

Fusion of TLS and RGB point clouds with TIR images for indoor mobile mapping

by L. Hoegner*, T. Abmayr**, D. Tosic*, S. Turzer**, U. Stilla*

**Technical University Munich, Photogrammetry and Remote Sensing, Arcisstrasse 21, 80333 Munich, Germany.*
Ludwig.hoegner@tum.de, draganaa.tosic@gmail.com, stilla@tum.de

***University of Applied Sciences Munich, Department of Geoinformation, Karlstrasse 6, 80333 Munich, Germany,*
thomas.abmayr@hm.edu, turzer@hm.edu

Abstract

Obtaining accurate 3D descriptions in the thermal infrared (TIR) is a quite challenging task due to the low geometric resolutions of TIR cameras and the low number of strong features in TIR images. Combining the radiometric information of the thermal infrared with 3D data from another sensor is able to overcome most of the limitations in the 3D geometric accuracy. In case of dynamic scenes with moving objects or a moving sensor system, a combination with RGB cameras and profile laserscanners is suitable. As a laserscanner is an active sensor in the visible red or near infrared (NIR) and the thermal infrared camera captures the radiation emitted by the objects in the observed scene, the combination of these two sensors for close range applications are independent from external illumination or textures in the scene. This contribution focusses on the fusion of point clouds from terrestrial laserscanners and RGB cameras with images from thermal infrared mounted together on a robot for indoor 3D reconstruction. The system is geometrical calibrated including the lever arm between the different sensors. As the field of view is different for the sensors, the different sensors record the same scene points not exactly at the same time. Thus, the 3D scene points of the laserscanner and the photogrammetric point cloud from the RGB camera have to be synchronized before point cloud fusion and adding the thermal channel to the 3D points.

1. Introduction

In building inspection geometric and radiometric properties are both important. For geometric accuracy, both point clouds from terrestrial laserscanners (TLS) and photogrammetric stereo reconstruction from RGB images can be used. In photogrammetry and computer vision a variety of methods are well developed for 3d reconstruction from ordered [2] and unordered [3,4] image sequences. These methods are limited to structured surfaces with features that can be detected as Homologous points through the sequences. As they operate in the visible spectrum, they are also dependent on the external lighting conditions. The detectors of features and descriptors of homologous points like SIFT [5], Förstner [6], and Harris [7] are based on radiometric similarity of homologous points. This is only valid, if the compared images are within the same spectral domain. In general, in the thermal infrared lines and edges do not show strong edges but appear blurred. The radiometric behaviour of features is different from the visible spectrum. These effects cause mismatches between features in the thermal infrared and visible domain and reduce the accuracy of object detection and extraction in infrared images. . A coregistration of images from the visible and thermal infrared domain based on segmentation has been introduced by Coiras et al. [8]. Park et al. [9] combine different spectral bands using so called transinformation.

In contrast to 3D reconstruction from images, methods based on runtime measurements with active sensors are independent from textures and corresponding points in several images. Laser scanners are recording weakly textured surfaces. The combination of laser scanners with cameras has already been introduced in industrial products. The sequential scanning principle delimitates laser scanners to static scenes. Dynamic scenes can be recorded with time-of-flight cameras recording depth values parallel for all elements of a detector matrix. The result is an intensity image in the near infrared and a depth image showing the distances of the observed object for every pixel of the image with a recording rate of several images per second [10]. Coregistration for TOF cameras and RGB images is done calculating the relative orientation in a bundle adjustment with homologous points [11] due to the fact, that the radiometric behaviour in near infrared and visible light is almost the same. Wang et al. [12] investigate foreground background separation from combining TOF depth values and RGB values, both recorded by one camera system. Due to the RGB camera, they learn a likelihood classification for foreground and background colors. In case of a thermal camera, the temperature of a person is known and so a fixed threshold can be used instead. In contrast to [12] a more complex geometric calibration has to be done for TOF and TIR cameras as to different optics are used and so a relative orientation has to be calculated [13].

A fusion of TIR images and point clouds from other sources allows calculating a more accurate 3D localisation of features detected in TIR images. To achieve this, a geometric calibration of the TIR camera is necessary that can be done following the strategy for cameras in the visible RGB [14], where the limited number of pixels compared to RGB cameras leads to lower accuracies for the intrinsic parameters and only a limited set of distortion parameters to be significant. In general, known 3D coordinates of the recordings are necessary for the fusion. This is done by GPS reference of the recordings or ground control points to generate a common coordinate system for all sensors. There is no external orientation via GPS or other systems available for indoor scenes. In such cases, simultaneous location and mapping algorithms (SLAM) are used to reconstruct the 3D scene and localise the sensor in the scene [15]. This leads to a relative coordinate system for every sensor. To fuse these coordinate systems, either the separately generated 3D point clouds of



the different sensors are coregistered or the different sensors have to be put on a common platform with known fixed lever arms between the sensors. In terms of this project, exploration of strategies for fusion of 3D point clouds acquired by laser scanner and stereo camera with thermal imagery was done. To complete the task, registration of indoor movements (localization) was done with help of a scanning laser range finder. Data from all the sensors were processed and manipulated in Matlab. Data from some of the sensors was also recorded directly with help of Matlab (Hokuyo laser range finder), whereas for recording data from Zoller und Fröhlich lasers scanner, as well as the thermal camera connected to it the software Z+F LaserControl was used. Point clouds and images of the ZED stereo camera were recorded with help of the ZED - Software development kit (ZED – SDK). The point clouds are fused to one point cloud that is afterwards extended by thermal infrared intensities from the thermal infrared images.

2. Geometric calibration of the cameras and the lever arms

As a reference system for the registration of all the sensors Z+F laserscanner data were used. The calibration of 2D laser range finder as well as ZED stereo camera to Z+F laser scanner (and therefore to each other) was done with help of targets, visible from every set of data (Figure 1). For calibrating the thermal camera to the ZED point cloud, an approach consisted of the Zhang's calibration method [17] and solving the so-called $AX = XB$ problem [18] was used. Angle increments of the vertical unit and the horizontal rotation of the scanner were used as fixed input parameters. Once the vertical position of the tilt-unit is fixed, the camera rotates on a circular path around the scanner.



Fig. 1: Targets used for sensor registration.

2.1 Definition of the used coordinate systems

For the explanation of the model, we will first mention the defined coordinate systems (Figure 2). An affine, orthogonal and right-handed coordinate system is denoted through the Quadruple

$$K: (\mathbf{a}, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) \quad (1)$$

with the translation \mathbf{a} and the basis vectors $\mathbf{e}_i \in \mathbb{R}^3$. Based on this notation, let the $\mathbf{k}_1 \in \mathbb{R}^3$, $\|\mathbf{k}_1\| = 1$ be the elevation rotation axis, and respectively, take $\mathbf{k}_3 \in \mathbb{R}^3$, $\|\mathbf{k}_3\| = 1$ for the azimuth rotation axis. If we assume that both axes intersect at $(\mathbf{k}_1, \mathbf{k}_3) = \mathbf{0}$, we can construct an affine, orthogonal and right-handed coordinate system of the laserscanner through

$$K^{(s)} = (\mathbf{0}, \mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \quad (2)$$

If we set the coordinate system as described, then the ideal sensor rotates the laser beam around the x-axis \mathbf{k}_1 . Additionally to this vertical rotation, the system rotates horizontally around the z axis \mathbf{k}_3 . In further notation, \mathbf{Z}_α describes the horizontal rotation matrix and \mathbf{X}_β the vertical rotation matrix.

The camera coordinate system is defined through

$$K^{(c)} = (\mathbf{w}_0, \mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3) \quad (3)$$

which describes the position of the camera with respect to the scanner. If the camera can be described as an ideal pinhole camera, the origin \mathbf{w}_0 is equal to the optical centre of the camera, the third vector \mathbf{w}_3 is orthogonal to the image plane, and \mathbf{w}_1 is parallel to the horizontal border of the image plane, and respectively \mathbf{w}_2 to its vertical border.

Based on this, a third coordinate system is necessary to describe the lever arm of the scanner and the camera through

$$K^{(u)} = (\mathbf{m}_0, \mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3) \quad (4)$$

which is called camera-tilt-unit coordinate system. This system holds the properties that the origin \mathbf{m}_0 is equal to the centre of the rotation axis of the tilt-unit and the direction of the first vector \mathbf{m}_1 is along the direction of the rotation axis.

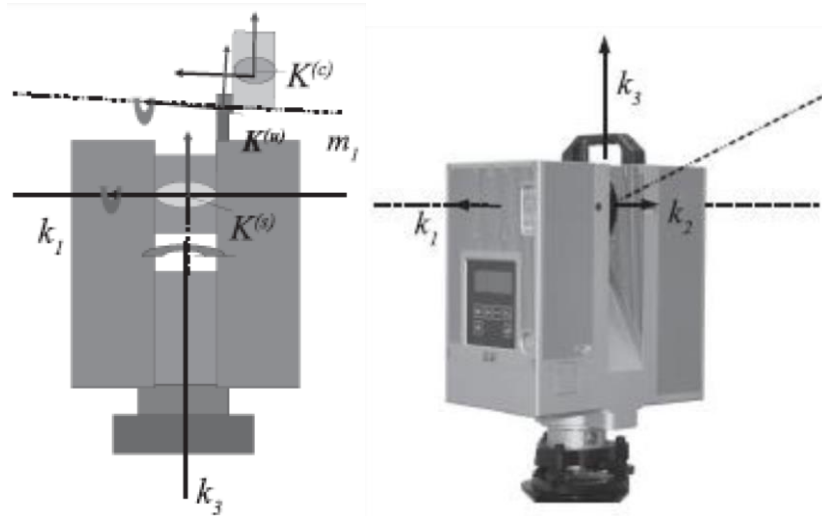


Fig. 2: Coordinate system of the polar measurement unit and tilt unit

Here, we use the Optris thermal camera with a tilt unit mounted on the laser scanner. Once the tilt unit and the camera are mounted on the scanner device and the relation between the sensors is calculated, the extrinsic parameters of this combination are then used to register thermal camera images with the ZED – point cloud (Figure 3). The angle increments of the vertical unit and the horizontal rotation of the scanner are assumed to be highly accurate and are used as inputs parameters. This significantly reduces the degrees of freedom for the external camera parameters. For modelling the intrinsic parameters of the camera, we use Zhang's camera model, which is based on the pinhole camera of perspective projection.

The overall projection from a point $X := (x, y, z)$ of the scanner coordinate system $K^{(s)}$ onto the pixel (u, v) in the image taken on horizontal scanner position α and vertical tilt-unit position β , can be written as concatenation of function as

$$\Xi_{\alpha, \beta}(X) = v_{k, \lambda} \circ \varphi_{S, C} \circ P \circ T_{\alpha, \beta}(X) \quad (5)$$

with

$$T_{\alpha, \beta}(X) = M X_{\beta} \tilde{M} Z_{\alpha} \quad (6)$$

where $T_{\alpha, \beta}$ defines the transformation from the scanner coordinate system $K^{(s)}$ into the camera coordinate system $K^{(c)}$. In particular Z_{α} describes the rotation of the scanner around its vertical axis around α into the scanner coordinate system $K_{\alpha}^{(s)}$ at position α ; Two rigid motions are needed to transform from there into the camera coordinate system $K^{(c)}$. \tilde{M} describes the basis change from $K_{\alpha}^{(s)}$ into the camera-tilt-unit coordinate system $K^{(u)}$; M describes the transformation from the camera-tilt-unit coordinate system $K^{(u)}$ to the camera coordinate system $K^{(c)}$. The perspective projection $P: \mathbb{R}^3 \rightarrow \mathbb{R}^2$ from the $K^{(c)}$ onto the normalized image plane is defined for $z \neq 0$ by

$$P(x, y, z) = \begin{pmatrix} x/z \\ y/z \end{pmatrix} \quad (7)$$

Set $\vec{c} := (c_x, c_y)$ and $\vec{s} := (s_x, s_y)$. Then is the mapping $\varphi_{\vec{c}, \vec{s}}: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ from normalized image coordinates to real image coordinates defined through:

$$\varphi_{\vec{c}, \vec{s}}(p_n, q_n) = c_x + s_x p_n, c_y + s_y q_n \quad (8)$$

The parameters \vec{c} describe the optical center of the camera and the scaling factors \vec{s} its focal length. Final mapping from undistorted image coordinates (p_r, q_r) into distorted coordinates, with consideration of the distortion is then

$$v_{\vec{k}, \vec{\lambda}}(p_r, q_r) = \begin{pmatrix} p_r \sum_{i=0}^2 k_i r^{2i} + 2\lambda_1 p_r, q_r + \lambda_2 (r^2 + 2p^2 r) \\ q_r \sum_{i=0}^2 k_i r^{2i} + 2\lambda_2 p_r, q_r + \lambda_1 (r^2 + 2q^2 r) \end{pmatrix} \quad (9)$$

where $\sqrt{r} = \sqrt{p_r^2 + q_r^2}$, $\vec{k} = (1, k_1, k_2)$ and $\vec{\lambda} = (\lambda_1, \lambda_2)$.

Considering that M and \tilde{M} are homogenous matrices and thus can be described through 6 parameter each, we get altogether 20 unknowns for solving equation 5. According to the setup, the rotation angles α and β define the actual position of the azimuth encoder of the scanner and the rotation angle of the tilt unit respectively and are assumed to be known. The parameters $\vec{k} = (1, k_1, k_2)$, $\vec{\lambda} = (\lambda_1, \lambda_2)$, $\vec{c} = (c_x, c_y)$ and $\vec{s} = (s_x, s_y)$ are camera-specific and are called intrinsic camera parameters, whereas M and \tilde{M} depend on the position of the camera and its tilt unit with respect to the scanner coordinate system and are called extrinsic camera parameters.

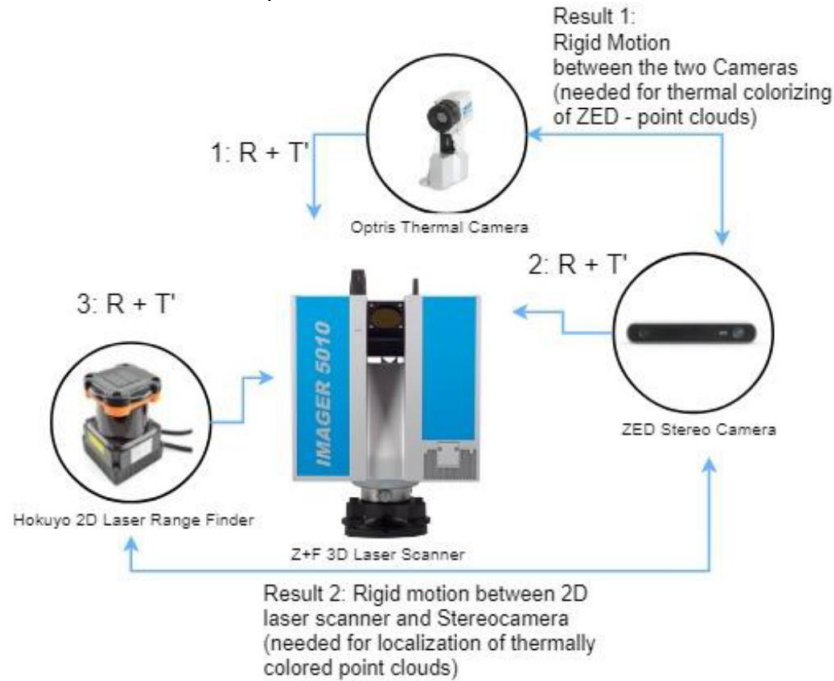


Fig. 3: Calibration of Devices to the common Coordinate System

2.2 Coregistration of laserscanner point cloud and camera images

We assume the intrinsic parameters of the TIR camera [14] and the RGB stereo camera [15] to be known. The relative orientation of the used sensors (laserscanner, TIR camera, RGB stereo camera, profile laserscanner) is assumed to be fixed by mounting all sensors onto a platform. The parameters of the orientation are determined by finding homologous image points in the intensity images of the laserscanner and the TIR and RGB images. As the relative orientation is fixed, manually chosen points are used to calculate the parameters of the orientations of the cameras in the coordinate system $K^{(s)}$ of the laserscanner. Following the notation introduced in chapter 2.1, each pixel (p, q) in the laserscanner intensity image can be assigned to a 3D coordinate $X_{pq} \in \mathbb{R}^3$. With equation 5 we get the corresponding pixel in the camera through

$$(i, j) = \Xi_{\alpha, \beta}(X_{pq}) \quad (10)$$

Applying equation 10 for all points $(p, q) \in S$ transforms an image I into the view of the laserscanner image S through

$$\widetilde{I}_{pq} := I_{ij} \quad (11)$$

and is called the backward projection of S . The combined dataset is then a coloured image of the laserscanner point cloud. In contrast, equation 6 is applied to transform X_{pq} into the camera coordinate system $K^{(c)}$ through $Y_{pq} := MX_{pq}M^T Z_{\alpha}$. If $d_{pq} := \|Y_{pq}\|$ is the distance of Y_{pq} to the origin $K^{(c)}$, then

$$D_{ij} = d_{pq} \quad (12)$$

is called forward projection of S for all points $(p, q) \in S$. In the camera image, intensity values for all projected pixels (p, q) are calculated for every band of the camera image by bilinear interpolation. These intensity values are transferred from the projected pixels back to the 3D points X_{pq} .

Combining the forward and backward projections directly connects the RGB stereo camera and the TIR camera in the coordinate system of the laserscanner. Like for the laserscanner points cloud, the point cloud of the RGB stereo camera can now be projected into the TIR images and thus thermal intensity values are interpolated for the photogrammetric stereo based point cloud.

3. Experiments

The experimental setup (Figure 4) uses a Z+F IMAGER® 5010, 3D laserscanner, a ZED stereo camera, a Hokuyo laser range finder UTM-30LX, and an Optris thermal bolometer camera with 382x288 pixels in the spectral range of 7.5 to 13 μm . The thermal camera is mounted directly on the laserscanner with a special mount and the ZED stereo camera and the Hokuyo range scanner are mounted onto a crossbar that is connected to the laserscanner.



Fig. 4: Sensor rig with all sensors. Relative orientation of the TIR and RGB camera and the laser range scanner with the Z+F laserscanner as origin of the coordinate system

The trajectory of the whole construction (Figure 4) was determined with help of Hokuyo 2D laser range finder. Prior to using it for the given task, a registration of this device with 3D laser scanner was done in similar manner – by using 3D object to mark the targets and enable them to be observed by 2D laser was done. The accuracy of this registration is shown in table 1. Four 3D point correspondences have been taken to evaluate the remaining orientation error. Figure 5 shows the profile scanned by the Hokuyo range finder with the 2D profile before (red) and after (green) coregistration with the point cloud of the laserscanner. Figure 6 shows the laserscanner point cloud and in red the coregistered profile. One can see that the profile fits the point cloud at the edge of the room and at the door.

Table 1. Accuracy of Hokuyo - Z+F Registration

Point (X,Y,Z)	Accuracy of Hokuyo – Z+F Registration [m]		
	X	Y	Z
P1	0.0073	-0.0358	-0.001
P2	-0.0229	-0.0078	0.0049
P3	0.0067	0.0201	0.0011
P4	0.0089	0.0235	-0.005

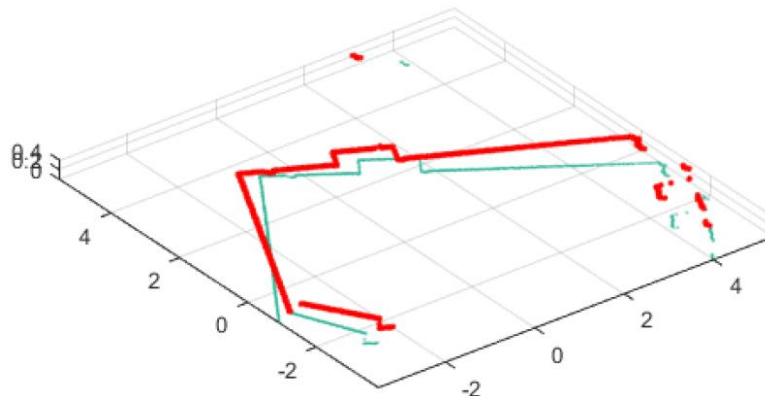


Fig. 5: Position of Hokuyo points in the coordinate system of the Z+F laserscanner. 2D profile with red: before the registration, green: after the coregistration.

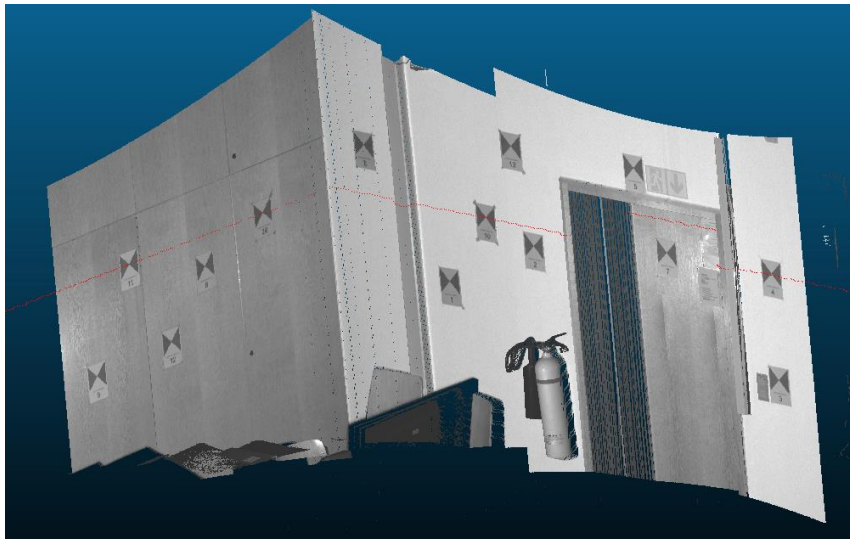


Fig. 6: Position of Hokuyo points in the coordinate system of the Z+F laserscanner. 3D coregistration: The red line shows the profile line of the laser range finder. One can see that the profile of the edge of the room and the doors with to the 3D point cloud.

For the trajectory determination, a known approach of Iterative Closest Point [19] was used to match each set of the scanned data (2D point cloud) with the previous scan and in this way tracking of the devices was depicted (Figure 7).

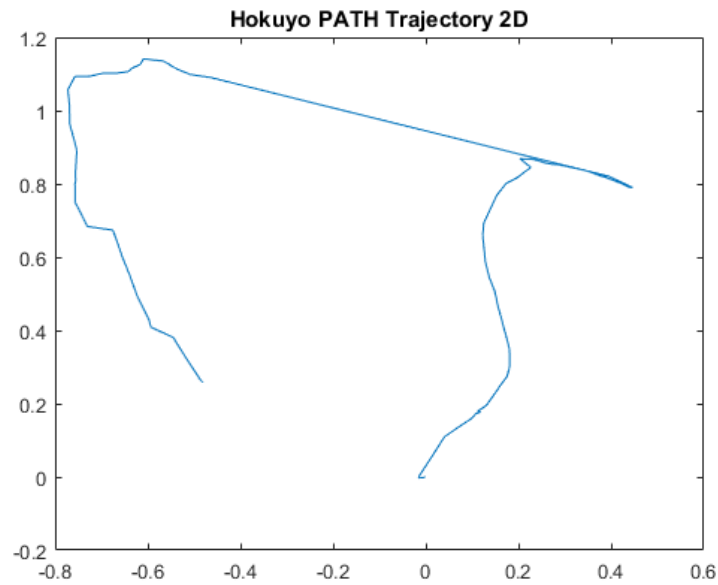


Fig. 7: Trajectory determined by 2D laser range finder.

Figure 8 shows a stereo image pair of the ZED camera and figure 9 the corresponding Optris TIR image. The control points are visible only in the stereo images but not in the thermal image. The installation channel is warm compared to the wall and visible in both cameras.



Fig. 8: left and right image of a stereo pair of the ZED camera. Control points included. One can see that the corner of the room is left of the installation channel.

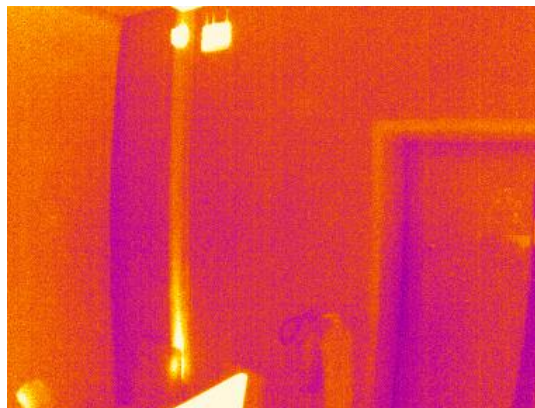


Fig. 9: Optris thermal image corresponding to the stereo image pair in fig. 7. One can see the warm installation channel and the WiFi device.

Figure 10 shows a part of the recorded scene with the point cloud generated from the stereo camera system. The left image shows the original RGB intensities of the ZED camera. The right image shows the intensity values by the interpolated thermal intensities of the coregistered thermal camera. One can see that the warm installation channel is located at the wall and not in the edge of the room. A direct radiometric feature matching would have registered the most prominent lines in the RGB and the TIR image, which would have been the installation channel in TIR and the edge of the room in RGB.

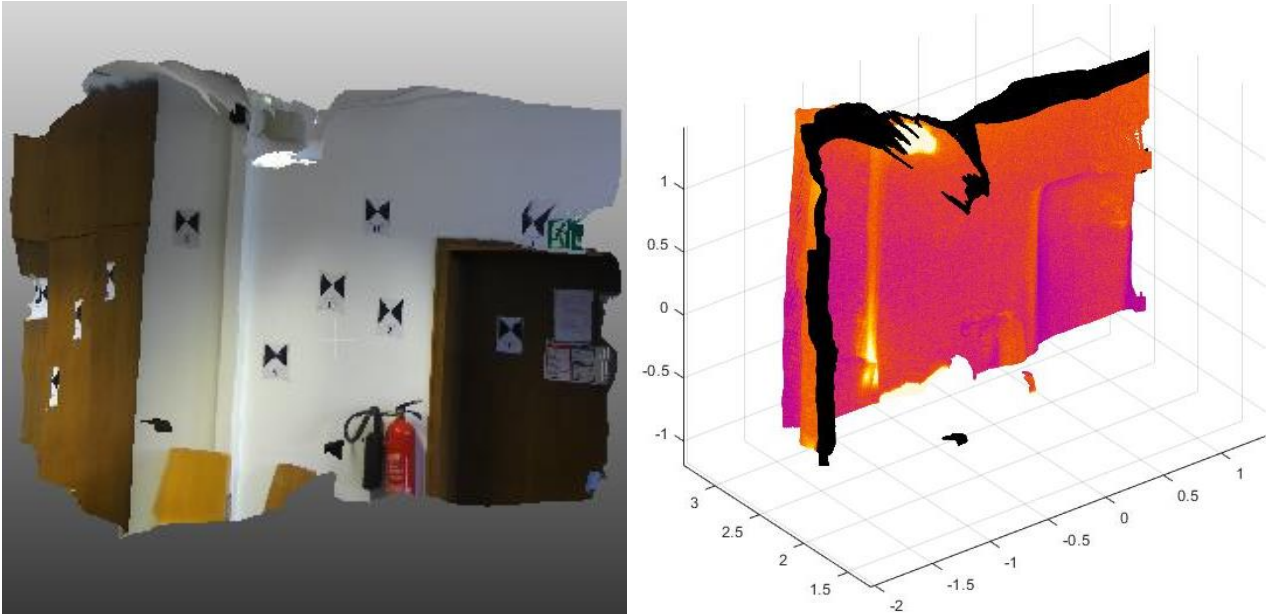


Fig. 10: ZED Stereo camera point cloud. Left: colored with RGB intensities of the ZED camera. Right: colored with Optris thermal image. One can see that the installation channel is warmer and correctly located on the wall and not in the edge of the room.

4. Conclusion

In terms of this project, we connected and registered four different devices (Zoller und Fröhlich 3D scanner, 2D laser rangefinder, ZED stereo camera and Optris thermal camera) with the goal to use this synchronized data for indoor thermal mapping of the point clouds. The laserscanner point cloud and its intensity image were used as master to connect the thermal images and the stereo image point cloud. One of the important points of this work should be, to skip the usage of huge datasets from the Z+F 3D laser scanner and use instead thermally colored point clouds acquired with ZED stereo camera for further work with our constructed platform. The laser range finder is used for localization purposes. Its accuracy is sufficient to replace the laserscanner for that task.

The next steps of our work will be the extension of the localization by including the stereo based point cloud into the location estimation. So far, the points cloud is only used for reconstruction and coregistration with the thermal images.

At the moment, only one thermal image is mapped onto the point clouds of the stereo camera and the laserscanner. For using the whole thermal image sequence, it will be necessary to include an algorithm to determine the geometric quality of every single TIR image for every point of the 3D point clouds to find for every 3D point the thermal image with the highest geometric resolution for intensity value extraction.

REFERENCES

- [1] Hoegner L., Stilla U. Automatic 3D reconstruction and texture extraction for 3D building models from thermal infrared image sequences. QIRT 2016, Gdansk, 04.-08. July 2016. [http://dx.doi.org/ 10.21611/qirt.2016.042](http://dx.doi.org/10.21611/qirt.2016.042).
- [2] Pollefeys M., Nistér D., Frahm J. M., Akbarzadeh, A., Mordohai P., Clipp B., Engels C., Gallup D., Kim S. J., Merrell P., Salmi C., Sinha S., Talton B., Wang L., Yang Q., Stewénius H., Yang R., Welch G., Towles H. Detailed real-time urban 3D reconstruction from video. *International Journal of Computer Vision* vol. 78(2-3), pp. 143-167, 2008.
- [3] Mayer H., Bartelsen J., Hirschmueller H., Kuhn A. Dense 3d reconstruction from wide baseline image sets. In: F. Dellaert, J.-M. Frahm, M. Pollefeys, L. Leal-Taixe and B. Rosenhahn (eds), *Outdoor and Large-Scale Real-World Scene Analysis. 15th International Workshop on Theoretical Foundations of Computer Vision*, Dagstuhl Castle, Germany, June 26 - July 1, 2011, Springer, Berlin, vol. 7474, pp. 285–304, 2012.
- [4] Snavely N., Seitz S.M., Szeliski R. Modeling the World from Internet Photo Collections. *International Journal of Computer Vision* vol. 80 (2), pp. 189-210, 2008.
- [5] Lowe D. G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*. Vol. 60(2), pp. 91-110, 2004.
- [6] Förstner W., Gülch E. A fast operator for detection and precise location of distinct points, corners and centres of circular features. *Proceedings of the ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data*, pp. 281-305, 1987.
- [7] Harris C., Stephens M. A combined corner and edge detector. *Proceedings of the 4th Alvey Vision Conference*, pp. 147-151, 1988.
- [8] Coiras E., Santamaria J., Miravet C. A segment-based registration technique for visual-IR images. *Optical Engineering*, vol. 39, pp. 282-289, 2000.
- [9] Park C., Bae K.-H., Choi S., Jung J.-H. Image fusion in infrared image and visual image using normalized mutual information. *Proceedings of SPIE*, vol. 6968, 69681Q-1-9, 2008.
- [10] Weinmann M., Jutzi B. A step towards dynamic scene analysis with active multi-view range imaging systems. In: Shortis M., Paparoditis N., Mallet C. (Eds.), *XXII ISPRS Congress, Technical Commission III. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 39 (Part B3), pp. 433-438, 2012.
- [11] Hastedt H., Luhmann T. Investigations on a combined RGB / Time-of-Flight approach for close range applications. In: Shortis M., Mills J. (Eds.), *XXII ISPRS Congress, Technical Commission V. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 39 (Part B5), pp. 333-338, 2012.
- [12] Wang L., Gong M., Zhang C., Yang R., Zhang C., Yang Y.H. Automatic Real-Time Video Matting Using Time-of-Flight Camera and Multichannel Poisson Equations. *International Journal of Computer Vision*, vol. 97(1), pp. 104-121, 2012.
- [13] Weinmann M., Leitloff J., Hoegner L., Jutzi B., Stilla U., Hinz S. Thermal 3D mapping for object detection in dynamic scenes External Link. *ISPRS annals*, vol. II-1, pp. 53-60, 2014.
- [14] Luhmann T., Ohm J., Piechel J., Roelfs T. Geometric calibration of thermographic cameras. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 38(5), pp. 411-416, 2010.
- [15] Hartley R, Zisserman A. *Multiple view geometry in computer vision*. Cambridge University Press, 2004, ISBN 0521540518.
- [16] Hirschmueller H. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30(2), pp. 328-341, 2008.
- [17] Zhang Z. Flexible Camera Calibration by Viewing a Plane from Unknown Orientations. *International Conference on Computer Vision*. Corfu, Greece, vol. 3(4), pp. 666 – 673, 1999.
- [18] Abmayr T. *Calibration and Registration Framework for Multisensor Panoramic Color Scanning*. Dissertation. Technische Universität München, 2009.
- [19] Rusinkiewicz S., Levoy M. Efficient variants of the ICP algorithm. In: *Proceedings of the IEEE 3rd International Conference on 3D Digital Imaging and Modeling*, pp. 145–152, 2001.