FUSION OF 3D POINT CLOUDS WITH TIR IMAGES FOR INDOOR SCENE RECONSTRUCTION

L. Hoegner[1,]*T. Abmayr[2], D. Tosic[1], S. Turzer[2], U. Stilla[1]

[1] Photogrammetry and Remote Sensing - Technical University of Munich
(ludwig.hoegner, stilla)@tum.de, draganaa.tosic@gmail.com
[2] University of Applied Sciences Munich, Department of Geoinformation
(thomas.abmayr, turzer)@hm.edu

Commission I

KEY WORDS: photogrammetric point cloud, terrestrial laserscanning, thermal images, indoor, 3D reconstruction

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 (Pollefeys et al., 2008) and unordered (Mayer et al., 2012; Snavely et al., 2008) 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 (Lowe, 2004), Förstner (Förstner and Gülch, 1987), and Harris (Harris and Stephens, 1988) 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 be introduced by Coiras et al. (2000). Park et al. (2008) combine different spectral bands using so called transinformation.

3D reconstruction and texture extraction in thermal infrared are applied for sets of images and ordered terrestrial image sequences (Hoegner and Stilla, 2015) or image sequences taken by a thermal camera mounted on a RPAS (Westfeld et al., 2015; Hoegner

and Stilla, 2018). Both 3D reconstruction and texturing are influenced by various conditions as the thermal radiation of façades depends on temperature differences between inside and outside, weather conditions, and materials. To overcome limitations in the 3D accuracy of thermal infrared based 3D points, a combination of thermal infrared cameras and 3D recording systems like laserscanners (Borrmann et al., 2013) and time-of-flight cameras (Hoegner et al., 2014) or photogrammetric point clouds (Hoegner et al., 2016; Hoegner and Stilla, 2018) is possible.

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 (Weinmann and Jutzi, 2012). Coregistration for TOF cameras and RGB images is done calculating the relative orientation in a bundle adjustment with homologous points (Hastedt and Luhmann, 2012) due to the fact, that the radiometric behaviour in near infrared and visible light is almost the same. Wang et al. (2012) 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 colours. 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 Wang et al. (2012) a more complex geometric calibration has to be done for TOF and TIR cameras as to different optics are

---

*Corresponding author

used an so a relative orientation has to be calculated (Weinmann et al., 2014).

As described in Hoegner et al. (2014), a combination of 3D point clouds and thermal infrared images can be used for building inspections as well. Time-of-flight-cameras have a very limited range for accurate measurements. Laserscanners are either fixed on the ground and thus only scan façades without roofs or are mounted on flying platforms. Depending on the paylod of a RPAS, two sensor can be mounted together with fixed relative orientation or one sensor has to be mounted after the other. In the second case, the limited accuracy in orientation and navigation of the RPAS leads to small differences in the recording positions even for the same planned flight path. Because of the low geometric resolution of TIR cameras compared to RGB cameras, a combination of RGB camera and TIR camera is useful, where the RGB camera provides the hight resolution images for the 3D reconstruction and the TIR camera the thermal information for the 3D points.

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 (Luhmann, 2003), 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 (Hartley and Zisserman, 2004). These leads to a relative coordinate system for every sensor. To fuse these coordinate systems, either the separately generated dense 3D point clouds (Hirschmueller, 2008) 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 SENSOR SYSTEM

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 (Zhang, 1999) and solving the so-called AX = XB problem (Abmayr, 2010) 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.

The laserscanner defines an affine, orthogonal and right-handed coordinate system $K^{(s)} = (0, k_1, k_2, k_3)$ with the perspective centre of the scanner in the origin and k_i defining the orthogonal axes of the coordinate system. If we set the coordinate system as described, then the ideal sensor rotates the laser beam around the x-axis k_1. Additionally to this vertical rotation, the system rotates horizontally around the z axis k_3. In further notation, Z_α describes the horizontal rotation matrix and X_β the vertical rotation matrix. The leverarm $K^{(u)} = (m_0, m_1, m_2, m_3)$ with the translation m_0 and the three orthogonal axes m_1, m_2, m_3 defines the transformation of the camera coordinate system $K^{(c)}$ to the scanner coordinate system. This allows to projecting the scanner point cloud into thermal images and RGB images of the stereo camera system for colouring the point cloud and to transfer the stereo camera based point cloud into the coordinate system of the scanner point cloud for coregistration. The coordinate systems are visualized in figure 1.
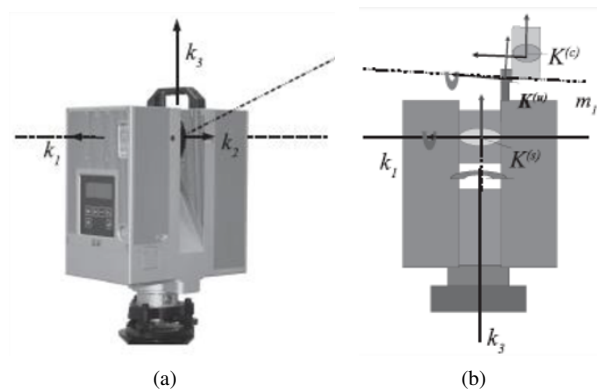

Figure 1. Defined coordinate systems.

The overall projection from a point $P = (X, Y, Z)$ of the scanner coordinate system $K^{(s)}$ onto the image point $p = (u, v)$ in the image taken on horizontal scanner position α and vertical tilt-unit position β, can be written as

$$\Xi_{\alpha,\beta}(X) = \nu_{k,\lambda} \dot{\delta}_{k,\lambda} \dot{P} \dot{T}_{\alpha,\beta}(X) \qquad (1)$$

where

$$T_{\alpha,\beta}(X) = M X_\beta \tilde{M} Z_\alpha \qquad (2)$$

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_α^c at position α. $\tilde{M}$ describes the basis change from K_α^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. Considering that M and $\tilde{M}$ are homogenous matrices, both can be described through 6 parameter each and together describe the exterior orientation of the camera.

The coordinate system K^c is now projected onto the image plane is by a 3D-to-2D projection matrix $P(x, y, z) = (x/z, y/z)$. The

projection on the image plane is defined by 8 parameters including the interior orientation of the camera, so that in total 20 unknowns have to be estimated to solve equation 1.

These parameters are estimated in a two step approach: The interior orientation of the camera is estimated by a geometric camera calibration following Luhmann et al. (2010). A calibration as shown in figure 2 is used both for the stereo camera and the thermal camera. The black dots are black painted metal skews that are heated up so that they are visible for the thermal camera. A set of 30 images with different viewing directions and rotations is taken to estimated the parameters of the interior orientation.
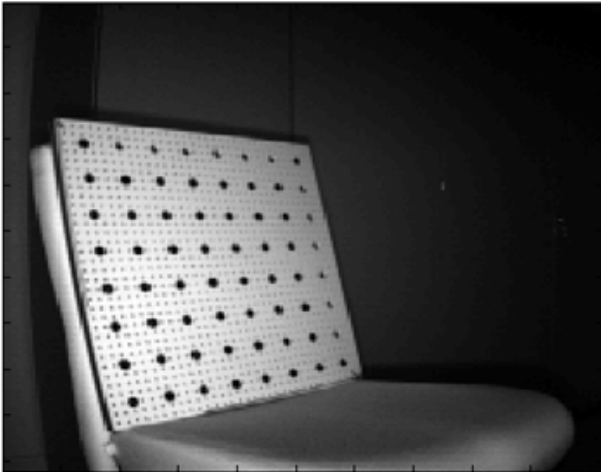

Figure 2. Calibration pattern as seen by the thermal camera.

The calibration of the sensor system is done using a calibration area with known 3D points. A wall with control points of the test site can be seen in figure 3. At first, the laserscanner is used in scan mode to generate a full 3D point cloud of the calibration area. Second, the thermal camera mounted on the laserscanner on an automatic swivel takes a set of pictures of the whole calibration area. From that set, the thermal camera is geometrically calibrated given the coordinates of the ground control points of the calibration area. The perspective centres of the laserscanner and the thermal camera are estimated and the lever arm of the thermal camera to the laserscanner is derived as the difference of the estimated perspective centres. For the stereo camera, the geometric calibration is done separately using the given calibration software. The camera is mounted fixed on the mobile mapping platform looking forward and records a part of the calibration area including a set of ground control points.

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. 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. From these points the perspective centre end the lever arm to the laserscanner are derived. Last, the profile laserscanner has to be included. This is done by fitting in the horizontal profile of the scanner with a given fixed height into the point cloud of the scanning laserscanner by a least squares matching.

3. DATA FUSION

We assume the intrinsic parameters of the TIR camera (Luhmann et al., 2010) and the RGB stereo camera (Hartley and Zisser-

Figure 3. Control points on one wall of the test site.

man, 2004) 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 (chapter 2.). Following the notation introduced in chapter 2., each pixel (p, q) in the laserscanner intensity image S can be assigned to a 3D coordinate X_{pq}. With equation 1 we get the corresponding pixel (i, j) of image I of the camera through

$$(i, j) = \Xi_{\alpha,\beta}(X_{pq} \tag{3}$$

Applying equation 3 for all points (p, q) of the laserscanner intensity image S transforms the image I into the view of the laserscanner image S through the backprojection of S by

$$\tilde{I_{pq}} := I_{ij} \tag{4}$$

The combined dataset is then a coloured image of the laserscanner point cloud. In contrast, equation 2 is applied to transform X_{pq} into the camera coordinate system $K^{(c)}$ through

$$Y_{pq} := MX_{\beta}\tilde{M}Z_{\alpha} \tag{5}$$

$||Y_{pg}||$ is the distance of Y_{pq} to the origin $K^{(c)}$. Then

$$D_{ij} = d_{pq} \tag{6}$$

is the forward projection of S for all points (p, q). 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.

4. EXPERIMENTS

The experimental setup (Fig. 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.
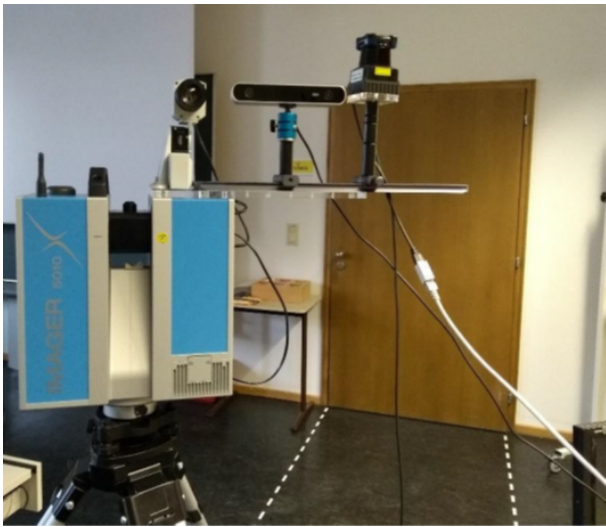

Figure 4. Sensor rig with all sensors. Relative orientation of the TIR and RGB camera and the laser range scanner with the Z+F laserscanner as origin of the coordinate system.

The trajectory of the whole construction (Fig. 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 objects to mark the targets and enable them to be observed by 2D laser was done. The remaining differences in the 3D coordinates of the control points $P1$ to $P4$ after the registration are shown in table 1.

Point accuracy in [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

Table 1. Remaining differences in the estimated 3D coordinates for four corresponsing control points P1 to P4.

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 of the Hokuyo. One can see that the profile fits to the point cloud at the edge of the room and at the door. The remaining differences (tab. 1) is mainly caused by the measurement accuracy and slightly different viewing directions from the lever arm.

For the trajectory determination, a known approach of Iterative Closest Point (Rusinkiewicz and Levoy, 2001) as 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 (Fig. 7).

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.
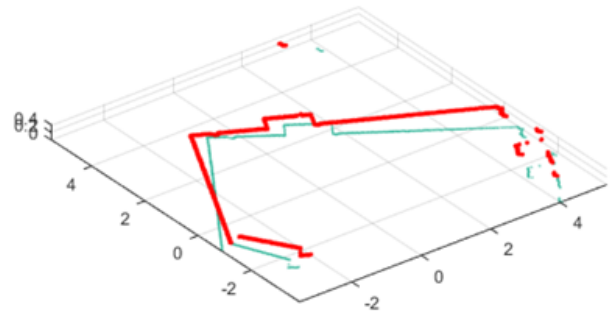

Figure 5. Position of Hokuyo points in the coordinate system of the Z+F laserscanner. 2D profile with red: before the registration, green: after the coregistration.
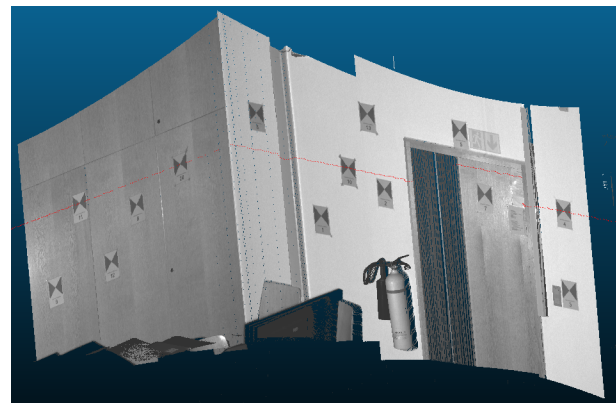

Figure 6. Position of Hokuyo points in the coordinate system of the Z+F laserscanner. 3D coregistration: The red line shows the profile line of the laser range finder. One can see that the profile of the edge of the room and the doors with to the 3D point cloud.
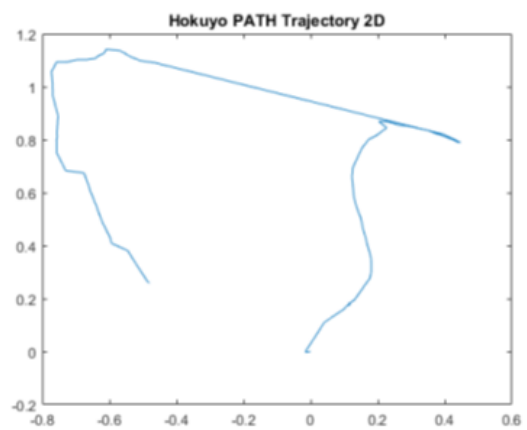

Figure 7. Trajectory determined by 2D laser range finder.

Figure 10 shows a part of the recorded scene with the point cloud generated from the stereo camera system. One can see the original RGB intensities of the ZED camera. Figure 11 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

Figure 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.
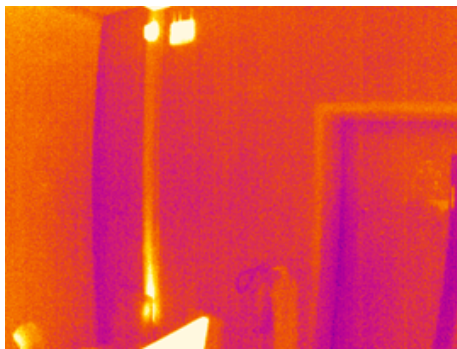

Figure 9. Optris thermal image cooresponding to the stereo image pair in fig. 7. One can see the warm installation channel and the WiFi device.

have been the installation channel in TIR and the edge of the room in RGB.

Figure 10. ZED Stereo camera point cloud colourized with RGB intensities of the ZED camera.

5. 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 colorized point clouds
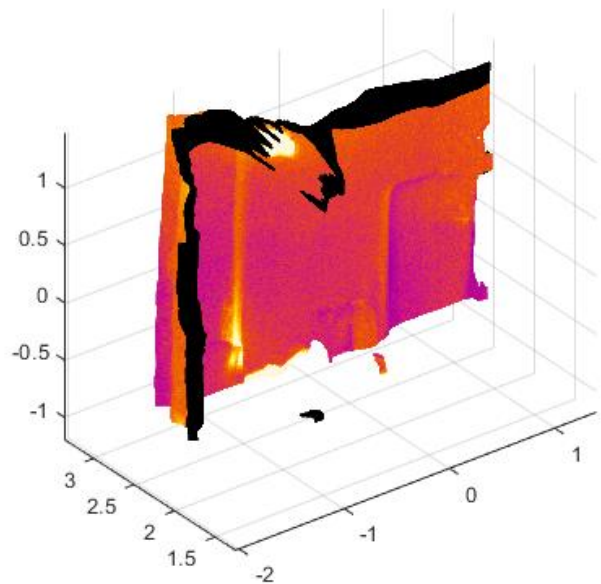

Figure 11. ZED Stereo camera point cloud colourized with Optris thermal image. One can see that the installation channel is warmer and correctly located on the wall and not in the edge of the room.

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

Abmayr, T., 2010. Calibration and Registration Framework for Multisensor Panoramic Color Scanning. Shaker, Aachen. ISBN 9783832288303.

Borrmann, D., Houshiar, H., Elseberg, J. and Nüchter, A., 2013. Vom kombinieren von 3d-modellen mit farb- und temperaturinformationen. *Photogrammetrie – Laserscanning – Optische 3D-Messtechnik. Beiträge der Oldenburger 3D-Tage 2013* pp. 246–253.

Coiras, E., Santamaría, J. and Miravet, C., 2000. Segment-based registration technique for visual-infrared images. *Optical Engineering* 39, pp. 282–289.

Förstner, W. and Gülch, E., 1987. A fast operator for detection and precise location of distinct points, corners and centers of circular features. In: *Proceedings of the ISPRS Intercommission Workshop on Fast Processing of Photogrammetric Data*, Vol. 1987, pp. 281–305.

Harris, C. and Stephens, M., 1988. A combined corner and edge detector. In: *Proc. Fourth Alvey Vision Conferenceg*, pp. 147–151.

Hartley, R. and Zisserman, A., 2004. Multiple View Geometry in Computer Vision. Cambridge University Presss. ISBN 0521540518.

Hastedt, H. and Luhmann, T., 2012. Investigations on a combined rgb / time-of-flight approach for close range applications. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 39(B5), pp. 333–338.

Hirschmueller, H., 2008. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30(2), pp. 328–341.

Hoegner, L. and Stilla, U., 2015. Building facade object detection from terrestrial thermal infrared image sequences combining different views. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* II-3/W4, pp. 55–62.

Hoegner, L. and Stilla, U., 2018. Mobile thermal mapping for matching of infrared images with 3d building models and 3d point clouds. *Quantitative InfraRed Thermography Journal* 15(2), pp. (in press).

Hoegner, L., Hanel, A., Weinmann, M., Jutzi, B., Hinz, S. and Stilla, U., 2014. Towards people detection from fused time-of-flight and thermal infrared images. In: *Photogrammetric Computer Vision 2014 - PCV2014*, Vol. XL-3, pp. 121–126.

Hoegner, L., Tuttas, S., Xu, Y., Eder, K. and Stilla, U., 2016. Evaluation of methods for coregistration and fusion of rpas-based 3d point clouds and thermal infrared images. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLI-B3, pp. 241–246.

Lowe, D., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2), pp. 91–110.

Luhmann, T., 2003. Nahbereichsphotogrammetrie: Grundlagen, Methoden und Anwendungen. Wichmann. ISBN 3879073988.

Luhmann, T., Ohm, J., Piechel, J. and Roelfs, T., 2010. Geometric calibration of thermographic cameras. *International Archives of Photogrammetry, Remote Sensing and Spacial Information Sciences* 38(5), pp. 411–416.

Mayer, H., Bartelsen, J., Hirschmueller, H. and Kuhn, A., 2012. 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*, Vol. 7474, Springer, Berlin, pp. 285–304.

Park, C., Bae, K.-H., Choi, S. and Jung, J.-H., 2008. Image fusion in infrared image and visual image using normalized mutual information.

Pollefeys, M., Nister, D., Frahm, J., Akbarzadeh, A., Mordohai, P., Clipp, B., Engels, C., Gallup, D., Kim, S., Merrell, P., Salmi, C., Sinha, S., Talton, B., Wang, L., Yang, Q., Stewénius, H., Yang, R., Welch, G. and Towles, H., 2008. Detailed real-time urban 3d reconstruction from video. *International Journal of Computer Vision* 78(2-3), pp. 143–167.

Rusinkiewicz, S. and Levoy, M., 2001. Effcient variants of the ICP algorithm. In: *Proceedings of the IEEE 3rd International Conference on 3D Digital Imaging and Modeling*, pp. 145–152.

Snavely, N., Seitz, S. M. and Szeliski, R., 2008. Modeling the world from internet photo collections. *Int. J. Comput. Vision* 80(2), pp. 189–210.

Wang, L., Gong, M., Zhang, C., Yang, R., Zhang, C. and Yang, Y.-H., 2012. Automatic real-time video matting using time-of-flight camera and multichannel poisson equations. *International Journal of Computer Vision* 97(1), pp. 104–121.

Weinmann, M. and Jutzi, B., 2012. A step towards dynamic scene analysis with active multi-view range imaging systems. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 39(B3), pp. 433–438.

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

Westfeld, P., Mader, D. and Maas, H., 2015. Generation of TIR-attributed 3d point clouds from uav-based thermal imagery. *PFG Photogrammetrie – Fernerkundung – Geoinformation* 2015(5), pp. 381–394.

Zhang, Z., 1999. Flexible camera calibration by viewing a plane from unknown orientations. In: *ICCV*.