

## Automatic 3D reconstruction and texture extraction for 3D building models from thermal infrared image sequences.

by L. Hoegner and U. Stilla

Technische Universitaet Muenchen – Photogrammetry and Remote Sensing, Arcisstrasse 21, 80333 Muenchen, Germany, [ludwig.hoegner@tum.de](mailto:ludwig.hoegner@tum.de), [stilla@tum.de](mailto:stilla@tum.de)

### Abstract

This paper discusses the accuracies of two methods for automatic image orientation, 3D reconstruction, and extraction of textures of building facades from thermal IR image sequences. The first method uses the image sequence and camera calibration information only to reconstruct the scene in model coordinates and coregisters this model to a given 3D building model to derive optimized orientation parameters. The second method directly includes the given 3D building model into the bundle adjustment of the image sequence orientation. The resulting optimized orientation parameters are used to project the 3D model into the images and extract façade textures. Both methods are evaluated on their geometric accuracy using experimental image sequences.

### 1. Introduction

Monitoring buildings in thermal infrared is mostly done manually from different viewing positions. Despite the outer conditions like sun, daytime, wind, and outside temperature, the measured radiance depends on the inside temperature, the emission coefficient and reflectance of the surface and surrounding objects increasing the radiance of a façade without an energy loss of the building. A step towards automatic analysis of thermal infrared images is done by using standard image segmentation techniques on the level of isolated images to detect leakages [1]. Big buildings and building complexes have to be recorded in a set of images that are either interpreted isolated or have to be merged to an image mosaic. In most cases, no 3D reference to a building model is included. To be able to monitor bigger building complexes an integration of the thermal infrared image information into a 3D building model is of a big interest: It becomes possible to analyze thermal building textures and locate leakages, windows, and other structures with a 3D reference to the building. Data from different sensors and recording times can be combined for change detection and the observation of dynamic processes.

The combination of a set of images into a common orientation is mainly done by two different main strategies: a direct georeferencing using ground control points and/or GPS/INS information of the position of the camera [2,3,4,5,6,7] or a relative orientation using homologous feature points found in the images [4,8,9,10,11,12,13,14]. After the orientation of the images, a 3D reconstruction and texture generation for facade elements is possible [14,15].

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 [16,17]. 3d reconstruction and texture extraction in thermal infrared are applied for sets of images and ordered terrestrial image sequences [18] or image sequences taken by a thermal camera mounted on a RPAS [19]. 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 time-of-flight cameras [20] is possible.

Terrestrial images [21] taken from a vehicle can be used for documentation of frontal faces visible from the street level, while airborne images taken from a remotely piloted aerial system (RPAS) or helicopter can capture roofs. [22,23]. Using oblique view images inner yards can also be covered [23]. Both ways require a known 3D building model to coregister different image data sets. [21] generate façade textures for the given 3D model from oriented fused image sequences, whereas [23] directly map single images onto the 3D model. This 3D referenced textures allow classical image processing algorithms to analyze 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 [24].

As described in [20], 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 contribution concentrates on the evaluation of possible accuracies for automatic 3D reconstruction and texture extraction from image sequences in the thermal infrared. The coregistration strategies presented follow the methods of Hoegner and Stilla [25]. Additionally, an evaluation of the accuracy of a point cloud based coregistration of images from the visible domain (RGB) and the thermal infrared domain (TIR) taken from a RPAS platform is introduced for scenarios where no given 3D model is present, but the model has to be extracted from the image data itself. This method is especially of interest for roof structures as they are mainly invisible from the street and in most cases not modelled in existing building models.

## 2. Methodology

In contrast to conventional IR inspection of buildings, for texturing the whole building or a building complex in dense urban areas and their narrow streets, the low resolution and the small field of view avoid a direct line matching of edges in the images and edges of the given building model. The proposed concept is based on the assumption that a 3D model of the recorded building is given containing 3D vertex coordinates and triangulated polygon surfaces with given texture coordinates. The coordinates should be given in a national coordinate system like Gauss-Krueger. GPS coordinates are often inaccurate and a direct line matching of the images and the projection of the model's edges fails because of the lack of visible façade edges in many of the images. The usage of continuous image sequences taken from a moving car or a RPAS allows performing a relative orientation of the images of a sequence to extract a 3D point cloud in relative model coordinates and a relative camera orientation for every image [11,12,13]. This relative oriented model is to be coregistered to given 3D building model information. This information can be a polygonal model or a 3D point cloud as well. Method one calculates the relative oriented 3D point cloud without preknowledge and then coregisters this to a given polygonal model by using recorded GPS information of the camera track and minimizing the distance of the 3D point cloud generated from the images and the given 3D building model façade planes. Method two includes the preknowledge of the existing polygonal model into the estimation of the camera orientations and the extraction of the 3D point cloud from the images directly. Method three calculates a 3D point cloud from the images including the GPS/INS orientations of the RPAS and coregisters this point cloud using ICP [26] to a point cloud generated from RGB images. In all cases, the resulting improved estimated orientation parameters of the TIR images are used to calculate the projection of the given 3D building model or the RGB based 3D point cloud into the TIR images. The resulting textures are cut out, combined from different images and transferred to the 3D building model.

### 2.1. Orientation of Image Sequences based on Feature Points

Mayer [11] has introduced an approach for wide-baseline image sequences. Mayer adopts Nistér's five-point-algorithm for relative orientation of image pairs and image triplets [10] for facade extraction and texturing from multiple views. Given the known inner orientation of the camera, SIFT features [27] are matched via cross-correlation. RANSAC [28] is used to choose SIFT features for the estimation of the fundamental matrix  $F$  and trifocal tensor  $T$  of triplets of adjacent images of the sequence. The found inliers are used for a robust bundle adjustment [8]. To orient the whole image sequence, the triplets are linked based on homographies and already known 3D points of the already oriented sequence part. Figure 1 shows an image out of a sequence with detected homologues points and estimated point movement in the image space.



**Fig. 1.** IR image with selected SIFT features, that have correspondences in the following image. Arrows show the moving direction of the points

## 2.2. Coregistration of the Image Sequence and the Building Model

As a result of the relative orientation of image pairs, all tracked homologue points are given 3D coordinates and all images are given a relative orientation in a model coordinate system. Using GPS coordinates of the recording camera as a priori estimation, the 3D point cloud is co-registered to the building model. For every 3D point, its normal vector is estimated from an estimated plane through the 3d points of its local neighborhood. Neighboring points with similar normal are supposed to be on the same façade. These points are grouped. It is assumed that the given GPS coordinates are accurate enough to do a first pre-orientation of the point cloud to the building model coordinate system. Point groups are assigned to the nearest façade of the building model with a similar surface normal. The distance of the point groups to their assigned façade planes is minimized in a bundle adjustment with seven parameters of the absolute orientation as unknowns: position X,Y,Z; the three angles around the axes, and a scale factor. The resulting image orientations are used for the projection of the 3D model into the image to extract the façade textures.

## 2.3. Integration of the Building Model into the Image Sequence Orientation

Like in method 1, SIFT features are detected and tracked through the image sequence. By introducing recorded orientation parameters from i.e. a GPS system, the images are transferred into the global coordinate system. This allows including pre-knowledge to the process. A 3D point reconstructed from homogenous pixels of a set of images that show a part of a building façade must be on or close to the model façade polygon. If it is not, it is either the result of a wrong point pair or it does not belong to the façade but i.e. to a tree or car standing in front of the building.

The object coordinate estimation from more than two homologues points is an error minimization problem that is formulated here as least mean squares error in a bundle adjustment. All homologues image points  $i$  of a 3d point  $k$  are given as observations  $(x_{ik}, y_{ik})$ . The 3D coordinates of the point are set as unknowns with an initial estimation from the homologues points of the first two images. Because of the limited accuracy of the image coordinates, the two projection rays of the homologues points will not intersect in 3D space. Instead the center point of the connection of the smallest distance of the two rays is chosen. The camera calibration parameters and the camera orientation are given as observations but with higher standard deviation compared to the image coordinates of the homologues points. This is done to allow a refinement of the GPS orientation parameters and to minimize the projection errors.

The model knowledge is included as additional observations where the distance of the 3D point to the closest surface of the model is minimized. In the bundle adjustment process, the first partial derivative of the calculations of the observations from the unknown derived for every unknown is set to zero. An image point  $(x_i, y_i)$  is calculated from an object point  $(X_i, Y_i, Z_i)$  using the collinearity equations

$$\begin{aligned} x_i - x_0 &= c_x \cdot \frac{a_{11}(X_i - X_0) + a_{21}(Y_i - Y_0) + a_{31}(Z_i - Z_0)}{a_{13}(X_i - X_0) + a_{23}(Y_i - Y_0) + a_{33}(Z_i - Z_0)} + dx' \\ y_i - y_0 &= c_x \cdot \frac{a_{12}(X_i - X_0) + a_{22}(Y_i - Y_0) + a_{32}(Z_i - Z_0)}{a_{13}(X_i - X_0) + a_{23}(Y_i - Y_0) + a_{33}(Z_i - Z_0)} + dy' \end{aligned} \quad (1)$$

where  $(X_0, Y_0, Z_0)$  is the camera position,  $(a_{11}$  to  $a_{33})$  is a 3x3 rotation matrix of the viewing direction of the camera, and  $(x_0, y_0, dx', dy')$  are the parameters of the geometric camera calibration.

Using the Hessian form of a plain

$$\vec{r} \cdot \vec{n} - d = 0 \quad (2)$$

a 3d point is on the plain if its space vector  $\mathbf{r} = (X_i, Y_i, Z_i)^T$  fulfills equation 2.

The weights for the observations are adapted in every iteration step of the bundle adjustment with respect to the remaining distance to the closest model façade as points with high distances are either wrong homologues points or 3d points that are not on the façade. As a result, the orientation parameters of the images are optimized so that the extracted 3d points fit best to the given 3D building model.

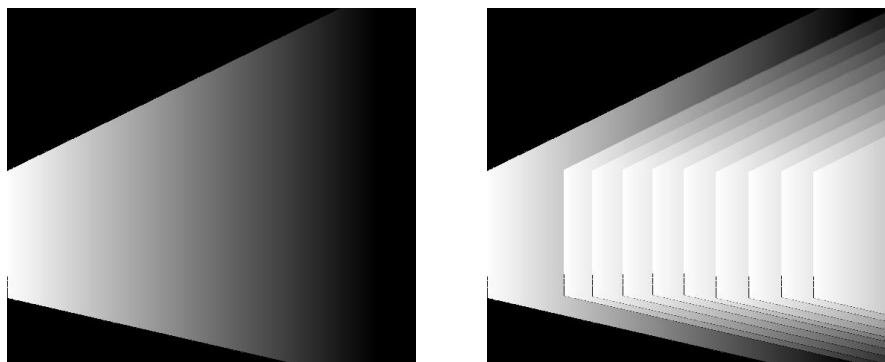
## 2.4. Coregistration of RGB and TIR Point Cloud

This method uses for both the RGB images and the TIR images 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 tracked SIFT features are introduced as observations and their 3d object coordinates as unknown in the collinearity equations (equ. 1). The resulted adjusted bundle block delivers corrected orientation parameters that are used for semi-global-matching [29]. Semi global matching

is, dependant on the recording configuration, not only applicable for RGB images [30], but also for TIR images [19]. The GPS/INS parameters used as input values for the bundle adjustment can be seen as accurate enough to assume that both the 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 Iterative Closest Point (ICP) [26] to coregister the TIR point cloud to the RGB point cloud.

## 2.5. Texture Extraction

The texture extraction part is split into three steps. In the first step, based on the results of the orientation, the images are projected onto the building model. This is done using equation 1 reverse and calculating the cross point of the projection ray of every pixel of every image with the model. In a second step, based on every projected image a partial texture for every visible surface of the model is generated. Because of the perspective projection the geometric resolution of an image onto a surface of the model might be not constant. In the examples introduced in section 3, an oblique forward looking view was chosen. In this case, the geometric resolution is as seen in figure 2a. Defining a unique texture size for all partial textures of a surface, one can see that for an ordered set of input images every following partial texture only shows parts of the surface, where it has a higher geometric resolution than the already combined partial textures (Fig. 2b). This allows a very simple combination strategy. For every surface every projected image is used for a bilinear interpolation of a partial texture with the same size in pixels. Then the pixels of the partial textures that are given a value from the interpolation – which means that this part of the surface was visible in the image – are copied in recording order into the final texture. A correlation is done for every added texture to remove remaining positioning errors. In the end, every pixel of the final texture is given the value from the input image with highest geometric resolution for this pixel.



**Fig. 2a.** Schematic view of the geometric resolution of a partial texture taken from an image in oblique forward looking view. white: high resolution, grey: low resolution, black: not visible in the image **2b.** Geometric resolution of a surface textures from a set of partial textures

## 3. Experimental setup

Two different scenarios have been used for the investigation. The terrestrial image sequences have been taken from a big building complex with different façade structures. These experiments have been taken using a SC3000 thermal infrared camera with a detector of 320x240 pixels and a temperature resolution of 0.02 K. Image sequences have been taken before sunrise. Figure 3 shows some image out of a sequence showing a part of a building complex. The sequence was recording with 50 frames per second. This allows a very good feature tracking (Fig. 1) due to only small movements between the images. For the 3D reconstruction every 10<sup>th</sup> image was taken to reduce the computational effort and guarantee a 3D base necessary for the initial 3D point estimation.

The aerial images have been taken from a small building complex that was still under construction. No 3D building was given here. 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 \citep{simmler2009, luhmann2010a} 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. 4).



**Fig. 3.** Four images taken out of a sequence of the SC3000 camera. Only parts of corners are visible.



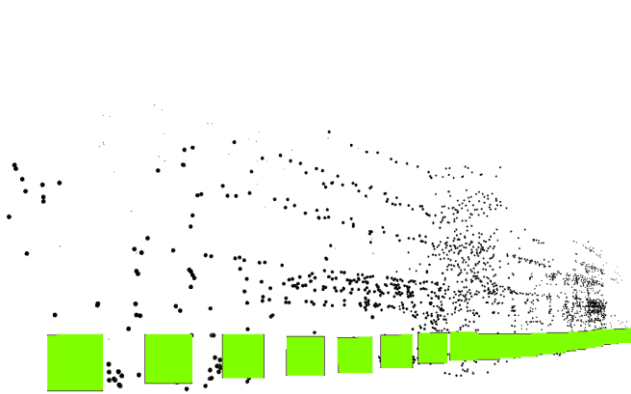
**Fig. 4.** Octocopter with Nex-7 camera mounted in front of the construction site.

#### 4. Results

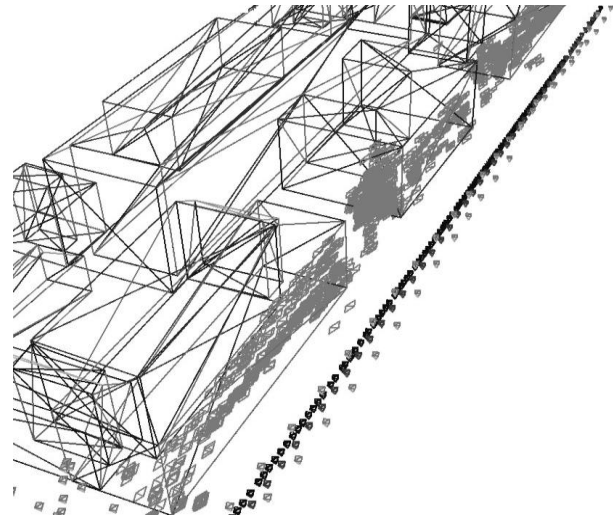
Four different scenes (a to d) have been recorded and processed with both methods for coregistration of an image sequence to a given 3D building model and one scene has been recorded for the point cloud based coregistration. Figure 5a shows the relative orientation of the resulting 3d points and camera positions of method 1. The tracked feature points are reconstructed as 3D points. Most of the points are grouped in rows along the windows of the façade. Figure 5b shows the 3d points and camera orientations matched with the building model in the bundle adjustment. Every small square represents a 3D point reconstructed from tracked feature points where the orientation of the square indicates the estimated surface plane in that point. Figure 6 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. Table 1a gives the accuracies of the relative orientation and table 1b the accuracies of the absolute orientation with bundle adjustment. Comparing the standard deviations for the estimation without (table 1a) and with preknowledge (table 1b), the unknowns including pre-knowledge of the building model show small improvements on the accuracy of the bundle adjustment. Whereas the improvements for scene a and c are quite small, the improvements for scene b and d are significant. Scene b and d have very high standard deviations in the relative orientation compared to scene a and c. This is caused by the homogenous structure of the façades (b) and the occlusions (d). Both scenarios show less homogeneous points that can be tracked through the sequence and more false correspondences. Including



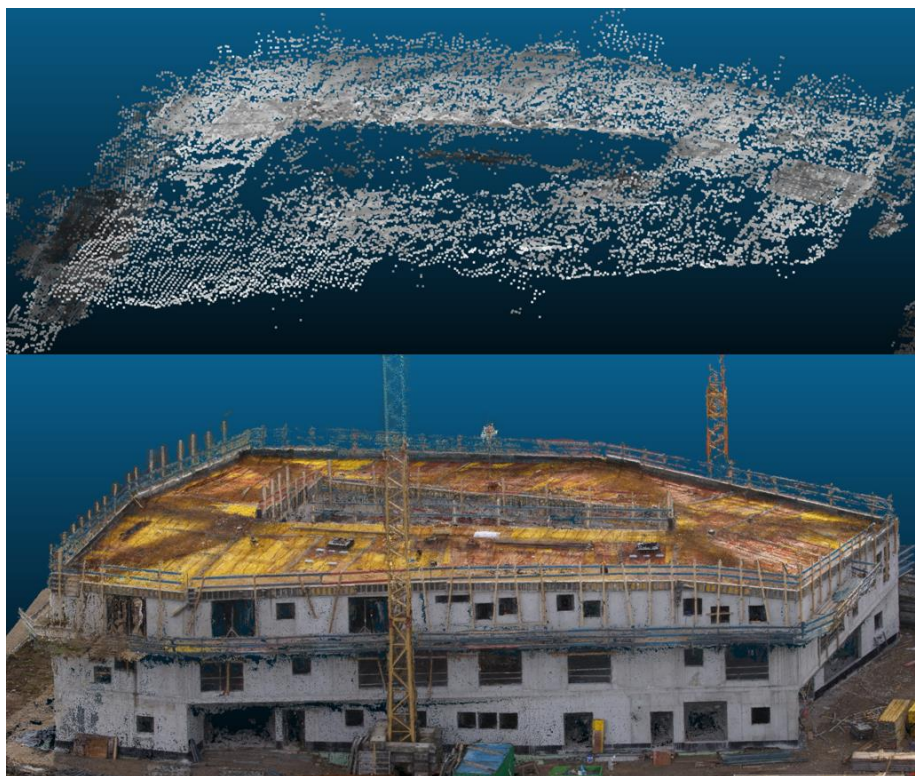
the building model allows to remove more false correspondences which helps in scene d. In scene b the known facade plane assists in estimating the movement of the features through the sequences.



**Fig. 5a.** Relative orientation: resulting point cloud and orientation from method 1



**Fig. 5b.** Bundle adjustment with building model: Building model with point cloud and orientations from method 2



**Fig. 6.** Point clouds extracted from images from a RPAS mounted TIR (top) and RGB (bottom) camera.

Table 2 shows the accuracies of the 3D point cloud reconstruction for the RPAS based RGB and TIR image sequence. One can see that the higher geometric resolution of the RGB images leads to a higher number of feature points per image and to a much higher number of densified points in the semi-global matching. The root mean squares error (RMS) for the optimized orientation parameters shows slightly better values for the TIR images. This is because of the lower geometric resolution that allows to keep a bigger part of the errors in the observed feature point coordinates (see mean reprojection error) instead of the orientation parameters.

relative orientation	a	b	c	d
$\sigma_{xy}[\mu\text{m}]$	0.3857	12.74	0.9486	14.71
$\sigma_{XYZ}[\text{m}]$	0.06855	0.3627	0.1068	0.3815
$\sigma_{X0}[\text{m}]$	0.1285	0.2512	0.08374	0.3486
$\sigma_{Y0}[\text{m}]$	0.04055	0.1856	0.2731	0.4502
$\sigma_{Z0}[\text{m}]$	0.02063	0.05744	0.06718	0.0815
$\sigma_{\omega0}[\text{gon}]$	0.03016	0.05230	0.6114	1.029
$\sigma_{\varphi0}[\text{gon}]$	0.06945	0.2041	0.09485	0.5176
$\sigma_{\kappa0}[\text{gon}]$	0.2259	0.4583	0.5834	1.923

**Table 1a.** Accuracy of the relative orientation without preknowledge

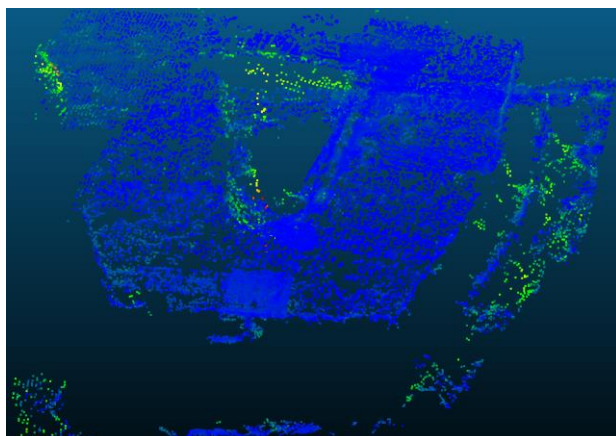
absolute orientation	a	b	c	d
$\sigma_{xy}[\mu\text{m}]$	0.3534	1.529	0.9185	1.405
$\sigma_{XYZ}[\text{m}]$	0.05944	0.1910	0.09583	0.1748
$\sigma_{X0}[\text{m}]$	0.04454	0.06776	0.04969	0.03390
$\sigma_{Y0}[\text{m}]$	0.02294	0.01750	0.1005	0.1972
$\sigma_{Z0}[\text{m}]$	0.009547	0.03065	0.06421	0.04909
$\sigma_{\omega0}[\text{gon}]$	0.02323	0.3148	0.5981	0.6936
$\sigma_{\varphi0}[\text{gon}]$	0.02196	0.08770	0.05319	0.4930
$\sigma_{\kappa0}[\text{gon}]$	0.2016	0.3806	0.5991	1.522

**Table 1b.** Accuracy of absolute orientation including preknowledge of the building model

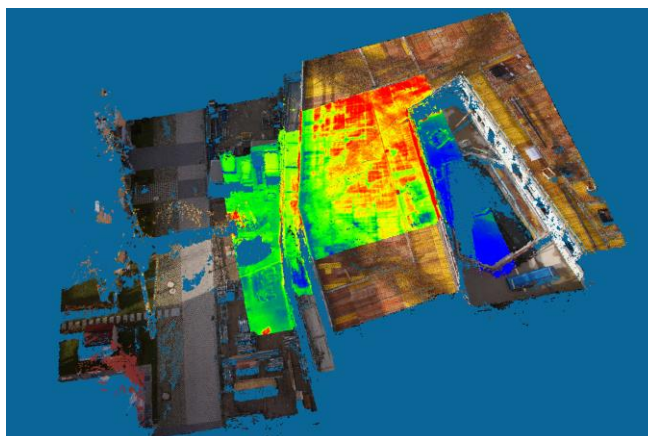
	TIR	RGB
Diff. to optimized interior orientation	1.97 %	5.08 %
RMS $X_0$ [m]	0.795129	0.320422
RMS $Y_0$ [m]	0.397178	0.334195
RMS $Z_0$ [m]	0.627111	0.585101
RMS $\omega_0$ [degree]	7.105335	16.816294
RMS $\varphi_0$ [degree]	7.597303	22.161192
RMS $\kappa_0$ [degree]	7.92476	15.828145
Median number 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 2.** 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

As the density of the RGB point cloud is much higher than for the TIR point cloud, it is assumed that for every TIR point cloud 3D point the closest 3D point of the RGB point cloud is the best fit. Doing a coregistration via ICP reduces the mean distance of a corresponding pair of 3D points of the two point clouds from 1.81998 meters to 0.108672 meters and the standard deviation of 3D coordinates from 0.589367 meters to 0.251902 meters. The distances are visualised in figure 7a, 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. Figure 7b shows a part of the RGB point cloud overlaid with the projected TIR intensities after the coregistration. Roof structures and scaffolds fit very well. For small objects on the ground some offsets remain.

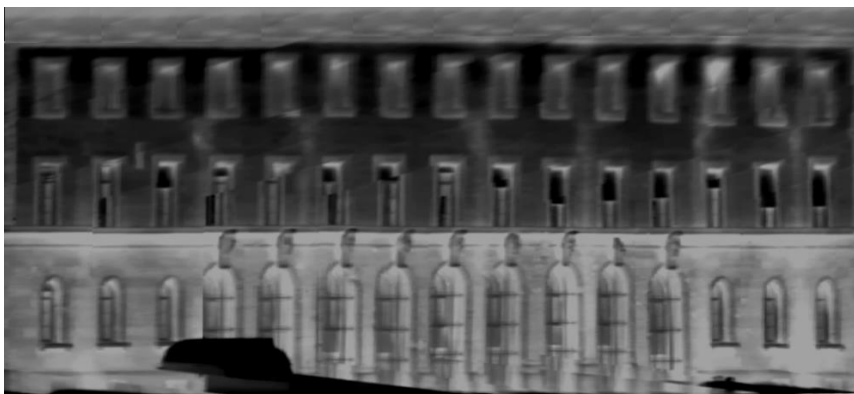


**Fig. 7a.** Distance of TIR points to the RGB point cloud after coregistration. Color from blue to red with increasing distance

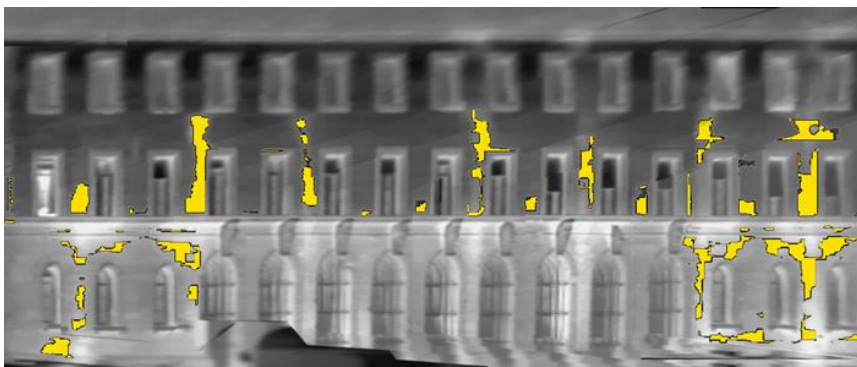


**Fig. 7b.** Overlay of RGB image and coregistered TIR image

In figure 8, a final texture composed of several extracted partial textures of a façade as described in [25] shows exemplarily the quality of the texture extraction and combination of the partial textures from all images of the sequence. The geometric resolution decreases from the first floor to the roof. This is due to the projection of the images onto the building model. Segmentation techniques are now applied on the façade textures to detect leakages (fig. 9). Image parts are identified that differ in their intensity from the mean intensity value. The windows are removed by searching for regular structures in the segmented blobs. The outlines of the windows are found using the method described by Michaelsen et al. [24]. Remaining segments that differ from the façade mean intensity are interpreted as leakages. They are shown in yellow in figure 9.



*Fig. 8. Surface texture generated from the complete image sequence*



*Fig. 9. Surface texture with detected leakages marked in yellow*

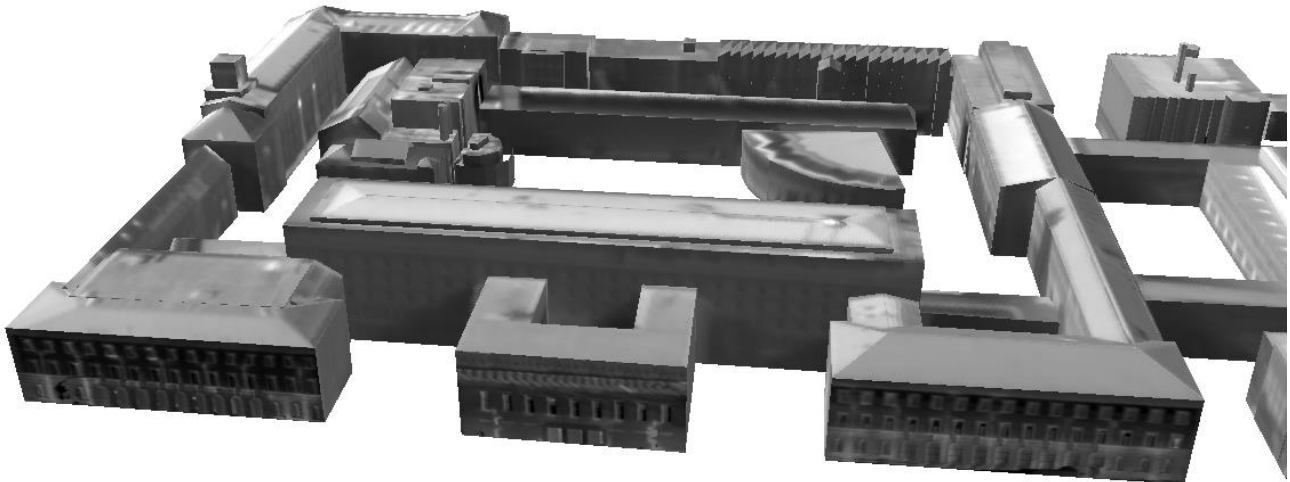
#### 4. Discussion and Outlook

Different aspects influence the quality of the extracted textures. The accuracy of the 3D reconstruction is limited compared to image sequences in the visual spectrum. One aspect is the lower image resolution that causes a higher discretization of the scene and reduces the number of details that can be matched within the sequence. On the other hand, the radiometric behavior of object in the thermal infrared leads to blurred edges and a low number of intensity changes and details. Homologous points are rarer than in the visual spectrum and worse locatable. A second aspect is the geometric calibration of the cameras. Because of the two limitations already mentioned, the calibration accuracy is also limited compared to cameras in the visual spectrum. The quality of the initial GPS positions is limited in urban scenarios caused by mirroring and occlusion of the GPS signals. The initial values have quite an influence on the final results as initial assignments of homologous points to facades influence the weightings in the iterative process. The quality of the building model itself is also a limiting factor. Building in level-of-detail 1 are only block models. In these models, the overhang at the roofs is not modeled. This causes two possible errors. If no overhang is modeled, the facades can either be set to the position of the footprint. Then, the position of the facades is correct, but the overhang is projected onto the facade. This causes misalignment of the objects on the facade like windows and heating pipes. If the building model is generated from aerial images, the roofs outline is taken as footprint and the position of the facades is incorrect in the model. In this case, the bundle adjustment tries to move the camera path which very often works but not necessarily. The number of visible elements in the thermal infrared influences the quality of the position refinement. On facades with only few objects or only repetitive patterns, the quality of the refinement is significantly reduced. If this falls together with bad initial GPS positions, the method can totally fail. The texture extraction is quite sensitive to errors in the



viewing direction estimation. The combination of relative orientation and 3D model knowledge reduces these errors significantly. Remaining errors between to partial textures.

The transfer from 2D image interpretation to a 3D reconstruction and texturing allows the combination of different sensors. Figure 10 shows a complex building model combining textures from terrestrial image sequences for the facades and textures from aerial images for the roofs. Terrestrial textures show a higher level of detail compared to aerial images and are easier to acquire. Recording facades from aerial images in urban areas is limited by occlusions from other buildings and flight limitations especially for RPAS flights as operation above traffic or persons is not allowed. Terrestrial image sequences can be recorded from a moving car.



**Fig. 10.** Building model with textures from terrestrial image sequences for the facades and textures from aerial images for the roofs.

#### Acknowledgment

Our work is part of the DFG (German Research Society) research projects “Enrichment and Multi-purpose Visualization of Building Models with Emphasis on Thermal Infrared Data” and “Construction progress control based on integrated point cloud interpretation and 4d building models”. The authors thank Dr. Clement, Dr. Schwarz and Mr. Kremer of Fraunhofer IOSB, Ettlingen, for their assistance during the recording campaign.

#### REFERENCES

- [1] M. Klingert, 2006. The usage of image processing methods for interpretation of thermography data ,17th International Conference on the Applications of Computer Science and Mathematics in Architecture and Civil Engineering, Weimar, Germany, 12-14 July 2006.
- [2] R. Haralick, H. Joo, C. Lee, X. Zhuang, V. Vaidya, M. Kim, 1989. Pose estimation from correspondence point data. SMC, 19 (6), pp. 1426-1446.
- [3] R. Horaud, B. Conio, O. Le Boulleux, B. Lacolle, 1989. An Analytic Solution for the Perspective 4-Point Problem. Computer Vision, Graphics and Image Processing, 1989 (47), pp. 33-44.
- [4] R. Haralick, C. Lee, K. Ottenberg, M. Nölle, 1994. Review and Analysis of Solutions of the Three Point Perspective Pose Estimation Problem. International Journal of Computer Vision, 13 (3), pp. 331-356.
- [5] L. Quan, L. Lan, 1999. Linear n-point camera pose determination. IEEE Transactions on Pattern Analysis and Machine Intelligence, 21 (8), pp. 774-780.
- [6] B. Triggs B, 1999. Camera Pose and Calibration from 4 or 5 known 3D Points. In: Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999 (1), pp. 278-284.
- [7] L. Zhi, J. Tang, 2002. A Complete Linear 4-Point Algorithm for Camera Pose Determinatio. MM Research Reprints, 21, pp. 239-249.
- [8] R.L. Hartley, A. Zisserman, 2004. Multiple View Geometry in Computer Vision”, Cambridge University Press, ISBN: 0521623049

- [9] V. Lepetit, J. Pilet, P. Fua, 2004. Point Matching as a Classification Problem for Fast and Robust Object Pose Estimation. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004 (2), pp. 244-250.
- [10] D. Nistér, 2004. An efficient solution to the five-point relative pose problem, In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26(6), 756–777
- [11] H. Mayer, 2007. 3D Reconstruction and Visualization of Urban Scenes from Uncalibrated Wide-Baseline Image Sequences. In: *Photogrammetrie – Fernerkundung – Geoinformation* 2007(3), pp. 167–176.
- [12] M. Pollefeys, D. Nistér, J.M. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, S.J. Kim, P. Merrell, C. Salmi, S. Sinha, B. Talton, L. Wang, Q. Yang, H. Stewénus, R. Yang, G. Welch, H. Towles, 2008. Detailed real-time urban 3D reconstruction from video. In: *International Journal of Computer Vision (IJCV)*, 78(2-3), 143-167.
- [13] M. Heinrichs, O. Hellwich, V. Rodehorst, 2008. Robust Spatio-Temporal Feature Tracking. In: Chen Jun, Jiang Jie and Wolfgang Förstner: Proc. of the XXI Congress of the Int. Society for Photogrammetry and Remote Sensing, Beijing, China, *International Archives of Photogrammetry and Remote Sensing* 37(B3a), pp. 51-56.
- [14] S. Reznik, H. Mayer, 2008. Implicit Shape Models, Self-Diagnosis, and Model Selection for 3D Façade Interpretation. In: *Photogrammetrie – Fernerkundung – Geoinformation*, 2008 (3), pp. 187-196.
- [15] S. Tuttas, U. Stilla, 2013. Reconstruction of facades in point clouds from multi aspect oblique ALS. *ISPRS Annals*, II-3/W3, pp. 91-96.
- [16] T. Luhmann, J. Ohm, J. Piechel, T. Roelfs, 2010. Geometric Calibration of Thermographic Cameras. In: Mills, J. P.; Barber, D. M.; Miller, P. E.; Newton, I. (Eds.), Proceedings of the ISPRS Commission V Mid-Term Symposium 'Close Range Image Measurement Techniques'. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 38 (Part 5), 411-416.
- [17] S. Lagüela, H. Gonzalez-Jorge, J. Armesto, P. Arias, 2011. Calibration and verification of thermographic cameras for geometric measurements. *Infrared Physics & Technology*, 54, pp. 92-99.
- [18] L. Hoegner, U. Stilla, 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.
- [19] P. Westfeld, D. Mader, H.G. Maas, 2015. Generation of attributed 3d point clouds from uav-based thermal imagery. In: *Photogrammetrie – Fernerkundung – Geoinformation*, 2015(5), pp. 381–394.
- [20] L. Hoegner, A. Hanel, M. Weinmann, B. Jutzi, S. Hinz, U. Stilla, 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.
- [21] L. Hoegner, H. Kumke, L. Meng, U. Stilla, 2007. Automatic extraction of textures from infrared image sequences and database integration for 3D building models. In: PFG Photogrammetrie Fernerkundung Geoinformation. Stuttgart: Schweizerbart'sche Verlagsbuchhandlung vol. 2007(6), pp. 459-468
- [22] U. Stilla, J. Kolecki, L. Hoegner, 2009. Texture mapping of 3D building models with oblique direct georeferenced airborne IR image sequences. *ISPRS Hannover Workshop 2009: High-resolution earth Imaging for geospatial information*. 38(1-4-7/W5)
- [23] D. Iwaszczuk, L. Hoegner, M. Schmitt, U. Stilla, 2012. Line based matching of uncertain 3d building models with IR image sequences for precise texture extraction. In: *Photogrammetrie – Fernerkundung – Geoinformation*, 2012(5), pp. 511–521.
- [24] E. Michaelsen, D. Iwaszczuk, B. Sirmacek, L. Hoegner, U. Stilla, 2012. Gestalt grouping on facade textures from ir image sequences: Comparing different production systems. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Science*, Vol. 39/B3, pp. 303–308.
- [25] L. Hoegner, U. Stilla. 2014. Automatic generation of façade textures from terrestrial thermal infrared image sequences. *QIRT 2014, Bordeaux, 07.-12. July 2014*.
- [26] S. Rusinkiewicz, M. Levoy, 2001. Efficient variants of the ICP algorithm. In: *Proceedings of the IEEE 3rd International Conference on 3D Digital Imaging and Modeling*, pp. 145–152.
- [27] D. Lowe, 2004. Distinctive Image Features from Scale-Invariant Keypoints. In: *International Journal of Computer Vision*, 60(2), pp. 91-110.
- [28] M. Fischler, R. Bolles, 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography", *Communications of the ACM* 24: 381–395
- [29] H. Hirschmueller, 2008. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30(2), pp. 328–341.
- [30] H. Mayer, J. Bartelsen, H. Hirschmueller, A. Kuhn, 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.