see Fig.1.

The parameters of the thermal camera and RGB-D camera are different, both for field of view and resolution, as shown in Fig.3. To know the temperature of each pixel in the RGB image, we take the following steps to perform alignment:

1) Resize thermal images, from 320×256 to 640×480 (same as the RGB image) by using bilinear interpolation.
2) Capture the same numbers of resized thermal and RGB images using the calibration tool; Detect feature points and register the pixel position of each feature points in the images.
3) Calculate the intrinsic parameters of the thermal and RGB-D cameras by applying the calibration algorithm from [23]; Compute the extrinsic parameters of the thermal camera relative to the RGB-D camera by using the OpenCV function *'stereoCalibration()'*.

The implemented algorithm is based on the method presented in [22]. The calibration tool is made of aluminum sheet with regular holes and covered with white paper (see Fig. 2); A heater strip is placed on the backside, which is used for heating the sheet and produce high contrast thermal image feature. Feature detections with RGB images is simply done using standard high contrast visible spectrum features, e.g., with black and white chessboard patterns. Feature points are located at the center of circles/holes, which can be easily matched with both cameras; The above mentioned Step 3 is then performed.

Due to the large difference in field of view (FoV) between the RGB and thermal images, as well missing depth information, some of the pixels in the RGB image cannot be matched with temperature data, e.g., the edge of RGB image. To solve this problem, we discard any RGB pixel without temperature information and cut the edge symmetrically and

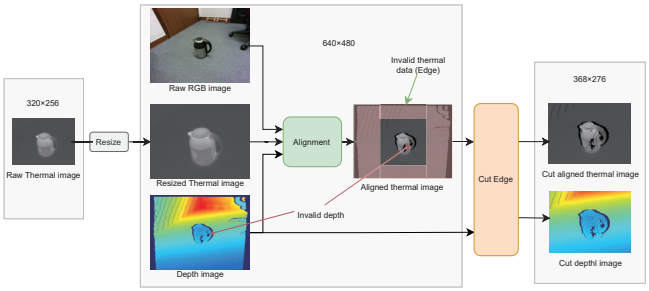
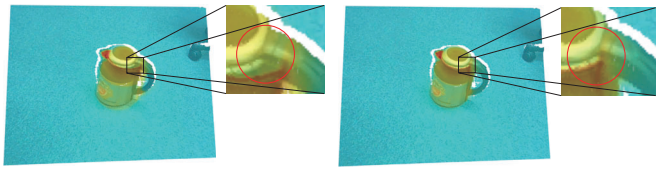


Fig. 3. Flowchart of the aligned thermal image to RGB image. The output is aligned with thermal-depth (T-D) information



(a) Before alignment (b) After alignment

Fig. 4. Point cloud shows the alignment of thermal information and RGB image. (a) Before alignment, RGB and thermal information cannot be well aligned. (b) After alignment, thermal information can correctly match with the RGB image. The result is clear in the red circle in the zoom-in image.

proportionally along the vertical axis. After the procedure above, the resolution of the RGB image becomes 368×276 and with relevant thermal data. Depth information is aligned using the default utilities provided by the RealSense camera. The detailed process is shown in Fig. 3

To test the cross-registration between different sensors, a comparison of the aligned thermal and RGB images is needed. This is to assess the result of alignment, and visualize the thermal information on the RGB image. However, one complication is to find a method that allows to present four channel information (viz., red, green, blue, and thermal) into the three color channels (red, green, blue). In paper [8], the authors utilize intensity-hue mapping to incorporate RGB images into thermal data. We follow a similar approach but covert the thermal images from gray scale (one channel) into JET color space (with three channels). The result of each pixel in the mapping image is shown in (1):

$$r_{rgb} = c_{rgb} + \omega(t_{rgb} - c_{rgb}), \quad 0 \leq \omega \leq 1, \qquad (1)$$

where r_{rgb} presents the result to each pixel, c_{rgb} and t_{rgb} are the pixel values in the RGB image and thermal image (JET color space), respectively; ω represents a weight interpolation between thermal and RGB images.

To visualize the accuracy of alignment, we simply generate point clouds of a kettle containing hot water using Open3D [24] without alignment and with alignment by setting $\omega = 0.5$, as shown in Fig. 4. Note that the thermal information at the edge of the kettle lid has a clear offset compared with the RGB image before alignment, while their overlap improves after alignment. This result proves that the algorithm we provide in [22] can align thermal information with the RGB image and depth information.

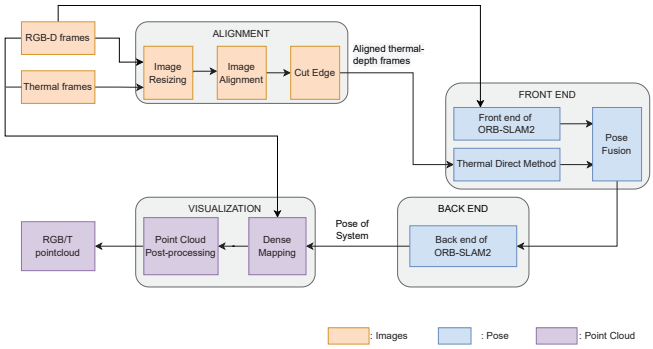


Fig. 5. The whole system contains three parts: thermal-depth information alignment, multi-spectrum pose estimation and visualization part. Note that the input of ORB-SLAM2 and the original RGB-D information becomes the original image that has a wider FoV containing more information.

B. Thermal-Assisted Pose Estimation

In this paper, we propose a "Thermal Direct Method" that uses the aligned thermal and depth information to enhance the accuracy of estimating the pose of a moving vision system. The basic idea behind our method is to use thermal information to deal with changes in illumination and shading, which is more robust than solely relying on RBG images [25]. Most commercial thermal cameras capture long wave infrared radiation emitted from objects, which provides a reliable method to capture objects and features under unstable illumination conditions. The overview flow chart of the whole system is shown in Fig. 5 and has three main functions: 1) Align RGB-D and thermal data and outputs an aligned thermal-color-depth image; 2) Estimating the pose by fusing the pose from ORB-SLAM2 and the Thermal Direct method; 3) Making dense point clouds and visualization. The input of the complete algorithm is the original RGB-D image and the aligned thermal image; The point cloud structure contains color, thermal and position information; Color and thermal modes can be switched on and off after generating the point cloud.

1) Thermal Direct Method: Assuming that objects have a constant temperature in consecutive thermal frames is reasonable because the temperature cannot change rapidly. Our approach is to obtain the translation of the pose of two consecutive thermal frames when the system is moving. Given the previous (denoted as $i-1$) and new (denote das i) aligned thermal frames, we first extract M ($M = 200$ in our program) good features to track (GFTT) [26] points from the thermal frame $i-1$. As Fig 6 shows, $p_{i-1,1}$ is one of the GFTT feature points in frame $i-1$, from the aligned thermal-RGB-D image. Because the movement of the system is not fast and the temperature is constant between frames, there must be a thermal pixel $p_{i,1}$ that has nearly the same temperature in frame $i-1$. Thus, we can build a function that minimizes the difference of temperature $E_{i,i-1}$ as follows:

$$\min E_{i,i-1} = \min \sum_{n=1}^{M} \|\boldsymbol{I}(p_{i-1,n}) - \boldsymbol{I}(p_{i,n})\|^2 \qquad (2)$$

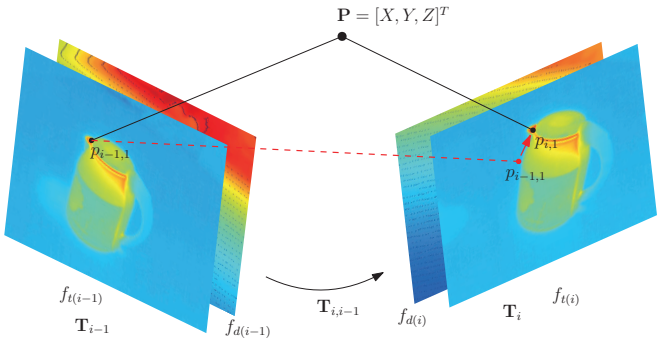


Fig. 6. The schematic diagram of the Thermal Direct method. the red arrow shows the shift of feature points between frames. The feature point $p_{i-1,1}$ can be translate to frame i utilizing translation matrix $T_{i,i-1}$, which is $p_{i,1}$. $p_{i-1,1}$ and $p_{i,1}$ share the same point position in 3D point $\boldsymbol{P}$ expressed as $[X, Y, Z]$ under camera coordinate at frame $i-1$. So the estimation of system pose change $T_{i,i-1}$ can be obtained by minimizing the gap of temperature between $p_{i-1,1}$ and $p_{i,1}$.

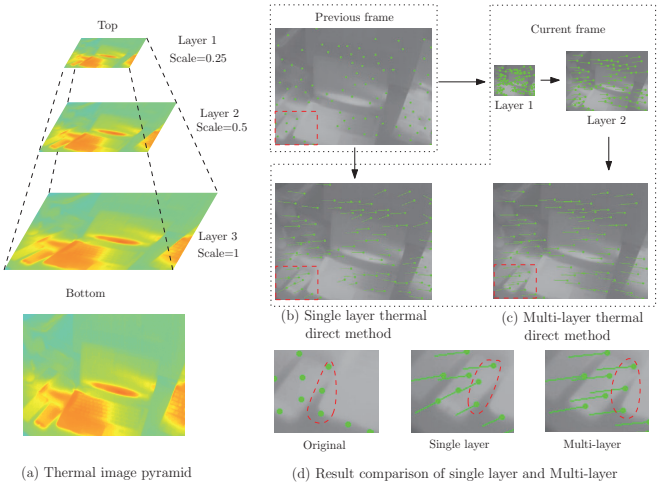


Fig. 7. Multi-layer thermal direct method. The thermal image is of a laptop on the table. (a) Thermal image pyramid of the laptop, the zoom-in scale increasing from top to bottom. (b) Result of single layer thermal direct method on two continuous thermal frames. Green dots and green line with dots shows the feature point and their shifting in previous and current frames respectively. (c) Shows thermal direct method result on multi-layer and presents the result of each layer. (d) comparison of single layer and multi-layer. Obviously, that result of multi-layer is better.

where $\boldsymbol{I}$ presents the temperature values shown in thermal frames.

Based on the camera model and aligned depth frames, we can project feature points $\boldsymbol{P}$ into 3D for both frames

$$p_{i-1,n} = \frac{1}{Z_{i-1,n}} \boldsymbol{\kappa P} \tag{3}$$

$$p_{i,n} = \frac{1}{Z_{i,n}} \boldsymbol{\kappa T P} \tag{4}$$

$$\boldsymbol{T} = exp((\boldsymbol{\xi})^{\wedge}) \tag{5}$$

The $\boldsymbol{T} \in \mathbb{R}^{4\times4}$ is a perspective projection matrix constructed with a rotation and translation $\boldsymbol{R}, \boldsymbol{t}$, and which can be convert to a Lie algebra $\mathfrak{se}(3)$ (represented as $\boldsymbol{\xi} \in \mathbb{R}^6$) by using the exponential mapping, for $\boldsymbol{\xi}^{\wedge}$ as an anti-symmetric matrix. The term κ denotes the intrinsic parameters matrix of the thermal camera. Note that thermal frames have been aligned with RGB frames, so we use the intrinsic parameters of the RGB frames with edge cut, which means κ is exactly known. With (3) and (4) known, the difference term $E_{i,i-1}$ becomes:

$$\min E_{i,i-1} = \min \sum_{n=1}^{M} \| \boldsymbol{I}(\frac{1}{Z_{i-1,n}} \boldsymbol{\kappa P}) - \boldsymbol{I}(\frac{1}{Z_{i,n}} \boldsymbol{\kappa T P}) \|^2 \tag{6}$$

Note that the above equation is a least-squares error that can be minimized by the Gauss-Newton method, with an updating function computed as:

$$\boldsymbol{T}^t = \boldsymbol{T}^{t-1} + \Delta \boldsymbol{T} \tag{7}$$

By denoting the pixel value in frame i at feature point n when iteration time is t as $\boldsymbol{I}^t_{i,n}$, then $E_{i,i-1}$ becomes:

$$E_{i,i-1} = \sum_{n=1}^{M} \| \boldsymbol{I}^t_{i-1,n} - \boldsymbol{I}^{t-1}_{i,n} - \frac{\partial \boldsymbol{I}^{t-1}_{i,n}}{\partial \Delta \boldsymbol{T}} \Delta \boldsymbol{T} \|^2$$

$$= \sum_{n=1}^{M} \| d_n + \boldsymbol{J}_n \Delta \boldsymbol{T} \|^2 \tag{8}$$

where d_n is the difference temperature value at point n between iterations, and $\boldsymbol{J}_n$ is the Jacobian matrix of d_n.

The Gauss-Newton function of the thermal direct method is shown in (9), where $\Delta \boldsymbol{T}$ is optimizing direction:

$$\boldsymbol{J}_n^T \boldsymbol{J}_n \Delta \boldsymbol{T} = -\boldsymbol{J}_n^T d_n \tag{9}$$

To obtain $\Delta \boldsymbol{T}$, we need to specify the Jacobian matrix. To this end, consider a pair of pixels in two consecutive frames and compute the difference value e. To obtain the minimum difference, the following derivative of e with respect to $\boldsymbol{T}$ is computed:

$$e(\boldsymbol{T}) = \boldsymbol{I}(p_{i-1}) - \boldsymbol{I}(p_i) = \boldsymbol{I}(\frac{1}{Z_{i-1}} \boldsymbol{\kappa P}) - \boldsymbol{I}(\frac{1}{Z_i} \boldsymbol{\kappa T P}) \tag{10}$$

$$\frac{\partial e}{\partial \boldsymbol{T}} = -\frac{\partial \boldsymbol{I}(p_i)}{\partial p_i} \frac{\partial p_i}{\partial \delta \boldsymbol{\xi}} \delta \boldsymbol{\xi} = -\boldsymbol{J} \delta \xi \tag{11}$$

where

$$\frac{\partial p_i}{\partial \delta \boldsymbol{\xi}} = \begin{bmatrix} \frac{f_x}{Z} & 0 & -\frac{f_x X}{Z^2} & -\frac{f_x XY}{Z^2} & f_x + \frac{f_x X^2}{Z^2} & -\frac{f_x Y}{Z} \\ 0 & \frac{f_y}{Z} & -\frac{f_y Y}{Z^2} & -f_y - \frac{f_y Y^2}{Z^2} & \frac{f_y XY}{Z^2} & \frac{f_y X}{Z} \end{bmatrix} \tag{12}$$

for f_x and f_y as the focal length parameters, which are known after calibration, $[X, Y, Z]^T$ denotes the position of $\boldsymbol{P}$, and the Jacobian matrix $\boldsymbol{J}$ can be computed as (11) which can be obtained by known parameters, then ΔT in equation 9 can be solved.

In our experiment, when two consecutive thermal frames have obvious differences, $E_{i,i-1}$ will be trapped in local minimum because of the typical non-convexity of thermal images. Thermal images lack of texture, when compared with RGB images [15], make them easier reach local minima. To solve this problem, we build an intuitive image pyramid scheme that create a multi-layer thermal direct method; The

key idea is zooming out the image with different scales from the smallest (top, $scale = 0.25$ in program) to the original scale (bottom, $scale = 1$ in program), and then using the thermal direct method for each layer from top to bottom. Fig 7(a) shows a laptop on the table, where the initial estimation translation value of the layer is the result of the upper layer, which means having a better initial value than the single layer. More layers of the image pyramid lead to more accurate results but increase the computation cost; We set the layer number $L_n = 3$ and the side length is twice than the upper until it is the same size as the original frame. In our study, we find that three layers are usually faster than a single layer because the former one iterate around nine times in total (around three times per layer) while the single layer case iterates around thirteen times.

Fig. 7(a)–(b) show the result of the single-layer thermal direct method and multi-layer thermal direct method; Fig 7(d) shows the enlarged image of a red rectangular area showing the comparison of the two methods above. We first convert to gray scale the thermal image to have better visualization of the result. The green dots in the preview frames are extracted using GFTT features. In the current frames, the green line shows the translation path and the green dots show the result of the translation of feature points. The Multi-layer leads to better results compared with the single layer, which is marked in red circles.

2) Pose Fusion: The front-end of ORB-SLAM2 and the thermal direct method work separately and estimate the pose independently in the front-end part; Hence, pose fusion is needed. The pose estimation of the ORB-SLAM2 part is directly influenced by the extracted ORB feature points. In our experiments, we found that the brightness of the environment significantly influences the quality of the ORB features while it affects little the thermal direct method. The method to obtain the estimated pose is based on dynamically adjusting the confidence of the pose estimation for each thread by the color richness of the RGB image.

Our methodology of pose fusion is as follows. First, capture an RGB frame and generate its histogram. Next, get the median value of the gray-scaled frame as m. Then, set a threshold t_{sh} and find the pixels in the range $(m \pm t_{sh})$, as follows:

$$n_{i,j} = \begin{cases} 1, & \text{if } I_{i,j} \in (m \pm t_{sh}) \\ 0, & \text{else} \end{cases} \tag{13}$$

Next, obtain the proportion p in that range by:

$$p = \frac{\sum_{i=1}^{u} \sum_{j=1}^{v} n_{i,j}}{u \times v} \tag{14}$$

for u and v as the height and width of the raw RGB images. Finally, the result of pose transformation presented as Lie algebra $\boldsymbol{\xi}$ can be obtained with:

$$\xi = \begin{cases} (1-p)\xi_{rgb} + p\xi_{thermal}, & p > 0.3 \\ \xi_{rgb}, & p \leq 0.3 \end{cases} \tag{15}$$

In our experiments, we noticed that thermal images have fewer textures than RGB images and the thermal direct method will lead to relatively lower pose estimation results compared with ORB feature points using RGB images. Thus, we consider that when p is smaller than 0.3, which means rich texture in RGB images and the feature point matching is good, we only utilize the RGB frames. When the tracking state of ORB-SLAM2 is LOST, the pose estimation is based on the thermal direct method.

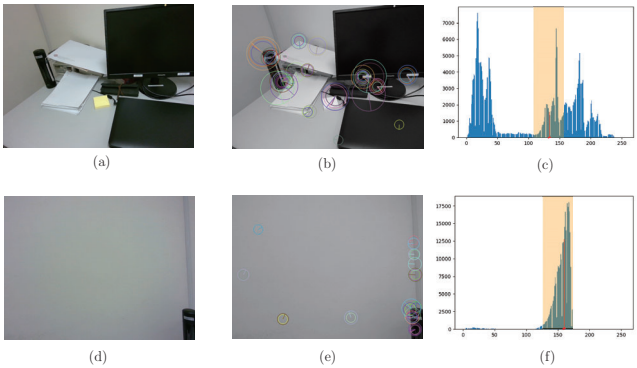


Fig. 8. (a) and (d): the raw image of good tracking and lost tracking state respectively. (b) and (e): Colored circles are the position of the ORB feature point with direction. (c) and (e): The image histogram of (a) and (d). The red point on X axis is the median m of the pixel value and the red bar show the position. The colored shade denote threshold of $m \pm t_{sh}$.

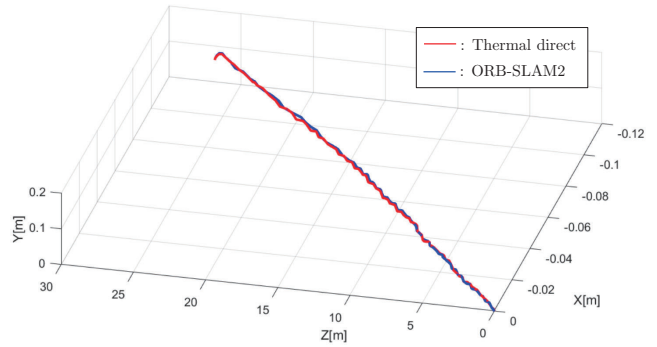
We extract the ORB feature in two images and set $t_{sh} = 15$, one for a good tracking state and one for tracking losses to extract the feature points and generate a histogram, as shown in Fig. 8. The center of the circles in (b) and (e) is the position of the feature points. Fig. (c) and (f) are the histograms of images (a) and (d), which also show the median m (the red point at the x axis) and threshold range (colored shade). In Fig. 8 (a), $m = 129$ and only has 16.17% pixels above the threshold, whereas for Fig. 8 (b), $m = 159$ and 89.06% pixels above the threshold. This shows that the images with more evenly distributed grayscale values generate more reliable feature points and lead to better tracking results.

IV. Experiments

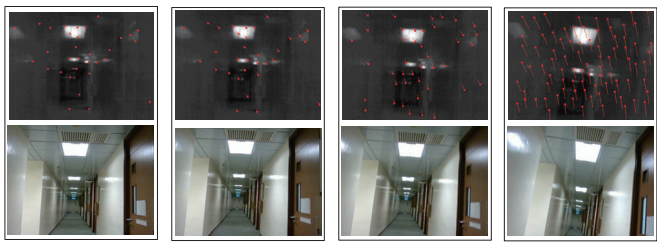
In this section, we first verify the feasibility of the thermal direct method by using our dataset containing aligned thermal and depth images. Then, we use our system to reconstruct 3D objects with both thermal and RGB information. Finally, we test our system and algorithm in different environments for large-scale dense mapping. In our experiments, the processor is a laptop with Intel Core i7-10870H CPU, 16G of RAM and Nvidia GeForce RTX 2060 GPU for real-time image processing.

A. Verification of the Thermal Direct Method

To assess the performance of the proposed approach, we investigate the relationship between thermal texture richness and the result of only using the thermal direct method. As (to the best of our knowledge) there is no publicly available dataset containing RGB, depth and thermal images with ground truth trajectories, we first perform an experiment of



(a) Trajectories in expriment 2.

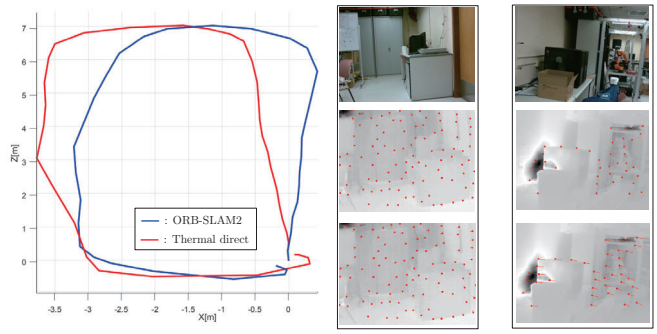


(b) Frames in experiment 2.

Fig. 9. Experiment 1: Results of Compare thermal direct method with ORB-SLAM2 of an indoor corridor. (a) shows the trajectories of a different method. (b) are some result frames of the thermal direct method.

comparing ORB-SLAM2 with the proposed thermal direct method, i.e., we take the result from ORB-SLAM2 as ground truth. We acquired two data sequences from different environments, then, calculate each of their trajectories; The description of datasets are as follows:

1) A corridor with various lamps on generating heat.
2) A laboratory room with various computer servers.



(a) Trajectory (b) Lack of image texture (c) Rich image texture

Fig. 10. Experiment 2: Demonstration of lack of texture in thermal images. (a) shows the trajectory of two methods. (b) and (c) is the tracking results of a lack of textures and rich texture.

For environment 1, we test the thermal direct method. The characteristic of this environment is that it has a regular heat source coming from the fluorescent lamp on the roof and electric control box next to an office). As Fig. 9(b) shows, as the temperature measurements are consistent, these heat sources provide stable feature points that are easy to tracking (red dots are the feature points and red lines show the transform between neighboring thermal frames). The trajectories of the thermal direct method and ORB-SLAM2, as shown in Fig. 9(a), indicate that our approach is reliable even with not very feature-rich thermal environments. For these tests, we also found that the mean frame distance with the proposed method is only 0.0714 meters, which is acceptable for most engineering applications.

For environment 2, we assess the tracking performance of the thermal direct method under the uneven distribution of thermal textures. We test the proposed approach in an underground laboratory room (infamously referred by the authors as "the dungeon") where thermal frames lack texture. As the Fig. 10(a) shows, the calculated trajectories are clearly different. One of the reasons for this result is that some of the thermal frames have poor image texture as the temperature of the background is almost uniform (shown in Fig. 10(b)); The tracking result is good when there are distinctive thermal features (shown in Fig. 10(c)). These problems can be potentially solved by fusing/combining the result of ORB-SLAM2 with that of the thermal direct method.

B. 3D Reconstruction with Dual Spectrum Vision

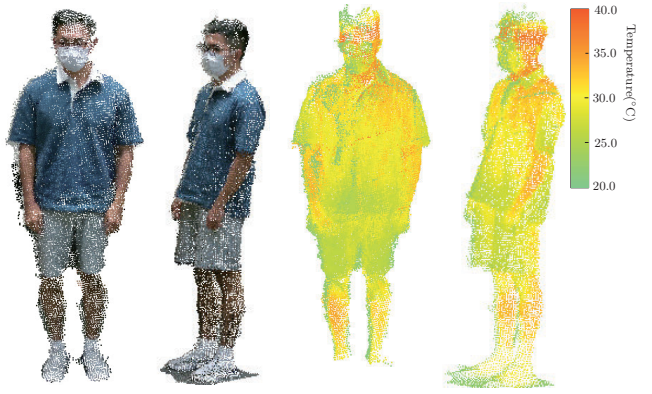
We generate a point cloud of a target object to demonstrate the result of our dual spectrum system and our proposed algorithm; The localization method fuses the ORB-SLAM2 with the thermal direct method. The whole system runs on ROS (Robot Operation System) and has four main nodes to perform the following:

1) Receive data from RGB-D and thermal cameras.
2) Process the received frames by aligning and trimming the edges.
3) Simultaneous localization, including ORB-SLAM2, thermal direct method and pose fusion.
4) Environment mapping, and generation of dense thermal and RGB point clouds.

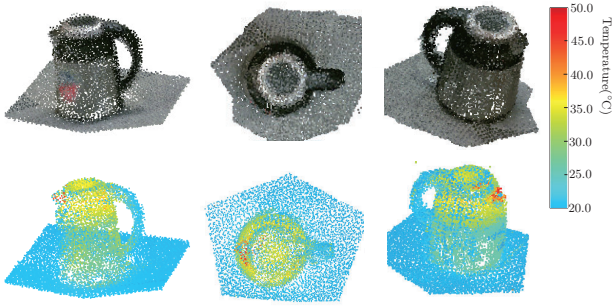
In the mapping node, we only add a new point cloud when a new keyframe is received in the ORB-SLAM2 thread, or after every five seconds. After processing by pose fusion, the location of the system is improved. To show the quantify of our method, we generate a point cloud for both visible-spectrum (RGB) and thermal images of different objects, as shown in Fig. 11. By visualizing objects as a point cloud in a dual spectrum manner (i.e., with both visible and thermal), we can easily sense the object's shape and thermal information, such as the forehead temperature at 36.7°C or the handle of a kettle at 20°C.

C. Mapping under Large Scale Environments

To further test the mapping result of our dual spectrum system, we conducted experiments for some large environments, including indoor and outdoor environments. The data was captured by hand-holding the system and walking along the environment. Fig. 12(a) shows the point cloud of a corridor from the top-down view. The upper image is for the visible spectrum and the lower shows the thermal point cloud. By visualizing the corridor using two spectrum point clouds, we can easily identify which place is warmer or colder. Fig. 12(b) depicts the point cloud of a side-view air

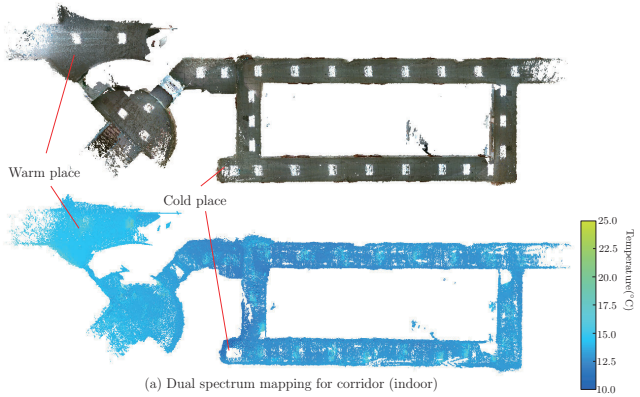


(a) Person, RGB and thermal information respectively



(b) Kettle, RGB and thermal information respectively

Fig. 11. Point clouds from different perspectives, including color and thermal information. (a)Front and side view of a person with the cloth. (b) front, top and side view of a kettle containing hot water.



(a) Dual spectrum mapping for corridor (indoor)

(b) Dual spectrum mapping for air conditioners (outdoor)

Fig. 12. 3D reconstruction for large-scale environments with visible and thermal information. (a) shows an indoor corridor and (b) is an air conditioner outdoor unit.

conditioner outdoor unit. As the figure shows, our system not only can present the color and shape/geometry of objects but also the temperature information. The ability to combine 3D reconstruction with temperature measurements is very useful in industrial applications. For example, the abnormal temperature can be sensed and located in an industrial plant, and all data can be recorded dynamically for later analysis.

V. Conclusions

In this paper, a dual spectrum real-time 3-dimensional reconstruction method is proposed by combining the thermal direct method and ORB-SLAM2 to reconstruct the environment with both visible and thermal information. We first use a special calibration board to obtain the extrinsic parameters of the thermal and RGB-D cameras. To perform this task with different fields of view from the cameras, we cut the edges of a larger FoV image and then apply the image fusion algorithm presented in [22]. After this image processing stage, we input the aligned thermal-depth information and RGB-depth information into the localization algorithm, which consists of the thermal direct method, ORB-SLAM2 and pose fusion thread. Finally, the dual spectrum point cloud is generated simultaneously using PCL library [27] for visualization.

In the experiment section, we demonstrate that our proposed thermal direct method is fairly accurate when the image texture is rich enough. Then, we test our whole system on 3D reconstruction. We generate the point cloud which contains thermal and color information. The results qualitatively demonstrate the accuracy of the system. Our approach can be used in industrial areas for monitoring the temperature of the machine, such as fire management and response. Our system can also reconstruct thermal and visible spectrum 3D models for intuitive thermal analysis. Its lightweight hand-held hardware and integration based on ROS makes it easy to be operated.

Our system combines thermal information in 3D mapping, which can be widely used in industrial monitoring and service applications. However, note that thermal information has not been used in global optimization and close-loop detection. Our future work includes adding thermal data into the back-end of SLAM to enhance the robustness under unstable illumination environments. In addition, we plan to deploy our system on mobile robots [28] or manipulators [29], [30] for practical service applications; Another interesting application domain of our multi-modal sensing technology is in robotic manipulation tasks that involve safe interactions with humans, e.g., in cosmetic dermatology [31], [32] or ultrasound scanning of tissues [33], [34].

References

[1] S. Vidas and P. Moghadam, "Heatwave: A handheld 3d thermography system for energy auditing," *Energy and buildings*, vol. 66, pp. 445–460, 2013.

[2] L. Hu, D. Navarro-Alarcon, A. Cherubini, M. Li, and L. Li, "On radiation-based thermal servoing: New models, controls, and experiments," *IEEE Transactions on Robotics*, vol. 38, no. 3, pp. 1945–1958, 2022.

[3] L. Hu, A. Duan, M. Li, A. Cherubini, L. Li, and D. Navarro-Alarcon, "Paint with the sun: A thermal-vision guided robot to harness solar energy for heliography," *IEEE Sensors Journal*, vol. 22, no. 18, pp. 18 130–18 142, 2022.

[4] C. Li, W. Xia, Y. Yan, B. Luo, and J. Tang, "Segmenting objects in day and night: Edge-conditioned cnn for thermal image semantic segmentation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 7, pp. 3069–3082, 2020.

[5] Y. Cao, B. Xu, Z. Ye, J. Yang, Y. Cao, C.-L. Tisse, and X. Li, "Depth and thermal sensor fusion to enhance 3d thermographic reconstruction," *Optics express*, vol. 26, no. 7, pp. 8179–8193, 2018.

[6] A. O. Müller and A. Kroll, "Generating high fidelity 3-d thermograms with a handheld real-time thermal imaging system," *IEEE Sensors Journal*, vol. 17, no. 3, pp. 774–783, 2016.

[7] J. Rangel, S. Soldan, and A. Kroll, "3d thermal imaging: Fusion of thermography and depth cameras," in *International Conference on Quantitative InfraRed Thermography*, vol. 3, 2014.

[8] S. Vidas, P. Moghadam, and S. Sridharan, "Real-time mobile 3d temperature mapping," *IEEE Sensors Journal*, vol. 15, no. 2, pp. 1145–1152, 2014.

[9] S. Vidas, P. Moghadam, and M. Bosse, "3d thermal mapping of building interiors using an rgb-d and thermal camera," in *2013 IEEE international conference on robotics and automation*. IEEE, 2013, pp. 2311–2318.

[10] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison *et al.*, "Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera," in *Proceedings of the 24th annual ACM symposium on User interface software and technology*, 2011, pp. 559–568.

[11] R. Jamiruddin, A. O. Sari, J. Shabbir, and T. Anwer, "Rgb-depth slam review," *arXiv preprint arXiv:1805.07696*, 2018.

[12] R. Mur-Artal and J. D. Tardós, "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras," *IEEE transactions on robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.

[13] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.

[14] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of rgb-d slam systems," in *2012 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, 2012, pp. 573–580.

[15] L. Chen, L. Sun, T. Yang, L. Fan, K. Huang, and Z. Xuanyuan, "Rgb-t slam: A flexible slam framework by combining appearance and thermal information," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 5682–5687.

[16] P. Liu, X. Yuan, C. Zhang, Y. Song, C. Liu, and Z. Li, "Real-time photometric calibrated monocular direct visual slam," *Sensors*, vol. 19, no. 16, p. 3604, 2019.

[17] J. Engel, T. Schöps, and D. Cremers, "Lsd-slam: Large-scale direct monocular slam," in *European conference on computer vision*. Springer, 2014, pp. 834–849.

[18] Y. Sun, W. Zuo, and M. Liu, "Rtfnet: Rgb-thermal fusion network for semantic segmentation of urban scenes," *IEEE Robotics and Automation Letters*, vol. 4, no. 3, pp. 2576–2583, 2019.

[19] S. S. Shivakumar, N. Rodrigues, A. Zhou, I. D. Miller, V. Kumar, and C. J. Taylor, "Pst900: Rgb-thermal calibration, dataset and segmentation network," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 9441–9447.

[20] J. H. Lee, J.-S. Choi, E. S. Jeon, Y. G. Kim, T. T. Le, K. Y. Shin, H. C. Lee, and K. R. Park, "Robust pedestrian detection by combining visible and thermal infrared cameras," *Sensors*, vol. 15, no. 5, pp. 10 580–10 615, 2015.

[21] Y.-S. Shin and A. Kim, "Sparse depth enhanced direct thermal-infrared slam beyond the visible spectrum," *IEEE Robotics and Automation Letters*, vol. 4, no. 3, pp. 2918–2925, 2019.

[22] T. Zhang, L. Hu, L. Li, and D. Navarro-Alarcon, "Towards a Multi-spectral RGB-IR-UV-D Vision System — Seeing the Invisible in 3D," in *IEEE Int Conf on Robotics and Biomimetics (ROBIO)*, 2021, pp. 1723–1728.

[23] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[24] Q.-Y. Zhou, J. Park, and V. Koltun, "Open3d: A modern library for 3d data processing," *arXiv preprint arXiv:1801.09847*, 2018.

[25] Q. Fu, H. Yu, X. Wang, Z. Yang, Y. He, H. Zhang, and A. Mian, "Fastorb-slam: Fast orb-slam method with descriptor independent keypoint matching," *arXiv preprint arXiv:2008.09870*, 2020.

[26] J. Shi *et al.*, "Good features to track," in *1994 Proceedings of IEEE conference on computer vision and pattern recognition*. IEEE, 1994, pp. 593–600.

[27] R. B. Rusu and S. Cousins, "3d is here: Point cloud library (pcl)," in *2011 IEEE international conference on robotics and automation*. IEEE, 2011, pp. 1–4.

[28] J. G. Romero, D. Navarro-Alarcon, E. Nuño, and H. Que, "A globally convergent adaptive velocity observer for nonholonomic mobile robots affected by unknown disturbances," *IEEE Control Systems Letters*, vol. 7, p. 85–90, 2023. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9807266

[29] P. Zhou, J. Zhu, S. Huo, and D. Navarro-Alarcon, "LaSeSOM: A latent and semantic representation framework for soft object manipulation," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 5381–5388, 2021.

[30] P. Zhou, R. Peng, M. Xu, V. Wu, and D. Navarro-Alarcon, "Path planning with automatic seam extraction over point cloud models for robotic arc welding," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 5002–5009, 2021.

[31] M. Muddassir, D. Gomez Dominguez, L. Hu, S. Chen, and D. Navarro-Alarcon, "Robotics meets cosmetic dermatology: Development of a novel vision-guided system for skin photo-rejuvenation," *IEEE/ASME Trans Mechatronics*, vol. 27, no. 2, pp. 666–677, April 2022.

[32] M. Muddassir, G. Limbert, and D. Navarro-Alarcon, "Development of a numerical multi-layer model of skin subjected to pulsed laser irradiation to optimise thermal stimulation in photorejuvenation procedure," *Comput. Methods Programs Biomed.*, vol. 216, p. 106653, 2022.

[33] A. Duan, M. Victorova, J. Zhao, Y. Sun, Y. Zheng, and D. Navarro-Alarcon, "Ultrasound-guided assistive robots for scoliosis assessment with optimization-based control and variable impedance," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 8106–8113, 2022.

[34] M. Victorova, M. K.-S. Lee, D. Navarro-Alarcon, and Y. Zheng, "Follow the curve: Robotic ultrasound navigation with learning-based localization of spinous processes for scoliosis assessment," *IEEE Access*, vol. 10, pp. 40 216–40 229, 2022.