EVALUATION OF METHODS FOR COREGISTRATION AND FUSION OF RPAS-BASED 3D POINT CLOUDS AND THERMAL INFRARED IMAGES

L. Hoegner *, S. Tuttas, Y. Xu, K. Eder, U. Stilla

Technische Universitaet Muenchen (ludwig.hoegner, sebastian.tuttas, yusheng.xu, konrad.eder, stilla)@tum.de

Commission III, WG III/2

KEY WORDS: photogrammetric point cloud, 3D building model, coregistration, fusion, thermal infrared

ABSTRACT:

This paper discusses the automatic coregistration and fusion of 3d point clouds generated from aerial image sequences and corresponding thermal infrared (TIR) images. Both RGB and TIR images have been taken from a RPAS platform with a predefined flight path where every RGB image has a corresponding TIR image taken from the same position and with the same orientation with respect to the accuracy of the RPAS system and the inertial measurement unit. To remove remaining differences in the exterior orientation, different strategies for coregistering RGB and TIR images are discussed: (i) coregistration based on 2D line segments for every single TIR image and the corresponding RGB image. This method implies a mainly planar scene to avoid mismatches; (ii) coregistration of both the dense 3D point clouds from RGB images and from TIR images by coregistering 2D image projections of both point clouds; (iii) coregistration based on 2D line segments in every single TIR image and 3D line segments extracted from intersections of planes fitted in the segmented dense 3D point cloud; (iv) coregistration of both the dense 3D point clouds from RGB images and from TIR images using both ICP and an adapted version based on corresponding segmented planes; (v) coregistration of both image sets based on point features. The quality is measured by comparing the differences of the back projection of homologous points in both corrected RGB and TIR images.

1. INTRODUCTION

Today, automatic 3d reconstruction and texture extraction is focused on high resolution images and image sequences from the visual spectrum. These methods are mainly using homologous points to link the images in a relative orientation and extract 3d coordinates for the homologous points (Hartley and Zisserman, 2004). Some of these methods also include a self-calibration for the camera (Agarwal et al., 2009; Longuet-Higgins, 1981; Maybank, 1993; Mayer et al., 2012). The resulting camera orientations and 3d points have to be transferred from the model coordinate system to the global coordinate system either using external GPS/INS orientation information and / or a matching with given 3d building models.

To find a good set of homologous points and remove outliers, the RANSAC algorithm (Fischler and Bolles, 1981) is used in most solutions, where random minimal sets of homologous points are used to estimate a relative orientation, that is checked against the other homologous points. A reduction to a minimum set of five corresponding points is possible for known interior orientation (Nister, 2004). This reduces the possibility of outliers in the minimum set of the RANSAC and increases the quality of the best relative orientation. An additional quality improvement is the extension of image pairs to image triplets (Hartley, 1997; Fitzgibbon and Zisserman, 1998; McGlone, 2004; Stewenius et al., 2005). Espacially for image sequences with a constant viewing direction, the trifocal tensor derived from image triplets increased significantly the stability of the relative orientation compared to image pairs using the fundamental matrix.

Recording image sequences of building facades and roofs most of the homologous points are on the facade plane. This has a significant influence on the quality of the relative orientation as the reconstruction of the interior orientation is not possible from

a planar scene (Maybank, 1993). For calibrated cameras with known interior orientation a planar scene has two solutions in general (Longuet-Higgins, 1986; Maybank, 1993). The geometrical correct solution can be found by searching the solution with all 3d points in front of both cameras. For planar scenes the homography is an alternative way of orienting image pairs (Hartley and Zisserman, 2004; Pollefeys et al., 2002).

Corrected relative orientations are used as input for dense matching techniques to generate a 3D point cloud not only for the tracked features used in the bundle adjustment, but for most of the pixels of the input images with correspondences in several images (Hirschmueller, 2008). Several works have used this two-step 3D point cloud extraction for image sets from RPAS mounted RGB cameras. ((Mayer et al., 2012; Tuttas et al., 2015).

There is quite a limited number of works on transferring methods for 3D reconstruction to the thermal infrared domain. A geometric calibration including principal point, focal length, and radial distortion parameters has been investigated by some groups (Simmler, 2009; Luhmann et al., 2010; Lagela et al., 2011). 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). Both 3d reconstruction and texturing are influenced by various conditions as the thermal radiation of facades 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) is possible.

Terrestrial images (Hoegner et al., 2007) taken from a vehicle can be used for documentation of frontal faces visible from the street level, while airborne images taken from an unmanned aerial ve-

---

*Corresponding author

hicle (UAV) or helicopter can capture roofs. Using oblique view images inner yards can also be covered (Iwaszczuk et al., 2012). Both ways require a known 3D building model to coregister different image data sets. (Hoegner et al., 2007) generate facade textures for the given 3D model from oriented fused image sequences, whereas (Iwaszczuk et al., 2012) directly map single images onto the 3D model. This 3D referenced textures allow classical image processing algorithms to analyse thermal infrared textures and extract objects under the surface like heating pipes and thermal leakages. First attempts for window extraction using grammars have already shown the potential of thermal infrared textures (Michaelsen et al., 2012).

As described in (Borrmann et al., 2013) and (Hoegner et al., 2014), a combination of 3D point clouds and thermal infrared images can be used for building inspections as well. Both methods show limitations in their applicability. Time-of-flight-cameras have a very limited range for accurate measurements. Laserscanners are either fixed on the ground and thus only scan facades without roofs or are mounted on flying platforms. On RPAS systems, you either need a RPAS with high payload or geometric accuracy and point density are quite limited compared to point clouds from RPAS mounted RGB camera systems.

This paper concentrates on a set of methods to coregister and fuse images taken with both a RGB and a thermal infrared (TIR) camera mounted on a RPAS. Instead of mounting both systems together with a fixed, calibrated relative orientation, here both cameras record the scene one after another in two fights following the same predefined flight path. Due to limited accuracies in both the onboard GPS/INS and the active position control of the RPAS, the exterior orientations of a corresponding image pair (one RGB and one TIR image for the same predefined recording orientation) show slightly different parameters so that a coregistration of RGB and TIR image is necessary. Section 2 introduces the investigated methods for coregistration and explaines the quality assessment. Section 3 describes the test data used for the evaluation and presents the results. The paper is closed with section 4 discussing the results.

2. METHODOLOGY

It is assumed that two datasets are given, one containing RGB high resolution images and one containing thermal infrared (TIR) images. Both sets are taken by camera systems mounted on a flying platform. One dataset is recorded after the other one. For every image of the RGB dataset there exists one corresponding image in the TIR dataset that was taken with the same exterior orientation. Due to the accuracies in the GPS/INS system and platform controls, there remain small differences in the exterior orientation of such an RGB-TIR image pair. Both cameras are assumed to be geometrically calibrated. It has to be taken into account that the ground sampling distance for RGB and TIR images is different and that the TIR image covers only a part of the area visible in the RGB image.

To remove remaining differences in the exterior orientation, different strategies for coregistering RGB and TIR images are discussed: (i) coregistration based on 2D line segments for every single TIR image and the corresponding RGB image. This method implies a mainly planar scene to avoid mismatches; (ii) coregistration of both the dense 3D point clouds from RGB images and from TIR images by coregistering 2D image projections of both point clouds (Hoegner et al., 2014; Urban and Weinmann, 2015); (iii) coregistration based on 2D line segments in every single TIR image and 3D line segments extracted from intersections of planes fitted in the segmented dense 3D point cloud (Iwaszczuk

et al., 2012); (iv) coregistration of both the dense 3D point clouds from RGB images and from TIR images using both ICP (Besl and McKay, 1992; Rusinkiewicz and Levoy, 2001; Chetverikov et al., 2005) and adapted version based on corresponding segmented planes (Hebel and Stilla, 2009); (v) coregistration of both image blocks based on SIFT features (Lowe, 2004). The quality is measured by comparing the differences of the back projection of homologous points in both corrected RGB and TIR images.

2.1 Coregistration of 2D line segments

The first method is working on the image level only. Every image pair consisting of one RGB and one TIR image is coregistered in 2D image space. It is assumed that the RGB image orientation ist correct and the TIR image can be registered to the RGB image. A coregistration of 2D line segments implies that the difference in the field of view and the exterior orientation of the RGB and the TIR image is small enough to ignore differences in the visibility and perspective of both images. This methods works isolated for every image pair and does not use any knowledge of the overall recorded scene. It is also assumed that prominent edges appear in the RGB image as well as in the TIR image. Using these conditions an affine transformation of the TIR image to the RGB image is estimated. As a pre-step, a preliminary scale factor is estimated for the TIR image based on the different ground sampling distances derived from the exterior orientation and camera specifications. For the proeccsing steps it is estimated from the exterior orientation and field of view, which part of the RGB image are possibly visible in the TIR image and its smaller field of view. Only this part is used for the processing.

At first, an edge detection based on the Sobel operator is done in both the scaled TIR and the cut RGB image followed by a non-maximum-suppression. The different radiometric appearance of edges is mainly eliminated by this step. Based on the first estimation of the scale factor of the TIR image and the corresponding part of the RGB image, assuming a limited error in the exterior orientation, the coregistration is done using iterative estimation of the best transformation parameters in a similarity metric (Styner et al., 2000). The algorithm results in an affine transformation of the TIR image to the RGB image.

For a 3D point cloud calculated from the RGB images, now an additional TIR channel can be added through the correspondence of every RGB pixel to an interpolated value of the TIR image. Due to overlap in the TIR images, there are parts of the point cloud, that are visible in more than one TIR image. For these parts, the mean intensity from the visible TIR images is calculated.

2.2 Coregistration of projected RGB point cloud and TIR image

The second strategy introduces 3D scene information derived from the RGB images generating a 3D point cloud. The TIR images are handled isolated and are connected to the point cloud generated from the RGB images.

The RGB images are coregistered using given GPS/INS orientation for every image and corresponding homologous points using SIFT features. A bundle block adjustment is done to refine the orientations of the images. Resulting orientations are used for semi-global-matching (Hirschmueller, 2008). The dense 3D point cloud is projected into the TIR images using the recorded orientations of the TIR images. The coregistration is done in image space as proposed in method 1 using iterative estimation of the best transformation parameters in a similarity metric (Styner et al., 2000). After minimizing the reprojection error, a bilinear

interpolation is done for the projected 3D points of th RGB point cloud in the TIR image to derive an additional TIR intensity per 3D point. Due to overlap in the TIR images, there are parts of the point cloud, that are visible in more than one TIR image. For these parts, the mean intensity from the visible TIR images is calculated.

2.3 Coregistration of 3D RGB and 2D TIR line segments

The third strategy uses the 3D object space to orient the RGB images in an image block and to correct their orientations. The TIR images are handled isolated and optimized one by one to the 3D point cloud from the RGB images. Every TIR image is processed as proposed in method i) to detect edges using Sobel operator for gradient detection and non-maximum-suppression followed by Hough transform to extract prominent edges. These line segments have to be coregistered to 3D lines extracted from the RGB point cloud.

A 3D point cloud is generated from the RGB images as mentioned in method 2. In the resulting dense point cloud, planes are estimated as proposed in Hebel and Stilla (2009). Intersection lines are calculated for adjacent planes. These 3D lines are projected into the TIR images using the TIR image orientation parameters recorded during the flight. This part follows Iwaszczuk et al. (2012), where 3D lines were taken from a given 3D polygonal building model. Here, the building model is replaced with the 3D point RGB point cloud and the plane intersection lines. It is assumed that the initial orientations of the TIR images are accurate enough to assign possible candidates for corresponding lines by searching for line segments with similar orientation and smallest distance in the image. In a bundle adjustment, the distances of these corresponding line pairs are minimized with the final TIR image orientation parameters as estimated unknowns. The bundle adjustment results in corrected TIR orientation parameters for every single TIR image adopted to the RGB based 3D point cloud.

As result, the 3D points of the RGB based point cloud get an additional TIR intensity that is interpolated from the TIR image. Like in method 1 and 2, due to overlap in the TIR images, there are parts of the point cloud, that are visible in more than one TIR image. For these parts, the mean intensity from the visible TIR images is calculated. For later quality comparisons, a 3D point cloud is calculated using the adjusted TIR image orientations.

2.4 Coregistration of 3D point clouds

The forth method uses for both the RGB images and the TIR images 3D the whole image block. For both the set of RGB images and the set of TIR images, a 3D point cloud is calculated. The recorded GPS/INS information is used as initial values for the unknown estimated orientation parameters. Homologous points using SIFT features are introduced as observations and their 3d object coordinates as unknown. The resulted adjusted bundle block delivers corrected orientation parameters that are used for semiglobal-matching (Hirschmueller, 2008). Semi global matching is, dependant on the recording configuration, not only applicable for RGB images (Mayer et al., 2012), but also for TIR images (Westfeld et al., 2015). The GPS/INS parameters used as input values for the bundle adjustment can be seen as accurate enough to assume that both th RGB point cloud and the TIR point cloud are very close together in object space and have almost the same scale. This allows to directly use Interative Closest Point (ICP) (Rusinkiewicz and Levoy, 2001) to coregister the TIR point cloud to the RGB point cloud.

The correction parameters for the ICP are used to correct the orientations of the TIR images. Like in method 1, 2, and 3, due to overlap in the TIR images, there are parts of the point cloud, that are visible in more than one TIR image. For these parts, the mean intensity from the visible TIR images is calculated.

2.5 Coregistration of image blocks using point features

So far, the four presented methods are either not using 3D information at all (method 1) or calculating the final 3D point coordinates only from the RGB images. In the fifth method, one 3D point cloud is generated from the whole set of RGB and TIR images. It is assumed that at least in urban scenes, there are enough structures that are visible both in RGB and TIR images, but show different radiometric behaviour. Keeping this assumption, detected feature points in RGB and TIR images based on geometric and not on radiometric behaviour should have the same or similar 3D object point.

Two types of tie points are used in the bundle adjustment: first the standard SIFT features as used in method 4 for coregistering all RGB images together and all TIR images together; second the geometric features used to connect the pairs of one RGB and one TIR image. Due to known calibration and orientation parameters the search space for corresponding features is quite small. After generating two bundle blocks as mentioned in method 4, RANSAC (Fischler and Bolles, 1981) is used with a minimum set of geometric RGB/TIR features to remove outliers. The already optimized SIFT features within the RGB image block and the TIR image block are used for quality calculation of the model generated by RANSAC. The resulting bundle block contains both the block of RGB images and the block of TIR images. Due to the much higher geometric resolution of the RGB image, the 3D point cloud is generated from the RGB images only. The resulting point cloud is projected into the TIR images to interpolated TIR intensities for all 3D points. There are parts of the point cloud, that are visible in more than one TIR image. For these parts, the mean intensity from the visible TIR images is calculated.

2.6 Quality measurements

The quality of the coregistration is evaluated both in 2D and 3D. The 2D evaluation investigates the remaining distances of 2D features between the RGB image and the corresponding point in the TIR image. The 3D analysis investigates the remaining differences in the 3D point clouds. As the RGB point cloud is very dense compared to the TIR images based point cloud, for every 3D point of the TIR point cloud the minimum distance to a point in the RGB point cloud is determined as it is assumed that the discretization errors in the position of the 3D points are small in the RGB point cloud compared to the 3D points in the TIR point cloud. Additionally, coordinates from ground control points are compared to their estimated 3D coordinates.

3. EXPERIMENTS

3.1 Dataset

RGB images were recorded with a Sony Nex-7 camera with 6000 x 4000 pixels on a 23,4 x 15,6 mm cmos chip. The TIR images were recorded with a FLIR Tau640 with 640 x 512 pixels on a bolometer chip. Both cameras were mounted at an AscTec Falcon 8 octocopter mount with stabilised orientation that is recorded additionally to the GPS position of the octocopter itself. Both cameras have been calibrated geometrically (Simmler, 2009; Luhmann et al., 2010) offline with fixed focal length. The

flight path that was used for both camera flights has been prepared before the flight with waypoints and camera orientation. As the main differences between the techniques are expected for scenes with different heights and occlusions, a small urban scene has been chosen for the test. A construction site with several construction objects around was recorded at first using the RGB camera followed by the same waypoint sequences with the TIR camera mounted (Fig. 1).

Figure 1. Octocopter with Nex-7 camera mounted in front of the construction site

The test site has been recorded with three overlapping stripes in two different height levels with nadir viewing cameras and at the north and east facade with oblique camera orientation for facade reconstruction. The west and south facades have not been recorded due to close high buildings.

3.2 Quality of the adjusted orientations and 3D point clouds

The bundle block adjustment for the TIR images and RGB images are collected in table 1. The left column shows the results for the TIR image block, the right shows the results for the RGB image block for the correction of the interior orientation, the mean squares error of the exterior orientation (RMS), number of 2D and 3D points, there reprojection error and the total number of points of the point clouds. The big difference in the resolution of the recorded images and the ground sampling distance reflects in the big difference in observed feature points between the TIR images and the RGB images. One remarkable fact is the big errors in the angels for the RGB block compared to the TIR block. This may be caused by the higher point density especially on the walls which leads to stronger geometric constraints to improve the exterior orientation.

Diff. to optimized interior orientation	1.97%	5.08%
RMS X_0 [m]	0.795129	0.320422
RMS Y_0 [m]	0.397178	0.334195
RMS Z_0 [m]	0.627111	0.585101
RMS ω_0 [degree]	7.105335	16.816294
RMS ϕ_0 [degree]	7.597303	22.161192
RMS κ_0 [degree]	7.92476	15.828145
Median number of keypoints / image	6810	37612
Median matches per image	1608	14170
Number of 2D keypoints	39409	1204638
Number of 3D points	17137	465868
Mean reprojection error [pixels]	0.218287	0.152046
Number of 3D densified points	41126	32313301

Table 1. Bundle block adjustment: The left column shows the results for the TIR image block, the right shows the results for the RGB image block. Interior orientation error, exterior orientation error (RMS), number of 2D and 3D points, their projection error and the total number of points of the point clouds

Figure 2 shows the resulting dense point clouds for the TIR images and the RGB images. It is obvious that the point density of the RGB point cloud is much higher because of the much higher geometric resolution on the ground.
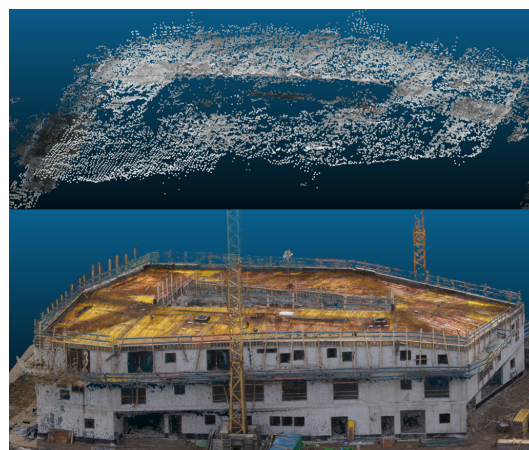

Figure 2. Dense point clouds from TIR images (top) and RGB images (bottom). The higher density of the RGB point cloud is clearly visible.

3.3 Coregistration

For method 1 and 2, the coregistration is done in the image domain. The TIR intensities are interpolated for the RGB pixels and a 3D point is calculated only from the RGB point cloud. The quality of the coregistration is given in table 2. From the table it can be seen, that the maximum distance to the corresponding point is not significantly reduced for method 1, but the median distance is minimized to sub-pixel accuracy. Most of the found edges are on the roof top and the affine transformation is optimized for that plane. For edges on the ground or on facades the distance stays quite big as the affine transformation parameters are not able to model the 3D depth of the scene. In the projection of the RGB point cloud into the TIR image in method 2 differences is the field of view have much less influence on the final result. That is why the results also is better for areas outside the roof plane. The remaining quite big distances are caused by different visibilities because of the different field of view and visibility of facade elements.

	before	img2img	cloud2img
Max dis [pixel]	82.06	76.53	15.93
Median dis [pixel]	26.37	0.8625	0.5283

Table 2. Coregistration results based on image coregistration in method 1 and 2. Quality for the projected lines is significantly better than for original images.

Figure 3 shows an image pair of RGB and TIR image overlayed in pink (RGB) and green (TIR) generated with method 1. One can see the prominent lines at the roof and the scaffold fitting very well. For the objects on the ground to the left of the scaffold, one can see a big offset as there were only few line segments in the TIR image compared to the roof and that is why the affine transformation was optimized for the roof only. In contrast, figure 4 shows the 3D point cloud from the RGB images. In one part of the point cloud, the RGB information is replaced with the added interpolated TIR intensity. The roof and scaffold fit very well. For small objects on the ground, there can be seen a remaining offset.

It has to be noticed that the image-to-image registration failed for about one third of the images. In a few images there are not enough line segments found in the TIR images. In a few other images, the parallel lines of the roof and the scaffolds and the

inner yard are misaligned and the registration failed. In method 2, the problem with few line segments in TIR images remains. The problem with the parallel line segments and especially the inner yard decreased because of a better initial position of the projected point cloud.
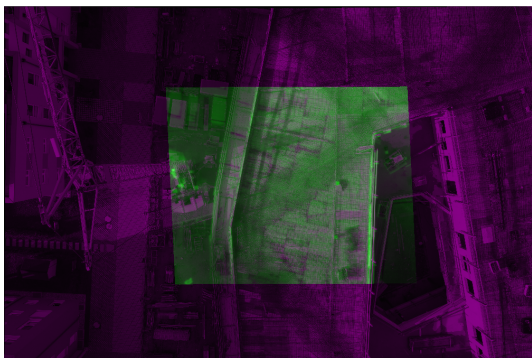

Figure 3. Overlay of RGB image (pink) and coregistered TIR image (green)
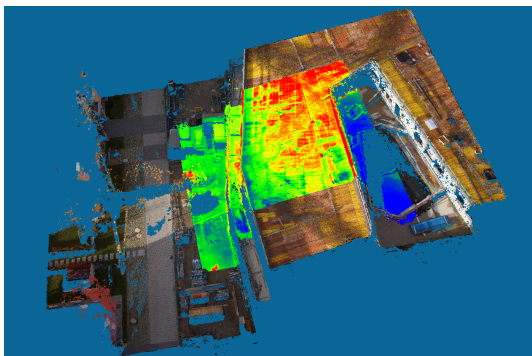

Figure 4. Overlay of RGB image and coregistered TIR image

For methods 3 to 5, the coregistration is done in the 3D space. Table 3 shows the minimum distances of the points of the TIR point cloud to the RGB point cloud. Method 3 extracting 3D line segments from the RGB point cloud lacks on the same basic problem as method 1 and 2 for image areas with only few line segments in the TIR image but performs better than SIFT for the median distance of the 3D points. This is mainly caused by the scene structure. As the roof covers relevant parts of most of the images and shows most of the long line segments, method 3 optimizes on the roof with its many points what decreases the median distance. Method 5 using SIFT is coregistering SIFT features all over the images and tries to optimize the orientation of the TIR images. This leads to slightly smaller errors on the ground but bigger errors on the roof. The best performing method is method 4 using ICP. The reason for this is the dense RGB point cloud and the similar visibility of the scene in the image sets. Thus, both point clouds cover almost the same surfaces and share the same scale as for both point clouds the GPS/INS orientations are used as initial parameters.

	before	ICP	SIFT	3D line
Max dis [m]	5.20715	2.75979	3.04624	2.91476
Median dis [m]	1.81998	0.108672	0.327250	0.285371
σ [m]	0.589367	0.251902	0.274852	0.327446

Table 3. Comparison of TIR point cloud and RGB point cloud before and after coregistration the 3D point clouds (ICP), the image blocks (SIFT), or the 3D line segments of the RGB model and 2D line segments of TIR images (3D line).

For the ICP point cloud coregistration the final RMS for 29120 3D points of the TIR point cloud that have been matched onto the RGB point cloud is 0.354372 meters. Table 4 gives the transformation matrix for the TIR point cloud to fit the RGB point cloud. The distances are visualised in figure 5, where blue is a distance close to zero and increasing up to five meters for red points. It can be seen, that bigger distances are mainly concentrated on the scaffolds and the inner yard. In both cases, small differences in the visibility caused by the different viewing angle and small differences in the recording orientations lead to these distances.

1.000	-0.006	-0.002	1.502
0.006	1.000	-0.008	6.201
0.002	0.008	1.000	-3.279
0.000	0.000	0.000	1.000

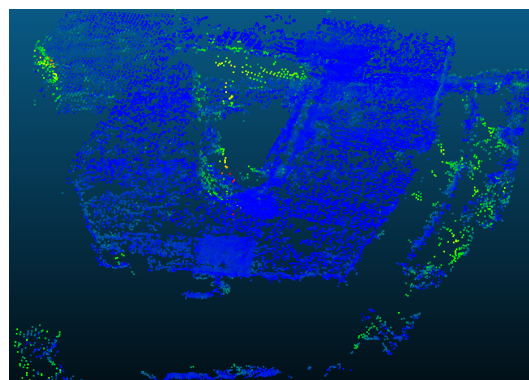Table 4. Transformation matrix to correct the TIR point cloud after ICP is calculated.

Figure 5. Distance of TIR points to the RGB point cloud after coregistration. Color from blue to red with increasing distance.

4. DISCUSSION AND OUTLOOK

In this paper some methods for coregistering RGB and TIR images from cameras mounted on an octocopter have been investigated on there quality. Methods optimizing a projection in 3D space (method 3) or minimizing distances in 3D space (methods 4 and 5) are show higher accuracies in areas with different depth than methods coregistering only in the image space. The image space coregistration on the other hand perform much faster than the 3D coregistrations. A possible use may be an online processing during flight to decide whester further images have to be taken or not. The high quality offline processing is then done based on the 3D coregistration techniques.

The flight path and recording positions and orientations have been optimized for the RGB camera. The TIR camera has a smaller field of view and much lower geometric resolution. For future campaigns, it seems to be useful to have smaller baseline between two adjacent images. This increases the overlap of the TIR images to strengthen the TIR bundle block. Also for the oblique facade views, the number of images for the TIR camera has to be extended. For method 4, further investigations will concentrate on sets of images, where all RGB image positions are also uased for the TIR image, but additional recording positions are only used for the TIR camera.

References

Agarwal, S., Snavely, N., Simon, I., Seitz, S. and Szeliski, R., 2009. Building rome in a day. In: *Proceedings of the International Conference on Computer Vision*.

Besl, P. and McKay, N., 1992. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14(2), pp. 239–256.

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

Chetverikov, D., Stepanov, D. and Krsek, P., 2005. Robust euclidean alignment of 3d point sets: the trimmed iterative closest point algorithm. *Image and Vision Computing* 23(3), pp. 299–309.

Fischler, M. and Bolles, R., 1981. Random sample consensus: A paradigm for model fit-ting with applications to image analysis and automated cartography. *Communications of the ACM* 24(6), pp. 381–395.

Fitzgibbon, A. and Zisserman, A., 1998. Automatic camera recoveryfor closed or open image sequences. *Lecture Notes in Computer Science*.

Hartley, R., 1997. Lines and points in three views and the trifocal tensor. *International Journal of Computer Vision* 22(2), pp. 125–140.

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

Hebel, M. and Stilla, U., 2009. Automatische koregistrierung von als-daten aus mehreren schrgansichten stdtischer quartiere. *PFG Photogrammetrie – Fernerkundung – Geoinformation* 2009(3), pp. 161–175.

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., 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., Kumke, H., Meng, L. and Stilla, U., 2007. Automatic extraction of textures from infrared image sequences and database integration for 3d building models. *PFG Photogrammetrie – Fernerkundung – Geoinformation* 2007(6), pp. 459–468.

Iwaszczuk, D., Hoegner, L., Schmitt, M. and Stilla, U., 2012. Line based matching of uncertain 3d building models with ir image sequences for precise texture extraction. *PFG Photogrammetrie – Fernerkundung – Geoinformation* 2012(5), pp. 511–521.

Lagela, S., Gonzlez-Jorge, H., Armesto, J. and Arias, P., 2011. Calibration and verification of thermographic cameras for geometric measurements. *Infrared Physics & Technology* 54, pp. 92–99.

Longuet-Higgins, H., 1981. A computer algorithm for reconstruction of a scene from two projections. *Nature* 203, pp. 133–135.

Longuet-Higgins, H., 1986. The reconstruction of a plane surface from two perspective projections. In: *Proceedings of the Royal Society of London. Series B, Biological Sciences*, Vol. 227(1249), pp. 399–410.

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

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

Maybank, S., 1993. Theory of Reconstruction from Image Motion. Springer Verlag. ISBN 0387555374.

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.

McGlone, C., 2004. Manual of Photogrammetry. ASPRS. ISBN 1570830711.

Michaelsen, E., Iwaszczuk, D., Sirmacek, B., Hoegner, L. and Stilla, U., 2012. Gestalt grouping on facade textures from ir image sequences: Comparing different production systemse. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Science*, Vol. 39/B3, pp. 303–308.

Nister, D., 2004. An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26(6), pp. 756–777.

Pollefeys, M., Verbiest, F. and Gool, L. V., 2002. Surviving dominant planes in uncalibrated structure and motion recovery. In: *Proceedings of the ECCV 2002. Part II*, Vol. 2351, Springer, Heidelberg, pp. 837–851.

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.

Simmler, C., 2009. Entwicklung einer Messanordnung zur geometrischen Kalibrierung von Infrarot-Kameras. Bachelorarbeit der Technischen Universitt Mnchen.

Stewenius, H., Schaffalitzky, F. and Nistr, D., 2005. How hard is 3-view triangulation really? In: *Proceedings of the Tenth IEEE International Conference on Computer Vision*, Vol. 2005(1), pp. 686–693.

Styner, M., Brechbuehler, C., Szkely, G. and Gerig, G., 2000. Parametric estimate of intensity inhomogeneities applied to mri. *IEEE Transactions on Medical Imaging* 19(3), pp. 153–165.

Tuttas, S., Braun, A., Borrmann, A. and Stilla, U., 2015. Validation of bim components by photogrammetric point clouds for construction site monitoring. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* II-3/W4, pp. 231–237.

Urban, S. and Weinmann, M., 2015. Finding a good feature detector-descriptor combination for the 2d keypoint-based registration of tls point clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* II-3/W5, pp. 121–128.

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.