

ALIGNMENT OF 3D BUILDING MODELS AND TIR VIDEO SEQUENCES WITH LINE TRACKING

D. Iwaszczuk^{a,b,*}, U. Silla^a

^a Technische Universität München, Photogrammetry & Remote Sensing, Arcisstr. 21, 80333 München, Germany

^b Technische Universität München, Institute of Astronomical and Physical Geodesy, Arcisstr. 21, 80333 München, Germany

Commission I, WG I/2

KEY WORDS: Registration, Aerial, Thermal, Infrared, Image, Sequences, Building, Model

ABSTRACT:

Thermal infrared imagery of urban areas became interesting for urban climate investigations and thermal building inspections. Using a flying platform such as UAV or a helicopter for the acquisition and combining the thermal data with the 3D building models via texturing delivers a valuable groundwork for large-area building inspections. However, such thermal textures are useful for further analysis if they are geometrically correctly extracted. This can be achieved with a good coregistrations between the 3D building models and thermal images, which cannot be achieved by direct georeferencing. Hence, this paper presents methodology for alignment of 3D building models and oblique TIR image sequences taken from a flying platform. In a single image line correspondences between model edges and image line segments are found using accumulator approach and based on these correspondences an optimal camera pose is calculated to ensure the best match between the projected model and the image structures. Among the sequence the linear features are tracked based on visibility prediction. The results of the proposed methodology are presented using a TIR image sequence taken from helicopter in a densely built-up urban area. The novelty of this work is given by employing the uncertainty of the 3D building models and by innovative tracking strategy based on a priori knowledge from the 3D building model and the visibility checking.

1. INTRODUCTION

Coregistration of multiple data is one of the main tasks in photogrammetry and remote sensing and often is an important step of data fusion needed for various applications. Depending on the application, many methods and strategies have been proposed in literature. In this paper, we present alignment of 3D building models and thermal infrared (TIR) image sequences taken from a flying platform.

1.1 Motivation

The interest in thermal images of urban scenes at various scales has been of increasing interest in the recent years. The applications range from urban climate observation and heat island detection in large scale satellite images (Weng, 2009), through urban districts inspections using airborne TIR imagery (Iwaszczuk et al., 2012), to building inspection using street view TIR images (Chandler, 2011).

Combining TIR images with three-dimensional (3D) geometries allows for the spatial reference of the thermal data and facilitates their interpretation. Thermal data can be combined with different kinds of 3D geometries: with Building Information Models (BIM) (Mikeleit and Kraneis, 2010), with 3D building models (Hoegner et al., 2007) and (Iwaszczuk et al., 2011), with 3D point clouds (Cabrelles et al., 2009), (Borrmann et al., 2012) and (Vidas et al., 2013) or with a point cloud and aerial photographs at the same time (Boyd, 2013). Combination with point clouds is carried out by assignment and interpolation of the measured temperature to the points. Using point cloud as spacial reference enables fast generation of results with a high level of detail and is appropriate for visual interpretation. 3D building models deliver more generalized and structured representation to support

automatic analysis. The TIR images can be fused with 3D building models by texturing, i.e. assignment of image sections to the polygons of the 3D building models. Such textures map thermal radiation of one building element (for example façade) on a georeferenced 3D polygon. Hence, they allow for geolocation of the outputs of further analysis in the 3D object space, for example for embedding of detected heat leakages in the 3D building model. Using a semantic 3D building model, also combination with other data and spatial queries are possible (Kaden and Kolbe, 2013).

To make such use of the thermal textures, large scale coverage and precise texture extraction are needed. Airborne oblique TIR videos taken from a flying platform such as UAV or helicopter allow for capturing complete building envelope including roofs and façades in inner yards and enable fast acquisition of entire cities. Precise texture extraction can be achieved based on a well coregistered data. The coregistration is made by 3D to 2D projection of the 3D building model using known exterior orientation parameters. Approximated exterior orientation parameters are usually derived from the navigation device with an accuracy of up to several meters for the position and up to few degrees for orientation, depending on the accuracy class of the device. However, these initial exterior orientation parameters are not sufficient for precise texture extraction in most cases, even using a high accuracy navigation. In addition, the mismatch between the projected 3D model and image structures can be caused by some other errors, such as error in boresight and lever-arm calibration or errors in the 3D model related to the creation technique and generalization. Therefore, the coregistration should be refined in a matching process which takes not only the uncertainty in the exterior orientation parameters, but also in the 3D model and in image features. Moreover, the advantage of the high frequency rate in the image sequences should be taken.

*Corresponding author, dorota.iwaszczuk@tum.de

1.2 Related work

The model-to-image matching problem for airborne imagery is frequently addressed in literature and many methods for solving the problem have been presented, also in the texturing context. In (Früh et al., 2004) line matching based on slope and proximity by testing different random camera positions is proposed. In this method the relation between the frames is not used. In (Hsu et al., 2000) and (Sawhney et al., 2002) existing 3D models are textured using a video sequence taking the small change of the perspective from frame to frame. They assume the camera pose to be known in the first frame of the sequence and predict the pose in the next frame. The correspondence between the frames is estimated using optical flow. Then they search for the best camera position by minimizing the disagreement between projected edges and edges detected in the image by varying the camera pose and try to maximize the integral of this field along the projected 3D line segment in the steepest descent method. This methodology employs the properties of image sequences, but it uses random SIFT for tracking and not features related to the buildings. Besides, it does not take the uncertainty of the 3D building models into account.

In some other researches, uncertainty is considered in different context. (Sester and Förstner, 1989) and (Schickler, 1992) introduce uncertainty in three model parameters (width, length and slope) for a simple case of roof sketches and integrate them in the adjustment, together with the uncertainties in 2 parameters of 2D lines detected in the image. (Luxen and Förstner, 2001) present a method for optimal estimate for the projection matrix with the covariance matrix for its entries using point and line correspondences. Using homogeneous coordinates, they represent 3D lines as join of two 3D points and the projection of these lines as projection planes. In doing so, the entries of the projection matrix for points of size 3x4 has to be calculated, avoiding calculation of the projection matrix for lines of size 3x6. In the adjustment model they introduce the uncertainty of the 2D points and lines.

(Heuel and Förstner, 2001) and (Heuel, 2002) also use homogeneous representation of geometric uncertain entities to match line segments, to optimally reconstruct 3D lines and finally group them. (Heuel, 2002) gives a very detailed and structured overview of representation of uncertain entities in 2D and 3D, such as points, lines and planes and geometric reasoning with them. He also presents the constructions using uncertain entities and appropriate error propagation. (Beder, 2004) and (Beder, 2007) uses the same representation for grouping of points and lines by statistical testing for incidence.

(Meidow et al., 2009a) and (Meidow et al., 2009b) collect, evaluate, discuss and extend various representations for uncertain geometric entities in 2D. Additionally, they provide a generic estimation procedure for multiple uncertain geometric entities with Gauss-Helmert model. They handle uncertain homogeneous vectors and their possibly singular covariance matrices by introducing constrains for the observations in addition to the conditions for the observations and parameters, and restrictions for the parameters.

1.3 Overview

In this paper we present a methodology for the alignment of 3D building models and oblique TIR image sequences taken from a flying platform. In Section 2., the methodology for line based coregistration in a single image is presented. Then, a concept for tracking linear features based on visibility prediction is presented in Section 3. In Section 4. experiments carried out with TIR image sequences taken from helicopter in a densely built up urban area are presented.

The novelty of this work is given by employing the uncertainty of the 3D building models in a line based matching and adjustment and by innovative tracking strategy based on a priori knowledge from the 3D building model and the visibility checking.

2. LINE-BASED COREGISTRATION OF 3D BUILDING MODELS AND TIR IMAGES

2.1 Finding correspondences

In the presented approach, line segments are selected to be used for matching. The corresponding image line segments are assigned to the model edges using an accumulator approach. The model edges are projected into the image using the initial exterior orientation taken from the navigation device and moved and rotated in the image creating a 3D accumulator space. For each position of the projected model in the image, the number of fitting line segments in the image is counted. Correspondences which voted for the accumulator cell with most line correspondences are used for the optimal pose estimation.

2.2 Optimal camera pose estimation

Estimation of the exterior orientation parameters in the projective space is formulated using the coplanarity of l_j , X_{1i} and X_{2i} , where X_{1i} and X_{2i} are the endpoints of a building edge corresponding to line segment l_j detected in the image (Fig. 1). In this Section, the index i refers to the edges of the 3D building model, and the index j to the line segments extracted in the image. Coplanarity of l_j , X_{1i} and X_{2i} is expressed as incidence of the projected points χ'_{1i} and χ'_{2i} with the line l_j . The projected points $\chi'_{1i} : \mathbf{x}'_{1i} = \mathbf{P}\mathbf{X}_{1i}$ and $\chi'_{2i} : \mathbf{x}'_{2i} = \mathbf{P}\mathbf{X}_{2i}$, where \mathbf{P} is the projection matrix. Then, the incidence conditions $\mathbf{l}_j^T \mathbf{x}'_{1i} = 0$ and $\mathbf{l}_j^T \mathbf{x}'_{2i} = 0$ write

$$\mathbf{l}_j^T \mathbf{P}\mathbf{X}_{1i} = 0, \quad (1)$$

$$\mathbf{l}_j^T \mathbf{P}\mathbf{X}_{2i} = 0. \quad (2)$$

These two equations are directly adapted in the Gauss-Helmert model as conditions

$$\mathbf{g}_1(\hat{\beta}, \hat{\mathbf{y}}) = \mathbf{l}_j^T \mathbf{P}\mathbf{X}_{1i}, \quad (3)$$

$$\mathbf{g}_2(\hat{\beta}, \hat{\mathbf{y}}) = \mathbf{l}_j^T \mathbf{P}\mathbf{X}_{2i} \quad (4)$$

for the observations and parameters.

Also in the projective space the uncertainty of the image features and 3D building model can be introduced. The covariance matrix for a 3D point X represented in homogeneous coordinates \mathbf{X} can be directly derived from the covariance matrix Σ_{XX} for the Euclidean representation \mathbf{X} of this point as

$$\Sigma_{\mathbf{X}\mathbf{X}} = \begin{bmatrix} \Sigma_{XX} & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix}. \quad (5)$$

However, due to redundancy in the homogeneous representation, the covariance matrix $\Sigma_{\mathbf{X}\mathbf{X}}$ is singular (Förstner, 2004) which leads to restrictions in the optimization. To solve this problem, all entities have to be spherically normalized¹ (Kanatani, 1996), so that $\mathbf{l}_j^s = N_s(\mathbf{l}_j)$, $\mathbf{X}_{1i}^s = N_s(\mathbf{X}_{1i})$ and $\mathbf{Y}_{1i}^s = N_s(\mathbf{Y}_{1i})$. In the rest of this Section, the index ^s is omitted assuming the homogeneous

¹The covariance matrix $\Sigma_{\mathbf{ll}}$ calculated as shown in eq. 10 is not singular but \mathbf{l} is also spherically normalized in order to avoid ambiguity of the homogeneous representation.

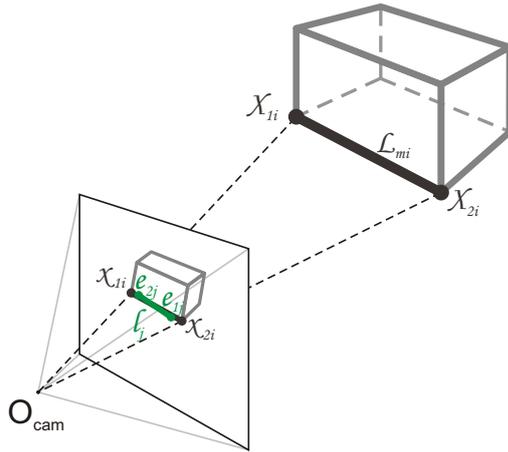


Figure 1: Assignment of 3D model edges and 2D line segments detected in the image

coordinates to be spherically normalized. This normalization has to hold also during the estimation. Therefore, also the constrains

$$\mathbf{c}_1(\hat{\mathbf{y}}) = \|\mathbf{l}_j\| - 1 \quad (6)$$

$$\mathbf{c}_2(\hat{\mathbf{y}}) = \|\mathbf{X}_{1i}\| - 1 \quad (7)$$

$$\mathbf{c}_3(\hat{\mathbf{y}}) = \|\mathbf{X}_{2i}\| - 1 \quad (8)$$

for the observations are needed.

To find optimal solution for the unknown parameters $\hat{\beta} = [\hat{X}_0, \hat{Y}_0, \hat{Z}_0, \hat{\omega}, \hat{\phi}, \hat{\kappa}]$, the optimization method for homogeneous entities presented in (Meidow et al., 2009a) and (Meidow et al., 2009b) is adapted for this functional model. For this purpose, the Lagrange function

$$\mathcal{L} = \frac{1}{2} \hat{\mathbf{v}}^T \Sigma_{\mathbf{y}\mathbf{y}}^{-1} \hat{\mathbf{v}} + \lambda_1^T \mathbf{g}_1(\hat{\beta}, \mathbf{y} + \hat{\mathbf{v}}) + \lambda_2^T \mathbf{g}_2(\hat{\beta}, \mathbf{y} + \hat{\mathbf{v}}) + \nu_1^T \mathbf{c}_1(\mathbf{y} + \hat{\mathbf{v}}) + \nu_2^T \mathbf{c}_2(\mathbf{y} + \hat{\mathbf{v}}) + \nu_3^T \mathbf{c}_3(\mathbf{y} + \hat{\mathbf{v}}) \quad (9)$$

is minimized, where λ and ν are the Lagrangian vectors. In contrast to (Meidow et al., 2009a) and (Meidow et al., 2009b), here the restriction for the estimated parameters $\mathbf{h}_1(\hat{\beta}) = 0$ is not needed, because the estimated parameters are defined directly as exterior orientation parameters $X_0, Y_0, Z_0, \omega, \phi, \kappa$. The observation vector for each pair of corresponding lines writes $\mathbf{y}_{ij} = [\mathbf{l}_j, \mathbf{X}_{1i}, \mathbf{X}_{2i}]^T$, where $\mathbf{l} = [a, b, c]^T$ is the homogeneous representation for the image line segment and $\mathbf{X}_{1i}, \mathbf{X}_{2i}$ is the homogeneous representation of the corners of the corresponding 3D building edge. The covariance matrix $\Sigma_{\mathbf{ll}}$ is assumed to be known as result of the line fitting or as result of error propagation knowing the covariance matrices of the end points of the detected line segment. This is done using

$$\Sigma_{\mathbf{l}_j \mathbf{l}_j} = \mathbf{S}(\mathbf{x}_{2j}) \Sigma_{\mathbf{x}_{1j} \mathbf{x}_{1j}} \mathbf{S}^T(\mathbf{x}_{2j}) + \mathbf{S}(\mathbf{x}_{1j}) \Sigma_{\mathbf{x}_{2j} \mathbf{x}_{2j}} \mathbf{S}^T(\mathbf{x}_{1j}), \quad (10)$$

where \mathbf{S} is the skew-symmetric matrix

$$\mathbf{S}(\mathbf{x}) = \begin{bmatrix} 0 & -w & v \\ w & 0 & -u \\ -v & u & 0 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} u \\ v \\ w \end{bmatrix}. \quad (11)$$

Switching from the Euclidean to the homogeneous representation for point χ in 2D or \mathcal{X} in 3D is usually effected by adding 1 as an additional coordinate (homogeneous part). Hence, for a 2D point in Euclidean space $\mathbf{x} = [u, v]^T$, the equivalent homo-

geneous representation is $\mathbf{x} = [u, v, 1]^T$, and for a 3D point in Euclidean space $\mathbf{X} = [U, V, W]^T$, the equivalent homogeneous representation is $\mathbf{X} = [U, V, W, 1]^T$. In many photogrammetric applications, particularly in aerial photogrammetry, the points are given in geodetic coordinate systems (for example, Gauss-Krüger, UTM), where the values for U and V is in order 10^6 . Computations with such inconsistent number can lead to numerical instability of the computations. To solve this problem, the homogeneous entities should be conditioned. Similar to the conditioning proposed by (Heuel, 2002), also here, the entities are conditioned before optimizing, by checking the condition

$$\max_{h_o} = \max_i \left(\frac{\|x_{hi}\|}{\|x_{oi}\|} \right) \geq f_{min} = 0.1, \quad (12)$$

where x_{hi} is the homogeneous and x_{oi} the Euclidean part of a homogeneous entity \mathbf{x}_i for point representation

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix} = \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} x_o \\ x_h \end{bmatrix}. \quad (13)$$

If $\max_{h_o} < f_{min}$ the conditioning factor is calculated as

$$f = \max_{h_o} \cdot f_{min} \quad (14)$$

In case of very large Euclidean part x_o compared to the homogeneous part x_h , f calculated as shown in eq. 14 can be smaller than the machine accuracy ϵ_h . Hence, if $f < \epsilon_h$ then f should be calculated as

$$f = \frac{f_{min}}{\max_i(\|x_{oi}\|)} \quad (15)$$

(Heuel, 2002). Next, each entity is conditioned using matrices

$$\mathbf{W}_x(f) = \begin{bmatrix} f \mathbf{l}_2 & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix} \quad (16)$$

for the 2D points,

$$\mathbf{W}_1(f) = \begin{bmatrix} \mathbf{l}_2 & \mathbf{0} \\ \mathbf{0}^T & f \end{bmatrix} \quad (17)$$

for the 2D lines and

$$\mathbf{W}_x(f) = \begin{bmatrix} f \mathbf{l}_3 & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix}, \quad (18)$$

so that the conditioned coordinates $\mathbf{x}^c, \mathbf{l}^c$ and \mathbf{X}^c are calculated as

$$\mathbf{x}^c = \mathbf{W}_x(f_{im}) \mathbf{x}, \quad (19)$$

$$\mathbf{l}^c = \mathbf{W}_1(f_{im}) \mathbf{l} \quad (20)$$

and

$$\mathbf{X}^c = \mathbf{W}_x(f_{mod}) \mathbf{X}, \quad (21)$$

where f_{im} is the conditioning factor for the 2D image entities and f_{mod} is the conditioning factor for the 3D entities. Conditioning entities changes also the transformation matrix. Here the transformation matrix is the projection matrix \mathbf{P} which can be reconditioned using

$$\mathbf{P} = \mathbf{W}(f_{im})^{-1} \mathbf{P}^c \mathbf{W}(f_{mod}). \quad (22)$$

3. LINE TRACKING

Most digital cameras, also the cameras operating in TIR domain, are able to capture image sequence with a relatively high frame rate. The frame rate 20-25 frames per second is available in low

and mid cost TIR cameras. Such frequency range enables acquisition with a very large overlap between the images. Accordingly, the position shift in the image space from frame to frame is in range of few pixels for most objects. Besides, the viewing angle does not change between the frames significantly. Hence, if the correct match is found in one frame, it is relatively easy to find correspondences in the next frame and calculate camera pose for this frame.

Accordingly, in an image sequence with a very large overlap between the frames the whole process of model-to-image matching does not have to be carried out for all frames. In order to reduce computational effort, so called *key-frames* are introduced in this thesis (Section 3.1) and selected lines are tracked from frame to frame (Section 3.2).

3.1 Key-Frame Concept

The main goals of the key-frame concept is to reduce computational effort on the one hand and to ensure the reliability of the calculated camera pose for each frame on the other hand. A key-frame is a frame in which the image-to-model matching and pose estimation are carried out as described in Section 2.. In a key-frame the correspondences are selected independently of the previous frame. In general, the key-frames can be:

- *pre-defined* or
- *dynamically selected* during the process.

In order to initiate the process, the first frame f_i , $i = 1$ is always a key-frame. In case of pre-defined key-frames, they appear in certain intervals. The interval size should be adjusted to the overlap between the frames. For image sequence with a very high overlap the interval can be higher as for frames with smaller overlap. If the overlap is not constant and not enough reliable correspondences with the model edges can be found, a dynamic selection of key-frames is applied.

Dynamic selection of key-frames is based on the current status of the reliability of matching and tracking. This reliability is due to two main conditions:

- *sufficient overlap* between the frames f_i and f_{i-1} ,
- *sufficient reliability* of the assignment in f_{i-1} .

In a video sequence, the sufficient overlap between frames f_i and f_{i-1} is given in most cases. Sometimes, however, when for example the camera is switched off for some time, the overlap can be too small to reliably track line segments from frame to frame. The reliability of the assignments depends on the number of selected correspondences and how much we believe that this assignment is correct. While the number of correspondences is very easy to measure, the belief is more difficult to express.

3.2 Predicted Tracking of Line Segments

Due to very small movements of the camera between the frames, line segments can be assumed to be only shifted by few pixels in the next frame. They can be tracked therefore using cross-correlation. Cross-correlation method is suitable for tracking in such application due to nearly invariant scale and viewing angle between two neighboring frames. Accordingly, the appearance of the tracked line segment and its surrounding will stay almost unchanged.

During the tracking some projected model edges cannot be visible all the time in the sequence. The information which model edge can be seen in a particular frame is derived from the model and approximated camera position. The fact that the model edge is seen or not can be considered as state of a particular model edge in each frame. For each model edge, following states are possible: alive/sound (fully visible), alive/injured (partially occluded), occluded and dead (out of field of view). Each model edge can change its state in an event. Such event may occur for each model edge among the image sequence. Tab. 1 presents these events including the change of the state caused by each event.

Table 1: Possible events and states for tracked lines (alive/sound - fully visible edge, alive/injured - partially occluded edge)

Event	Possible states before the event	Possible states after the event
Birth	\emptyset	Alive/sound Alive/injured
Appearing	Occluded	Alive/sound Alive/injured
Injury	Alive/sound	Alive/injured
Healing	Alive/injured	Alive/sound
Disappearing	Alive/sound Alive - injured	Occluded
Death	Alive/sound Alive/injured Occluded	Dead

First event which occur for a model edge is birth. It is the moment, when the model edge is visible in the image for the first time. After birth, the model edge can have one of two states alive/injured or alive/sound. Alive/injured means that the edge appears only partially in the frame or is partially occluded. This will be the most common case directly after the birth of the edge, because it happens very rarely that an entire edge appears at once (no edge in frame f_{i-1} , entire edge in frame f_i). Such case would directly result in alive/sound, which means fully visible edge. An alive/injured edge can become alive/sound in the healing event. Vice versa, alive/sound edge can become alive/ injured, it gets partially occluded by an object or part of this edge is not seen anymore in the current frame. Such an event is called injury. If the edge gets completely occluded by an object such event is called disappearing and the state after this event is occluded. Disappearing can occur for alive/sound or alive/injured edges. The opposite of the disappearing event is appearing. It happens when an occluded edge becomes alive/sound or alive/injured. The last possible event is death of the edge. It happens if the whole edge is not anymore seen in the current frame, it means it is out of the field of view. Death can happen to an alive/sound, alive/injured or occluded edge.

By defining the states of the model, it is determined for which model edges, a corresponding image line segments should be searched. Correspondences can be found only for alive edges. Injury is the only state which can be expressed with level of injury, it means how much of the edge is occluded. Highly injured edges are skipped when searching for correspondences.

4. EXPERIMENTAL RESULTS

4.1 Data description

For our experiments we used a test dataset captured in a densely built city area The thermal images were taken with IR camera

AIM 640 QLW FLIR with a frame rate 25 images per second, which was mounted on a platform carried by 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 chip is 640 x 512 pixels. The 3D building model was created semi-automatically using commercial software for 3D building reconstruction from aerial images.

4.2 Results

Exemplary results of the alignment are presented in Fig. 2. In order to reduce computational effort, first lines have been pre-selected using a buffer around each projected model edge (Fig. 2a). Then, the preliminary correspondences are reduced using the outlier detector - the accumulator approach (Fig. 2b). Finally, the selected correspondences were used for optimal pose estimation. The model projection using estimated exterior orientation parameters is presented in Fig. 3.

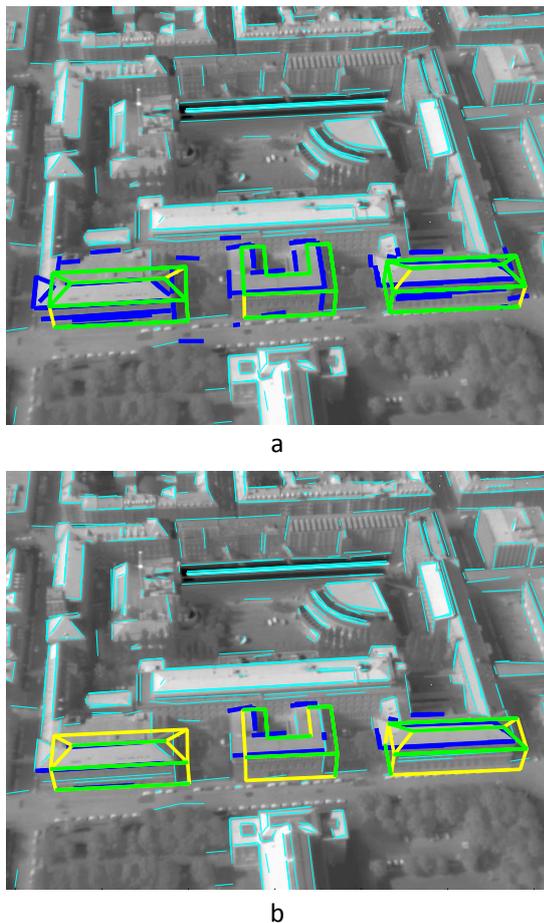


Figure 2: Exemplary result on model-to-image matching: a) pre-selected correspondences using a buffer around each projected model, b) correspondences selected using the outlier detector. Color coding: green - model edges with found correspondences, yellow - model edges without correspondences, blue - image line segments with correspondences, cyan - image line segments without correspondences

To evaluate the method and to investigate the sensitivity of the method with respect to changes in the initial exterior orientation parameters, one frame was selected. Using normally distributed random numbers with mean $\mu = 0$ and standard deviation $\sigma_{XYZ} = 1$ m, $\sigma_{\omega\phi\kappa} = 0.1^\circ$, the initial exterior orientation parameters were degraded. For every randomly degraded

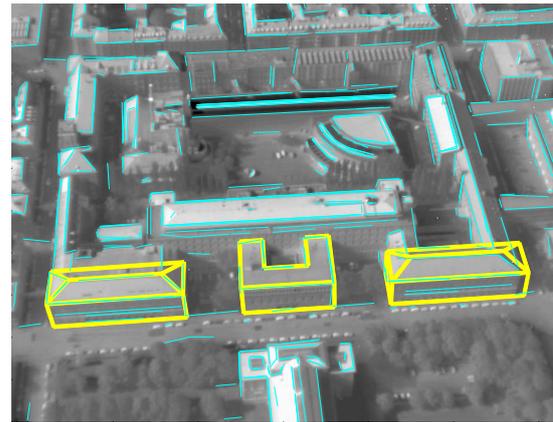


Figure 3: 3D building model (yellow) projected into the image with estimated exterior orientation parameters

Table 2: 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 $\sigma = [\sigma_{XYZ}, \sigma_{\omega\phi\kappa}]^T$, where $\sigma_{XYZ} = 1$ m, and $\sigma_{\omega\phi\kappa} = 0.1^\circ$

Successfully matched samples when downgrading the exterior orientation with normally distributed numbers using mean $\mu = 0$ and standard deviation				
σ	3σ	4σ	5σ	7σ
96%	87%	82%	83%	75%

set of exterior orientation parameters, the matching was carried out. Then, such tests were successively repeated with multiple of σ_{XYZ} and of $\sigma_{\omega\phi\kappa}$. For each set of multiplied σ_{XYZ} and $\sigma_{\omega\phi\kappa}$, 100 normally distributed random values were generated and used for downgrading the initial exterior orientation parameters (Tab. 2).

To test the implemented tracking, pre-defined key-frames were used. The interval between the *key-frames* was set to 3, 5 and 7. The first frame was always defined as a *key-frame*. Exemplary results on tracking are presented in Fig. 4 and Fig. 5. In these figures, sections of four following frames are shown. In the lower right corner of each image section, the ID of corresponding frame was plotted, in order to establish a link between Fig. 4 and Fig. 5. In the presented example, the interval between the *key-frames* was set to 3, hence frames #13141 (initial frame f_i with $i = 1$) and #13144 are *key-frames*, while frames #13142 and #13143 are normal frames.

Fig. 4 presents the projected model: in green - tracked model edges and in yellow - model edges projected with estimated parameters. Fig. 5 shows the image line segments corresponding to the edges in the current frame (cyan) and the image line segments tracked as correspondences from the previous frame (blue).

Applying the presented tracking strategy, every projected model edge in frame f_i , where $i > 1$ can get two types of correspondences with the image edges:

1. Assigned correspondences (with extracted edges)
2. Tracked correspondences (virtual, with tracked edges)

Virtual (tracked) correspondences can be helpful when not enough new correspondences are found in the current frame. However,

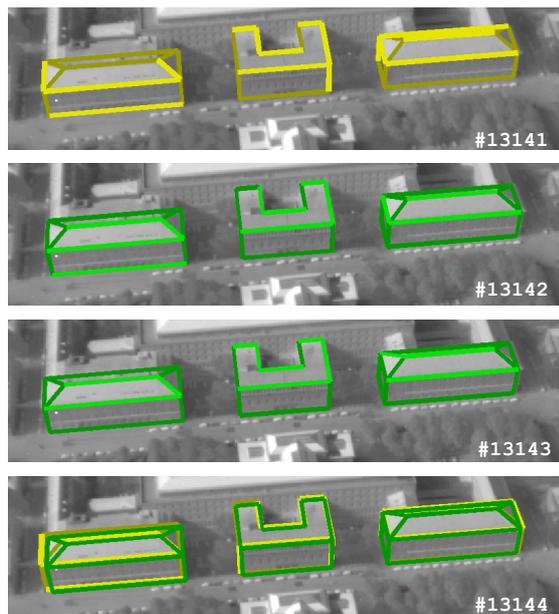


Figure 4: Image sections from a sequence of four images with two key-frames with projected 3D building model. Color coding: bright yellow - model lines with correspondences projected after parameter estimation, dark yellow - model lines without correspondences projected after parameter estimation, bright green - tracked model lines with correspondences, dark green - tracked model lines without correspondences

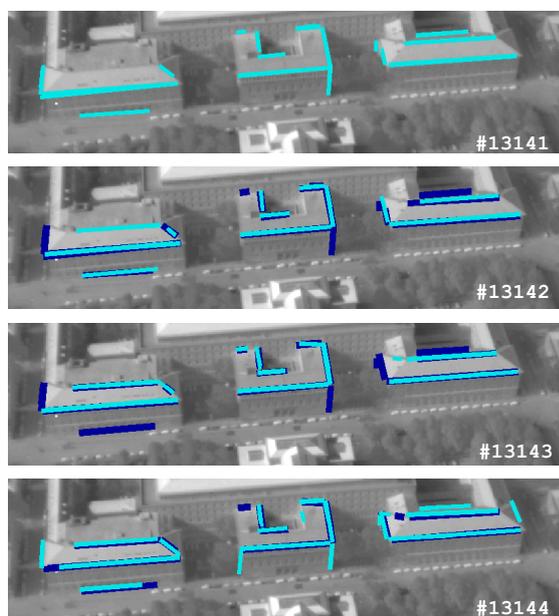


Figure 5: Image sections from a sequence of four images with two key-frames and plotted image line segments with correspondence. Color coding: cyan - image line segments detected in current frame corresponding to a model edge, blue - image line segments tracked as correspondences from the previous frame

they are not needed, if the a new correspondence was found for a certain edge. Hence, in each frame, a verification of correspondences is carried out. It was tested, whether there was a new correspondence which was equivalent to the tracked correspondence. This case occurs when the tracked image edge and newly assigned image edge are equal. This is tested using three conditions:

- the middle points of the line segments are close to each other,
- their are of similar length and
- their are almost incident.

The first two conditions are carried out by setting a threshold. For the third condition, statistical test are implemented. If for only few correspondences where found in the current frame, then the missing correspondences are extended with the virtual correspondences. Hence, the virtual correspondences are also used for tracking in the next frame.

The results on the verification are presented in Fig. 6. Also here, the frame IDs are plotted in the lower right corner. The image section from frame #13141 is missing in this figure. The reason for it is that #13141 was the initial frame, so not verification could be carried out. Two line were verified as the same lines if the distance between the middle points was smaller than 7 pix , the length ratio grater 70%, and the statistical test confirmed their incidence with the significance level $\alpha = 0.01$. The virtual correspondences were added to the current correspondence when less then 30% model edges got a corresponding image line segment.

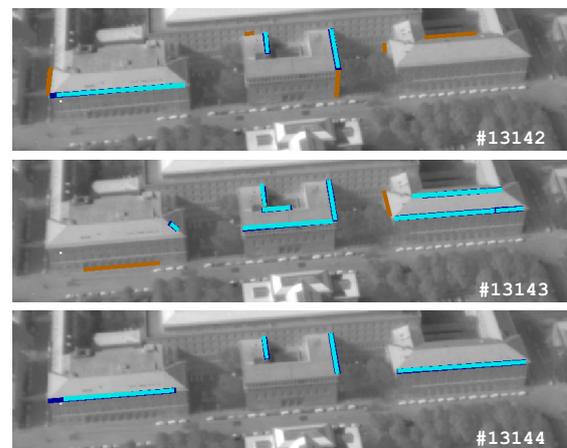


Figure 6: Verification of the edge correspondence. Color coding: cyan - image line segments detected in current frame corresponding to a model edge, blue - verified virtual correspondences with correspondences in the current frame, dark orange - virtual correspondences which were added to the correspondence list and used for tracking in the next frame

In order to assess the accuracy of the tracking, the model edges were also tracked into the *key-frames*. As a measure for this assessment, the distance between the tracked and projected model edges after estimation was used. For each corresponding pair of tracked and projected model edges, the area between was calculated and divided by the length of the model edge. This value was considered to be the average distance between those two edges. This distance was summed up and averaged among whole frame, and then stored as the quality value per frame. Tab. 3 shows analysis of these values stored per frame, dependent on the pre-defined interval between the *key-frames*.

Table 3: Quality measure for tracking expresses with average distance between the tracked and projected model edges. Here analysis of this value per frame.

Interval between the <i>key-frames</i>	3 frames	5 frames	7 frames
Average distance	1.6 <i>pix</i>	2.4 <i>pix</i>	3.1 <i>pix</i>
Maximum distance	2.2 <i>pix</i>	3.1 <i>pix</i>	4.0 <i>pix</i>
Minimum distance	0.9 <i>pix</i>	1.4 <i>pix</i>	0.8 <i>pix</i>

5. DISCUSSION AND OUTLOOK

Line based model-to-image matching has a 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.

Taking the uncertainty of image lines and of the building model into account allows using statistical analysis based on uncertainty, such as statistical test and robust estimation with outlier detector. Also a better fit between the building model and the image structures is achieved.

By tracking the line segments assigned to the 3D model from frame to frame the search area is restricted and the time needed for calculation reduced. Up to now, the experiments on line tracking have been conducted with pre-defined key-frames. In the future, more attention should be paid to dynamically selected key-frames and to the criteria for reliability of the coregistration in a single frame. Depending on this reliability, the next frame in the sequence can be set to a key-frame (in case of low reliability) or to standard frame (in case of high reliability).

The other idea to select the *key-frames* is based on the importance of a frame. Because finally the TIR images are used for texturing, the usefulness for this purpose should be considered. Due to the large number of images in the sequence, each face of the 3D model can be seen many times. However, in certain frame, the quality of the extracted texture of particular face, would be better than in other frames. Frames, which deliver high quality textures for many or for most important face, will be favored to be key-frames.

ACKNOWLEDGEMENTS

The authors would like to thank Fraunhofer IOSB, Ettlingen, for providing thermal images.

REFERENCES

Beder, C., 2004. Fast statistically geometric reasoning about uncertain line segments in 2d- and 3d-space. In: C.E.Rasmussen, H.H.Bülthoff, B.Schölkopf and M.A.Giese (eds), Proceedings of the DAGM Symposium, Springer, pp. 375–382.

Beder, C., 2007. Grouping Uncertain Oriented Projective Geometric Entities with Application to Automatic Building Reconstruction. PhD thesis, Photogrammetry Department, Institute for Geodesy and Geoinformation, Bonn University, Germany, Nussallee 15, 53115 Bonn, Germany.

Borrmann, D., Elseberg, J. and Nüchter, A., 2012. Thermal 3d mapping of building façades. In: Proceedings of the 8th Conference on Intelligent Autonomous Systems. Jeju Island, Korea.

Boyd, G., 2013. Zeroing in on energy savings with thermal imaging. *Photogrammetric Engineering & Remote Sensing* 79(4), pp. 313–316.

Cabrelles, M., Galcera, S., Navarro, S., Lerma, J. L., Akasheh, T. and N., H., 2009. Integration of 3d laser scanning, photogrammetry and thermography to record architectural monuments. In: Proceedings of the 22nd CIPA Symposium, Kyoto, Japan.

Chandler, D., 2011. The big picture on energy loss. MIT news. <http://web.mit.edu/newsoffice/2011/ir-scanning-energy-0316.html> Access on July 30, 2012.

Förstner, W., 2004. Uncertainty and projective geometry. In: E. Bayro-Corrochano (ed.), *Handbook of Computational Geometry for Pattern Recognition, Computer Vision, Neurocomputing and Robotics*, Springer, pp. 493–535.

Früh, C., Sammon, R. and Zakhor, A., 2004. Automated texture mapping of 3d city models with oblique aerial imagery. In: Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'04).

Heuel, S., 2002. Statistical Reasoning in Uncertain Projective Geometry for Polyhedral Object Reconstruction. PhD thesis, Photogrammetry Department, Institute for Geodesy and Geoinformation, Bonn University, Germany, Nussallee 15, 53115 Bonn, Germany.

Heuel, S. and Förstner, W., 2001. Matching, reconstructing and grouping 3d lines from multiple views using uncertain projective geometry. In: Proceedings of Computer Vision and Pattern Recognition.

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

Hsu, S., Samarasekera, S., Kumar, R. and Sawhney, H., 2000. Pose estimation, model refinement, and enhanced visualization using video. In: Proceedings of Computer Vision and Pattern Recognition, Vol. I, pp. 488–495.

Iwaszczuk, D., Hoegner, L. and Stilla, U., 2011. Matching of 3d building models with ir images for texture extraction. In: U. Stilla, P. Gamba, C. Juergens and D. Maktav (eds), Proceedings of JURSE 2011 - Joint Urban Remote Sensing Event, pp. 25–28.

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

Kaden, R. and Kolbe, T. H., 2013. City-wide total energy demand estimation of buildings using semantic 3d city models and statistical data. In: Proc. of the 8th International 3D GeoInfo Conference, 28.-29. 11. 2013 in Istanbul, Turkey, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. II-2/W1.

Kanatani, K., 1996. Statistical optimization for geometric computation : theory and practice. Machine intelligence and pattern recognition, Elsevier, Amsterdam, Lausanne, New York.

Luxen, M. and Förstner, W., 2001. Optimal camera orientation from points and straight lines. In: B. Radig and S. Florczyk (eds), Proceedings of the DAGM 2001, pp. 84–91.

Meidow, J., Beder, C. and Förstner, W., 2009a. Reasoning with uncertain points, straight lines, and straight line segments in 2d. *ISPRS Journal of Photogrammetry and Remote Sensing* 64(2), pp. 125–139.

Meidow, J., Förstner, W. and Beder, C., 2009b. Optimal parameter estimation with homogeneous entities and arbitrary constraints. In: J. Denzler and G. Notni (eds), *Pattern Recognition (Symposium of DAGM)*, Springer, pp. 292–301.

Mikeleit, V. and Kraneis, D., 2010. Planen in fünf dimensionen. *GIS Business* 2010(08), pp. 58–59.

Sawhney, H. S., Arpa, A., Kumar, R., Samarasekera, S., Aggarwal, M., Hsu, S. and Nister, D. Hanna, K., 2002. Video flashlights: real time rendering of multiple videos for immersive model visualization. In: *Proceedings of the 13th Eurographics workshop on Rendering*, pp. 157–168. Pisa, Italy.

Schickler, W., 1992. Feature matching for outer orientation of single images using 3-d wireframe controlpoints. In: *International Archives for Photogrammetry*, Washington, DC, pp. 591–598.

Sester, M. and Förstner, W., 1989. Object location based on uncertain models. In: H. Burkhardt, K. Hähne and B. Neumann (eds), *Muster Erkennung 1989, 11. DAGM-Symposium*, Springer-Verlag London, pp. 457–464. ISBN:3-540-51748-0.

Vidas, S., Moghadam, P. and Bosse, M., 2013. 3d thermal mapping of building interiors using an rgb-d and thermal camera. In: *Proceedings of IEEE International Conference on Robotics and Automation*.

Weng, Q., 2009. Thermal infrared remote sensing for urban climate and environmental studies: methods, applications, and trends. *ISPRS Journal of Photogrammetry and Remote Sensing* 64(4), pp. 335–344.