HAL
open science

Registration of RGB and thermal point clouds generated by structure from motion

Trong Phuc Truong, Masahiro Yamaguchi, Shohei Mori, Vincent Nozick, Hideo Saito

To cite this version:

Trong Phuc Truong, Masahiro Yamaguchi, Shohei Mori, Vincent Nozick, Hideo Saito. Registration of RGB and thermal point clouds generated by structure from motion. Multi-Sensor Fusion for Dynamic Scene Understanding, Oct 2017, Venice, Italy. 10.1109/ICCVW.2017.57 . hal-01625092

HAL Id: hal-01625092

https://hal.science/hal-01625092

Submitted on 27 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Registration of RGB and thermal point clouds generated by structure from motion

Trong Phuc Truong[1], Masahiro Yamaguchi[1], Shohei Mori[1], Vincent Nozick[2], Hideo Saito[1]
[1]Graduate School of Science and Technology
[2]Japanese French Laboratory for Informatics, CNRS, UMI 3527
[1,2]Keio University, Kanagawa, Japan
[1]{ttphuc, yamaguchi, mori, saito}@hvrl.ics.keio.ac.jp [2]vincent.nozick@u-pem.fr

Abstract

Thermal imaging has become a valuable tool in various fields for remote sensing and can provide relevant information to perform object recognition or classification. In this paper, we present an automated method to obtain a 3D model fusing data from a visible and a thermal camera. The RGB and thermal point clouds are generated independently by structure from motion. The registration process includes a normalization of the point cloud scale, a global registration based on calibration data and the output of the structure from motion, and a fine registration employing a variant of the Iterative Closest Point optimization. Experimental results demonstrate the accuracy and robustness of the overall process.

1. Introduction

Originally developed for military purpose, infrared thermography became a common tool for numerous other fields. In the last decades, thermal imaging has been employed in applications such as infrastructure and electrical systems monitoring, human detection, breast cancer diagnostic, and see-through smoke or fog environment [25, 2]. In most of these cases, 2D thermal images are considered, exploiting the facility to install a thermal camera, as well as the non-invasive and non-destructive system of recording the temperature regardless of the ambient light. By applying 3D reconstruction photogrammetric techniques on infrared thermal (IRT) images, such as structure from motion (SfM), it is possible to take advantage of both thermal and geometric properties. This way, scene understanding can be enhanced, for example to study energy efficiency of the building sector, or to perform 3D object recognition and classification. In this paper, we propose an automated registration method for RGB and IRT point clouds generated by SfM from visual and infrared sequences. The obtained aligned model fuses visible colors, thermal, and depth information.

1.1. Related Work

Current state-of-the-art techniques to generate 3D RGB-thermal model rely on merging information from different type of sensors. In general, these methods can be classified in two different approaches.

The first approach is to map the RGB and IRT information to a point cloud reconstructed using time-of-flight technologies. Borrmann *et al*. present a mobile platform equipped with a 3D laser scanner, an RGB camera and a thermal camera to create 3D thermal models [4]. The mobile robot is able to autonomously collect the data, then map thermal and color information onto the 3D data without any scale ambiguity given sensors calibration data. A similar method is proposed in [7], where the sensors are instead placed on a wearable backpack comprising of five 2D laser range scanners, two optical cameras and two infrared cameras. In [26], the authors present a registration method using a range camera that is able to simultaneously provide both range and intensity images. Given 2D correspondences between the intensity and thermal images, the range camera can assign a 3D point to each 2D matches. The thermal point cloud is then derived by applying the *Efficient Perspective-n-Point* algorithm to these 3D/2D correspondences. Considering the cost of 3D laser scanners, some systems use low-cost RGB-D camera (Microsoft Kinect). They also generate RGB point clouds on which the thermal information is added. A system using and RGB-D camera and a single additional thermal camera is presented in [24], where after computing the poses corresponding to each RGB and IRT image, raycasting is used to map RGB and IRT intensities to the voxels reconstructed from the range sensor. Even though depth cameras and range scanners can provide accurate point cloud with low processing costs, their precision and range can be limited when operated outdoors depending on the technology

employed [1, 13].

The second approach is an image-based point cloud reconstruction. In [15], the authors present a semi-automatic framework to generate RGB-IRT 3D model achieved through image stitching and surface reconstruction techniques. This method needs an operator to define the temperature interval for each dataset and to verify every matches between thermal images as well as the matches between RGB to thermal images. This task is decisive for the image stitching and registration process. In [10], the authors solve the matching problem using specific built-in digital lenses that can capture simultaneously both RGB and thermal images. However, their registration pipeline still includes a multi-view stereo process to optimize the camera poses. Several RGB-IRT point cloud registration methods for Remotely Piloted Aircraft System (RPAS) are evaluated in [12]. The most accurate proposed method relies on the on-board GPS/INS of the RPAS to generate the point clouds from RGB and IRT image inputs. Then, under the assumption that the two point clouds are very close together and have the same scale, the *Iterative Closest Point* (ICP) algorithm is directly used for the registration.

1.2. Motivations and Methodology

In each aforementioned method, the generation of the thermal point cloud is relying on the extrinsic parameters of the sensors retrieved during a calibration step, *i.e.*, the 3D IRT point cloud cannot be reconstructed without using other sensors. To the best of our knowledge, the whole SfM pipeline has never been directly employed on thermal image sequence, even though state-of-the-art local detectors and descriptors can be also employed to find correspondences in the latter [18, 14]. We propose a framework using SfM directly on visible and thermal image sequences. Hence, unlike previous existing methods, the point cloud reconstruction is completely independent from the registration process.

The proposed system to align the RGB and IRT point clouds is based on the constant relative position and orientation between the two cameras. This condition is fulfilled by fixing the two cameras on a stereo rig. The relative pose is beforehand determined by a calibration procedure using a chessboard. Then, two independent RGB and IRT point clouds are generated using SfM from image sequence inputs. Since the scale of the two reconstructions are arbitrary, a normalization procedure must be first performed before the registration. Next, a rigid body transformation is defined to achieve a global registration considering that the two cameras are fixed on a stereo rig. Finally, a variant of the *Iterative Closest Point* algorithm is applied to perform a fine registration overcoming errors from previous steps. An overview of the framework is presented in Figure 1.

The output of the presented method is two aligned RGB and thermal point cloud which leads to more informative scene representation. Our main contributions include:

- Providing a two-step algorithm for RGB and IRT point cloud registration that can recover from an inaccurate calibration or poses estimation.
- Proposing a way to normalize independent point clouds output from SfM for a stereo rig.
- Presenting an enhanced calibration method for the thermal camera using a chessboard pattern.
- Demonstrating the possibility to generate thermal point cloud only using the SfM pipeline.

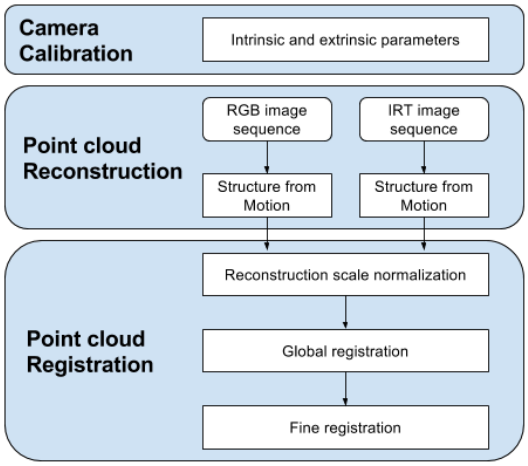


Figure 1. Overview of the proposed RGB and thermal 3D reconstruction and registration method.

2. Camera Calibration

This section details the camera calibration procedure to obtain the intrinsic and extrinsic parameters of the RGB and the thermal camera.

2.1. Calibration Setup

The calibration of the RGB and IRT cameras follows the procedure using a planar pattern [6, 28]. We use an enhanced version of the usual calibration chessboard adapted to perform with both RGB and IRT cameras. Indeed, the difference of emissivity of the black and white regions of a regular chessboard is insufficient for the thermal camera to have stable corners to perform the calibration. A common method to increase the temperature difference between the two regions is to heat the pattern with a flood lamp [17, 20]. Nevertheless, Vidas *et al*. [23] pointed out the struggle to detect crisp corners for accurate calibration and the difficulty to execute it.

To overcome these issues, we propose a simple yet accurate calibration method using a modified chessboard and a

corners refinement step. A conductive rubber tape is placed on the black parts of the printed chessboard pattern coupled with a low heat capacity support to increase the thermal radiation. Then, instead of using a flood lamp that produces a non-uniform heating [20], we cool down the calibration board. The Figure 2 shows the calibration chessboard captured simultaneously by both cameras.

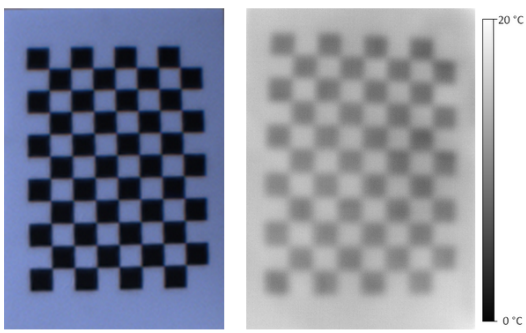


Figure 2. Calibration chessboard simultaneously captured by the RGB camera (left) and the thermal camera (right).

2.2. Radial Distortion

Due to the blur on the detected chessboard on the thermal image, regular calibration methods that handle lens distortion correction are unlikely to perform well. Thus, we first consider the rectification of the radial distortion of the thermal images.

We select a very usual 4th order polynomial as the distortion model to correct the effect of distortion. More precisely, we use the method proposed by Devernay and Faugeras [9], derived from the plumb line approach. They assume the distortion center to be equal to the image center and the pixels to be square. This approximation leads to satisfactory results that still can be enhanced afterwards with a global optimization scheme during the calibration process.

In practice, this process requires some patches of points supposed to be aligned. Since the lens distortion correction is most likely to be performed only when the lenses are changed or modified (zoom in / out), we select these points manually. Note that straight lines in thermal images are not hard to find. A radial distortion correction result is depicted in Fig. 3.

2.3. Camera Parameters Estimation

Once the radial distortion is corrected, both RGB and IRT cameras can be calibrated using automatic tools.

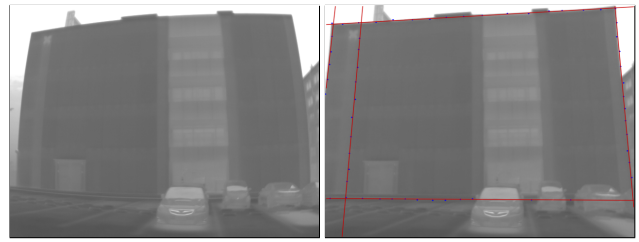


Figure 3. Lens distortion correction on the thermal image using [9].

2.3.1 Camera Model

We use the pinhole camera model for both cameras. The camera sensor is assumed to be zero-skewed with squared pixels. Given a scene points $\mathbf{X} \in \mathbb{P}^3$ that projects to an image point $\mathbf{x} \in \mathbb{P}^2$, the camera projection matrix [11] can be expressed by

$$\mathbf{x} = \mathbf{K}\left[\mathbf{R} \mid \mathbf{t}\right]\mathbf{X}, \quad \text{with } \mathbf{K} = \begin{bmatrix} \alpha & 0 & x_0 \\ 0 & \alpha & y_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1}$$

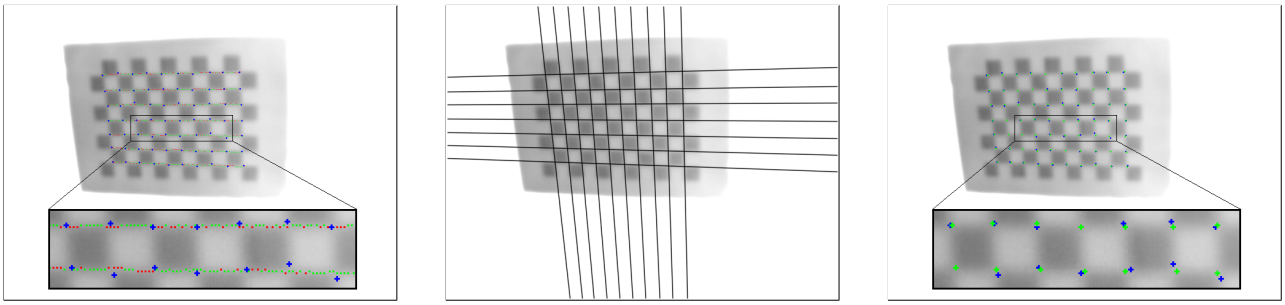
where $\mathbf{K}$ is the intrinsic parameters matrix, defined by the focal length α expressed in pixel unit and the principal point (x_0, y_0). $\left[\mathbf{R} \mid \mathbf{t}\right]$ is the extrinsic parameters matrix, which contains the rotation and the translation that relates the world coordinate system to the camera coordinate one.

2.3.2 Chessboard Corners Refinement

The chessboard corners are most of the time difficult to locate in a thermal image due to low difference of temperature between the black and white areas, causing a thermal blur effect. Therefore, we perform a refinement step from an approximative position of the corners computed with a common chessboard detection algorithm dedicated for visible image [6].

This refinement step is inspired by the method proposed by De la Escalera and Armingol [8] based on chessboard line intersections optimization. However, instead of using the Hough transform to detect the line, we estimate the lines from the initial approximative corners.

Each line of the chessboard is first estimated using a least square fit. Assuming a Gaussian noise on the chessboard corner detection due to the thermal blur, the fitted line can still be used as a guideline to accurately find the chessboard edges. These edges can be represented by points identified by computing the maximum of the image intensity gradient in an orthogonal direction to the fitted line. An example of additional detected line points is represented in Fig. 4 (a), where outliers are discarded using RANSAC line fitting. Then, the horizontal and vertical line coefficients are optimized to obtain the same vanishing point.



(a) (b) (c)

Figure 4. Chessboard detection refinement. (a) Initial blue point are fitted to robustly extract more points according to the image gradient. The inliers (green) and outliers (red) of the final RANSAC line fitting are also represented. (b) Intersection of the RANSAC and vanishing point optimized fitted lines. (c) Refined corners (green) compared to initially estimated corners (blue).

The final chessboard edges are finally computed using the intersections of the optimized lines, as shown in in Fig. 4 (b). A comparison of the chessboard corners before and after refinement is depicted in Fig. 4 (c).

2.3.3 Global Optimization

Given a set of 3D to 2D point correspondences over multiple frames, the intrinsic and extrinsic parameters can be computed using the closed form solution introduced by Zhang [28]. The resulting intrinsic parameters are used as an initial estimate for a global optimization to refine the calibration as well as the lens distortions correction. This non-linear process (computed with Levenberg-Marquardt optimizer) minimizes the reprojection error function defined as follows

$$\sum_i \sum_j \left\| \mathbf{x}_{ij} - \hat{p}(\mathbf{K}, \mathbf{D}, \mathbf{R}_i, \mathbf{t}_i, \mathbf{X}_j) \right\|^2 , \qquad (2)$$

where $\mathbf{x}_{ij}$ is the position of the jth point in image i and $\hat{p}(\mathbf{D}, \mathbf{K}, \mathbf{R}_i, \mathbf{t}_i, \mathbf{X}_j)$ is the projection of the 3D points $\mathbf{X}_j$ in image i. In other words, the function $\hat{p}$ first projects the point $\mathbf{X}_j$ from Eq. (1) and then applies the lens distortion correction (section 2.2) on the projected point.

3. Point Cloud Reconstruction

This section will present how the point clouds are generated from image sequences and its filtering process.

3.1. Structure From Motion and Multi-view Stereo

Although visible image and thermal image share different physical properties, one common characteristic is the similarity of the shape of an object. In a 3D sense, it means that the two point clouds can be related using edges and surface of the structure. Consequently, we propose a point cloud alignment method based on the sparse 3D reconstruction since only the strong distinguishable features, such as edges and corners, will be reconstructed. Nevertheless, we also consider the dense reconstruction as it can be used for a visualization purpose.

To reconstruct the sparse and dense 3D structure, the open source software COLMAP [21, 22] is used. It is a general-purpose structure from motion and multi-view stereo pipeline. The sparse reconstruction differs from the usual incremental reconstruction process notably by using geometric verification to improve the robustness of the initialization, a next best view planning, a robust triangulation method and a more efficient bundle adjustment parametrization.

3.2. Point Cloud Filtering

Since thermal images have less stable features compared to the visible ones, we filter the point cloud by performing a statistical analysis on the neighborhood of each point as proposed in [19]. We compute the average distance from a point to all its K nearest neighbors, and repeat this operation for all points. Then, assuming a Gaussian noise on the reconstructed point cloud, the points with a mean distance farther than the interval defined by the global distances mean and standard deviation are considered as outliers. This way, we are able to filter noisy points arising from false matches between thermal images features.

4. Point Cloud Registration

After having reconstructed the 3D point clouds from the RGB and IRT image sequences, the registration operation can be achieved. This section will cover how to define the rigid body transformation that aligns the IRT point cloud to the RGB one after having normalization. Then, the fine registration step using sparse ICP will be explained.

4.1. Normalization of the 3D Reconstruction

In the incremental SfM pipeline, the scale of the reconstruction is fixed by selecting two frames from the image

sequence and normalizing the length of their baseline [3]. If the normalization was performed with different pair of frames for the two independently reconstructed 3D point clouds, there will be a scale ambiguity between them.

We define the reconstruction scale s_r as the factor that the IRT point cloud must be scaled with to obtain the same scale as the RGB one. This unknown scale s_r can be estimated by analyzing the computed trajectories of the two cameras defined by the set of camera poses $\mathbf{C}^{rgb}$ and $\mathbf{C}^{irt}$, both estimated using SfM. In fact, since they are both subject to the same rigid motion in the real world, the distance between two different camera poses i and $i+n$ in $\mathbf{C}^{rgb}$ should be equal of the distance between its counterpart in $\mathbf{C}^{irt}$, in the case of a pure translation. This condition is expressed as

$$d(\mathbf{C}_i^{rgb}, \mathbf{C}_{i+n}^{rgb}) = d(\mathbf{C}_i^{irt}, \mathbf{C}_{i+n}^{irt}), \tag{3}$$

where $d(\mathbf{C}_i, \mathbf{C}_{i+n})$ is the distance between the camera position $\mathbf{C}_i$ and $\mathbf{C}_{i+n}$. As mentioned above, this constraint only holds for pure translation of the camera rig, *i.e.* the relative rotation between the ith and $i{+}n^{th}$ camera pose must be the identity matrix. Thus, the camera poses must be first clustered by their orientation so that Eq. (3) can be used within a cluster of cameras related with pure translations. We perform the clustering by constructing a histogram where the bin width corresponds to a specified maximal angular deviation of the camera orientations. The angular deviation is computed as the sum of difference of the Euler angles. By denoting $\mathbf{V}$ as the set of computed clusters, we can use the N_c largest clusters of camera poses $\mathbf{V}_i$ to recover the reconstruction scale s_r as follows

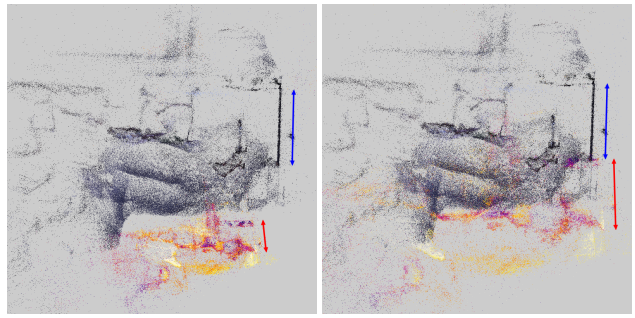
$$s_r = \frac{\sum_i^{N_c} d_v(\mathbf{V}_i^{rgb})}{\sum_i^{N_c} d_v(\mathbf{V}_i^{irt})}, \tag{4}$$

where $d_v(\mathbf{V}_i)$ is the sum of distance between every pair of camera position in the ith cluster. N_c is chosen such that the impact of inaccurate pose estimation is reduced. As a result, the IRT point cloud can be scaled by s_r to match the scale of the RGB one. We can note that the IRT camera positions in $\mathbf{C}^{irt}$ are also affected by s_r. The new scaled set of IRT camera pose is denoted as $\tilde{\mathbf{C}}^{irt}$.

Since $s_r \approx 1$ in practice, for clarity purpose, we illustrate the scale normalization in Fig. 5 with two point clouds generated by different SfM softwares leading to more distinctive arbitrary reconstruction scales.

4.2. Global Registration

The global registration can be performed by applying a rigid body transform to the IRT point cloud since the two point clouds have the same scale after the normalization.



(a) (b)

Figure 5. Normalization of the IRT point cloud generated with COLMAP and the RGB one with VisualSFM [27] for the dataset *Shovel*. The length of the wall for both point clouds is represented (a) before normalization and (b) after normalization.

The relationship between the poses of the two cameras as well as the ambiguity of their relative translation are considered to define this transformation.

4.2.1 Rigid Body Transformation

The coarse registration process is based on the fixed relative pose between the two cameras, which was computed apart during the calibration step. Let N_p be the number of pair of RGB-IRT camera poses $(\mathbf{C}^{rgb}, \mathbf{C}^{irt})$ estimated using SfM. Given an RGB and an IRT point cloud with the same scale after normalization, we can perform the alignment of these two point cloud by applying to the IRT point cloud the rigid body transformation $\mathbf{G}(s_t)$ described as

$$\begin{aligned} \mathbf{G}(s_t) &= \frac{\mathbf{W}^{rel}(s_t)}{N_p} \sum_i^{N_p} \mathbf{C}_i^{rgb} (\tilde{\mathbf{C}}_i^{irt})^{-1} \\ \text{with } \mathbf{W}^{rel}(s_t) &= \begin{bmatrix} \mathbf{R}^{rel} & s_t \mathbf{t}^{rel} \\ 0 & 1 \end{bmatrix} \end{aligned} \tag{5}$$

where s_t is the relative translation scale to be found, and $\mathbf{W}^{rel}(s_t)$ is the rigid transformation defined by the relative rotation $\mathbf{R}^{rel}$ and position $\mathbf{t}^{rel}$ retrieved during the camera rig calibration. An interpretation of the rigid body transformation $\mathbf{G}(s_t)$ is depicted in Fig. 6 for i^{th} IRT camera pose. Nevertheless, the IRT point cloud will only have the correct orientation, as the translation in $\mathbf{W}^{rel}(s_t)$ is defined with an arbitrary scale. By modifying s_t, the point cloud is translated in the direction defined by $\mathbf{t}_{rel}$. Consequently, to perform the correct rigid body transformation, we can estimate the relative translation scale by maximizing the number of overlapping points between the RGB and IRT point clouds.

It can be noted that in Eq. (5), the transformation between the IRT camera to the RGB camera needs to be optimized with the N_p pairs to reduce the impact of the pose

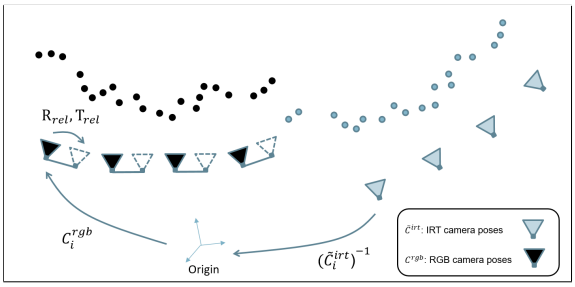


Figure 6. Interpretation of the rigid body transform defined by the computed camera poses and the relative pose between the two cameras.

error from the structure from motion algorithm. We used the arithmetic mean of the estimated poses for simplicity but it could be further improved by considering the mean in the rotation group for the orientation of the cameras [16].

4.2.2 Relative Translation Scale Estimation

By denoting $\tilde{\mathbf{X}}_i^{irt}$ as the homogeneous vector of the ith point of the normalized IRT point cloud, and $\mathbf{X}_j^{rgb}$ as the homogeneous vector of the jth point of the original RGB point cloud, we propose a scoring function which determines how well the RGB and IRT point clouds overlap. The main concept of our approach is to locally measure, within a certain radius, the distances from one point of a point cloud to its neighboring points belonging to the other point cloud. By generalizing this process to all points of the first point cloud, the function to be maximized can be computed as

$$\begin{aligned} &\arg\max_{s_t} \sum_i \sum_j f\big(\mathbf{G}(s_t)\tilde{\mathbf{X}}_i^{irt}, \mathbf{X}_j^{rgb}\big), \\ &\text{with } f(\mathbf{x}_1, \mathbf{x}_2) = \begin{cases} \dfrac{1}{1+d^2} & \text{if } d \leq R_{th} \\ 0 & \text{if } d > R_{th} \end{cases}, \end{aligned} \tag{6}$$

where the function f computes a score based on the Euclidean distance d between two 3D points ($\mathbf{x}_1$, $\mathbf{x}_2$) and a threshold R_{th}.

Since only the global structure of the two points cloud will be similar due to the properties of each camera, the determination of R_{th} is important. If R_{th} is too big, even though there is a weight based on the distance, many distant neighbors j will be included into the score of the considered point i leading to false maximum of the total score. On the other hand, if R_{th} is too small, the lack of local overlapping will also impact the score. Furthermore, it can be noted that every 3D reconstruction is computed with an arbitrary scale, thus R_{th} should be chosen accordingly.

For these reasons, by denoting $S(x_c, R_{th})$ as the interior points of the sphere centered in x_c of radius R_{th}, we propose to define R_{th} as the maximum radius of the sphere so that it does not cover more than 5% of the RGB point cloud when centered at any point of the latter. In other words, the following equation must hold

$$\begin{aligned} &\forall x_c \in \mathbf{X}^{rgb}: \\ &\frac{Card(\{x \mid x \in \mathbf{X}^{rgb} : x \in S(x_c, R_{th})\})}{Card(\mathbf{X}^{rgb})} \leq 0.05 \end{aligned} \tag{7}$$

This way, we can ensure that the global and the local area of the two point clouds are covered while overcoming the arbitrary scale.

On Fig. 7, we show the score of the dataset *Facade* in function of the relative translation scale. The result of the coarse registration using the rigid body transformation corresponding to its maximum is depicted in Fig. 8.

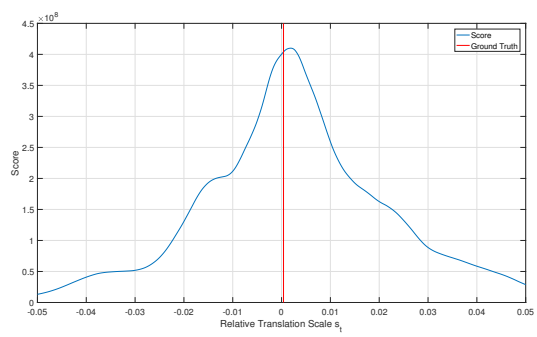


Figure 7. Score in function of the scale s_t (blue) to find and the ground truth (red) of the dataset *Facade*.

4.3. Local Registration

Once the rigid body transform is applied on the normalized IRT 3D reconstruction, the two point clouds will not always be perfectly aligned. There will still be errors that can arise from unreliable calibration, simplification of the model, inaccurate scale estimation, *etc*. To reduce these errors, the *Sparse Iterative Closest Point* (SICP) algorithm proposed by Bouaziz *et al*. [5] is applied. This variant of the well-known *Iterative Closest Point* (ICP) algorithm solves the issues related to outliers and missing data by formulating the registration optimization using sparsity inducing norms. Estimating an optimal rigid alignment for noisy and incomplete geometry is important in our application since the RGB and IRT point cloud may be completely different in certain areas due to the nature of the cameras.

The traditional two-step optimization of the ICP algorithm using the l_2 norm is reformulated using l_p norms, where $p \in [0, 1]$ as follows [5]

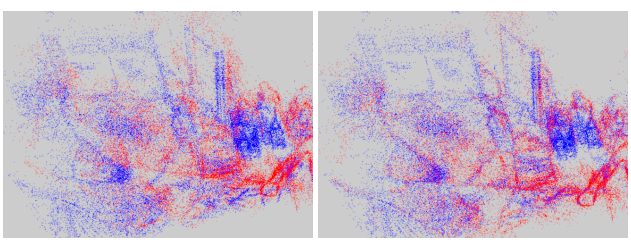
(a) (b) (c)

Figure 8. Global registration of the dataset *Facade* using the optimal relative translation scale. (a) RGB point cloud (b) IRT pointcloud (c) Alignment result (red) IRT point cloud (blue) RGB point cloud.

$$
\begin{aligned}
&1)\ \arg\min_{\mathbf{Y}} \sum_i \left\| \mathbf{L}\mathbf{G}(s_t)\tilde{\mathbf{X}}_i^{irt} - y_i \right\|_2^p \ \text{for } y_i \in \mathbf{X}^{rgb} \\
&2)\ \arg\min_{\mathbf{L}} \sum_i \left\| \mathbf{L}\mathbf{G}(s_t)\tilde{\mathbf{X}}_i^{irt} - y_i \right\|_2^p ,
\end{aligned}
\tag{8}
$$

where $\mathbf{L}$ is a rigid body transformation that registers the RGB and IRT point cloud, and $\mathbf{Y}$ is a set of point in $\mathbb{R}$ that has the same number of element as $\mathbf{X}^{rgb}$. Each point y_i in Y represents the closest point in $\mathbf{X}^{rgb}$ to the transformed point $\mathbf{L}\mathbf{G}(s_t)\tilde{\mathbf{X}}_i^{irt}$. In Eq. (8), the l_p norm can be interpreted as a penalty associated to the residual, *i.e.*, residuals with higher value will have less impact on the optimization problem when p is small. This way, large amount of outliers can be robustly handled.

An example of the local registration effect is depicted in Fig. 9, where the errors due to an inaccurate extrinsic calibration are corrected after using the SICP algorithm.



(a) (b)

Figure 9. Fine (blue) RGB and (red) IRT point cloud registration of the dataset *desk* using the SICP algorithm. (a) Before the algorithm (b) after the algorithm.

5. Experimental Results

In this section, we demonstrate applications using the registration of RGB and IRT point clouds and show its accuracy by using the projection of the aligned point clouds. Three different datasets including outdoor and indoor scenes are presented: *Facade*, *Shovel*, and *Desk*. The datasets have been captured with a optris PI 640, as the thermal camera, and a PointGrey Flea3, as the visible camera.

5.1. Dense Thermal to Visible Image Projection

Using the same transformation employed to align the sparse IRT point cloud to the RGB one, we can also register the dense thermal reconstruction to the latter. By projecting the aligned IRT point cloud to an RGB camera knowing its parameters, it is possible to superimpose thermal information on a visible image. We show this projection for the datasets *Facade* and *Shovel* in Fig. 10, where the RGB point cloud has been omitted for clarity purpose.

5.2. Multi-Sensor Image Synthesis

Another application using the transformation computed during the sparse point cloud registration is to align dense RGB and IRT 3D reconstructions. The Figure 11 shows the projection of both dense point clouds to a virtual camera for the dataset *Desk*. This way, a new multi-sensor (RGB-IRT) image can be synthesized.

6. Discussion

In this paper, we propose a method to register RGB and thermal point cloud generated by SfM. The proposed algorithms are evaluated on three different datasets containing indoors and outdoors environment, where the accuracy of the registration is illustrated with projections of the aligned point clouds. Even though the SfM framework is easy to perform, a limitation is that it requires sufficient detected features. This is especially difficult for thermal images as the temperature difference can be weak in some regions. Image normalization method coupled with low features detection threshold can be used to increase the number of features. However, this will lead to more noise in the 3D reconstructions. Still, the proposed method is able to register two incomplete or noisy point clouds as long as there are pair of RGB-IRT camera poses computed simultaneously. Our future work includes a real-time variant of the proposed algorithm based on SLAM techniques, and the possibility to improve a 3D model at night by registering a thermal point cloud to the incomplete RGB one.

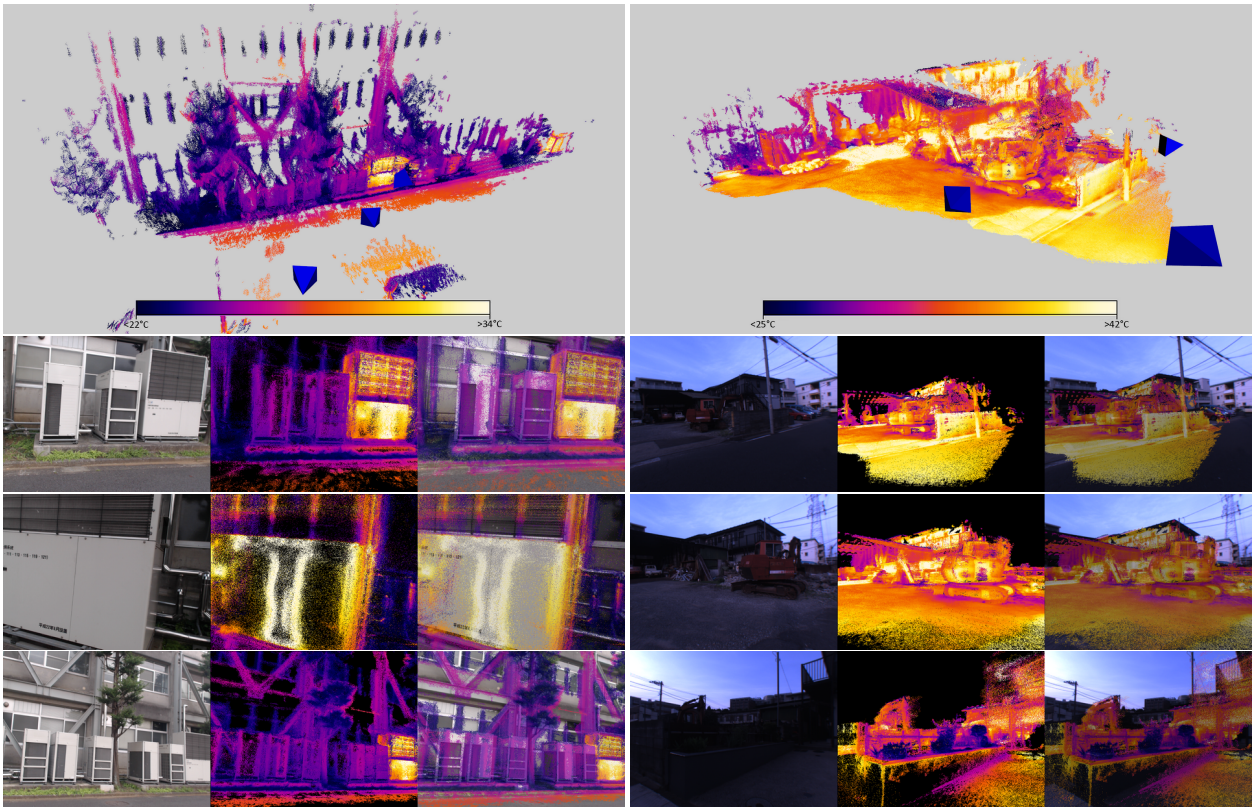


Figure 10. (Top) Dense thermal point cloud of the dataset (first column) *Facade* and (second column) *Shovel*. (Each row from left to right) Visible image, projected dense thermal point cloud on the RGB camera, and fused visible image and projected thermal information.

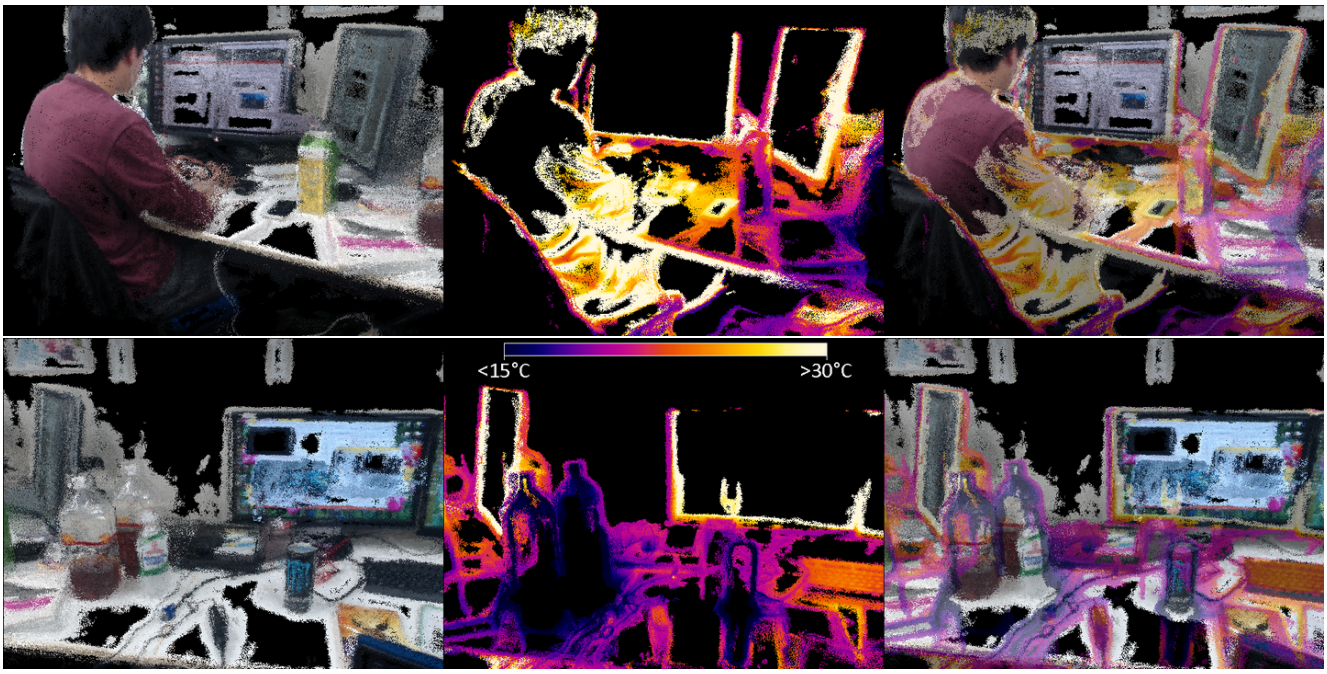


Figure 11. Example of multi-sensor image synthesis for the dataset *Desk*. (Each row from left to right) Projected dense RGB point cloud, projected dense IRT point cloud, and fused RGB and thermal information.

References

[1] G. Alenyà, S. Foix, and C. Torras. Using tof and rgbd cameras for 3d robot perception and manipulation in human environments. *Intelligent Service Robotics*, 7(4):211–220, 2014.

[2] N. Arora, D. Martins, D. Ruggerio, E. Tousimis, A. J. Swistel, M. P. Osborne, and R. M. Simmons. Effectiveness of a noninvasive digital infrared thermal imaging system in the detection of breast cancer. *The American Journal of Surgery*, 196(4):523–526, 2008.

[3] C. Beder and R. Steffen. Determining an initial image pair for fixing the scale of a 3d reconstruction from an image sequence. In *Joint Pattern Recognition Symposium*, pages 657–666. Springer, 2006.

[4] D. Borrmann, A. Nüchter, M. Đakulović, I. Maurović, I. Petrović, D. Osmanković, and J. Velagić. A mobile robot based system for fully automated thermal 3d mapping. *Advanced Engineering Informatics*, 28(4):425–440, 2014.

[5] S. Bouaziz, A. Tagliasacchi, and M. Pauly. Sparse iterative closest point. In *Computer graphics forum*, volume 32, pages 113–123. Wiley Online Library, 2013.

[6] G. Bradski and A. Kaehler. *Learning OpenCV: Computer vision with the OpenCV library*. " O'Reilly Media, Inc.", 2008.

[7] J. Cramer. Automatic generation of 3d thermal maps of building interiors. *ASHRAE transactions*, 120:C1, 2014.

[8] A. De la Escalera and J. M. Armingol. Automatic chessboard detection for intrinsic and extrinsic camera parameter calibration. *Sensors*, 10(3):2027–2044, 2010.

[9] F. Devernay and O. Faugeras. Straight lines have to be straight. *Machine vision and applications*, 13(1):14–24, 2001.

[10] Y. Ham and M. Golparvar-Fard. An automated vision-based method for rapid 3d energy performance modeling of existing buildings using thermal and digital imagery. *Advanced Engineering Informatics*, 27(3):395–409, 2013.

[11] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[12] L. Hoegner, S. Tuttas, Y. Xu, K. Eder, and U. Stilla. Evaluation of methods for coregistration and fusion of rpas-based 3d point clouds and thermal infrared images. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 41, 2016.

[13] R. Horaud, M. Hansard, G. Evangelidis, and C. Ménier. An overview of depth cameras and range scanners based on time-of-flight technologies. *Machine Vision and Applications*, 27(7):1005–1020, 2016.

[14] J. Johansson, M. Solli, and A. Maki. An evaluation of local feature detectors and descriptors for infrared images. In *ECCV Workshops (3)*, pages 711–723, 2016.

[15] S. Lagüela, J. Armesto, P. Arias, and J. Herráez. Automation of thermographic 3d modelling through image fusion and image matching techniques. *Automation in Construction*, 27:24–31, 2012.

[16] M. Moakher. Means and averaging in the group of rotations. *SIAM journal on matrix analysis and applications*, 24(1):1–16, 2002.

[17] S. Prakash, P. Y. Lee, T. Caelli, and T. Raupach. Robust thermal camera calibration and 3d mapping of object surface temperatures. *SPIE Proceedings: ThermoSense XXVIII*, 6205:62050J, 2006.

[18] P. Ricaurte, C. Chilán, C. A. Aguilera-Carrasco, B. X. Vintimilla, and A. D. Sappa. Feature point descriptors: Infrared and visible spectra. *Sensors*, 14(2):3690–3701, 2014.

[19] R. B. Rusu and S. Cousins. 3d is here: Point cloud library (pcl). In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 1–4. IEEE, 2011.

[20] P. Saponaro, S. Sorensen, S. Rhein, and C. Kambhamettu. Improving calibration of thermal stereo cameras using heated calibration board. In *Image Processing (ICIP), 2015 IEEE International Conference on*, pages 4718–4722. IEEE, 2015.

[21] J. L. Schönberger and J.-M. Frahm. Structure-from-motion revisited. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[22] J. L. Schönberger, E. Zheng, M. Pollefeys, and J.-M. Frahm. Pixelwise view selection for unstructured multi-view stereo. In *European Conference on Computer Vision (ECCV)*, 2016.

[23] S. Vidas, R. Lakemond, S. Denman, C. Fookes, S. Sridharan, and T. Wark. A mask-based approach for the geometric calibration of thermal-infrared cameras. *IEEE Transactions on Instrumentation and Measurement*, 61(6):1625–1635, 2012.

[24] S. Vidas, P. Moghadam, and M. Bosse. 3d thermal mapping of building interiors using an rgb-d and thermal camera. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 2311–2318. IEEE, 2013.

[25] M. Vollmer, M. Klaus-Peter, et al. *Infrared thermal imaging: fundamentals, research and applications*. John Wiley & Sons, 2010.

[26] M. Weinmann, J. Leitloff, L. Hoegner, B. Jutzi, U. Stilla, and S. Hinz. Thermal 3d mapping for object detection in dynamic scenes. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2(1):53, 2014.

[27] C. Wu. Towards linear-time incremental structure from motion. In *3DTV-Conference, 2013 International Conference on*, pages 127–134. IEEE, 2013.

[28] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, 22(11):1330–1334, 2000.