PFG 2012 / 5, 0511 – 0522
Stuttgart, October 2012

Article

Line based Matching of Uncertain 3D Building Models with IR Image Sequences for Precise Texture Extraction

Dorota Iwaszczuk, Ludwig Hoegner, Michael Schmitt & Uwe Stilla, München

Keywords: line matching, infrared images, building model, texture

Summary: Thermal building textures are used for the detection of damaged or weak spots in the insulation of building hulls. These textures can be extracted from directly geo-referenced oblique airborne infrared (IR) image sequences by projecting a 3D building model into the images. However, the direct geo-referencing is often not sufficiently accurate and the projected 3D model does not match the structures in the image. Thus we present a technique with the main goal of finding the best fit between the 3D building model and the IR image sequence. For this purpose we correct exterior orientation via line based matching. We assign image lines to projected model lines based on the distance and angle between them. The maximum distance and maximum angle between assigned lines is given by the uncertainties in the projected lines, which is derived from the uncertainties in the 3D building model and from the uncertainties in the camera position and orientation by error propagation. Then we use the random sample consensus (RANSAC) to remove incorrect correspondences. The correspondences selected by RANSAC are adjusted using the least squares method. In the adjustment we consider both uncertainties in the model and in the image features. To evaluate the presented method we test it running the algorithm among the set of images and visually assess the improvement.

Zusammenfassung: *Linienbasierte Zuordnung von unsicheren 3D-Gebäudemodellen mit IR Bildsequenzen zur präzisen Texturgewinnung.* Thermische Gebäudetexturen werden für die Detektion von Schwach- und Schadstellen in der Isolation von Gebäudehüllen eingesetzt. Solche Texturen können aus den direkt georeferenzierten, infraroten (IR) Luftbildsequenzen gewonnen werden, indem das 3D Gebäudemodell ins IR Bild projiziert wird. Das direkte Georeferenzieren ist jedoch oft nicht genau genug und das projizierte 3D Modell stimmt nicht mit den Bildstrukturen überein. Deswegen wird hier eine Technik präsentiert, mit dem Ziel, die beste Anpassung zwischen dem IR Bild und dem 3D Gebäudemodell zu finden. Dafür korrigieren wir die Parameter der äußeren Orientierung mittels linienbasierter Zuordnung. Die Bildlinien werden den projizierten Modelllinien aufgrund der Entfernung und aufgrund des Winkels zwischen ihnen zugeordnet. Die maximale Distanz und der maximale Winkel zwischen zugeordneten Linien ergeben sich aus der Unsicherheit der projizierten Linien, die aus den Unsicherheiten der Kameraposition und Orientierung und aus den Unsicherheiten der 3D-Gebäudemodelle mittels Fehlerfortpflanzung berechnet werden. Danach verwenden wir RANdom SAmple Consensus (RANSAC), um die fehlerhaften Korrespondenzen auszusortieren. Die von RANSAC ausgewählten Korrespondenzen werden mit der Methode der kleinsten Quadrate ausgeglichen. Bei der Ausgleichung werden zwei Unsicherheiten berücksichtigt: die des 3D Gebäudemodells sowie die der Bildmerkmale. Um die präsentierte Methode zu evaluieren, wird der Algorithmus in einer Bildsequenz angewandt und die Verbesserung visuell beurteilt.

www.schweizerbart.de
DOI: 10.1127/1432-8364/2012/0135
1432-8364/12/0135 $ 2.75

1 Introduction

Thermal inspections of buildings contribute to the detection of damaged and weak spots in the building structure. Three dimensional (3D) spatial referencing of the captured images helps the interpretation of the data, especially for large area inspection using images taken by a mobile mapping system. Façades seen from the street level can by captured by a camera mounted on a vehicle (Hoegner et al. 2007) and the roofs are imaged from a flying platform. Using multi aspect airborne oblique images, the missing walls in inner yards are captured. The spatial reference is achieved by combining infrared images with 3D building models via texture mapping. The existing 3D building models can be projected into the infrared (IR) images and the building textures can be extracted. For the projection the exterior and interior orientation parameters of the camera need to be known. These parameters can be determined directly from the navigation device and camera system calibration parameters (camera calibration, boresight and lever-arm calibration). Unfortunately, the direct geo-referencing is often not sufficiently accurate and the model does not match the structures in the image. To refine the registration a model-to-image matching should be carried out.

2 Related Work

In literature the model-to-image matching problem for airborne imagery is frequently addressed and many methods for solving the problem have been presented. Früh et al. (2004) propose line matching based on slope and proximity by testing different random camera positions. However, as Ding & Zakhor (2008) mentioned, this method requires high computational effort. Hsu et al. (2000) search for the best camera position by minimizing the disagreement between projected features and features detected in the image. Other authors propose methods for coarse orientation which use vanishing points (Ding & Zakhor 2008, Förstner 2010). These methods lead to faster results, but they assume so called "Manhattan scenes", where many horizontal and vertical lines can be detected in the image. In some works (Vosselman 1992, Eugster & Nebiker 2009) relational matching is applied, which does not only consider the agreement between an image feature and a model feature, but also takes the relations between features into account.

Methods can also be differentiated based on which image features they use for matching. Some authors propose points (Ding & Zakhor 2008, Avbelj et al. 2010), but most works consider lines as more natural for building structures and use them for co-registration (Debevec 1996, Früh et al. 2004, Schenk 2004, Eugster & Nebiker 2009). In some papers hybrid methods employing points and lines at the same time are presented (Zhang et al. 2005, Tian et al. 2008).

Only few authors take the uncertainty of the applied 3D models into account. Usually the models used for these kinds of matching are stored in a parameterized form (Sester & Förstner 1989, Lowe 1991, Debevec et al. 1996, Vosselman 1998), which is very useful for 3D reconstruction, because the parameterized models represent simple buildings or building primitives. However, reconstructed building models are frequently modelled by polyhedra and stored in a format supporting polyhedral models, e.g. CityGML. Thus, for matching images with existing 3D building models other methods should be developed.

In our research we present a method for matching polyhedral 3D building models with thermal IR imagery. Because we do not use the parametric model representation, we cannot directly apply the stochastic model proposed by Sester & Förstner (1989). In contrast to Hoegner et al. (2007) we do not use terrestrial image sequences but instead match images taken by a camera mounted on a flying platform. We extend our previous work (Avbelj et al. 2010) as we do not use the intersection points but apply a line based matching and consider both, the uncertainties in the 3D building model and those in the image features. The goal of this technique is an improvement of the camera position, so that the best fit between the 3D building model and the image structure is achieved.

3 Line based Matching

In the presented research the main goal is to find the best fit between the existing 3D building model and the IR image sequence. We assume a calibrated camera system with GPS/INS navigation, known interior orientation of the IR camera, and known lever-arm and bore-sight parameters. Accordingly, the camera position and orientation are determined quite accurately. This allows the projection of the 3D building models into the image using the collinearity equations. Unfortunately the projected 3D model often does not match the structures in the image. On one hand this is a consequence of the remaining error after the camera calibration and the camera pose (position and orientation) determination. On the other hand the mismatch can be related to the rolling shutter effect. Rolling shutter occurs in cameras with a line-wise readout system when the photographed objects or the camera are moving. Such systems are often used in IR cameras, thus the rolling shutter effect is particularly noticeable in IR images taken by a moving camera and can lead to a "shrinking" or "stretching" effect on the imaged objects. Another reason for the mismatches can be the unmodelled lens distortions of the IR camera or unmodelled vibrations of the camera. To reduce all these remaining errors we propose a matching procedure between the 3D models and each IR image frame.

Our method is based upon the least squares method. We use the stochastic model to embed the uncertainties of the extracted image lines and the uncertainty of the 3D building models. The uncertainty of the image lines is given by the uncertainty of the extraction process of the lines, and the uncertainty of the 3D building models is a result of the inaccurate extraction process and generalization. In this research we use 3D building models generated from aerial images. Therefore we assume the roofs to be more reliable, because they were directly measured during the 3D reconstruction. Besides, the radiometric properties of the ground (sidewalks and streets) in thermal IR are similar to the properties of the walls, so that the edges between have very low contrast and often cannot be extracted.

3.1 *Line Parameterization*

Polyhedral 3D building models are stored as a set of polygons defined by lines (edges) and points (corners). Lines are natural building structures, which can be detected in the image. To use these lines for co-registration a mathematical description of a line is needed. Typically a line in 3D is described by a direction vector $\boldsymbol{v}$ and a point $\boldsymbol{P}$. For this description any point $\boldsymbol{P}$ belonging to the line can be used, thus there is more than one set of parameters describing one line. To solve this problem Roberts (1988) introduced a line representation which is unique and unambiguous. This line representation was discussed, varied and applied in photogrammetric context by Schenk (2004).

Roberts' line representation is based on two orientation parameters (α, θ) and two positional parameters (X_s, Y_s). The azimuth α and zenith θ can be deduced from the spherical coordinates of vector $\boldsymbol{v}$. (X_s, Y_s) are coordinates of the intersection point with the plane $X'Y'$, where $X'Y'Z'$ is the rotated original coordinate system XYZ, so that the Z'-axis is parallel to the line (Fig. 1a). All equations required to calculate these parameters are given by Schenk (2004) and Meierhold et al. (2008). Each point on the line can be expressed as

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} X_s \cos\alpha\cos\theta - Y_s \sin\alpha + t\cos\alpha\sin\theta \\ X_s \sin\alpha\cos\theta + Y_s \cos\alpha + t\sin\alpha\sin\theta \\ -X_s \sin\theta + t\cos\theta \end{pmatrix}. \tag{1}$$

As we can see, all lines, including vertical and horizontal ones, are defined using (1). This parameterization uses four parameters, which is the number of degrees of freedom of a 3D line. We use this representation of lines to express the edges of the 3D building model.

Similarly, we searched for a 2D line representation which uses the minimal number of parameters and is defined in all cases. For this purpose the representation with angle γ and distance p can be used:

$$x\cos\gamma + y\sin\gamma - p = 0\,, \tag{2}$$

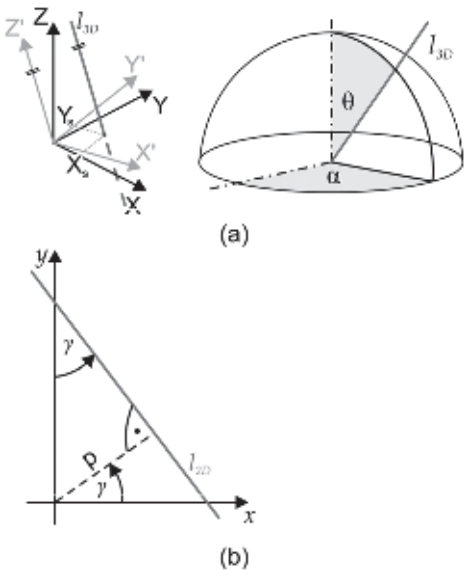

Fig. 1: Parametrization of a line – graphical representation a) in 3D space using 4 parameters; b) in 2D space using 2 parameters.

p denotes the shortest distance from the line to the origin of the coordinate system, and γ denotes direction angle of the normal vector to the line (Fig. 1b).

3.2 *Assignment of Corresponding Lines*

Assignment of corresponding lines is carried out in the 2D image space. The model lines are projected into the image using coarse exterior orientation parameters obtained by direct georeferencing, and for each model line potentially corresponding image lines are found. In this work we apply an assignment based on relative position and orientation. We calculate a buffer S_i around every projected visible model line segment. The width $2{\cdot}\Delta p$ of S_i is given by $\Delta p = 3{\cdot}\sigma_p$, where σ_p is the uncertainty of the parameter p (distance) of the projected model line, calculated by propagating the uncertainty of the camera position and the uncertainty of the model. For all image line segments within the buffer we calculate the angle difference $\Delta\gamma_{ij}$ (Fig. 2). All model line segments for which $\Delta\gamma_{ij}$ is smaller than a threshold $\Delta\gamma_{max}$ are accepted as correspondences. $\Delta\gamma_{max}$ is calculated as $\Delta\gamma_{max} = 3{\cdot}\sigma_\gamma$, where σ_γ is the uncertainty of the parameter (angle) γ of the projected model line, which is calculated by propagating the uncertainty of the camera position and the uncertainty of the model. Formally we can write these conditions as:

$$l_j \rightarrow l_{mi} \quad \text{if } \{e_{1j}, e_{2j}\} \subseteq S_i \wedge |\Delta\gamma_{ij}| \leq \Delta\gamma_{\max} \quad (3)$$

$$\Delta\gamma_{ij} = |\gamma_i - \gamma_j|, \quad (4)$$

where l_{mi} denotes the i^{th} model line, S_i is the search space for the i^{th} model line (buffer around l_{mi}), l_j denotes the j^{th} image line, e_{1j} and e_{2j} are the end points of the j^{th} image line, γ_i is the γ-parameter of the representation given by (2) for the projection of the i^{th} model line and γ_j is the γ-parameter for the j^{th} image line.

This search for correspondences is applicable in our case because we assume to know the exterior orientation of the camera from the GPS/INS path precisely enough for the projected model lines to be only a few pixels away from their corresponding image features. To ensure a reliable assignment a visibility checking algorithm is also required, so that hidden edges are not projected into the image and no correspondences for them are assigned.

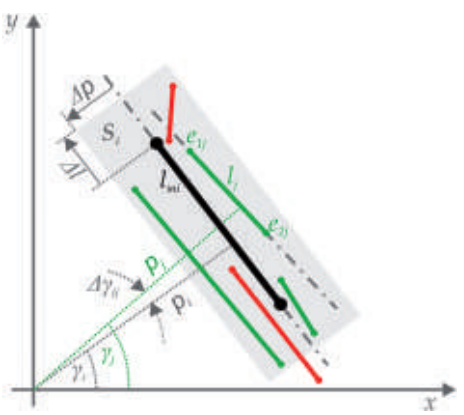

Fig. 2: Assignment of correspondences. l_{mi} is the i^{th} model line. For every l_{mi} (in black) a search space S_i (the buffer around l_{mi}) is defined. If both end points e_{1j} and e_{2j} of image line are within S_i and the angle difference $\Delta\gamma_{ij}$ between l_{mi} and l_j is smaller then a threshold, then the image line l_j is accepted as a correspondence for l_{mi} (in green). Otherwise l_j is rejected as a correspondence for l_{mi} (in red).

3.3 *Line Based Least Squares Adjustment*

The correspondence between 3D coordinates and their perspective projection into the image is given by the collinearity equations. The collinearity equations can be combined with (1) and the parameters *m* and *d* from the line representation

$$l: y = mx + d \tag{5}$$

can be calculated. Detailed equations needed to express *m* and *d* in terms of camera position parameters are given by MEIERHOLD et al. (2008). The authors also mention the problem of vertical image lines, which cannot be expressed using (4) and propose to change the line representation to

$$l: x = m'y + d', \text{ where } m' = \frac{1}{m},\ d' = \frac{d}{m}. \tag{6}$$

The problem during the adjustment is that some lines can change from being non-vertical to being vertical in the iteration process, so that the Jacobian matrix would have to be re-designed. To avoid this problem we use (2), express γ and p in terms of camera orientation parameters and use them as observations.

$$\gamma = \arctan\left(-\frac{1}{m}\right), \quad p = \frac{d}{\sqrt{m^2+1}} \tag{7}$$

For the adjustment we use the least squares method with the model:

$$\gamma_j + \hat{v}_{\gamma j} = f(\hat{X}_0, \hat{Y}_0, \hat{Z}_0, \hat{\omega}, \hat{\phi}, \hat{\kappa}, \alpha_i, \theta_i, X_{Si}, Y_{Si}), \tag{8}$$

$$p_j + \hat{v}_{pj} = f(\hat{X}_0, \hat{Y}_0, \hat{Z}_0, \hat{\omega}, \hat{\phi}, \hat{\kappa}, \alpha_i, \theta_i, X_{Si}, Y_{Si}, c, x_0, y_0). \tag{9}$$

In (8) and (9) the hat "^" is used for unknowns to be estimated. **f** denotes the functional model derived from the modified collinearity equations (7) based on the exterior orientation of the camera (X_0, Y_0, Z_0, ω, φ, κ), the interior orientation parameters (c, x_0, y_0) and the parameters (α_i, θ_i, X_{Si}, Y_{Si}) of the 3D lines. The interior orientation parameters and the parameters of the 3D lines are assumed to be constant. $\hat{v}_{\gamma j}$ and $\hat{v}_{pj}$ denote the corrections we seek to minimize in the least squares context:

$$\left(\sum \hat{v}_{\gamma j}{}^2 + \sum \hat{v}_{pj}{}^2\right) \rightarrow \min. \tag{10}$$

Writing the model in a vector form, where

$$\mathbf{b} = [\gamma_1, p_1, \gamma_2, p_2, ..., \gamma_n, p_n]^T, \tag{11}$$

$$\hat{\mathbf{v}} = [\hat{v}_{\gamma 1}, \hat{v}_{p1}, \hat{v}_{\gamma 2}, \hat{v}_{p2}, ..., \hat{v}_{\gamma n}, \hat{v}_{pn}]^T, \tag{12}$$

$$\hat{\mathbf{x}} = [\hat{X}_0, \hat{Y}_0, \hat{Z}_0, \hat{\omega}, \hat{\phi}, \hat{\kappa}] \tag{13}$$

we get:

$$\mathbf{b} + \hat{\mathbf{v}} = \mathbf{f}(\hat{\mathbf{x}}) = \mathbf{f}(\dot{\mathbf{x}} + \Delta\hat{\mathbf{x}}), \quad \hat{\mathbf{x}} = \dot{\mathbf{x}} + \Delta\hat{\mathbf{x}}. \tag{14}$$

Here **b** denotes the observation vector for n-correspondences, $\hat{\mathbf{x}}$ denotes the vector of estimated unknowns, and $\dot{\mathbf{x}}$ denotes the vector of approximated values for unknowns. Applying a first order approximation using a Taylor series the Jacobian matrix **A** is calculated as

$$\mathbf{A} = \frac{\partial \mathbf{f}(\dot{\mathbf{x}})}{\partial \dot{\mathbf{x}}}. \tag{15}$$

Then $\Delta\hat{\mathbf{x}}$ and $\hat{\mathbf{v}}$ are estimated using

$$\Delta\hat{\mathbf{x}} = \left(\mathbf{A}^T\mathbf{P}_{bb}\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{P}_{bb}\left(\mathbf{b} - \mathbf{f}(\dot{\mathbf{x}})\right), \tag{16}$$

$$\hat{\mathbf{v}} = \mathbf{A}\Delta\hat{\mathbf{x}} - \left(\mathbf{b} - \mathbf{f}(\dot{\mathbf{x}})\right). \tag{17}$$

where $\mathbf{P}_{bb}$ is the weighting matrix and

$$\mathbf{P}_{bb} = \mathbf{Q}_{bb}{}^{-1} = \sigma_0^2 \cdot \mathbf{C}_{bb}{}^{-1}. \tag{18}$$

$\mathbf{Q}_{bb}$ denotes the weight coefficient matrix, $\mathbf{C}_{bb}$ denotes the covariance matrix and σ_0^2 denotes the variance factor.

3.4 *Uncertainty of the Building Model*

The uncertainty of the building model is related to the inaccuracies of creation and generalization. Many building models are created from aerial imagery, where the roof corners and the height to the ground are measured. Hence often the roof overlap is not modelled, and the wall edges are less accurate (less reliable) than the roof edges. We assume differ-

ent accuracy values for the roof points and for the wall/ground points in the model, which are presented graphically as ellipses in Fig. 3. The *Z*-coordinate is assumed to be less accurate than *X*- and *Y*-coordinates. In Fig. 3 it can be observed that in the case of oblique airborne images not only the *X*- and *Y*-components, but also the *Z*-component of the uncertainty has a strong influence on the position error of the projected point.

The model uncertainty can be embedded into the model presented in section 3.3 extending the vector of unknowns with the model line parameters:

$$\hat{\mathbf{x}} = [\hat{X}_0, \hat{Y}_0, \hat{Z}_0, \hat{\omega}, \hat{\phi}, \hat{\kappa}, \hat{\alpha}_i, \hat{\theta}_i, \hat{X}_{Si}, \hat{Y}_{Si}] \,. \tag{19}$$

Also the functional model **f** (14) has to be extended with equations for the model line parameters as follows:

$$\begin{aligned} \alpha_i + \hat{v}_{\alpha i} &= \hat{\alpha}_i \\ \theta_i + \hat{v}_{\theta i} &= \hat{\theta}_i \\ X_{Si} + \hat{v}_{Xsi} &= \hat{X}_{Si} \\ Y_{Si} + \hat{v}_{Ysi} &= \hat{Y}_{Si} \end{aligned} \tag{20}$$

Accordingly the Jacobian matrix **A** is extended with *4n* observations and *4n* unknowns. Unfortunately the uncertainties of the parameters of the 3D line models (α_i, θ_i, X_{Si}, Y_{Si}) are not directly known. Usually the position accuracy of the 3D building models created from aerial imagery is given for building corners.

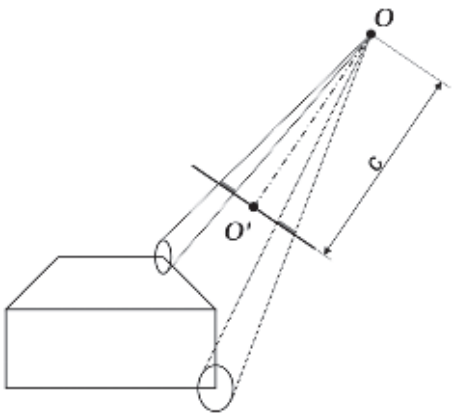

Fig. 3: Projection of model point uncertainty into the image in oblique geometry.

Therefore the uncertainty in the parameters of the 3D line models (α_i, θ_i, X_{Si}, Y_{Si}) has to be calculated using error propagation law:

$$\mathbf{C}_{yy} = \mathbf{F}\mathbf{C}_{xx}\mathbf{F}^T \,. \tag{21}$$

$\mathbf{C}_{xx}$ denotes the covariance matrix for 3D coordinates of the corners, **F** is the Jacobian for the function transforming the *XYZ* coordinates to line parameters α, θ, X_S, Y_S, and $\mathbf{C}_{yy}$ denotes the covariance matrix for the line parameters. The uncertainty of X_S and Y_S depends on the coordinate system. Therefore all calculations should be carried out in the local coordinate system.

3.5 *Reliability and Uncertainty of Image Lines*

Edge extraction is carried out using the Canny algorithm. The edges are detected with different extraction parameters. By varying the minimum edge strength required for a feature to be accepted as an edge, different results are achieved. Lowering this parameter results in multiple detections, because low-contrast edges are also included. However, these edges are less reliable as building edges. Setting a high minimum edge strength means "stronger" edges can be detected, but there may not be a sufficient amount to fit the model. Thus we propose edge detection with varying minimum edge strength. As a result we get three sets of detected edges. We then combine all three edge sets using the minimum strength to weight them.

Additionally, we assume that building edges correspond to long image lines. Short lines more likely correspond to other objects or to noise. Therefore we also use the length of the detected lines for weighting. The weights of the lines are calculated as follows:

$$g_j = \frac{1}{2}\left(\frac{l_j}{d_{\max}} + \frac{a_j}{255}\right) \tag{22}$$

where g_j denotes the weight for the j^{th} image line, l_j denotes the length of the j^{th} line, $a_j \in [0,255]$ denotes the threshold for the minimum edge strength used for the extraction of the j^{th} line and d_{max} is the length of the diagonal

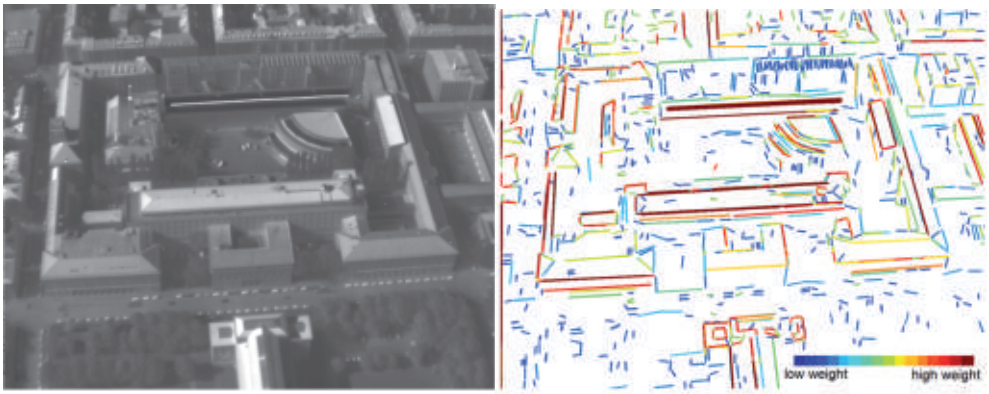

Fig. 4: An exemplary IR frame (a) and the weighting of the lines extracted in this frame (b). Three different minimum edge strengths were used for the extraction.

of the entire image. An exemplary weighting of the lines is presented in Fig 4.

We use weighting to identify the reliable edges and do not consider those with low weights for the assignment of correspondences.

For the image features the position uncertainty of the end points is given. Thus, we calculate the $\mathbf{P}_{bb}$ matrix for γ and p using error propagation as shown in (21).

4 Elimination of Incorrect Correspondences Using RANSAC

The method of assigning correspondences described in section 3.2 allows for the selection of multiple image lines corresponding to one model line. This leads to many incorrect correspondences, which have to be eliminated or reduced. For this purpose we apply the random sample consensus – RANSAC (Fischler & Bolles 1981). From the set of all hypothetical correspondences selected in the procedure from section 3.2, we randomly select three correspondences from different parts of the model and calculate exterior orientation parameters without redundancy. We then check how many of the remaining correspondences fit the randomly estimated exterior orientation parameters. This procedure is repeated k-times, and k is calculated as

$$k = \frac{\log(1-0.99)}{\log(1-(1-\varepsilon)^r)}, \tag{23}$$

where r is the number of necessary observations, $\varepsilon \in (0,1)$ is the outliers rate and the probability that RANSAC makes at least one error free selection is 99 %. We estimate ε as

$$\varepsilon = N - N_{Mod}, \tag{24}$$

where N is the number of hypothetical correspondences selected by the assignment algorithm and N_{Mod} is the number of model lines which have at least one assigned image line. The RANSAC algorithm results in new exterior orientation parameters and a set of correct correspondences. These data are taken as the input for the adjustment procedure described in section 3.3.

5 Data Description

For our experiments we used a test dataset captured in a densely built-up area in the centre of Munich, Germany. The thermal images were taken using the IR camera AIM 640 QLW FLIR with a frame rate of 25 images per second. The camera was mounted on a platform carried by a helicopter. The flying height was approximately 400 m above ground level. The camera was forward-looking with an oblique view of approximately 45°. The size

of the images is 640 × 512 pixels. According to the flying height, camera orientation and camera parameters, the ground resolution of the IR images varies from about 0.5 m in the foreground to about 1.4 m in the background of the oblique images.

For geo-referencing we use data acquired by an Applanix POS AV 510 GPS/INS system. To correct the INS drift a Kalman filter is applied and an extended bundle adjustment is carried out (Kolecki et al. 2010). The corrected exterior orientation parameters are used for the model projection.

The 3D building models were created semi-automatically using commercial software for 3D building reconstruction from nadir view aerial images and stored in the *CityGML* format. In this format the buildings are stored face-wise as sequences of 3D points. Accordingly, all the lines are stored twice (once in each of the two faces intersecting at that line), and points even occur three or more times. We therefore reorganize the model to reduce double lines. We collect all points (corners) and remove repetitions. Then we store lines and faces as references to the points.

6 Results

We carried out tests with the data described in section 5. An exemplary frame with the image lines (cyan, blue and magenta) and the projected model (yellow and green) is presented in Fig. 5a. The model lines which have correspondences are printed in green and the model lines without correspondences in yellow. The image lines without correspondences are presented in cyan; image lines which were assigned to model lines, but refused by RANSAC are depicted in blue, and image lines which were accepted by RANSAC and taken as input for the adjustment are depicted in magenta. Fig. 5b shows the same frame with the projected model before (red) and after (green) adjustment.

In Fig. 5b an improvement of the position of the projected model can be observed. Projected building structures match the IR image very well. Thanks to this, thermal building textures can be extracted precisely. Nevertheless, accuracy of the exterior orientation parameters estimation is not very high. For the exemplary frame from Fig. 5 the standard de-

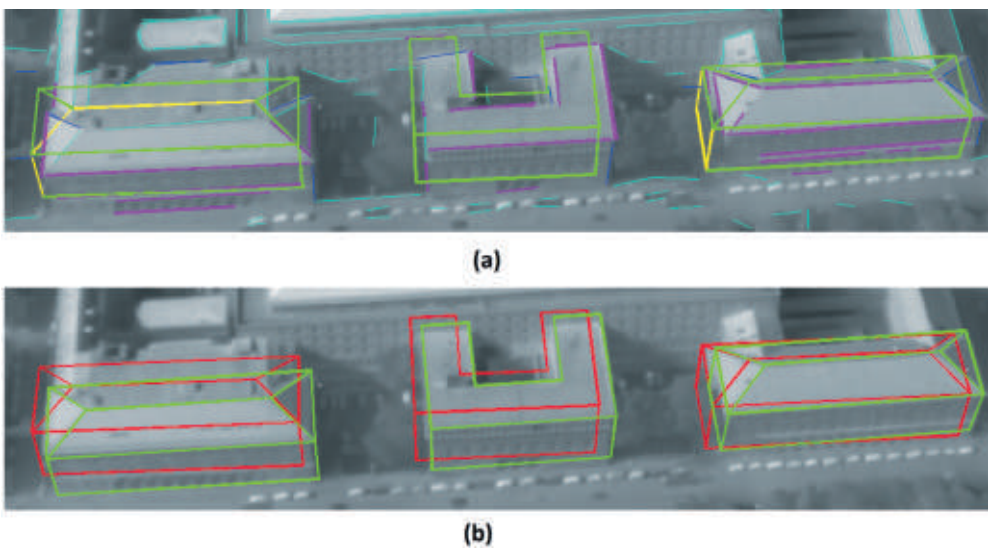

Fig. 5: Exemplary frame with extracted lines and a projected model: (a) before matching, where cyan = image lines without correspondences, blue = image lines with correspondences but refused by RANSAC, magenta = accepted image lines, green = model lines with correspondences, yellow = model lines without correspondences; and (b) after matching, where red = 3D building model projected with initial exterior orientation parameters, green = 3D building model projected with adjusted exterior orientation parameters.

viations are $\sigma_X = 3.8$ m, $\sigma_Y = 6.5$ m, $\sigma_Z = 4.5$ m, $\sigma_\omega = 0.60°$, $\sigma_\varphi = 0.69°$, and $\sigma_\kappa = 0.39°$. This is related to the low resolution of the images and the low accuracy of the extracted edges.

To evaluate the method and to investigate the sensitivity of the method with respect to changes in the initial exterior orientation we selected one frame and carried out a test in which we generated normally distributed random numbers with mean $\mu = 0$ and standard deviation $\sigma_{xyz} = 1$ m, $\sigma_{\omega\varphi\kappa} = 0.1°$ and used them to degrade the initial exterior orientation parameters. We successively repeated the test with increasing σ_{xyz} and $\sigma_{\omega\varphi\kappa}$. Results of this investigation are presented in the first row of Tab. 1. For comparison we conducted the same test without RANSAC (second row of Tab. 1).

Tab. 1: Percentage of successfully matched samples with downgraded initial exterior orientation. **σ** denotes the standard deviation used for the generation of normally distributed random numbers, with $\boldsymbol{\sigma} = \sigma_{xyz}, \sigma_{\omega\varphi\kappa}$, where $\sigma_{xyz} = 1$ m, and $\sigma_{\omega\varphi\kappa} = 0.1°$. First row: results using RANSAC; second row: results based on the assignment of correspondences without outlier detection.

Successfully matched samples when downgrading the exterior orientation with normally distributed numbers using mean μ = 0 standard deviation (%)					
	σ	3**σ**	4**σ**	5**σ**	7**σ**
(1)	96	68	65	61	46
(2)	98	43	20	16	0

Tab. 1 shows that our method works for well geo-referenced images. The search for correspondences proposed in section 3.2 delivers good results if the camera position and rotations are known with an accuracy of 1–2 m and 0.1–0.2°, respectively. As we mentioned in section 3.2, the size of the search space S_i and the angular threshold are calculated based on the expected displacement and rotation of the projected model lines, which are in turn calculated from the propagation of errors in the 3D building model and the exterior orientation parameters. If we downgrade the initial exterior orientation by 4σ, we get the width parameter of the buffer $\Delta p = 3 \cdot \sigma_p$ in range of about 40 pixels and the angular threshold $\Delta\gamma = 3 \cdot \sigma_\gamma$, which corresponds to about 8°. The values are so high, because σ_p and σ_γ are calculated as propagation of the uncertainty of the camera position and orientation and the uncertainty of the 3D building model. This leads to many incorrect correspondences with $\varepsilon \approx 85$ %. In case of downgrading with 7σ we get $\varepsilon \approx 93$ %. The typical least squares adjustment method cannot cope with such a large number of outliers. However, using RANSAC the algorithm is robust even if the camera is "shifted" from the initial position by few meters (e.g. Fig. 6). For 7σ we still get 46 % successfully matched samples.

7 Discussion and Future Work

Line based model-to-image matching has high potential for co-registration of buildings with oblique airborne images. Edges are the most representative features for building structures and can be easily detected in the image using standard image processing algorithms. Considering the uncertainty of image lines and of the building model, as proposed in this paper, a better fit between the building model and the image structures is achieved. However, esti-

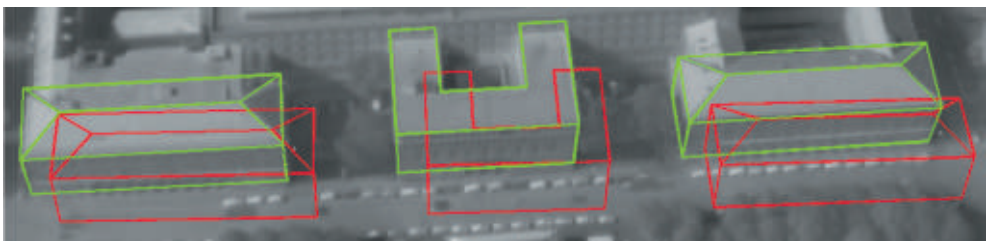

Fig. 6: Exemplary frame with the 3D building models projected with exterior orientation parameters downgraded by normally distributed numbers with μ and 7σ (red) and projected 3D building models after adjustment (green).

mated exterior orientation should not be understood as the true position of the camera, but only as parameters needed for a better initial projection in order to obtain a sufficient number of correct correspondences. In the future, neighbouring frames should also be employed in the adjustment to improve the relative orientation of the sequence. We expect that this will stabilize the model-to-image matching, in particular by reducing the "jumping" of the camera.

The uncertainty of the building models allows the calculation of the corrections for the model parameters. However, these corrections cannot be applied to improve the geometry of the model, because in the presented adjustment no constrains for planes are implemented. Therefore, some coplanar 3D building lines which originally belonged to one plane can be non-coplanar after adjustment. Nevertheless, the corrected position in the image can be used for texture extraction.

Applying the RANSAC algorithm we get good results not only for very well geo-referenced data but also for mid-quality geo-referencing. However, further improvements in the search for candidate correspondences are needed in case of low-quality input. Some already existing methods, e.g. vanishing points or relational matching, can be applied to find coarse exterior orientation parameters.

As an alternative or complement to RANSAC an outlier detector within the adjustment can be implemented using the cofactor matrix for the corrections Q_{vv}. However, only incorrect image lines can be efficiently detected. Errors in the model lines are difficult to recognize because their redundancy components are low, and therefore the influence of the errors on the correction for each parameter is very small.

In IR images depicting urban scenes it is often difficult to extract building edges on the ground. Often pavement appears similar to walls. Accordingly, in some frames a very good fit between the 3D building models and roof structure was achieved, while in the walls some remaining displacements occurred.

In the future more attention should be paid to the rolling shutter effect. This phenomenon should also be taken into consideration in the adjustment. Typically the IR cameras read out the measured radiation row-wise. If the camera is moving each row is read at a different point in time, and therefore at a different position, which should be accounted for in the geometrical model used for estimation.

Acknowledgements

The authors would like to thank FGAN-FOM, Ettlingen, for providing images of the flight campaign.

References

Avbelj, J., Iwaszczuk, D. & Stilla, U., 2010: Matching of 3D wire-frame building models with image features from infrared video sequences taken by helicopters. – International Archives of Photogrammetry, Remote Sensing and Spatial Geoinformation Sciences **38** (3B): 149–154.

Debevec, P.E., Taylor, C.J. & Malik, J., 1996: Modelling and rendering architecture from photographs: a hybrid geometry- and image-based approach. – **23rd** Annual Conference on Computer Graphics and Interactive Techniques: 11–20.

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).

Eugster, H. & Nebiker, S., 2009: Real-time Georegistration of Video Streams from Mini or Micro UAS using Digital 3d City Models. – **6th** International Symposium on Mobile Mapping Technology, Presidente Prudente, Sao Paulo, Brazil.

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

Förstner, W., 2010: Optimal vanishing point detection and rotation estimation of single image from a legoland scene. – International Archives of Photogrammetry, Remote Sensing and Spatial Geoinformation Sciences **38** (3A): 157–162.

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

Hoegner, L., Kumke, H., Meng, L. & Stilla, U., 2007: Automatic extraction of textures from in-

frared image sequences and database integration for 3D building models. – PFG **2007** (6): 459–468.

Hsu, S., Samarasekera, S., Kumar, R. & Sawhney, H.S., 2000: Pose Estimation, Model Refinement, and Enhanced Visualization Using Video. – Proceedings of CVPR **00 I:** 488–495.

Kolecki, J., Iwaszczuk, D. & Stilla, U., 2010: Calibration of an IR camera system for automatic texturing of 3D building models by direct georeferenced images. – EuroCOW 2010: on CD.

Lowe, D.G., 1991: Fitting Parameterized Three-Dimensional Models to Images. – IEEE Transactions on Pattern Analysis and Machine Intelligence **13** (5): 441–450.

Meierhold, N., Bienert, A. & Schmich, A., 2008: Linebased referencing between images and laser scanner data for image-based point cloud interpretation in a CAD environment. – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences **37** (B5): 437–444.

Roberts, K.S., 1988: A new representation for lines. – IEEE Proceedings of Computer Vision and Pattern Recognition: 635–640.

Schenk, T., 2004: From point-based to feature-based aerial triangulation. – ISPRS Journal of Photogrammetry and Remote Sensing **58:** 315–329.

Sester, M. & Förstner, W., 1989: Object Location Based on Uncertain Models. – Burkhardt, H., Höhne, K.H. & Neumann, B. (eds.): Mustererkennung **1989** (11): 457–464, DAGM-Symposium, Springer-Verlag London, UK, ISBN 3-540-51748-0.

Tian, Y., Gerke, M., Vosselman, G. & Zhu, Q., 2008: Automatic edge matching across an image sequence based on reliable points. – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences **37** (B3b): 657–662.

Vosselman, G., 1992: Relational matching. – Lecture Notes in Computer Science, Springer.

Vosselman, G., 1998: Interactive Alignment of Parameterised Object Models to Images. – International Archives of Photogrammetry and Remote Sensing **32** (3/1): 272–278.

Zhang, Y., Zhang, Z., Zhang, J. & Wu, J., 2005: 3D Building Modelling with Digital Map, LIDAR Data and Video Image Sequences. – Photogrammetric Record **20** (111): 285–302.

Address of the Authors:

Dorota Iwaszczuk, M.Sc., Dipl.-Inf. Ludwig Hoegner, Dipl.-Ing. Michael Schmitt & Univ.-Prof. Dr.-Ing. Uwe Stilla, Technische Universität München, Photogrammetry and Remote Sensing, D-80333 München, Tel.: +49-89-289-22637,-22680,-22672,-22671, Fax: +49-89-289-23202, e-mail: {dorota.iwaszczuk}{ludwig.hoegner}{michael.schmitt}{stilla}@bv.tum.de

Manuskript eingereicht: März 2012
Angenommen: Juni 2012