ACCURACY INVESTIGATION ON IMAGE-BASED CHANGE DETECTION FOR BIM COMPLIANT INDOOR MODELS

T. Meyer[1,2,*], A. Brunn[2], U. Stilla[1]

[1] Photogrammetry and Remote Sensing, Technical University of Munich (TUM), Munich, Germany
[2] University of Applied Sciences Wuerzburg-Schweinfurt (FHWS), Wuerzburg, Germany

Commission IV, WG 5

KEY WORDS: Image Orientation, BIM, Indoor Model, Line Features, Level of Accuracy Specification.

ABSTRACT:

Construction progress documentation is currently of great interest for the AEC (Architecture, Engineering and Construction) branch and BIM (Building Information Modeling). Subject of this work is the geometric accuracy assessment of image-based change detection in indoor environments based on a BIM. Line features usually serve well as geodetic references in indoor scenes in order to solve for camera orientation. However, building edges are never perfectly built as planned and often geometrically generalized for BIM compliant representation. As a result, in this approach, line correspondences for image-to-model co-registration are considered as statistically uncertain entities as this is essential for dealing with metric confidences in the field of civil engineering and BIM. We present an estimation model for camera pose refinement which is based on the incidence condition between model edges and corresponding image lines. Geometric accuracies are assigned to the model edges according to the Level of Accuracy (LOA) specification for BIM. The approach is demonstrated in a series of tests using a synthetic image of an indoor BIM. The effects of varying edge detection accuracies on the estimation are investigated as well as the effects of using model edges with different geometric quality by adding Gaussian noise to the synthetic observations, each within 100 simulation runs. The results show that the camera orientation can be improved with the presented estimation model as long as the BIM compliant references meet the conditions of LOA 30 or higher ($\sigma < 7.5\,mm$).

1. INTRODUCTION

In the AEC (Architecture, Engineering and Construction) industry, there is a great demand for automated and accurate scene capturing and surveying, mainly for the purpose of construction progress monitoring in the context of BIM (Building Information Modeling). With BIM, the construction project is based on a so-called digital twin, which represents the planned state of the object in its various phases of construction. In order for the construction management to be able to react to possible deviations from the planning at an early stage, a regular comparison of the as-planned and as-built project status must be carried out.
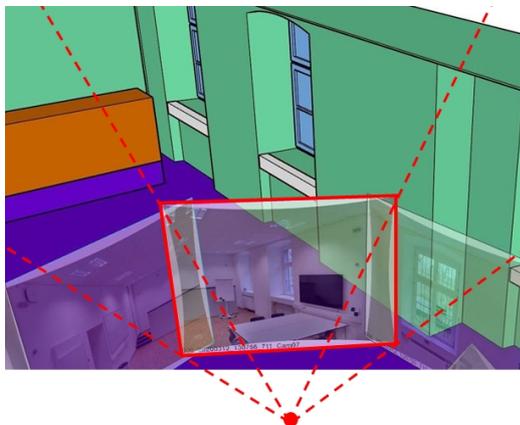

Figure 1. Image oriented in the BIM system showing both modeled BIM objects and those not included in the model.

---

* Corresponding author

With the increasing development of BIM, the demand on its geometric quality increases. This is especially true for the modeling of detailed structures in building interiors, as they continue to be maintained beyond the construction phase as a central database for many applications such as location based services and the internet of things. Image based measurement techniques can meet general requirements of accuracy and efficiency and therefore, are becoming more and more popular in this field. A camera's orientation is required in order to determine the spatial correspondence between an image and a 3D model (Fig. 1). However, the estimation of a camera's pose within a BIM's reference system is indoors much more difficult than outdoors. The lack of direct geo-referencing (GNSS denied environment) as well as occlusions and ambiguities due to general indoor structures are challenging circumstances for image based measurement techniques within a building's interior and related accuracy requirements.
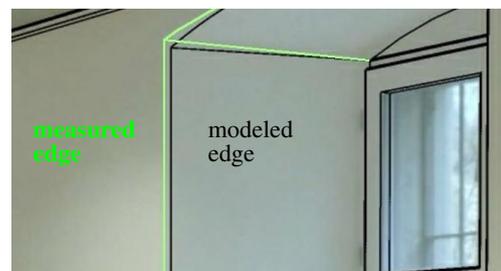

Figure 2. Discrepancy between measured edge (as-built) and modeled edge (as-planned)

Line features usually serve well as geodetic references in indoor scenes in order to estimate the camera orientation. They

are sufficiently available in man-made environments (and corresponding BIMs) and can be detected easily by common image processing algorithms. However, building edges are never perfectly built as planned and often geometrically generalized for BIM-compliant representation (Fig. 2). As a result, line correspondences for image-to-model co-registration must be considered as statistically uncertain entities as this is essential for dealing with metric confidences in the field of civil engineering and BIM. In this contribution, we present an estimation model
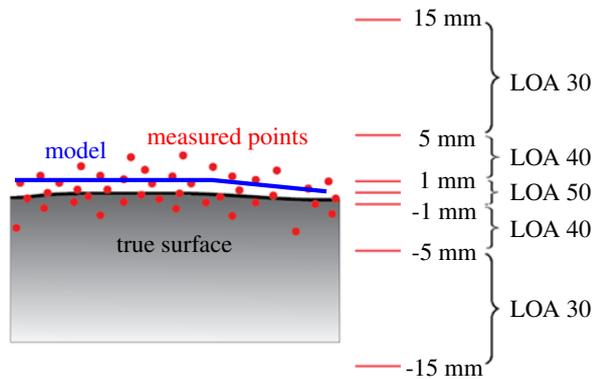

Figure 3. Level of Accuracy (LOA) (USIBD, 2020)

for camera pose refinement based on line references, given a coarse orientation (e.g. from IMU (Inertial Measurement Unit) sensors and visual odometry). The stochastic model includes the individual uncertainty information of BIM related objects according to the Level of Accuracy (LOA) specification for BIM (USIBD, 2020) (Fig. 3). The approach will support future developments of image based change detection in indoor environments with available BIM. According to compliance checking in BIM, we focus on structural and topological changes in the geometry of BIM compliant 3D models.

1.1 Related Work

Our motivation comes from the need for image-based building documentation and change detection that is, on the one hand, specifically focused on the interior and, on the other hand, takes into account BIM related accuracies and quality requirements. Nevertheless, our approach is based on some related research areas.

1.1.1 Image-based Construction Progress Documentation for BIM
BIM compliant construction progress documentation based on images and computer vision technologies has become a researched topic in recent years. Most existing methods consider outdoor observations. In Tuttas et al. (2017) the as-built state of a construction site is detected by photogrammetric outdoor surveys. Knowledge on the construction processes is inferred from BIM objects and SfM (Structure from Motion) methods enable 3D building elements to be located. Hoegner et al. (2016) co-register image sequences with a BIM for the documentation of geometric and radiometric changes of a construction site. They use model information as fictitious observations within the adjustment. An a priori sensor pose is given from GNSS and IMU (outdoors). Other authors contribute valuable approaches for image based indoor applications, however, they do not consider BIM compliant accuracies or high requirements on geometric quality. Han and Golparvar-Fard (2015) solve camera orientation by manually selected tie points. Relevant

model elements are back projected into the images in order to segment image regions for material classification and Kropp et al. (2018) increase automation for inspections on interior construction states by image sequences which are registered with a BIM to derive rich information about single construction tasks. In contrast, Brunn and Meyer (2016) showed that sub millimeter accuracies can be achieved with a multi camera system in close range indoor applications using precise geodetic control points for proper camera calibration and pose estimation.

1.1.2 Model-to-Image Matching based on Lines
Edge detection is a basic task in image processing, which is why line features are generally well suited to be matched to building models. However, many approaches also mostly focus on outdoor applications. Li-Chee-Ming and Armenakis (2014) use vertical lines from UAV-borne video streams in urban scenes, that are matched to extracted line features from synthetic images of 3D building models. Iwaszczuk et al. (2017) present a line based model-to-image matching for texturing 3D building models with airborne oblique view thermal infrared images and Verykokou and Ioannidis (2016) use vanishing point detection techniques for the rough estimation of the exterior orientation parameters of oblique aerial images.

1.1.3 Uncertainty of Matching Features
Usually, image features as well as geometric primitives from 3D models should supposed to be statistically uncertain because of unavoidable inaccuracies in measurement and modeling (generalization). Therefore, these uncertainties have to be considered in the representation and estimation of such entities. For fitting parameterized 3D models to images Lowe (1991) presents general methods that take account of inherent inaccuracies in the image measurements. Heuel and Förstner (2001) combine protective geometry and statistical methods by representing points, lines and planes and their uncertainties in homogeneous vectors. They show how errors in the construction of new elements are propagated. Building up on that, Meidow et al. (2009) developed a generic estimation model for homogeneous entities and multiple arbitrary constrains and Iwaszczuk and Stilla (2017) present a similar approach for camera pose optimization for the purpose of automated texture mapping, in which a line-based model-to-image matching is developed. Although these techniques have not been presented directly for the case of construction progress documentation, they can be used for this purpose. These fairly generic techniques can be adapted to the indoor case and extended with BIM specific accuracies resp. uncertainties.

1.2 Own Contribution and Paper Structure

In contrast to previous work, the main idea of this project is to make use of any initial BIM compliant information and therefore, the edges of known BIM objects which are identified in the images will be used for camera pose refinement. We focus on interior scenes and pay particular attention to the individual geometric quality of reference objects as metric tolerances have to be met in construction. In our approach, the individual accuracy information of BIM objects and corresponding model representations is taken into account. We investigate the accuracy of sensor orientation prior to 3D reconstruction and change detection. It is basic work for image based applications in building interiors for BIM as it enables geometrically and semantically correct image interpretation for subsequent processing steps.

In the following sections, we present an estimation model for camera pose refinement based on line references, given a coarse orientation (e.g. from IMU sensors and visual odometry). The

approach is based on the incidence condition between model edges and corresponding image lines in the projective space. This relation serves as a constraint in a Gauss-Helmert model. Our functional model is designed reflecting the mutual relation between the observations, using the homogeneous Plücker matrix for 3D line representation and the unknown parameters of the projection matrix. Furthermore, the stochastic model includes the individual uncertainty information of BIM related objects and facilities according to the LOA specification for BIM. The optimization approach is demonstrated in a series of tests using a synthetic image of an indoor BIM. The effects of varying edge detection accuracies on the estimation are investigated as well as the effects of using model edges with different geometric qualities by adding Gaussian noise to the synthetic observations, each within 100 simulation runs.

1.3 Method Overview

For the change detection application, a stereo camera system with an IMU will be used to easily obtain the relative orientation, the scale and an image based 3D point cloud by SfM and MVS (Multi View Stereo). Indoor environments can be characterized by their large number of straight edges e.g. on windows, doors, walls and furniture. 3D lines as geometric primitives can be used to describe the building interior in an abstract way. This brings advantages over unstructured 3D point clouds in terms of matching and co-registration with a corresponding 3D model. Therefore, the approach of Hofer et al. (2015) will be integrated in the process in order to generate an abstract 3D line model. With the results of a semantic image segmentation, those 3D line segments will be further extended with semantic information, similar to what is described by Iwaszczuk et al. (2017). The coarse absolute orientation of an image sequence in the BIM's reference system will be determined from matching corresponding semantic 3D line segments (Fig. 4).

3D scene reconstruction will also be supported by the available BIM information in an object based approach. Finally the change detection can be realized by the comparison with the last version of the BIM before it is updated with the new information. Altogether, this will result in a positive feedback loop because the higher the quality of the initial BIM, the faster and more precisely it can be updated again. The overall concept is depicted in figure 5.
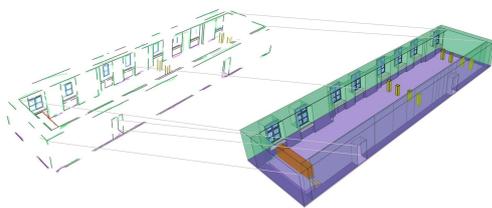

Figure 4. Assignment of semantic 3D line segments to corresponding model edges.

1.4 Projection of homogeneous Points and Lines

A two dimensional straight line can be represented in the homogeneous form $ax + by + c = 0$ with the homogeneous vector:

$$\mathbf{l} = [a, b, c]^T = [\cos(\theta), \sin(\theta), -d]^T \qquad (1)$$

where θ and d are the parameters of the Hessian normal form. A 3D point with Euclidean coordinates X, Y and Z has the ho-
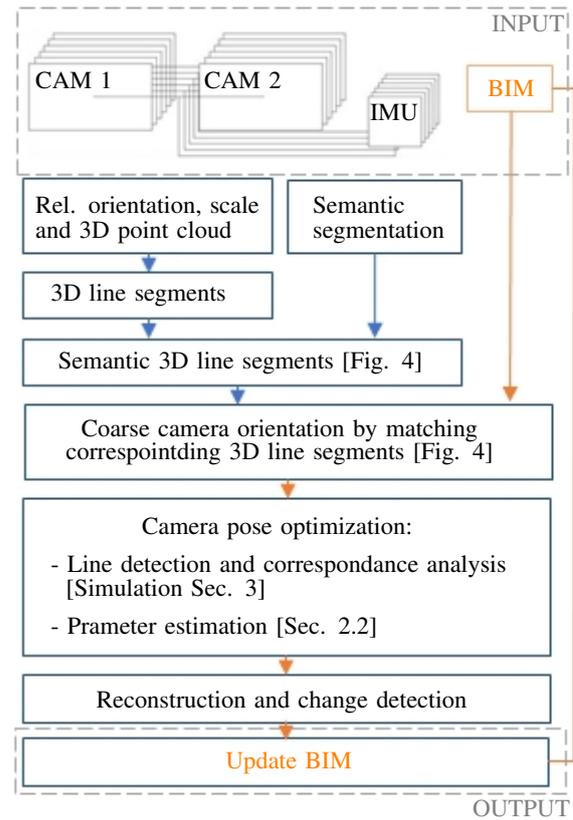

Figure 5. Method overview

mogeneous coordinates:

$$\mathbf{X} = [TX, TY, TZ, T]^T = [U, V, W, T]^T \qquad (2)$$

A 3D line can be represented by the homogeneous Plücker matrix in the projective space:

$$\Gamma(\mathbf{L}) = \mathbf{X}\mathbf{Y}^T - \mathbf{Y}\mathbf{X}^T \qquad (3)$$

where $\mathbf{X}$ and $\mathbf{Y}$ represent the endpoints of the line.

Mapping an object point $\mathbf{X}$ to an image point $\mathbf{x}$ with a pin hole camera is achieved by multiplication with the projection matrix P:

$$\mathbf{x} = \mathrm{P}\,\mathbf{X} \qquad (4)$$

with calibration matrix K, rotation matrix R, identity matrix I and the 3D coordinates of the projection center X^0, P results from:

$$\mathrm{P} = \mathrm{K}\,\mathrm{R}\,[\mathrm{I}|-X^0] \quad \text{with} \quad \mathrm{K} = \begin{bmatrix} c_x & s & x_0 \\ 0 & c_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \qquad (5)$$

2. METHODOLOGY

2.1 Camera Pose Optimization in the BIM's Reference System

The exterior orientation parameters of a camera ($X_0, Y_0, Z_0,$ ω, ϕ, κ) are implicitly available in the projection matrix P. The approximate projection matrix gets optimized in the BIM's reference system thru the observation of straight lines in the im-

age ($\mathbf{l}_i$) and corresponding 3D model edges ($\mathbf{L}_i$). The incidence condition of corresponding line features is used as input for the formulation of a functional model within a generic estimation model for homogeneous entities according to Meidow et al. (2009).

2.1.1 Functional Model

The estimation is based on a Gauss Helmert model with constraints. The corrected observations ($\hat{\mathbf{y}} = \mathbf{y} + \hat{\mathbf{v}}$) and the estimated unknown parameters ($\hat{\boldsymbol{\beta}}$) have to fulfill certain conditions. The G-conditions $\mathbf{g}(\hat{\mathbf{y}}, \hat{\boldsymbol{\beta}}) = 0$ describe the relation between the observations and the parameters. The H-restrictions $\mathbf{h}(\hat{\boldsymbol{\beta}}) = 0$ concern only the parameters and the C-constraints $\mathbf{c}(\hat{\mathbf{y}}) = 0$ are imposed on the observations alone. The incidence condition of a 3D model edge, which is projected into the image with an observed corresponding straight line in the image serves as G-condition. It represents two independent constraints (Förstner and Wrobel, 2016):

$$\Gamma(\mathbf{L}) \, \mathrm{P}^T \, \mathbf{l} = \mathbf{0} \qquad (6)$$

where $\mathbf{0}$ is a 4 x 1 zero vector.

For the handling of singular covariance matrices, the homogeneous observation entities as well as the parameters must be spherically normalized (normalized to 1). Therefore, the H-restrictions for the parameters is:

$$\mathbf{p}^T \mathbf{p} - 1 = 0 \qquad (7)$$

where $\mathbf{p} = \mathrm{vec}(\mathrm{P})$, a column vector with the reshaped elements of P.

Respectively the C-constraints for the observations result in:

$$\mathbf{l}_j^T \mathbf{l}_j - 1 = 0 \qquad \mathbf{X}_i^T \mathbf{X}_i - 1 = 0 \qquad \mathbf{Y}_i^T \mathbf{Y}_i - 1 = 0 \qquad (8)$$

The optimal solution for $\boldsymbol{\beta}$ is given by the minimum of the weighted squared residuals subject to the given constraints. This is achieved by minimizing the Lagrange function with the Lagrangian vectors $\boldsymbol{\lambda}, \boldsymbol{\mu}$ and $\boldsymbol{\nu}$:

$$
\begin{aligned}
L = \frac{1}{2} \, \hat{\boldsymbol{v}}^T \, \boldsymbol{\Sigma}_{\mathbf{yy}}^+ \, \hat{\boldsymbol{v}} &+ \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{y} + \hat{\boldsymbol{v}}, \hat{\boldsymbol{\beta}}) + \boldsymbol{\mu}^T \mathbf{h}(\hat{\boldsymbol{\beta}}) \\
&+ \boldsymbol{\nu}_1^T \mathbf{c}_1(\mathbf{y} + \hat{\boldsymbol{v}}) + \boldsymbol{\nu}_2^T \mathbf{c}_2(\mathbf{y} + \hat{\boldsymbol{v}}) + \boldsymbol{\nu}_3^T \mathbf{c}_3(\mathbf{y} + \hat{\boldsymbol{v}})
\end{aligned}
\qquad (9)
$$

The corrections for the observations and parameters are calculated in an iterative procedure according to Meidow et al. (2009). In every iteration, the Jacobians are calculated:

$$\mathbf{A} = \frac{\partial \mathbf{g}(\mathbf{y}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}, \mathbf{B}^T = \frac{\partial \mathbf{g}(\mathbf{y}, \boldsymbol{\beta})}{\partial \mathbf{y}}, \mathbf{C}^T = \frac{\partial \mathbf{c}(\mathbf{y})}{\partial \mathbf{y}}, \mathbf{H}^T = \frac{\partial \mathbf{h}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$$

Additionally, in every iteration (τ) the residuals of the constraints and the auxiliary variable $\mathbf{a}$:

$$\mathbf{g}_\tau = \mathbf{g}(\mathbf{y}^{(\tau)}, \boldsymbol{\beta}^{(\tau)}), \qquad \mathbf{h}_\tau = \mathbf{h}(\boldsymbol{\beta}^{(\tau)}), \qquad \mathbf{c}_\tau = \mathbf{c}(\mathbf{c}^{(\tau)})$$

$$\mathbf{a} = \mathbf{B}^T \mathbf{C}(\mathbf{C}^T \mathbf{C})^{-1}(\mathbf{C}^T(\mathbf{y} - \mathbf{y}^{(\tau)}) + \mathbf{c}_\tau) - \mathbf{B}^T(\mathbf{y} - \mathbf{y}^{(\tau)}) - \mathbf{g}_\tau$$

The normal equation system is solved by using $\mathbf{a}$ and matrix LU decomposition in order to receive the corrections for the estimated parameters:

$$
\begin{bmatrix} \mathbf{A}^T \boldsymbol{\Sigma}_{\mathbf{gg}}^{-1} \mathbf{A} & \mathbf{H} \\ \mathbf{H}^T & \mathbf{0} \end{bmatrix}
\begin{bmatrix} \widehat{\Delta\boldsymbol{\beta}} \\ \boldsymbol{\mu} \end{bmatrix} =
\begin{bmatrix} \mathbf{A}^T \boldsymbol{\Sigma}_{\mathbf{gg}}^{-1} \mathbf{a} \\ -\mathbf{h}_\tau \end{bmatrix}
\qquad (10)
$$

With the covariance matrix of the contradictions: $\boldsymbol{\Sigma}_{\mathbf{gg}} = \mathbf{B}^T \boldsymbol{\Sigma}_{\mathbf{yy}} \mathbf{B}$. Finally, the residuals:

$$\hat{\mathbf{v}}^{(\tau)} = -\boldsymbol{\Sigma}_{\mathbf{yy}} \mathbf{B} \boldsymbol{\lambda} - \mathbf{C}(\mathbf{C}^T \mathbf{C})^{-1}(\mathbf{C}^T(\mathbf{y} - \mathbf{y}^{(\tau)}) + \mathbf{c}_\tau) \qquad (11)$$

with:

$$\boldsymbol{\lambda} = \boldsymbol{\Sigma}_{\mathbf{gg}}^{-1}(\mathbf{A}\widehat{\Delta\boldsymbol{\beta}} - \mathbf{a}) \qquad (12)$$

2.1.2 Stochastic Model

The limited accuracy of the reference lines such as the inherent uncertainty of straight line detection have to be considered in the estimation. The homogeneous representations of 2D and 3D lines are therefore extended with individual stochastic information using covariance matrices.
A 3D point with Euclidean coordinates $X = [X, Y, Z]^T$ has the Euclidean covariance matrix:

$$
\Sigma_{XX} = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} & \sigma_{XZ} \\ \sigma_{XY} & \sigma_Y^2 & \sigma_{YZ} \\ \sigma_{XZ} & \sigma_{YZ} & \sigma_Z^2 \end{bmatrix}
\qquad (13)
$$

The homogeneous representation of a 3D point $X \to \mathbf{X}$ is:

$$\mathbf{X} = [X_E, X_h]^T = [U, V, W, T]^T = [\lambda X, \lambda Y, \lambda Z, \lambda]^T \qquad (14)$$

where $X_E = [U, V, W]^T$ is the Euclidean and $X_h = T$ is the homogeneous part. It follows the homogeneous covariance matrix $\Sigma_{\mathbf{XX}}$:

$$
\Sigma_{\mathbf{XX}} = \lambda^2 \begin{bmatrix} \sigma_X^2 & \sigma_{XY} & \sigma_{XZ} & 0 \\ \sigma_{XY} & \sigma_Y^2 & \sigma_{YZ} & 0 \\ \sigma_{XZ} & \sigma_{YZ} & \sigma_Z^2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \lambda^2 \begin{bmatrix} \Sigma_{XX} & \mathbf{0} \\ \mathbf{0}^T & 0 \end{bmatrix}
\qquad (15)
$$

For converting a homogeneous vector back in a Euclidean representation $\mathbf{X} \to \mathbf{X}^e$, the Jacobian J^e at $\mathbf{X}$ is needed, as the normalization is a non-linear function:

$$\mathbf{X}^e = \frac{\mathbf{X}}{X_h} = \begin{bmatrix} X \\ 1 \end{bmatrix} \qquad (16)$$

$$J^e(\mathbf{X}) = \frac{1}{X_h} \begin{bmatrix} I_3 & -\frac{1}{X_h} X_E \\ \mathbf{0}^T & 0 \end{bmatrix} \qquad (17)$$

$$\Sigma_{\mathbf{X}^e \mathbf{X}^e} = J^e(\mathbf{X}) \, \Sigma_{\mathbf{XX}} \, J^e(\mathbf{X})^T \qquad (18)$$

A two dimensional straight line with the homogeneous vector $\mathbf{l} = [a, b, c]^T$ has the homogeneous covariance matrix $\Sigma_{\mathbf{ll}}$:

$$
\Sigma_{\mathbf{ll}} = \begin{bmatrix} \sigma_a^2 & \sigma_{ab} & \sigma_{ac} \\ \sigma_{ab} & \sigma_b^2 & \sigma_{bc} \\ \sigma_{ac} & \sigma_{bc} & \sigma_c^2 \end{bmatrix}
\qquad (19)
$$

If an uncertain straight line is given by its homogeneous parameters, the parameters of the Hessian form can be derived form: $((\mathbf{l}, \Sigma_{\mathbf{ll}}) \to ([\theta, r], \Sigma_{hh}))$

$$\begin{bmatrix} \theta \\ r \end{bmatrix} = \begin{bmatrix} \arctan2(b, a) \\ -\frac{c}{\sqrt{a^2+b^2}} \end{bmatrix} \qquad (20)$$

The Jacobian J_{lh} is needed at $\mathbf{l}$:

$$J_{lh}(\mathbf{l}) = \frac{1}{s^3} \begin{bmatrix} -bs & as & 0 \\ ac & bc & -s^2 \end{bmatrix} \quad \text{with} \quad s = \sqrt{a^2 + b^2} \qquad (21)$$

The covariance matrix Σ_{hh} results from:

$$\Sigma_{hh} = J_{hl}(\mathbf{l}) \, \Sigma_{\mathbf{ll}} \, J_{hl}(\mathbf{l})^T \qquad (22)$$

2.2 Parameter Estimation

For each pair of corresponding lines, the vector with the observations is $\mathbf{y} = [\mathbf{l}_i, \mathbf{X}_i, \mathbf{Y}_i, ...]^T$ and the related covariances respectively $\Sigma_{\mathbf{yy}} = Diag[\Sigma_{\mathbf{l}_i \mathbf{l}_i}, \Sigma_{\mathbf{X}_i \mathbf{X}_i}, \Sigma_{\mathbf{Y}_i \mathbf{Y}_i}, ...]$. The elements of P are as $\mathbf{p}$ in $\boldsymbol{\beta}$.

2.2.1 Conditioning
As the homogeneous entities relate to Euclidean BIM objects, their coordinates are expressed with respect to the BIM's reference coordinate system. A big difference between the Euclidean and the homogeneous part causes the calculation to be numerical instable, which is why conditioning as proposed by Förstner and Wrobel (2016) is applied:

$$\mathbf{x}_c = \mathrm{T}_{2D}\,\mathbf{x} \quad \text{and} \quad \mathbf{y}_c = \mathrm{T}_{2D}\,\mathbf{y} \qquad (23)$$

For 2D points T_{2D} is composed with the centroid coordinates μ_x and μ_y and the maximum distance to the centroid s_{max}:

$$\mathrm{T}_{2D} = \begin{bmatrix} 1 & 0 & -\mu_x \\ 0 & 1 & -\mu_y \\ 0 & 0 & s_{max} \end{bmatrix} \qquad (24)$$

The procedure is analogous to 3D points.
Straight lines are conditioned and re-conditioned using:

$$\mathbf{l}_c = \mathrm{L}_{2D}\,\mathbf{l}, \quad \mathbf{l} = \mathrm{L}_{2D}^{-1}\,\mathbf{l}_c \quad \text{with} \quad \mathrm{L}_{2D} = (\mathrm{T}_{2D}^{-1})^T \qquad (25)$$

Conditioning is applied on the observations. For the projection matrix results:

$$\mathrm{P}_c = \mathrm{T}_{2D}\,\mathrm{P}\,\mathrm{T}_{3D}^{-1} \qquad (26)$$

2.2.2 Spherical Normalization
According to (7) and (8) the observations and initial parameters have to be spherically normalized prior to the adjustment: $\mathbf{l}_c^s := \mathbf{l}_c/|\mathbf{l}_c|$ and $\Sigma_{\mathbf{ll}} = \mathrm{J}_c\,\Sigma_{\mathbf{ll}}\,\mathrm{J}_c^T$. With the Jacobian J_c:

$$\mathrm{J}_c(\mathbf{l}) = \frac{1}{|\mathbf{l}_c|}\left(I_3 - \frac{\mathbf{l}_c\,\mathbf{l}_c^T}{\mathbf{l}_c^T\,\mathbf{l}_c}\right) \qquad (27)$$

In the same way it is done for the parameters and 3D model points: $\mathbf{X}_c^s := \mathbf{X}_c/|\mathbf{X}_c|$ with $\Sigma_{\mathbf{XX}} = \mathrm{J}_c\,\Sigma_{\mathbf{XX}}\,\mathrm{J}_c^T$ and $\mathbf{p}_c^s := \mathbf{p}_c/|\mathbf{p}_c|$.

In the following sections we assume the homogeneous coordinates to be conditioned and spherically normalized and omit the indices c and s.

2.2.3 Jacobians
The A-Matrix includes the first derivatives of the G-constraints according to the unknown parameters of $\mathbf{p}$. As eq. 6 corresponds to 4 constraints of which 2 are chosen, the matrix has as twice as many rows as observed line correspondences and the number of columns corresponds to the number of parameters:

$$\mathbf{A} = \left[\begin{pmatrix} \mathrm{x}_{1i} & 0 \\ 0 & \mathrm{x}_{2i} \end{pmatrix} \begin{pmatrix} \mathrm{l}_{1i}\,\mathbf{Y}_i^T & \mathrm{l}_{2i}\,\mathbf{Y}^T & \mathrm{l}_{3i}\,\mathbf{Y}^T \\ \mathrm{l}_{1i}\,\mathbf{Y}_i^T & \mathrm{l}_{2i}\,\mathbf{Y}^T & \mathrm{l}_{3i}\,\mathbf{Y}^T \end{pmatrix} \right.$$
$$\left. - \begin{pmatrix} \mathrm{y}_{1i} & 0 \\ 0 & \mathrm{y}_{2i} \end{pmatrix} \begin{pmatrix} \mathrm{l}_{1i}\,\mathbf{X}_i^T & \mathrm{l}_{2i}\,\mathbf{X}^T & \mathrm{l}_{3i}\,\mathbf{X}^T \\ \mathrm{l}_{1i}\,\mathbf{X}_i^T & \mathrm{l}_{2i}\,\mathbf{X}^T & \mathrm{l}_{3i}\,\mathbf{X}^T \end{pmatrix}, ... \right]^T \qquad (28)$$

The B-Matrix includes the first derivatives of the G-constraints (6) according to the observations $\mathbf{l}_i$, $\mathbf{X}_i$ and $\mathbf{Y}_i$. It is a diagonal matrix where each main diagonal element has 2 rows (2 constraints in (6) and 11 columns (11 elements in the homogeneous vectors $\mathbf{l}$, $\mathbf{X}$ and $\mathbf{Y}$).

$$\mathbf{B} = \mathrm{diag}\big[((\mathbf{X}_i\,\mathbf{Y}_i^T - \mathbf{Y}_i\,\mathbf{X}_i^T)\,\mathrm{P}^T, \mathbf{Y}_i^T\,\mathrm{P}^T\,\mathbf{l}_i\,I - \mathbf{Y}_i\,\mathbf{l}_i^T\,\mathrm{P},$$
$$\mathbf{X}_i\,\mathbf{l}_i^T\,\mathrm{P} - \mathbf{X}_i^T\,\mathrm{P}^T\,\mathbf{l}_i\,I), ...\big] \qquad (29)$$

The C-Matrix includes the first derivatives of the C-constraints (7) according to the observations $\mathbf{l}_i$, $\mathbf{X}_i$ and $\mathbf{Y}_i$. The derivatives for each observation triple form diagonal sub matrices of dimension (3,11), which in turn are written to the main diagonal matrix.

$$\mathbf{C} = \mathrm{diag}\big[\mathrm{diag}[2\,\mathbf{l}_i^T, 2\,\mathbf{X}_i^T, 2\,\mathbf{Y}_i^T], ...\big] \qquad (30)$$

The H-Matrix includes the first derivatives of the H-constraints according to $\mathbf{p}$. It has one row and 12 columns.

$$\mathbf{H} = [2\,\mathbf{p}^T] \qquad (31)$$

2.3 Euclidean Interpretation

The initial projection matrix is optimized during the adjustment. The homogeneous matrix, as well as the uncertain observations in the form of 3D points and 2D straight lines, are thereby conditioned and spherically normalized. For the application of the results in engineering practice, the output data should be interpreted Euclidean.

2.3.1 Projection Matrix
After the adjustment, an estimated projection matrix ($\hat{\mathrm{P}}_c^s$) is available, which is also conditioned and normalized. It is reconditioned with (26).

From the estimated projection matrix, the improved parameters of the exterior orientation can be derived by decomposing the matrix: $\mathrm{P} = [A|a] = [\mathrm{K\,R}| - \mathrm{K\,R}\,X^0]$ (Förstner and Wrobel, 2016).

The projection center X^0 is obtained from: $X^0 = -\mathrm{A}^{-1}\,a$. Factorizing A by QR decomposition, which expresses a matrix A as multiplication of an orthogonal matrix Q and an upper triangular matrix R, results in the rotation matrix R. A should have a positive determinant: $\bar{\mathrm{A}} = sign(|\mathrm{A}|)\,\mathrm{A}$. The inverse $\bar{\mathrm{A}}$ is decomposed: $[\bar{\mathrm{R}}^T, \bar{\mathrm{K}}^{-1}] = qr(\bar{\mathrm{A}}^{-1})$.

The sign s (here: $s = +1$) of principal distance needs to be specified for the calculation of the diagonal matrix D:

$$\mathrm{D} = \mathrm{diag}(sign(\mathrm{diag}(\mathrm{K})))\mathrm{diag}([s, s, +1])$$

Finally, the rotation matrix R results from: $\mathrm{R} = \mathrm{D}\,\bar{\mathrm{R}}$.

2.3.2 Corrected Observations
After the adjustment, Euclidean interpretation of uncertain, homogeneous, conditioned and spherically normalized 3D points and 2D lines is achieved through Euclidean normalization and re-conditioning (table 1).

3. EXPERIMENTS

The estimation model presented in the previous section for the optimization of an approximately known camera pose was tested with synthetic data as ground truth for accuracy assessment.

	3D point	**2D line**
State after adjustment	$\mathbf{X}_c^s, \Sigma_{\mathbf{X}_c^s \mathbf{X}_c^s}$	$\mathbf{l}_c^s, \Sigma_{\mathbf{l}_c^s \mathbf{l}_c^s}$
1. Euclidean normalization	$\mathbf{X}_c, \Sigma_{\mathbf{X}_c \mathbf{X}_c}$ with $J^e(\mathbf{X}_c^s)$	$\mathbf{l}_c^e, \Sigma_{\mathbf{l}_c \mathbf{l}_c}$ with $J^e(\mathbf{l}_c^s)$
2. Re-conditioning	$\mathbf{X}, \Sigma_{\mathbf{XX}}$ with T_{3D}	$\mathbf{l}, \Sigma_{\mathbf{ll}}$ with L_{2D}
3. Euclidean interpretation	$\mathbf{X}^e, \Sigma_{\mathbf{X}^e \mathbf{X}^e}$ with $J^e(\mathbf{X})$	$[\theta, r]^T, \Sigma_{hh}$ with $J_{hl}(\mathbf{l})$

Table 1. Euclidean interpretation of uncertain 3D points and 2D lines
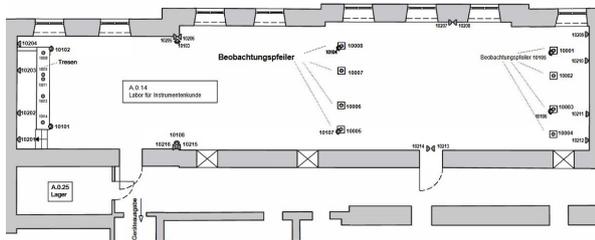

Figure 6. Floor plan of the test environment

Figure 7. BIM compliant indoor model of the test environment

The basis is a real test environment. In the following, the reference model as well as the test setup and the obtained results are presented.

3.1 Test Environment and Reference Model

A laboratory for measurements with geodetic reference network is used as test environment (Fig. 6). It has a local coordinate system whose main axes are aligned with the walls of the building. The network is realized by permanently installed mini prisms and enables the stationing of tacheometers in the lower millimeter range. The laboratory serves as a test environment for validating image measurement techniques and achievable accuracies with respect to the detection of typical objects and component classes inside buildings.

To obtain precise geodetic survey data, the laboratory was captured by a 3D laser scanner in a high resolution. The registration of the single scans was done by control points. The global, absolute measurement accuracy in the reference system is 1.5 millimeters and standard deviation of the relative orientation of the individual scans is in the sub millimeter range when cloud-to-cloud adjustment is performed. The resulting point cloud served as the basis for BIM compliant modeling. The modeling accuracy depends on the respective object class and is based on the recommendations according to the LOA specification. For example, interior doors and windows are assigned to accuracy level 30 or 40.

3.2 Virtual Camera and Data Input

A virtual view of the interior model was rendered with 36 mm sensor width, 1600 x 1200 pixels and 30 mm focal length. The projection center is located at position $\tilde{X}^0 = [43.907, 28.847, 8.053]^T$ in the reference system and the camera system is rotated by angles $\tilde{\omega} = -67.318°$, $\tilde{\phi} = 1.8°$ and $\tilde{\kappa} = 29.498°$ with respect to the BIM system. The true projection matrix ($\tilde{\mathrm{P}}$)

was derived from the parameters of the distortion-free virtual camera according to (5).

In the next step, 36 model points ($\mathbf{X}_i$) were projected with $\tilde{\mathrm{P}}$ and (4) into the synthetic image in order to obtain $\tilde{\mathbf{x}}_i$. For the simulation of inaccuracies in the detection of edges in the image, Gaussian noise of varying magnitude is added to the Euclidean part of the true pixels ($\tilde{\mathbf{x}}_i$) in the further course. Noisy pixels ($\mathbf{x}_i$) result with the normally distributed random number N: $\mathbf{x}_i = \tilde{\mathbf{x}}_i + \sigma_{\mathbf{x}_i} N$. Then, the uncertain image points ($\mathbf{x}_i, \Sigma_{\mathbf{x}_i \mathbf{x}_i}$) were joined to lines by join operation and variance propagation to obtain the uncertain image lines corresponding to the model edges ($\mathbf{l}_i, \Sigma_{\mathbf{l}_i \mathbf{l}_i}$) as observations. A standard deviation in translation of $0.15\,m$ and in rotation of $1°$ was then set for the external orientation of the camera for artificial degradation (σ_{cam}). This resulted in the initial projection matrix P which is used in the adjustment as approximate initial solution. Figure 8 a) shows the synthetic image with the model edges projected into it using $\tilde{\mathrm{P}}$ (green) and P (red).

The estimation of the projection matrix is to be done several times with different imprecise model reference. For this purpose, Gaussian noise is also added to the model points ($\mathbf{X}_i$) in several simulations. The determination of the standard deviation is based on the LOA specification: $\mathbf{X}_i = \mathbf{X}_i + \sigma_{\mathbf{X}_i} N$. The LOA are: $2\sigma = 15\,cm$ (LOA 10), $2\sigma = 5\,cm$ (LOA 20), $2\sigma = 15\,mm$ (LOA 30), $2\sigma = 5\,mm$ (LOA 40), $2\sigma = 1\,mm$ (LOA 50) (USIBD, 2020).

4. RESULTS

P is put in the estimation model as initial value. The actual accuracy of the reference model does not matter here when using synthetic data. For LOA 10 to LOA 50, 100 simulations each were calculated with image lines of varying degrees of noise ($\sigma_\mathbf{l} = 0.0$, $\sigma_\mathbf{l} = 0.5$, $\sigma_\mathbf{l} = 1.0$, $\sigma_\mathbf{l} = 1.5$ and $\sigma_\mathbf{l} = 3.0$ pixel). After the adjustment, the parameters of the exterior orientation are only implicitly available in the vector with the estimated parameters ($\hat{\boldsymbol{\beta}}$). $\hat{\boldsymbol{\beta}}$ relates to $\hat{\mathbf{p}}$ - the spherically normalized column vector containing the conditioned elements of $\hat{\mathrm{P}}$. In order to derive the exterior orientation parameters, $\hat{\mathbf{p}}$ is reshaped to the homogeneous 3x4 matrix $\hat{\mathrm{P}}$ and re-conditioned with (26). Decomposition results in the coordinates of the projection center (X^0) and the rotation matrix R (with the rotation angles ω, ϕ and κ). The detailed results are shown in table 2.

5. DISCUSSION

The results show that the exterior orientation of the camera is optimized, provided the reference model has sufficient geometric quality. In the presented example using synthetic data, the camera pose is improved as long as the reference model can be assigned to at least LOA 30 ($\sigma < 7.5\,mm$) and the standard deviation of the image points (for image line construction) for this camera is less than 3 pixels. While very good solutions are achieved with LOA 50 model accuracy, practically this demanding level of accuracy is very rarely achieved in real projects. In practical applications, however, it can be assumed that
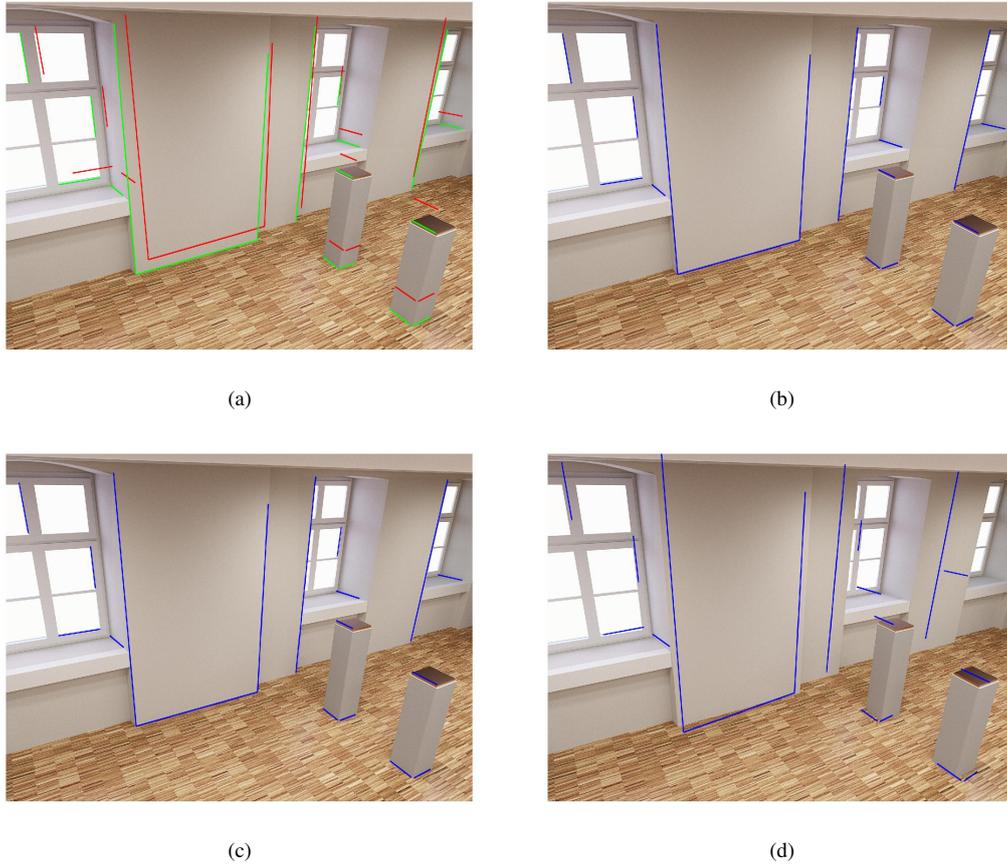
(a)

(b)

(c)

(d)

Figure 8. Synthetic image with projected 3D edges using a) $\tilde{P}$ (green) and P (red) b) $\hat{P}$, $\sigma_l = 1.0$ and LOA 50 model c) $\hat{P}$, $\sigma_l = 1.0$ and LOA 30 model d) $\hat{P}$, $\sigma_l = 1.0$ and LOA 10 model

LOA 50 ($\sigma = 0.5\,mm$)						
Orientation parameter	σ_{cam}	Empirical standard deviation (s) from n=100 simulations				
		σ_l=0.0 [px]	σ_l=0.5 [px]	σ_l=1.0 [px]	σ_l=1.5 [px]	σ_l=3.0 [px]
X_0 [m]	0.15	0.003	0.015	0.026	0.045	0.086
Y_0 [m]	0.15	0.002	0.010	0.020	0.030	0.072
Z_0 [m]	0.15	0.002	0.007	0.014	0.023	0.045
ω [°]	1	0.031	0.137	0.251	0.464	0.901
ϕ [°]	1	0.015	0.060	0.123	0.183	0.366
κ [°]	1	0.026	0.105	0.222	0.319	0.667
LOA 30 ($\sigma = 7.5\,mm$)						
Orientation parameter	σ_{cam}	Empirical standard deviation (s) from n=100 simulations				
		σ_l=0.0 [px]	σ_l=0.5 [px]	σ_l=1.0 [px]	σ_l=1.5 [px]	σ_l=3.0 [px]
X0 [m]	0.15	0.052	0.054	0.058	0.070	0.113
Y0 [m]	0.15	0.041	0.040	0.048	0.053	0.066
Z0 [m]	0.15	0.026	0.030	0.033	0.041	0.060
α [°]	1	0.398	0.473	0.571	0.588	0.980
β [°]	1	0.203	0.241	0.284	0.310	0.452
γ [°]	1	0.372	0.442	0.506	0.487	0.767
LOA 10 ($\sigma = 7.5\,cm$)						
Orientation parameter	σ_{cam}	Empirical standard deviation (s) from n=100 simulations				
		σ_l=0.0 [px]	σ_l=0.5 [px]	σ_l=1.0 [px]	σ_l=1.5 [px]	σ_l=3.0 [px]
X0 [m]	0.15	0.607	0.970	0.959	0.820	0.653
Y0 [m]	0.15	0.441	1.048	0.591	0.530	0.516
Z0 [m]	0.15	0.489	0.604	0.716	0.747	0.434
α [°]	1	5.353	6.636	4.863	5.087	5.138
β [°]	1	3.201	3.778	4.986	4.691	3.346
γ [°]	1	4.718	5.551	6.650	4.940	4.778

Table 2. Results on pose estimation with different image line quality and different LOA from 100 simulations each

essential interior elements such as doors, windows and walls are modeled at least LOA 30 compliant and are thus suitable as geometric references. In addition to the results shown, simulations were also calculated for LOD 20 and LOD 40, which also confirm this conclusion. While realtime matching is not our primary purpose, we focous on camera pose optimization. For maximum accuracy requirements (in the range of millimeters), relative rather than absolute accuracy could be used for parameter estimation. In this way, local changes would be precisely detectable in specially defined subsystems within the overall BIM. The *GeoPose* (OGC) standard will support this approach by its concept of frame transform chains for any real or digital object's pose. It encapsulates sufficient information to transform pose geometry in any other frame in its associated frame chain. The uncertainty in the geographic position can be integrated.

6. CONCLUSION AND OUTLOOK

We have shown that a coarse camera orientation can be improved to be accurate enough to reveal statements about relevant changes between the model and the actual state. BIM compliant indoor models with usual geometric quality are generally sufficient and useful as geodetic reference for change detection applications and construction progress documentation.

The presented estimation model will be used in further experiments in the test environment with real data from different camera sensors and varying geometric conditions (e.g. spatial distribution of line references). In addition, further tests regarding feature matching and pose estimation in other environments are planned to investigate whether difficult conditions such as high architectural symmetry could effect the results. In real-world use, it can be assumed that the BIM compliant reference model is faulty, regardless of the LOA definition. The as-planned state will likely not fully match the as-built state. Since outliers, miss-matches and deviations are to be detected in order to update the digital twin properly, a robust estimation has to be realized including an option for re-weighting observations. Additionally, vanishing point detection will be integrated to further support the process of camera orientation indoors.

Image based 3D reconstruction and change detection in the context of BIM is aimed. Therefore, the next step is to further validate the image measurement and reconstruction accuracy with optimized camera orientation. Also, the effect of the available image resolution on the method will be investigated.

In this context, also investigating state of the art sensors and technologies for visual localization within the interior of buildings for single images and image sequences in a visual-BIM-based SLAM approach is planned. These include the *Iphone 12 Pro* (Apple) with LiDAR sensor (if applicable with indoor localization infrastructures using beacons), the Intel *RealSense* depth camera including an IMU and stereo SLAM technology and Microsoft's *HoloLens2*.

ACKNOWLEDGMENT

References

Brunn, A., Meyer, T., 2016. Calibration of a Multi-Camera Rover. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B5, 445–452.

Förstner, W., Wrobel, B. P., 2016. *Photogrammetric Computer Vision*. 11, Springer International Publishing, Cham.

Han, K. K., Golparvar-Fard, M., 2015. Appearance-based material classification for monitoring of operation-level construction progress using 4D BIM and site photologs. *Automation in Construction*, 53, 44–57.

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

Hoegner, L., Tuttas, S., Xu, Y., Eder, K., Stilla, U., 2016. Evaluation of Methods for Coregistration and Fusion of RPAS-based 3d Point Clouds and thermal infratred Images. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 241–246.

Hofer, M., Maurer, M., Bischof, H., 2015. Line3D: efficient 3D Scene Abstraction for the Built Environment. *Computer Vision and Image Understanding (CVIU)*.

Iwaszczuk, D., Koch, T., Stilla, U., 2017. Innenraumrekonstruktion aus semantisch angereicherten 3D Punkten und Linien. *37. Wissenschaftlich-Technische Jahrestagung der DGPF in Würzburg*, 129–137.

Iwaszczuk, D., Stilla, U., 2017. Camera pose refinement by matching uncertain 3D building models with thermal infrared image sequences for high quality texture extraction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 132, 33–47.

Kropp, C., Koch, C., König, M., 2018. Interior construction state recognition with 4D BIM registered image sequences. *Automation in Construction*, 86, 11–32.

Li-Chee-Ming, J., Armenakis, C., 2014. Feasibility Study for Pose Estimation of Small UAS in Known 3D Environment Using Geometric Hashing. *Photogrammetric Engineering & Remote Sensing*, 80(12), 1117–1128.

Lowe, D. G., 1991. Fitting parameterized three-dimensional models to images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(5), 441–450.

Meidow, J., Förstner, W., Beder, C., 2009. Opimal Parameter Estimatin with Homogeneous Entities and Arbitrary Constraints. *Pattern recognition. Proceedings : 31st DAGM Symposium*, 292–301.

Tuttas, S., Braun, A., Borrmann, A., Stilla, U., 2017. Acquisition and Consecutive Registration of Photogrammetric Point Clouds for Construction Progress Monitoring Using a 4D BIM. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 85(1), 3–15.

USIBD, 2020. https://usibd.org/

Verykokou, S., Ioannidis, C., 2016. Exterior Orientation Estimation of oblique Aerial Imagery using Vanishing Points. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B3, 123–130.