MATCHING OF 3D WIRE-FRAME BUILDING MODELS WITH IMAGE FEATURES FROM INFRARED VIDEO SEQUENCES TAKEN BY HELICOPTERS OR UAVS

J. Avbelj, D. Iwaszczuk, U. Stilla

Photogrammetry & Remote Sensing, Technische Universitaet Muenchen (TUM)

Commission III, WG III/5

KEY WORDS: Infrared, Image, Sequences, Feature, Matching, Texture, Mapping, Urban

ABSTRACT:

3D building models are helpful for interpretation of the image contents. Registering them with the images a matching of image features and corresponding structures in the 3D model is required. In this work a method to register a building model and extracted features detected in IR image sequences taken by helicopter or UAVs is presented. Polygons of the building model are projected to the images, using coarse orientation parameters obtained by GPS/INS. By the point-to-point matching of projected polygons and features extracted from the image the values of the projection matrix are improved. Two different point types, i.e. Förstner points and intersection points, are extracted in infrared (IR) images and used for matching. For experiments a sample data set from a dense urban area is used and presented results are evaluated.

1. INTRODUCTION

1.1 Motivation

Analysis of images taken by mobile platforms from urban areas is an important field of research. It has significant meaning for development of applications, e.g. urban planning, simulations, change detection, subsidence of surface, urban heat island detection, building inspection. Urban areas appear different compared to rural areas due to defined boundaries of different land use, less vegetation, more and higher buildings, more complex road networks and regular man-made structures. Focusing on urban areas, there are many existing city models of different Level of Detail (LOD) which are often represented with textures in visual domain. City models are helpful for interpretation of the image contents (Stilla et al, 2000). Some structures appear in visual (VIS) and infrared (IR) domain, whereas some, e.g. heating systems, thermal leakages, can only be captured by IR data. For instance, urban heat islands can be observed in IR domain in small scale, whereas inspection of buildings requires larger scale. The thermal information of facades and roofs is important for building inspection and energy saving. Texturing existing 3D building models with IR images enrich the model database and enables further analysis. The regular structures of buildings should be used to find correspondences between IR images and models.

1.2 Related work

Textures can be extracted from the images of facades and roofs automatically, if the position and orientation - exterior orientation (ExtOri) parameters - of the camera are known. Approximated ExtOri parameters can be obtained from a GPS/INS system integrated with the camera. Using these parameters the projection of the 3D building models should be carried out to identify the region of the image corresponding to a specific face. Unfortunately, the ExtOri from a GPS/INS system is often not close enough to gain a model with high quality textures.

To solve this problem many researchers tried to estimate the pose automatically with methods of image processing. For this purpose many algorithms matching the 3D model with images were developed. In Frueh et al. (2004) Canny edge detector and recursive endpoint subdivision algorithm is used to identify the line segment in the image. Then projection of the model into the image from a random pose of the camera is carried out. For this pose a rating based on line matching is calculated and this procedure is repeated. The pose with the highest rating is chosen for texture mapping. It should be mentioned that the computational effort of such a method is very high. Hsu et al. (2000) also uses projection of the 3D model line segments and align them to oriented image gradient energy pyramids. The inter-frame pose prediction is calculated using tracked features. Edges belonging to the same entity in object space are often extracted incompletely and inaccurately in a single image. Thus, Tian et al. (2008) suggests an edge matching using a geometric constrains based on reliable points chosen in quality analysis through the sequence.

Some other authors propose application of vanishing points for estimation of the rotation of the camera. In Cham et al. (2010) a method for camera pose estimation in Manhattan-world scenes (i.e. consisting of piece-wise planar surfaces with dominant directions) is presented. Lee et al. (2002) describes a method for pose estimation using two or three vanishing points.

In Ding et al. (2008) a method for registration of aerial images with 3D building models is introduced. Vertical vanishing points combined with GPS data are used for rough estimation of extrinsic parameters of a calibrated camera. Then, the coarse parameters are refined using matching of orthogonal corners.

In previous papers (Stilla et al., 2009; Kolecki el al., 2010) we addressed the problems which are given by vibrating platforms such as UAVs or helicopters. Aim of our previous work was direct geo-referencing with GPS/INS data and system calibration.. In this contribution we present an approach to refine the ExtOri parameters applying a matching of 3D wire-

frame building models with image features from infrared video sequences.

In Section 2 the proposed method is described. Firstly, different types of features appropriate for IR images are investigated (Section 2.1). Secondly, a point-to-point matching algorithm is presented (Section 2.2). Then, in Section 3 the method is applied on test dataset, using sub-model of the building model to evaluate the matching algorithm. Finally, in Section 4 efficiency and usage of applied method is discussed.

2. METHOD

An automatic texturing process requires geo-referenced data and a very precise co-registration between images and faces of the model. Before texturing, polygons of the building model can be projected to the images, using coarse orientation parameters. By a matching process of projected polygons and features extracted from the image the values of the projection matrix should be improved. Aim of this work is to investigate automatic feature extraction in infrared images to improve orientation parameters of the sensor.

2.1 Investigation of features extraction in IR images

Different types of features (points, lines, segments) are investigated due to the specific properties of IR images. In this pre-survey standard feature detectors, e.g. Sobel and Canny edge detectors and Förstner operator are used. Most buildings have rectangular faces, with straight roof ridges and even building details, e.g. chimneys, windows, are often describable by simple geometric forms. For this reasons buildings can be good represented by straight line segments, instead of arcs or other curves. In Fig. 1 extracted straight edges are represented in orange colour. It can be easily noticed that almost all extracted lines represent edges on the roofs.

For matching extracted line segments and building model lines projected into images, a correlation function is needed. Computational effort of this matching method is high and additionally, extracted line segments often do not finish at edge points. Therefore, point to point matching algorithm is in favour.

Two different point types are extracted: Förstner points and intersection points of extracted straight edge segments. Förstner points extracted with commercial software MVTec Halcon are detected in two steps. This procedure results in so called "junction Förstner points" and "area Förstner points". "Junction points" are interpretable as intersection of image edges (illustrated as red crosses in Fig. 1) and should be used for matching with projected building model. "Area points" represent change of colour or brightness regarding their surrounding and are for this reason excluded.

Intersection points of extracted straight edge segments (intersection points) are calculated using intermediate step of extracting straight edge segments (illustrated as orange lines in Fig. 1). Distances d between all end points of all extracted straight edge segments are calculated. Maximum distance limit between edge points d_{max} in pixels is set. Large limit d_{max} would consequence in higher number of extracted intersection points of which most of them could not be physically interpretable and would furthermore cause more wrong matches between model and extracted points. For all pairs of end points which fulfil the condition $d \leq d_{max}$, an intersection angle α of belonging edge

segments is calculated. The intersection angle should be between limits α_{min} and $\alpha_{max} = \pi - \alpha_{min}$ where $\alpha \in [0, \pi]$. Intersection angle between nearby buildings edges are seldom very acute or very obtuse. An angle limits α_{min} and α_{max} prevents the algorithm to result in points defined by two almost parallel lines. Almost parallel lines should be avoided because they are most likely a result of their multiple edge response, i.e. bad feature detection, not very accurate detection of two parallel edges or in parallel edges of a same building on different heights which appear very near in images acquired in oblique view.

For all pairs of edge points that fulfil both of the conditions: $d \leq d_{max}$ and $\alpha_{min} \leq \alpha \leq \alpha_{max}$ intersection points are calculated. These intersection points of extracted straight edge segments are illustrated by the centres of the yellow circles in Fig. 1.
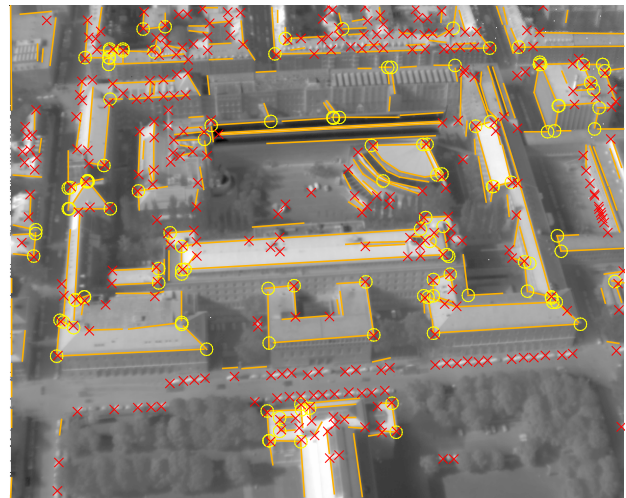

Figure 1. Extracted features; red cross – Förstner point, orange line segments – extracted straight segments and centres of yellow circles - extracted intersection point

2.2 Point-to-point matching

Establishing correspondences between model points projected into the IR image and extracted points is based on spatial relation among them. Three aspects should be considered.

- First of all, for each model point a circular search space with radius R in pixels is defined. Different cases of relations among extracted points and model points occur. More than one extracted point can correspond to one model point, but not vice versa. A case, when one extracted point corresponds to multiple model points occurs when there are short lines present in a model. Short lines are considered model lines of size of few pixels when they are projected into the image.

- Secondly, a sub model is derivated from a whole building model. In visual (VIS) domain the boundary lines between objects in an image are easily recognized. The precision with which a certain boundary line can be detected depends on spatial and radiometric resolution as well as length of detected edge. In IR domain acquired images appear to be blurred. Two neighbouring points acquired in IR domain have similar radiometric value and the boundaries are not as exact as is VIS domain. Blurred appearance of objects in IR images depends more on materials of observed objects and their physical characteristics than on the colour and illumination. In images in IR domain, differences in radiometric values

between low parts of facades and pedestrian areas are small, therefore only roof edges of model are considered.

- Thirdly, model is projected with coarse orientation parameters into an image with extracted points. Matching algorithm, basing on closeness between extracted points and model points, is applied. Size of a search space for correspondence depends on accuracy of coarse orientation parameters. Co-registered points are adjusted with least mean square (LMS) algorithm and corrected orientation parameters are calculated.

If coarse orientation parameters are not close enough to the correct values, matching algorithm should be applied iteratively. In first iteration coarse orientation parameters are used for projection of model into the image and in second iteration adjusted orientation parameters are used for projection, etc. For every subsequent iteration, a search space for correspondence between extracted and model points R is automatically reduced. LMS adjustment is very sensitive to outliers; therefore reducing R in iterations decreases a number of outliers. Small number of mismatches (outliers) according to correct matches will move a model in correct direction but not on exact position. By reducing R in next iteration, some mismatches can be awoided.

3. EXPERIMENTS

3.1 Data overview

3.1.1 The thermal images and geometry of acquisition: The thermal images were taken in dense urban area with high resolution IR camera AIM 640 QLW FLIR with the frame rate 25 images per second, which was mounted on a platform carried by helicopter. The flight height was approximately 400 m above ground level. The camera was forward looking with oblique view of approximately 45° (pitch angle). Resolution of images is 640 x 512 pixels. A helicopter flew 4 times over the test area, i.e. main campus of Technische Universitaet Muenchen (TUM), recording a sequence of IR images (Fig. 2). From a sequence 4 stripes were cut out.

Figure 2. Flight trajectory over test area TUM
(Hebel & Stilla 2007)

Helicopter was equipped with the GPS/INS Applanix POS AV 510. GPS antenna is mounted on the helicopter cockpit and

registers position with frequency of 1 Hz. The GPS measurements are used to correct the IMU drift within Kalman filter procedure (Grewal et al, 2007).

3.1.2 Extended system calibration: Only approximate Inertial orientation (IntOri) parameters and distortion values of IR camera were known. A system calibration was done solving extended bundle adjustment with camera self calibration (Stilla et al., 2009). Camera optics parameters that were estimated and corrected are: camera constant, principal point coordinates, radial and tangential distortions coefficients, y axis scaling and skewness in x and y direction. Kolecki et al. (2010) extended system calibration to estimate also leverarm and boresight corrections, position systematic error and its change in time. Applied corrections improve projection of a building model into IR images and only few pixels misalignments are present (Fig. 3). Kolecki et al. (2010) address an effect of platform vibration, which could be observed in angular measurements of IMU unit over time.

Figure 3. Projected model into a frame of a IR image using corrected INS/GPS data (Stilla et. al. 2009)

3.1.3 A 3D building model: A 3D wire model in level of detail 2 (LOD2) was produced with semi-automated method in photogrammetric software INJECT. The model was produced from aerial images and is showing the main campus of TUM. Due to generalization some inaccuracies in position of buildings edges occur. The position accuracy was estimated on a sample of buildings extracted from a model up to 1 m.

3.2 Test

Decision which type of feature(s) is appropriate for a given task is a first step of image registration. Different features can be extracted, e.g. contours, points, straight line segments, regions. A feature detecting algorithm should assure good detection with low probability to fail to detect existing edge or detecting non-existing edge. Detected edges should be as close to the real edges and an edge should be represented with only one detected edge.

Extracted straight edge segments and contours often do not finish at edge points of buildings, but they are a good representation of the roof building edges. Feature detectors are local operators on image which detect changes in radiometric values. Longer extracted line segments are more reliable as shorter to represent a building edge, because the step change of

radiometric value in image is detected in larger surroundings. For the same reason intersection points calculated from straight edge segments are more reliable as directly extracted points. In Förstner points extraction on given dataset results in relatively high number of extracted points but they usually do not lie on the corners of the buildings. Many details, e.g. cars, small roof structures are extracted using Förstner operator. Further, only few of extracted Förstner points correspond to the model. Extraction of points as intersection of the edge segments gives smaller number of extracted points. However, in contrast to Förstner point extraction, a relatively high number of extracted points correspond to the model.

Extraction of smaller structures and details in image, which are usually not relevant for searching the correspondence with the model, can be avoided using straight edge detection. A threshold value for minimum length of extracted lines is applied. In Fig. 4 extracted straight edge segments are classified in 3 groups on the basis of length in pixels. Threshold values for length of lines should be chosen regarding to the relative size of building edges in an image.
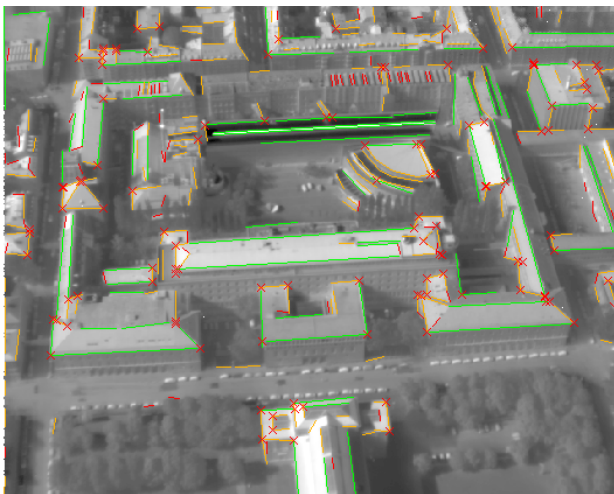

Figure 4. Extracted straight edge segments with traffic light colour coding. In red: short, not reliable lines (8-12 pixels); in orange: middle length lines (12-32 pixels); in green: the longest, most reliable (>32 pixels), R = 12 pix

From each of four stripes, images from the beginning, middle and end of stripes were taken. Two point extraction algorithms were used, i.e. Förstner points and intersection points. Firstly the method was tested on the sub-model of one building and two buildings and afterwards applied on the whole model. Secondly, for a sub-model of one building a subset of 95 images from stripe #4 is taken and extended evaluation of adjusted ExtOri parameters is carried out. Coarse ExtOri parameters are changed and the efficiency of method is observed.

3.3 Results

Applying methodology on sub-model of one building and on sample of 12 images gives poor result using Förstner point extraction as well as point extraction via straight edge detection. Testing the methodology on same image set and model of two buildings has shown some additional problems with wrong correspondence between model and extracted points. A relative position of sub-model in image significantly influences possibility of successful matching algorithm; the inspected building should not be occluded and should lie in the middle of

the image. Radial and tangential distortions have bigger influence on border of image; the upper part of image is more deformed and buildings appear relatively smaller to the ones on the bottom of the image due to oblique view. Other evident problems when searching for correspondence between extracted points and sub-model are model incorrectness and poor feature detection due to small edge response. Matching over the sequence with whole model, or sub-model is with this methodology not efficient on given dataset. Further investigation on a different 3D building model would be required to better evaluate the method.

An example of successful correction of ExtOri parameters is shown in Fig. 5. A sub–model of building 1 is projected into the image no. #13170, 4th flight. Corrected coordinates (the corrections applied are described in section Data description) are moved for 5 meters in position (X, Y, H direction), rotation angles are equal to corrected ExtOri parameters. With these moved ExtOri parameters a sub-model is projected into the image (Fig. 5A,). The matching algorithm was applied in 3 iterations For 1st projection of sub-model, there is one mismatch - encircled with red (Fig. 5A), which does not appear in the further iterations. Initial input ExtOri parameters for 1st iterations (Fig. 5 A, B) are corrected ExtOri parameters moved for 5 m; in 2nd iteration (Fig. 5 C, D) initial parameters are adjusted parameters from 1st iteration, etc. In second column movement of model is presented: in i[th] iteration in dark blue and i[th]-1 iteration in light blue colour, where iteration i =1,2,3 and i-1 is projection of sub-model with moved ExtOri parameters. R is reduced in iteration, with initial value R = 15 pixels.
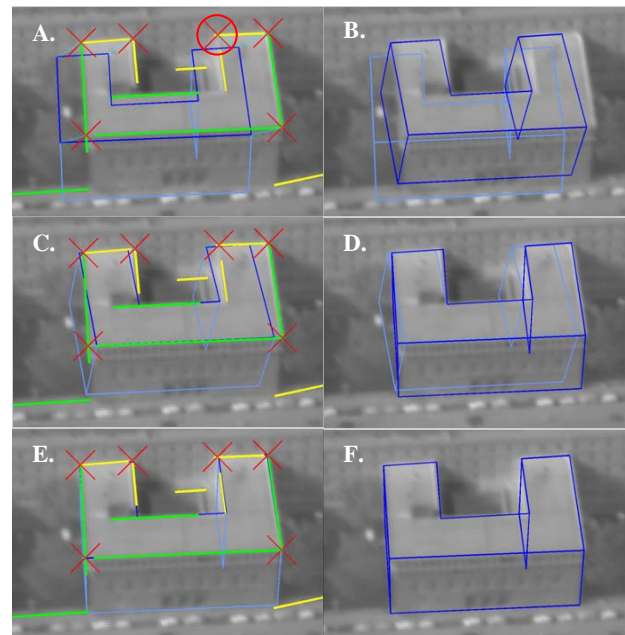

Figure 5. Efficient matching. In A.,C.,E. extracted lines and intersection points registered with the projected model,are presented. Red encircled point is wrong matched. In B.,D., and F. in light blue is projected model with initial ExtOri parameters and in dark blue with adjusted ExtOri. In F. Sub-model fits better to contour of the building.

In Fig. 6 and example of wrong matching is shown. Building 1 is projected into image #13164, the coordinates are moved for 5 meters in position, angular values are same as corrected ExtOri parameters. Extracted points encircled with red are connected to the wrong building corner, tho they are correctly extracted.

Points encirceled with yellow are correctly extracted but due to the large misalignement they are not connected to any model points. Points which are only marked with red cross are correctly extracted and mached to the model. Nevertheless only 4 points are registered in the initial projection of sub-model into the image, of which one is mismatch. In this case the method results in worsening the coarse ExtOri parameters which can be observed in Fig 6 F, where sub-model does not fit to the contour of the building.
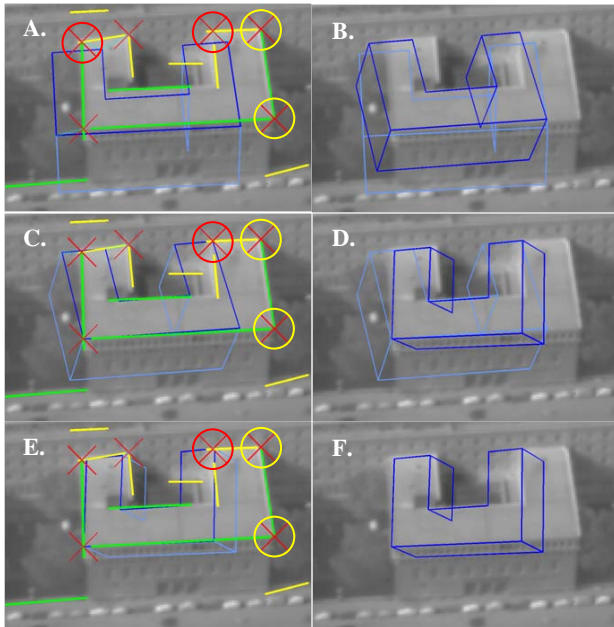

Figure 6. Unsuccessful matching. Red encircled intersection points are wrong connected to the model, yellow encircled are not connected, but correctly extracted. In F sub-model does not fit to the contour of the building.

3.4 Evaluation

Evaluation of algorithm was checked manually on a dataset of sub-model of one building, stripe #4 images 13141-13235. The efficiency of method, correctness and completeness of extraction and matching algorithm were surveyed. For feature extraction, point extraction via straight edge detection was used.

3.4.1 Efficiency: The method is efficient if the model re-projected with adjusted ExtOri parameters is on better position than with not-adjusted ExtOri parameters. Efficiency is calculated as quotient between all correctly moved rotated projections of model into the image and number of all tested images, that is 95 for this case. The initial, i.e. corrected ExtOri parameters were moved for 1 to 5 meters with 1 m increment in position, (X,Y and Z direction), than the roll, pitch and yaw initial values were moved for 10' to 50' with increment of 10'. The results are presented in Tab. 1. The algorithm is 100% efficient on a given dataset when ExtOri parameters were moved for 1 or 2 m, and up to 20'. However with 3 m in position or 40' in angular accuracy of coarse ExtOri parameters the efficiency is still over 90 %. The matching algorithm is very sensitive to angular changes.

Translation	dX=dY=dZ [m]				
	+1	+2	+3	+4	+5
Efficiency	1.00	1.00	0.92	0.65	0.16
Angular shift	droll=dpitch=dyaw[']				
	+10	+20	+30	+40	+50
Efficiency	1.00	1.00	0.97	0.93	0.61

Table 1. Efficiency of method

Efficiency of refinement of the corrected ExtOri parameters (Stilla et al., 2009; Kolecki el al., 2010) demands more close evaluation of method, because corrected ExtOri parameters are very close to real values. Position inaccuracies of detected features have greater influence on efficiency of the method when coarse ExtOri parameters have higher accuracy. Localisation inaccuracies in feature detection cause smaller rotation and/or translation of projected sub-model with adjusted ExtOri parameters compared to projection with initial, corrected ExtOri parameters. The effectiveness of algorithm was evaluated using 3 values:

- "Good", algorithm is efficient
- "Minor change", re-projected sub-model with adjusted ExtOri parameters appears to be a bit rotated or moved according to initial position of projected sub-model and image data
- "Bad", sub-model projected with initial ExtOri parameters is on better position than with adjusted ExtOri parameters, larger rotation and/or translation is present according to initial projection of sub-model and image. Results are in Tab. 2.

	Efficiency	
corrected ExtOri parameters	Better	0.64
	Minor change	0.22
	Worse	0.14

Table 2. Efficiency of method using corrected ExtOri parameters

3.4.2 Completeness and correctness of extraction: For the point extraction using straight edge detection for used extraction parameters the completeness, i.e. quotient between number of correctly extracted points and all model points, is estimated to nearly 75 %. Correctness of extraction, quotient between correctly extracted points and all extracted points, is estimated to nearly 88 %. Extracted features with Förstner operator have significantly lower correctness of extraction. This is a consequence of the fact that many of extracted Förstner points represent small structures that are not present in building models.

3.4.3 Completeness and correctness of matching algorithm: Corectness and completeness of matching algorithm were checked for two cases, firstly using corrected coordinates as initial values and secondly with moved coordinates for 3 m in position and 30' in angular values of ExtOri parameters. The matching and adjustment was made for each image in 3 iteratons. Evaluation the matching algorithm when moved coarse ExtOri parameters are used, it is seen, that completeness and correctness of matching is increasing in first 3 iterations. The corrected ExtOri parameters are very near the correct value, therefore no obvious change in values for correctness and completeness of matching is seen with higher iterations.

	Corectness of matching		
iteration	1	2	3
Corrected	0.904	0.904	0.904
Moved	0.870	0.874	0.903
	Completeness of matching		
iteration	1	2	3
Corrected	0.904	0.904	0.904
Moved	0.870	0.874	0.903

Table 3: Correctness and completeness of matching algorithm.

4. DISCUSSION

The tests of the algorithm showed that the adjustment of ExtOri parameters using correspondence between Förstner points and the model does not improve them. This is caused by not precise position of extracted points. In addition, many Förstner points do not represent corners, but other objects e.g. cars which appear in oblique view images close to the building corners. In contrast, adjustment based on intersection points of line segments gives good results when the inspected building is in the centre or lower part of an image. The coarse ExtOri parameters should be known with high accuracy, better than 4 m in position and 0.5° in orientation. With DGPS measurements better than 4 m accuracy of position is expected. However the problem remains with measurements of roll, pitch and yaw angles.

The problem with relative position of sub-model in the image can suggest that the distortions obtained during the system calibration are not accurate and presented method is not robust against this inaccuracy.

The presented method is applicable for a refinement of the position and orientation obtained in the global system calibration of the whole sequence. It can improve the texture mapping for single building or building part caused by not modelled vibrations of the system (Kolecki et al., 2010) or small inaccuracies in the model. However, on the given dataset the method is does not improve the ExtOri significantly when applied on a whole model.

5. CONCLUSION AND FUTURE WORK

In this work an approach to correct the exterior orientation (ExtOri) parameters of an IR camera mounted on a flying platform was presented. This method bases on a point-to-point matching of the image with a wire frame building model.

Extracting intersection points from straight edge segments can be extended using ellipse or rectangular search space for intersecting the line segments instead of distance between endpoints. In this way, direction and length of line are weighted. Further more points extracted from longer lines can be applied with higher weight in LMS adjustment. The matching could be made using first detecting features in lower resolution images and in iterations with higher resolution of the same image. Besides, RANSAC could be used to avoid the influence of mismatches to the re-estimated ExtOri parameters.

ACKNOWLEDGMENTS

REFERENCERS

Cham T J, Ciptadi A, Tan W C, Pham1 M T, Chia L T (2010) Estimating Camera Pose from a Single Urban Ground-View Omnidirectional Image and a 2D Building Outline Map. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, USA

Ding M, Zakhor A (2008) Automatic registration of aerial imagery with untextured 3D LiDAR models, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Anchorage, Alaska, June 2008

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)

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

Hsu S, Samarasekera S, Kumar R, Sawhney H S (2000) Pose Estimation, Model Refinement, and Enhanced Visualization Using Video, Proceedings of CVPR 00, Hilton Head Is., SC, vol. I, pp.488-495, June 2000.

Kolecki J, Iwaszczuk D, Stila U (2010) Calibration of an IR camera system for automatic texturing of 3D building models by direct geo-referenced images. Eurocow 2010, Spain

Stilla U, Soergel U, Jaeger K (2000) Generation of 3D-city models and their utilization in image sequences. International archives of photogrammetry and remote sensing, vol. 33, part B2, 518-524.

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)

Tian Y, Gerke M, Vosselman G, and Zhu Q (2008) Automatic Edge Matching Across an Image Sequence Based on Reliable Points. In: J. Chen, J. Jiang and W. Förstnner (Editors), the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science Beijing, pp. 657-662