

# CALIBRATION OF AN IR CAMERA SYSTEM FOR AUTOMATIC TEXTURING OF 3D BUILDING MODELS BY DIRECT GEO-REFERENCED IMAGES

J. Kolecki<sup>ab</sup>, D. Iwaszczuk<sup>a</sup>, U. Stilla<sup>a</sup>

<sup>a</sup> Photogrammetry and Remote Sensing, Technische Universitaet Muenchen (TUM)- (stilla)@bv.tum.de

<sup>b</sup> Faculty of Mining Surveying and Environmental Engineering, AGH University of Science and Technology, Cracow, (DAAD scholarship fellow at TUM)- kolecki@agh.edu.pl

**KEY WORDS:** Direct geo-referencing, IR images, boresight & leverarm estimation, texture mapping, 3D building models

## ABSTRACT:

In the last decade the capability of thermal IR-cameras has stepwise increased while the prices have gone down. This enables to use IR technology in wide field of applications. Infrared (IR) cameras can be used to capture the IR radiation of buildings. By analyzing data of thermal sensors it is possible to detect the areas with the highest loss of heat. Nowadays in most cases thermal images are analyzed only qualitatively. However, without analyzing the geometry of the data acquisition or not dealing with the geometry of the captured scene, it is difficult to combine thermal data with other imagery or semantic information stored in the spatial data base. Aim of our work is the automated texture mapping of existing 3D building models with images recorded by IR cameras. In this way, 2D images are connected with a 3D GIS database and can be further analyzed together with the information about the geometry of objects. For texture mapping the external orientation of the image is required. In general a direct geo-referencing is possible using data from inertial navigation system (INS) and GPS. For this purpose the camera system including GPS/INS has to be calibrated. In order to utilize the GPS/INS data correctly, the parameters of the boresight and the leverarm have to be estimated. An accurate estimation of internal orientation parameters have to be taken into account. A determination of distortion was also performed. As the acquisition of the images was conducted from a helicopter platform which was vibrating with a high frequency and amplitude, some additional problems occurred. Finally the texture mapping of selected parts of the Technische Universitaet Muenchen 3D building model was performed.

## 1. INTRODUCTION

Energy and climate changes are big topics in near future. In the European countries buildings consume 40 % of the energy. 47% of the consumed energy in buildings is used for heating. Much effort is required for reducing the energy loss. Inspection and monitoring of buildings contribute in further development saving energy. Thermal infrared (IR) imaging can be used to detect areas with highest loss of heat. In the last decade the capability of thermal IR-cameras has stepwise increased while the price has decreased. This enables the usage of IR technology in a wide field of applications. Nowadays, in most cases thermal images are analyzed qualitatively only. However, without considering the geometry of the data acquisition or not dealing with the geometry of the captured scene, we are unable to combine thermal data with other imagery or semantic information stored in a spatial data base. To create this spatial correspondence between IR-images and existing 3D building models geo-referencing of the images is necessary. Accordingly, the 3D model can be projected into the image and for each surface of the model a region of the image can be selected for texture.

For texturing the entire hull of buildings, images should be taken from different viewing direction. Terrestrial images can be used for texturing of façades (Lee et al., 2002, Hoegner et al., 2007), airborne imagery is required for roofs. The manual acquisition of terrestrial imagery for larger parts of a 3D city model is often a time consuming process. For automation of this process IR cameras mounted on a mobile platform, e.g. a vehicle, can be used to capture image sequences (Hoegner & Stilla, 2009). However, terrestrial images taken from vehicle in

most cases do not allow to capture façades from inner yards or building rings. In general, missing textures from roofs and inner yards can be captured by airborne platforms.

Although aerial, nadir views seem to be proper for flat roofs they do not provide a sufficient resolution for façades especially when low-resolution thermal IR camera has to be used. In recent years oblique imagery and projected building models (Stilla et al. 2000) are considered to be the reasonable solution for texture mapping of buildings particularly in densely built-up areas (Frueh et al. 2004, Grenzdoerffer et al., 2008, Stilla et al., 2009).

For geo-referencing of aerial images, the exterior orientation (EO) parameters have to be determined. Rough knowledge of position and orientation of the sensor can be directly determined using GPS. On the other hand, an inertial navigation system (INS) provides good short-term accuracy, but in a longer time a systematic drift occurs. However, the combination of GPS and INS allows to avoid the INS drift and to bridge any loss of satellite signal by GPS (Yastikli & Jacobsen, 2005).

During the past years many papers covering the topic of direct geo-referencing have been published. Yastikli & Jacobsen (2005) discuss the performance of direct an integrated sensor orientation. Skaloud & Legat (2008) present necessary transformations by geo-referencing in national coordinates. Geodetic datum transformations as well as additional geodetic transformations are discussed. Orientation corrections and corrections of geometrical distortions are introduced. Eugster & Nebiker (2008) projected simple building models into images taken from an unmanned aerial vehicle (UAV) using

measurements from a low cost and low accuracy IMU and GPS. They observed misalignments and suggested to improve the projection by a matching process. Such a matching process using IR imagery was shown in Stilla et al. (2000). Frueh et al. (2004) used high resolution aerial images and presents an approach to automated texture mapping of existing 3D city model. Using an approximation of the EO lines are projected and matched with the image. Random differences of the camera pose are used for repeating the process and finding the best pose which is used for texture mapping.

Some others address the problems of systematic errors resulting from the IMU misalignment and shift with respect to the camera. Eugster & Nebiker (2007), dealing with low cost IMUs installed on UAVs, proposes the incorporation of three misalignment angles (boresight parameters) as the additional parameters of bundle adjustment. The measurement of the coordinates of the lever arm vector before the flight is assumed to be sufficient. Cramer (2002) indicates the need of including leverarm values to the bundle block adjustment too. However, his work is mainly related to high quality IMUs used in high-accuracy demanding photogrammetric missions.

Aim of our work is the automated texture mapping of existing 3D building models with images recorded by IR cameras. In texturing with IR images a high accuracy of geo-referencing is necessary. Compared to aerial imagery (Frueh et al., 2004) IR imagery shows in general a lower contrast and a lower resolution which results in the much more difficult matching process based on edges. The IR image sequence was taken additionally and simultaneously to a laser scanning during a flight campaign (Hebel & Stilla, 2007). The IMU was mainly required for the laser scanner and a boresight calibration was carried out according to it. The interior orientation of the used IR camera was unknown as well as the shift and misalignment of the camera coordinate system related to the IMU coordinate system.

In our previous work (Stilla et al., 2009) a system calibration method based on simple comparison of bundle adjustment results and IMU measurements were presented. In this paper we propose an extended bundle adjustment with camera self calibration calculated for representative set of images. In addition for a one flight direction control points were measured in every image. Accordingly, calculated and measured EO parameters were compared and the differences were discussed. Considering the calculated corrections to the IMU observations a texture mapping was carried out.

## 2. GEO-REFERENCING OF THE SEQUENCE

### 2.1 Sequence geometry

During the flight campaign the helicopter flew four times above the test area TUM recording a sequence of oblique images. Afterwards four subsequences (stripes) representing each of the four flights were cut out. The helicopter trajectory and the stripes' directions are shown in Figure 1. The intersection of the trajectory is located above the center of the test area.



Figure 1. Flight trajectory over test area TUM (Hebel & Stilla 2007)

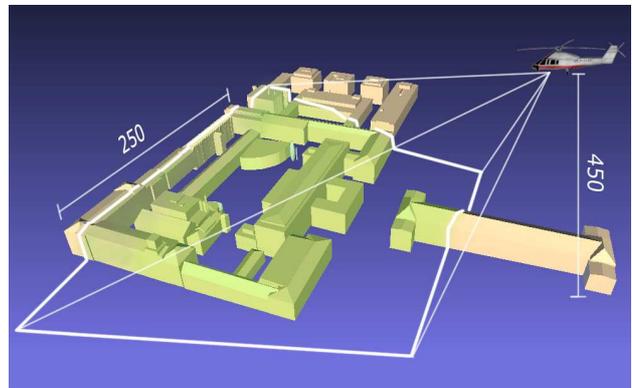


Figure 2. Approximate geometry of the single scene

Each stripe consists of 128 images (frames). Figure 2 shows the geometry of acquisition during flight of stripe #4. The inclination angle of the optical axis is approximately 45°. Images were captured with the AIM 640 QLW FLIR infrared camera with the frame rate of 25 images per second. The image resolution is 640 x 512 pixels (Figure 3).



Figure 3. An IR image (#13200) from the sequence (#4)

## 2.2 GPS and inertial data

Helicopter was equipped with the GPS/INS Applanix POS AV 510 system. Besides position and attitude/heading data system registers GPS time of the measurement and the current speed (about 160 km/h) of the helicopter. The WGS 84 coordinates are referred to the IMU center. The frequency of the IMU output is 200 Hz. The GPS coordinates are measured each second (1Hz) and are used to correct the IMU drift within the Kalman filter procedure (Grewal et al, 2007). Initial projections of the TUM 3D model in the IR image using GPS/INS EO (exterior orientation) parameters have shown about 8° misalignment of the camera and IMU coordinate systems only in the pitch rotation. Roll and yaw misalignment was few times smaller. The effects of platform vibration are observable in the plots showing angle over time. However, no such effect could be seen in the coordinate plots (compare figures 7 and 8). The frequency of vibrations is about 10 Hz and the amplitude is sometimes higher than 0.5° especially for roll and yaw angles. Pitch vibration has much lower amplitude.

## 3. TEXTURE MAPPING

In texture mapping for every surface of the 3D model one 2D texture image should be generated. To identify the region of the image corresponding to a specific surface of the model the projection of the model should be applied. In case of image sequences every single model surface is visible in many images, but only one texture with the best quality should be chosen. The quality of a texture depends on the orientation in relation to the viewing direction, the distance from the camera and possible occlusions.

To achieve the same resolution for every texture, its size is calculated from the size of the corresponding model face and the resolution coefficient that indicates the texture resolution in pixel per meter. After that 3D coordinates of every pixel are calculated. Then these 3D pixel points are projected into the image plane of the input IR image and they values are bilinear interpolated from the four neighbouring pixel values. The resulting textures are images given in pixel coordinates.

The 3D building model including links to the textures have been stored in *CityGML* format (OGC, 2008) which supports geometry and semantics of the model. This property of *CityGML* can help to interpret projected IR data with regard to surface type, such as roof, wall or ground surface. An example of a textured model using just one frame of the sequence is shown in Figure 4.

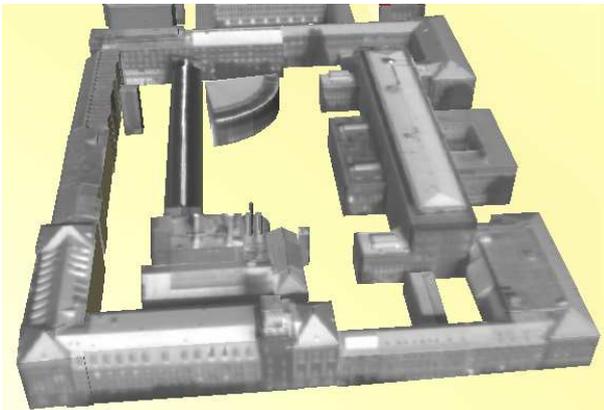


Figure 4. Textured 3D building model in *CityGML* format.

## 4. SYSTEM CALIBRATION

### 4.1 Model formulation

The system calibration was done solving the extended bundle adjustment with camera self calibration. More than 130 ground control points were measured with an accurate aerial stereo model. Control points were located mainly on the roof corners because only such details could be identified in the thermal infrared images. As only the approximate camera constant was known, its inclusion into the set of model parameters was necessary. Other parameters associated with the camera optics were: principal point coordinates, radial and tangential distortion coefficients, y axis scaling and axes skewness. Because the distance between camera projection centre and the IMU coordinate origin was unknown, the leverarm parameters has to be also be estimated. However, correct determination of boresight parameters is certainly going to have more significant importance. Assuming a 500 m distance between helicopter and TUM buildings (Figure 2), a 1° misalignment of camera and IMU results in nearly 9 m position error.

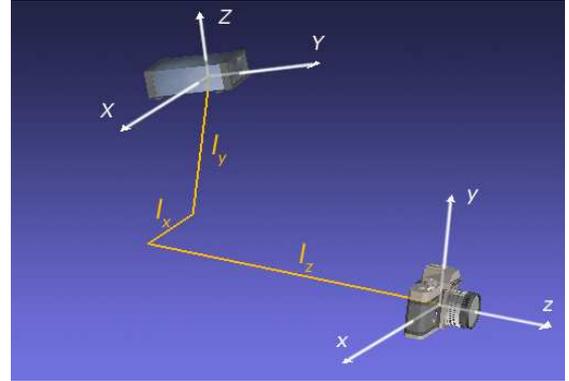


Figure 5. Leverarm vector ( $l$ ) and IMU misalignment

The leverarm vector should be given in the camera coordinate system as it stays constant during the whole flight (Figure 5). In addition the systematic GPS error could have stronger influence on the measured camera coordinates. In contrast to the leverarm values the vector of GPS error has to be expressed in the global coordinate system.

$$\bar{X}_k^{obs} + \bar{v}_{X_k} = \bar{X}_{0k} - R_k \bar{l} - \bar{D} - \bar{c}_X t \quad (1)$$

The coordinate observation equation (1) has several parameters modeling IMU position errors. The  $X_k^{obs}$  is the vector of observed coordinates of  $k^{th}$  image. To use this coordinates as a position of the projection centre we have to take lever arm ( $l$ ) and position systematic error ( $D$ ) vectors into account. The  $c_X$  is the vector of parameters describing the slow linear change of  $D$  in time ( $t$ ).  $R_k$  stands for the rotation matrix of  $k^{th}$  photo. The equations of angular observations (2) written in algebraic form have three boresight parameters ( $e$ ) respectively for roll ( $\Phi$ ), pitch ( $\theta$ ) and yaw ( $\psi$ ) angles. The modeling of gyro drift is also performed ( $c_\phi, c_\theta, c_\psi$ ).

$$\begin{aligned}
\phi_k^{obs} + v_{\phi_k} &= \phi_{0_k} - e_{\phi} - c_{\phi}t \\
\theta_k^{obs} + v_{\theta_k} &= \theta_{0_k} - e_{\theta} - c_{\theta}t \\
\psi_k^{obs} + v_{\psi_k} &= \psi_{0_k} - e_{\psi} - c_{\psi}t
\end{aligned} \quad (2)$$

## 4.2 Model application

The observation equations (1) and (2) are included into the model to find parameters which compensate systematic errors of the measurement system. The bundle adjustment is going to be calculated only for few photos selected from all four stripes. Very high weights should be assigned to the observed EO parameters to keep their residuals ( $v$ ) small. In this way the equations (1) and (2) can be regarded as conditions. The calculated parameters ( $l, e, D, c$ ) will be subsequently used to correct GPS/INS observations for all photos we want to use for texture mapping. The common parameters for four stripes are going to be found. The parameters of interior orientation and additional parameters compensating image distortion will be calculated.

The formulated model also appeared to be useful for analyzing the dependencies between the GPS/INS measurements and calculated exterior orientation parameters of images captured during longer period of flight. Ground control points have been measured on all photos from stripe #4 using automatic point transfer algorithms. Then the bundle adjustment was calculated treating the GPS/INS measurements as the low accurate observations and as the approximate EO parameters. The additional parameters ( $l, e, D, c$ ) were excluded to accommodate model to the simple bundle adjustment case. Subsequently the lever arm constraint was applied to see how this restriction influences the calculated angles. Some interesting dependencies concerning influence of the helicopter vibrations on the bundle adjustment results were observed.

## 5. RESULTS

### 5.1 System calibration

In order to determine all of the modelled parameters ( $l, e, D, c$ ), the bundle adjustment with inclusion of equations (1) and (2) was calculated using 11 images. The RMSE for GCP were 0.52, 0.47 and 0.50 meters respectively for X, Y and Z coordinate. The RMSE for fiducial x and y coordinates was 1.58 and 0.96 pixels. This is about three times higher then for adjustment without constraints (1) and (2). This shows that considerable tensions in the network must have appeared.

The calculated parameters are used to project the 3D model lines into the image for visual assessment. An example is shown in figure 6. Most of the lines fit to the edges of structures in the IR image. Some lines (see red arrows) show a misalignment which may result from geometric inaccuracies of the 3D model. In general 3D city models designed in LOD2 (OGC, 2008) may have accuracy up to 2 m in position and in height. In a small number of images the misalignments seem to be systematic which couldn't be compensated by the formulated functional model. A video stream of the projected lines together with the background image shows oscillating misalignments which seems to be a strongly connected to the helicopter vibrations

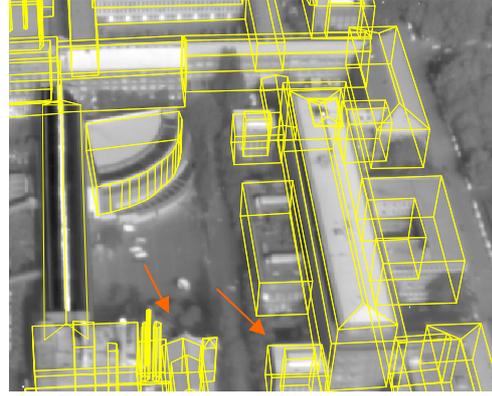


Figure 6. Projection of the model lines into the image using calculated parameters. Fragment of image from stripe #1.

### 5.2 Vibrations

In addition to system calibration the comparison between calculated and measured EO parameters of images from the whole stripe #4 was made. Figure 7 shows a plot of Z coordinate over time. It can be seen that the differences between calculated coordinate of two subsequent images have a fluctuation up to 2 meter. Plots of X and Y coordinates show similar effects.. We do not assume that this amplitudes are not realistic for vibrations of helicopter..

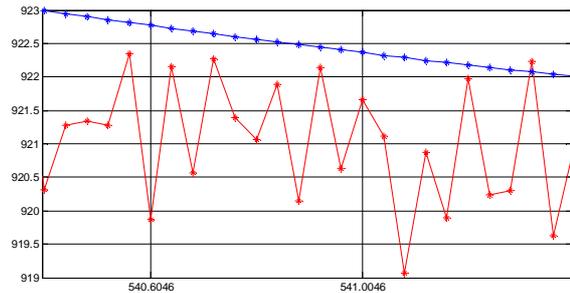


Figure 7. Z coordinate over time plot for fragment of stripe #4: blue - measured, red - calculated.

The calculated angles show that the bundle adjustment captures the general trend of changes. Figures 8 and 9 show the results for yaw and pitch angle for stripe #4. The plots of angle values have been shifted down to fit them together with the difference plot. In the figure 9 we see that the changes of calculated pitch angles are more rapid then for measured ones. Applying the leverarm constraint ( $l$  parameter), which makes the coordinate differences almost constant for the whole stripe, suppresses the angle jumps in general. This effect is different for each angle, and strongest for the pitch (Figure 10).

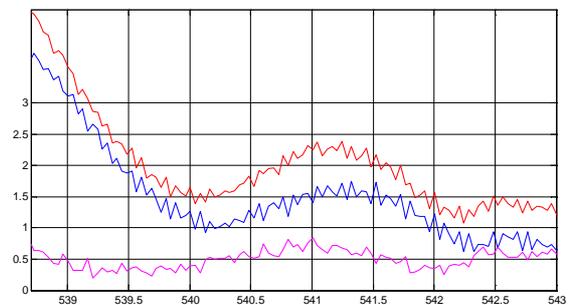


Figure 8. Yaw angle over time plot for stripe #4: blue - measured, red - calculated, magenta - difference.

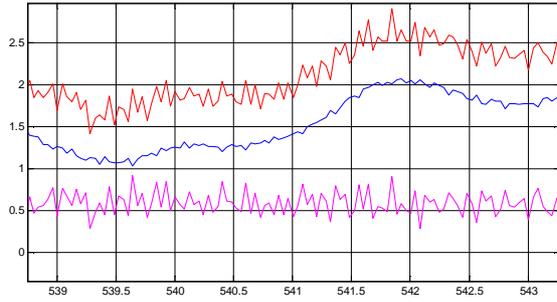


Figure 9. Pitch angle over time plot for stripe #4: blue - measured, red - calculated, magenta - difference.

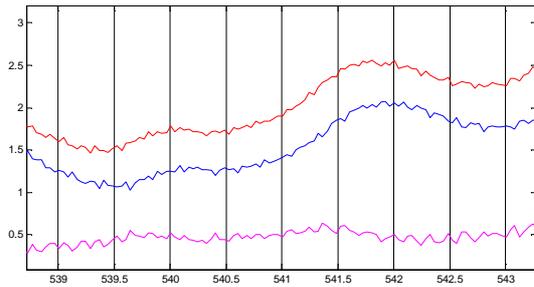


Figure 10. Pitch angle over time plot for stripe #4 after applying leverarm constraint: blue - measured, red - calculated, magenta - difference.

Presented results show the interesting bundle adjustment response on the camera vibrations. The strong vibrations result in rapid changes of all calculated EO parameters. The jumps of coordinates from image to image are about 1.5 m. This corresponds in our case (Figure 2) to terrestrial effect of angle change of about  $0.2^\circ$  so we can assume that the vibrations influence all EO parameters almost equally. Inclusion of leverarm condition cancels the jumps of calculated position and suppresses the differences of angles. At the same time the adjustment statistics are getting slightly worse. There is no evident explanation why the bundle adjustment converges to the solution with high differences of positions of two subsequent projection centres and not to the solution with smoother changes.

In addition to quick oscillations of angular differences observed in figures 8, 9 and 10 we can see some long term changes of yaw differences (Figure 8). This effect is more a less irregular and couldn't be compensated by  $e$  and  $c$  parameters of the model. This explains the systematic differences of projected model mentioned in 5.1.

## 6. DISCUSSION

Presented results show that the parameters calculated by the extended bundle adjustment allow to align the TUM 3D model with the thermal images. For a lot of IR photos the achieved model-to-image match allows texture mapping without any additional corrections. However projected model lines do not coincident perfectly with some of the image edges. In general observed differences are up to 3 pixels and could be caused by three factors. To begin with, for some small parts of the model there is no good match for any image from all of the four stripes. Such mismatches are most likely results of the 3D model errors. Besides the regional misalignment, caused by model errors, some systematic mismatches observed in the

whole image occur. Such inconsistencies result from longer-term gyro drift oscillations like these observed in figure 8. Such drift changes could not be modelled by the proposed method. The third undesirable effect appears as a rapid vibration of projected model lines. This effect is mainly due to helicopter vibration which in addition haven't been reflected in the GPS/INS position measurement. Further we will try to give some other reasons and possible solutions to this phenomenon.

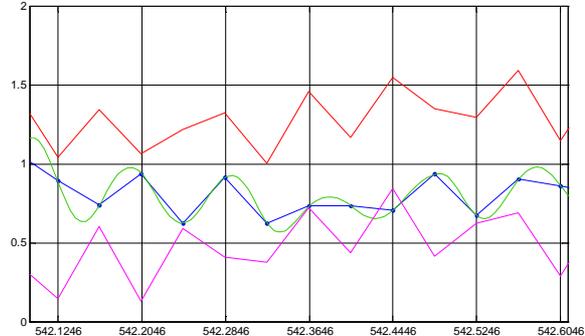


Figure 11. Yaw angle over time plot for stripe #4 fragment: blue - measured, green - IMU 200 Hz measurement, red - calculated, magenta - difference.

Figure 11 shows a fragment of the yaw measurement compared with adjustment results (red). Yaw values assigned to images are samples (blue) of the high frequency (200 Hz) measurement performed by IMU. The assignment of the measured value is done using the comparison of image acquisition time and IMU measurement time. If the helicopter flight had been smooth, then the possible delay of registered time stamp would have had no effect on the final value of the gyro measurements. However when vibration has frequency of 10 Hz and the amplitude of almost  $0.5^\circ$ , such delay impacts stored EO values.

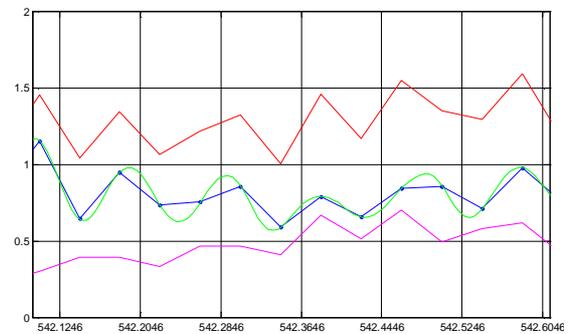


Figure 12. Yaw angle over time plot after applying 0.02 s time delay: blue - measured, green - IMU 200 Hz measurement, red - calculated, magenta - difference.

Analysing figure 11 we see that shift in the time of image acquisition of about 0.02 s effects in a better match of the red and blue lines. We can observe (Figure 12) that after performing such operation, the difference plot becomes much smoother. The effects observed in yaw plot are also truth for roll and pitch angle. Time delays which should be applied to these angles are respectively 0.007 s and - 0.01 s. Calculated time differences are different for each angle. This phenomenon could be the result of camera vibrations which are independent to IMU vibrations.

Not investigated and covered in this paper is also a rolling-shutter effect. In contrast to a frame shutter the rolling shutter

works differently, in that the photodiodes (pixels) do not collect light at the same time. All pixels in one row of the imager collect light during exactly the same period of time, but the time light collection starts and ends is slightly different for each row. The top row of the imager is the first one to start collecting the light and is the first one to finish collecting.

The researches presented in this paper show that aligning thermal IR images with 3D building model was a very complex task. The correction of measured EO parameters using calculated parameters gave satisfying results. If no perfect image to model match is possible, then some additional procedures like manual alignment should be applied. Obtained results give in all cases an excellent initial solution for automatic image to 3D model alignment based on edge and corner detection algorithms, if only such details could be detected in thermal IR imageries.

#### ACKNOWLEDGMENTS

The authors would like to thank FGAN-FOM, Ettlingen, for providing image of the flight campaign and K. Eder (TUM-FPF) for providing ground control information.

#### REFERENCERS

Cramer M. (2002) Experiences on Operational GPS/Inertial System Calibration in Airborne Photogrammetry, *Journal GIS-Geoinformationssysteme*, 6/2002, Wichmann Verlag, Germany, pp. 37-42.

Eugster H., Nebiker S. (2007) Geo-registration of Video Sequences Captured from Mini UAVs – Approaches and Accuracy Assessment, 5<sup>th</sup> International Symposium on Mobile Mapping Technology - Padua, Italy

Frueh C, Sammon R, Zakhor A (2004) Automated Texture Mapping of 3D City Models With Oblique Aerial Imagery, *Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'04)*.

Grenzdorffer, J.G., Guretzki M., Friedlander I. (2008) Photogrammetric Image Acquisition and Image Analysis of Oblique Imagery. In: *The Photogrammetric Record*, 23(124), pp. 372–386.

Grewal, MS, Weill, LR, Andrews AP (2007) *Global Positioning Systems, Inertial Navigation, and Integration*, WILEY-INTERSCIENCE, A John Wiley & Sons, Inc., Publication, pp. 382-424.

Hebel M., Stilla U. (2007) Automatic registration of laser point clouds of urban areas. *PIA07-Photogrammetric Image Analysis 2007. International Archives of Photogrammetry, Remote Sensing and Spatial Geoinformation Sciences*, Vol 36 (3/W49A): 13-18

Hoegner L, Kumke H, Meng L, Stilla U (2007) Automatic extraction of textures from infrared image sequences and database integration for 3D building models. *PFG Photogrammetrie Fernerkundung Geoinformation*. Stuttgart: Schweizerbart'sche Verlagsbuchhandlung. 2007(6): 459-468

Hoegner L, Stilla U (2009) Thermal leakage detection on building facades using infrared textures generated by mobile mapping. *Joint Urban Remote Sensing Event (JURSE 2009)*. IEEE

Lee S C, Jung S K, Nevatia R (2002), Automatic pose estimation of complex 3D building models. 6<sup>th</sup> IEEE Workshop on Applications of Computer Vision, Orlando, USA. (148-152)

Open Geospatial Consortium (OGC) Inc. (2008): *OpenGIS® City Geography Markup Language (CityGML) Implementation Specification*. Version 1.0.0, 2008

Skaloud J, Legat K (2008), Theory and reality of direct geo-referencing in national coordinates. *ISPRS Journal of Photogrammetry & Remote Sensing* 63 (2008). 272-282

Stilla U, Kolecki J, Hoegner L (2009) Texture mapping of 3D building models with oblique direct geo-referenced airborne IR image sequences. *ISPRS Hannover Workshop 2009: High-resolution earth Imaging for geospatial information*. 38(1-4-7/W5)

Stilla U, Sörgel U, Jäger K (2000) Generation of 3D-city models and their utilisation in image sequences. *International Archives of Photogrammetry and Remote Sensing*. Vol. 33, Part B2, 518-524

Yastikli N, Jacobsen K (2005), Direct sensor orientation for large scale mapping – potentials, problems, solutions. *The Photogrammetric Record* 20(111), September 2005: 274-284