

HYBRID ORIENTATION OF AIRBORNE LIDAR POINT CLOUDS AND AERIAL IMAGES

P. Glira^{a,b*}, N. Pfeifer^a, G. Mandlburger^{a,c}

^a TU Vienna, Department of Geodesy and Geoinformation, Vienna, Austria
(philipp.glira, norbert.pfeifer, gottfried.mandlburger)@geo.tuwien.ac.at

^b Austrian Institute of Technology (AIT), Vienna, Austria

^c University of Stuttgart, Institute for Photogrammetry, Stuttgart, Germany

KEY WORDS: orientation, calibration, strip adjustment, aerial triangulation, hybrid adjustment

ABSTRACT:

Airborne LiDAR (Light Detection And Ranging) and airborne photogrammetry are both proven and widely used techniques for the 3D topographic mapping of extended areas. Although both techniques are based on different reconstruction principles (polar measurement vs. ray triangulation), they ultimately serve the same purpose, the 3D reconstruction of the Earth's surface, natural objects or infrastructure. It is therefore obvious for many applications to integrate the data from both techniques to generate more accurate and complete results. Many works have been published on this topic of data fusion. However, no rigorous integrated solution exists for the first two steps that need to be carried out after data acquisition, namely (a) the lidar strip adjustment and (b) the aerial triangulation. A consequence of solving these two optimization problems independently can be large discrepancies (of up to several decimeters) between the lidar block and the image block. This is especially the case in challenging situations, e.g. corridor mapping with one strip only or in case few or no ground control data. To avoid this problem and thereby profit from many other advantages, a first rigorous integration of these two tasks, the hybrid orientation of lidar point clouds and aerial images, is presented in this work.

1. INTRODUCTION

The requirements for capturing topography are becoming more and more demanding, due to the need to provide data more efficiently, with shorter update cycles, at higher resolution and improved accuracy, and with more detail. This can be seen in different applications like precision agriculture (Mulla, 2013) and forestry (Holopainen et al., 2014), in urban planning (Xiao and Zhan, 2009), or in geomorphological analysis (Jaboyedoff et al., 2010).

Airborne acquisition covers a wide range of scales and extents, especially since the spreading of low flying unmanned vehicles. Given the nadir and optionally oblique viewing directions from airborne platforms makes them ideally suited for area-wide capturing for most of the topography, including natural as well as artificial objects upon the terrain. Due to the complimentary strengths of lidar and photos, more and more sensor combinations and integrated sensors, comprising a laser scanners and cameras became available (Toschi et al., 2018; Mandlburger et al., 2017).

Combining geo-data acquired by different sensors requires that they are in the same coordinate system. Within this article we concentrate especially on data of airborne laser scanners and aerial cameras, acquired from a common platform. A naive approach to achieve this geo-referencing is to process each data stream independently, using the same GNSS/INS (Global Navigation Satellite System/Inertial Navigation System) data or by using ground control measurements (e.g. points or planes) acquired in the same datum. In this article we suggest a more rigorous integration. It builds upon previous work in which orientation of airborne lidar strips with compensation of time dependent trajectory errors was presented (Glira et al., 2016) as well as approaches to integrate strip adjustment and bundle block adjustment by object space tie points (Mandlburger et al., 2015). Here, we go one step further and formulate a hybrid airborne lidar and photo orientation method in which a common trajectory for both sensors is

modeled, in addition to the previously suggested iteratively determined correspondences in object space. In addition to sufficient overlap in object space, the approach therefore requires exact recording of the photo's mid-exposure times.

Apart from presenting this rigorous model, we are also investigating, if local roughness is a feasible measure to select locations of correspondence between image and lidar.

1.1 Coordinate systems and notation

The coordinate systems (CS) used in this work are denoted by the following indices (cf. Figure 1):

- *s*: scanner CS
- *c*: camera CS
- *i*: INS CS, sometimes also referred to as body CS
- *n*: navigation CS (*x* = north, *y* = east, *z* = nadir)
- *e*: ECEF (earth-centered, earth-fixed) CS

Definitions of these coordinate systems can be found in Bäumker and Heimes (2001). A vector \mathbf{v} in the coordinate system with the index *a* is denoted by \mathbf{v}^a . A rotation matrix from the *a*-system to the *b*-system is denoted by R_a^b . In the hybrid adjustment, many different sensors and data types are used. To better differentiate an index within a specific group from the index of a coordinate system, we surround it with square brackets. For instance, the transformation of the *t*-th object point from the *n*-system to the *e*-system is denoted by $\mathbf{x}_{[t]}^e = R_n^e \mathbf{x}_{[t]}^n$.

2. HYBRID ORIENTATION METHOD

The aim of the hybrid adjustment is to simultaneously optimize the relative orientation and absolute orientation (georeference) of the lidar and image data. The sensor orientations are optimized by minimizing the discrepancies (a) within the overlap area of flight strips and/or images and (b) with respect to ground truth data, if available. The measurement process is thereby rigorously modelled using the original measurements of the sensors (i.e. scanner: polar measurements, camera: image coordinates) and the

*Corresponding author

flight trajectory of the aircraft. This way, systematic measurement errors can be corrected where they originally occur. Both, laser scanners and cameras, can be fully re-calibrated by estimating their interior calibration and mounting parameters (lever arm, boresight angles). Systematic measurement errors of the flight trajectory can be corrected individually for each flight strip. For highest accuracy demands or if a low-quality GNSS/INS navigation solution is used, time-dependent errors can be modelled by natural cubic splines. The methodological framework of the hybrid adjustment was adapted from the ICP algorithm (Besl and McKay, 1992). Consequently, correspondences are established iteratively and on a point basis to maintain the highest possible resolution level of the data. We present four different strategies for the selection of correspondences within the overlap area of point clouds. It was shown in Glira (2018) that the hybrid adjustment leads to many synergistic effects. The three major advantages are (a) the inherent optimization of the relative orientation between lidar and image data, (b) an increased block stability (avoiding block deformations, e.g. bending), and (c) an improved determinability of the parameters.

2.1 Mathematical foundation and parameter model

In this section, we describe the equations that form the core of the hybrid adjustment. These equations relate the measurements of the sensors on the aircraft (laser scanner(s), camera(s), GNSS, INS) to the observed object points on the ground (Figure 1). We will later use these equations in section 2.2 to establish the correspondences and thereby formulate the adjustment's observations. In the case of lidar point clouds, the relation between sensor measurements and ground points is given by the *direct georeferencing equation*. In the case of aerial images the relation is given by the *direct georeferencing equation* and the *collinearity equations*. For the correction of the common GNSS/INS flight trajectory we propose four different correction models.

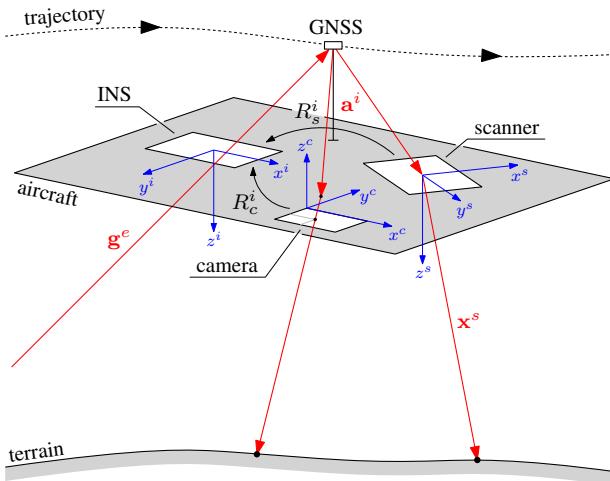


Figure 1: Schematic representation of a minimal set of sensors on an airborne platform.

Direct georeferencing of lidar point clouds The direct georeferencing equation is used to generate georeferenced point clouds from the measurements of a lidar multi-sensor system. This requires three types of input data (Hebel and Stilla, 2012; Skaloud and Lichten, 2006): (a) the polar measurements of the scanner, (b) the flight trajectory of the aircraft, and (c) the mounting calibration parameters. Combining all these measurements, the coordinates of an object point $[t]$, measured by a laser scanner $[l]$ at time t are given by

$$\mathbf{x}_{[t]}^e(t) = \mathbf{g}^e(t) + R_n^e(t) R_i^n(t) \left(\mathbf{a}_{[l]}^i + R_{s[l]}^i \mathbf{x}_{[t]}^s \right) \quad (1)$$

whereby

- $\mathbf{x}_{[t]}^s$ is a 3-by-1 vector with the coordinates of the laser point $[t]$ in the s -system. Generally, these coordinates can be expressed as a function of three polar elements, i.e. the range $\rho_{[t]}$ and the angles $\alpha_{[t]}$ and $\beta_{[t]}$:

$$\mathbf{x}_{[t]}^s(t) = \mathbf{x}_{[t]}^s(\rho_{[t]}, \alpha_{[t]}, \beta_{[t]}) \quad (2)$$

- $R_{s[l]}^i$ is a 3-by-3 rotation matrix describing the rotation from the s -system to the i -system, i.e. from the coordinate system of the laser scanner $[l]$ to the coordinate system of the INS. This rotation is usually denoted as *(boresight) misalignment* and is expressed through three Euler angles:

$$R_{s[l]}^i = R_{s[l]}^i(\alpha_{1[l]}, \alpha_{2[l]}, \alpha_{3[l]}) \quad (3)$$

- $\mathbf{a}_{[l]}^i$ is a 3-by-1 vector describing the positional offset between the phase centre of the GNSS antenna and the origin of the s -system. This vector is usually denoted as *lever-arm*:

$$\mathbf{a}_{[l]}^i = [a_{x[l]}^i \quad a_{y[l]}^i \quad a_{z[l]}^i]^T \quad (4)$$

Together, the lever-arm and the (boresight) misalignment angles form the *mounting calibration parameters*.

- $R_i^n(t)$ is a 3-by-3 rotation matrix describing the rotation from the i -system to the n -system, i.e. the local horizon system. This rotation constitutes the first (angular) part of the trajectory data. It can be estimated from the GNSS/INS measurements and is parametrized by three Euler angles roll ϕ , pitch θ , and yaw ψ :

$$R_i^n(t) = R_i^n(\phi(t), \theta(t), \psi(t)) \quad (5)$$

- $R_n^e(t)$ is a 3-by-3 rotation matrix describing the rotation from the n -system to the e -system. This rotation is not observed, but is a function of the longitude λ and latitude φ of the antenna's phase center position $\mathbf{g}^e(t)$:

$$R_n^e(t) = R_n^e(\lambda(t), \varphi(t)) \quad (6)$$

- $\mathbf{g}^e(t)$ is a 3-by-1 vector describing the position of the GNSS antenna in the e -system as second (translational) part of the trajectory data:

$$\mathbf{g}^e(t) = [g_x^e(t) \quad g_y^e(t) \quad g_z^e(t)]^T \quad (7)$$

In order to re-calibrate the laser scanner(s), equation (1) is extended by some calibration parameters. For this we formulate $\mathbf{x}_{[t]}^s$, according to equation (2), as a function of the polar elements $\rho_{[t]}$, $\alpha_{[t]}$, and $\beta_{[t]}$:

$$\mathbf{x}_{[t]}^s = \begin{bmatrix} \rho_{[t]} \cos \alpha_{[t]} \sin \beta_{[t]} \\ \rho_{[t]} \sin \alpha_{[t]} \\ \rho_{[t]} \cos \alpha_{[t]} \cos \beta_{[t]} \end{bmatrix}^s \quad (8)$$

For each polar coordinate two calibration parameters are introduced, an offset (bias) and a scale parameter. This yields to three offset parameters ($\Delta \rho_{[l]}$, $\Delta \alpha_{[l]}$, $\Delta \beta_{[l]}$) and three scale parameters ($\varepsilon_{\rho_{[l]}}$, $\varepsilon_{\alpha_{[l]}}$, $\varepsilon_{\beta_{[l]}}$) which are defined by

$$\rho_{[t]} = \Delta \rho_{[l]} + \rho_{0[t]} \cdot (1 + \varepsilon_{\rho_{[l]}}) \quad (9)$$

$$\alpha_{[t]} = \Delta \alpha_{[l]} + \alpha_{0[t]} \cdot (1 + \varepsilon_{\alpha_{[l]}}) \quad (10)$$

$$\beta_{[t]} = \Delta \beta_{[l]} + \beta_{0[t]} \cdot (1 + \varepsilon_{\beta_{[l]}}) \quad (11)$$

where the original scanner's measurements are denoted by $\rho_{0[t]}$, $\alpha_{0[t]}$, and $\beta_{0[t]}$. Usually only a subset of these six parameters is estimated by adjustment. The parameter selection mainly depends on the construction type of the scanner and (due to correlations) on the chosen trajectory correction model.

Direct georeferencing of aerial images The exterior orientation of an image is defined by the position of the projection center of the camera and the rotation of the image with respect to the object coordinate system. Under the assumption that the exposure time t of an image is known, the exterior orientation of an image can be directly derived from (a) the flight trajectory of the aircraft and (b) the mounting calibration parameters. We denote this type

of images as *coupled images*, as their exterior orientation is coupled to the flight trajectory. For a coupled image $[i]$ captured at time t , the projection center and the rotation matrix are given by:

$$\mathbf{x}_{0[i]}^e(t) = \mathbf{g}^e(t) + R_n^e(t) R_i^n(t) \mathbf{a}_{[c]}^i \quad (12)$$

$$R_{c[i]}^e(t) = R_n^e(t) R_i^n(t) R_{c[c]}^i \quad (13)$$

In addition to the entities already introduced in equation (1), we further specify:

- $\mathbf{x}_{0[i]}^e(t)$ is a 3-by-1 vector with the coordinates of the projection center of the image $[i]$ in the e -system:

$$\mathbf{x}_{0[i]}^e(t) = [X_{0[i]}^e(t) \quad Y_{0[i]}^e(t) \quad Z_{0[i]}^e(t)]^T \quad (14)$$

- $\mathbf{a}_{[c]}^i$ is a 3-by-1 vector describing the positional offset between the GNSS antenna and the projection center of the camera. This vector is denoted as *lever-arm*:

$$\mathbf{a}_{[c]}^i = [a_{x[c]}^i \quad a_{y[c]}^i \quad a_{z[c]}^i]^T \quad (15)$$

- $R_{c[i]}^e$ is a 3-by-3 rotation matrix describing the rotation from the c -system to the e -system. Thus, this rotation matrix describes the three-dimensional rotation, or attitude, of the camera with respect to the object coordinate system, defined as e -system in this work. It is parametrized through three Euler angles $\omega_{[i]}, \varphi_{[i]}, \kappa_{[i]}$:

$$R_{c[i]}^e = R_{c[i]}^e(\omega_{[i]}, \varphi_{[i]}, \kappa_{[i]}) \quad (16)$$

- $R_{c[c]}^i$ is a 3-by-3 rotation matrix describing the rotation from the c -system to the i -system, i.e. from the camera to the INS. In analogy to the laser scanner case, this rotation is denoted as (*bore sight*) misalignment and is parametrized through three Euler angles:

$$R_{c[c]}^i = R_{c[c]}^i(\beta_{1[c]}, \beta_{2[c]}, \beta_{3[c]}) \quad (17)$$

In the hybrid adjustment the mounting calibration parameters of the camera $R_{c[c]}^i$ and $\mathbf{a}_{[c]}^i$ are estimated. Through these parameters, the images are tightly coupled to the GNSS/INS trajectory. However, if the image residuals show systematic patterns, more flexibility for the exterior orientation of each image might be necessary. For this, additional correction parameters for the exterior orientation of an image are introduced, i.e. three correction parameters $\Delta X_{0[i]}^e, \Delta Y_{0[i]}^e, \Delta Z_{0[i]}^e$ for the position of the projection center and three correction parameters $\Delta \omega_{[i]}, \Delta \varphi_{[i]}, \Delta \kappa_{[i]}$ for the rotation of an image. These parameters *must* be observed through fictional observations to honor their zero expectation and to keep the coupling to the trajectory intact.

In practice, the time stamps t of the images are often unknown or not sufficiently accurate, or no GNSS/INS trajectory is available, e.g. if the imagery was collected independently from the lidar point clouds. In such cases, the direct georeferencing equation can not be used. Instead, the six elements of the exterior orientation of the images – that is $X_{0[i]}^e, Y_{0[i]}^e, Z_{0[i]}^e, \omega_{[i]}, \varphi_{[i]}, \kappa_{[i]}$ – can directly be estimated by adjustment. We denote these type of images in the following as *loose images*.

Collinearity equations The collinearity equations relate the 2D image coordinates with the 3D object coordinates of a single point. They can be written for an object point $[t]$, which was observed in an image $[i]$ taken by a camera $[c]$ as:

$$\begin{aligned} \bar{x}_{[i][t]}^c &= x_{0[c]}^c - c_{[c]}^c \frac{r_{11}(X_{[t]}^e - X_{0[i]}^e) + r_{21}(Y_{[t]}^e - Y_{0[i]}^e) + r_{31}(Z_{[t]}^e - Z_{0[i]}^e)}{r_{13}(X_{[t]}^e - X_{0[i]}^e) + r_{23}(Y_{[t]}^e - Y_{0[i]}^e) + r_{33}(Z_{[t]}^e - Z_{0[i]}^e)} \\ \bar{y}_{[i][t]}^c &= y_{0[c]}^c - c_{[c]}^c \frac{r_{12}(X_{[t]}^e - X_{0[i]}^e) + r_{22}(Y_{[t]}^e - Y_{0[i]}^e) + r_{32}(Z_{[t]}^e - Z_{0[i]}^e)}{r_{13}(X_{[t]}^e - X_{0[i]}^e) + r_{23}(Y_{[t]}^e - Y_{0[i]}^e) + r_{33}(Z_{[t]}^e - Z_{0[i]}^e)} \end{aligned}$$

whereby

- $\bar{x}_{[i][t]}^c, \bar{y}_{[i][t]}^c$ are the undistorted image coordinates of object point $[t]$ in image $[i]$
- $x_{0[c]}^c, y_{0[c]}^c$ are the coordinates of the principal point of camera $[c]$

- $c_{[c]}^c$ is the principal distance of camera $[c]$
- $X_{0[i]}^e, Y_{0[i]}^e, Z_{0[i]}^e$ are the coordinates of the projection center of image $[i]$
- r_{ij} are the elements of the rotation matrix $R_{c[i]}^e$
- $X_{[t]}^e, Y_{[t]}^e, Z_{[t]}^e$ are the coordinates of the object point $[t]$

In most of the cases it is necessary to extend the collinearity equations with additional parameters, e.g. to deal with distorted imagery or to account for distortions due to cartographic projections (Kraus, 1997, p. 280). In this work, we extend the collinearity equations by *image distortion parameters* according to the model introduced in Brown (1971). Thereby, a common approach is to use three radial distortion coefficients $K'_{n[c]}$ and two tangential distortion coefficients $P'_{n[c]}$.

Trajectory correction parameters The trajectory of the aircraft is assumed to be estimated in advance by the integration of GNSS and INS measurements in a Kalman filter (Kalman, 1960). As a result, the original position and orientation estimates are given, together with their precision, as a function of the flight time t (cf. equations (5) and (7)). The original flight trajectory forms the basis for the direct georeference of lidar strips (equation (1)) and aerial images (equation (12)). However, Skaloud et al. (2010) pointed out that GNSS and INS measurements are strongly affected by external influences (e.g. satellite constellation, flight maneuvers) and consequently their accuracy can not be assumed constant over time. This, in turn, leads to *time-dependent errors* of the estimated trajectory, which should be corrected by adjustment.

For coping with high accuracy expectations, we proposed a flexible Spline trajectory correction model (TCM) in Glira et al. (2016). It uses natural cubic splines with a constant segment length Δt in time domain as correction functions. Depending on the accuracy demand, less complex TCM models (e.g., bias, linear, or quadratic) to compensate systematic errors of the original flight, also detailed in Glira et al. (2016), might be suitable.

Table 1 gives a summary of the parameters that *can* be estimated by adjustment. It should be noted that depending on the assembly of the sensors, the flight configuration, and the terrain geometry, some of these parameters may be completely correlated and therefore not estimable. Besides that, *model overfitting* should be avoided, i.e. only the parameters that are needed for modeling systematic errors should be estimated.

2.2 Correspondences

We have discussed the adjustment's parameter model in the previous section. To estimate these parameters and simultaneously improve the georeference of the lidar strips and the aerial images, various types of correspondences are used. These correspondences are established between the following input data types:

- **lidar strips (STR):** given by the measurements of the scanner, the trajectory of the aircraft, and priors (approximate values) for the mounting calibration.
- **aerial images (IMG):** tie point observations of images either coupled to the trajectory by a time stamp and the mounting calibration (*coupled images*), or with priors for the exterior orientation (*loose images*).
- **control point clouds (CPC):** datum-defining point clouds with known coordinates in object space (e -system), e.g. point clouds from terrestrial laser scanning (TLS), a digital elevation model (DEM) (as point cloud) from an earlier flight campaign, or single (widely isolated) points from total station or GNSS measurements. These points are matched with the lidar strips only and consequently do not need to be identifiable in the images.

Parameters			
category	name		#
laser scanner	misalignment angles	$\alpha_{1[l]}, \alpha_{2[l]}, \alpha_{3[l]}$	$3L$
	lever-arm components	$a_x^i[l], a_y^i[l], a_z^i[l]$	$3L$
cameras	range offset (bias)	$\Delta\rho_{[l]}$	L
	range scale	$\varepsilon_{\rho[l]}$	L
images	angle offsets (biases)	$\Delta\alpha_{[l]}, \Delta\beta_{[l]}$	$2L$
	angle scales	$\varepsilon_{\alpha[l]}, \varepsilon_{\beta[l]}$	$2L$
cameras	misalignment angles	$\alpha_{1[c]}, \alpha_{2[c]}, \alpha_{3[c]}$	$3C$
	lever-arm components	$a_x^i[c], a_y^i[c], a_z^i[c]$	$3C$
interior orientation	2D coordinates of principal point	$x_{0[c]}^c, y_{0[c]}^c$	$2C$
	principal distance	$c_{[c]}^c$	C
coupled images	image distortion parameters	$K'_{n[c]}, P'_{n[c]}$	$C(n_r + n_t)$
	correction of 3D coordinates of projection centers	$\Delta X_{0[i]}^e, \Delta Y_{0[i]}^e, \Delta Z_{0[i]}^e$	$3I$
	correction of rotation angles	$\Delta\omega_{[i]}, \Delta\varphi_{[i]}, \Delta\kappa_{[i]}$	$3I$
	3D coordinates of projection centers	$X_{0[i]}^e, Y_{0[i]}^e, Z_{0[i]}^e$	$3I$
loose images	rotation angles	$\omega_{[i]}, \varphi_{[i]}, \kappa_{[i]}$	$3I$
	3D coordinates of image tie points	$X_{[t]}^e, Y_{[t]}^e, Z_{[t]}^e$	$3T$
trajectory	x correction coefficients	$\Delta g_x^e[s]$	S
	y correction coefficients	$\Delta g_y^e[s]$	S
	z correction coefficients	$\Delta g_z^e[s]$	S
	roll correction coefficients	$\Delta\phi_{[s]}$	S
rotation	pitch correction coefficients	$\Delta\theta_{[s]}$	S
	yaw correction coefficients	$\Delta\psi_{[s]}$	S
	datum	datum correction parameters	$\Delta g_x^e, \Delta g_y^e, \Delta g_z^e$
			3

for $s \in \mathcal{S} = \{1, \dots, s, \dots, S\}$, $l \in \mathcal{L} = \{1, \dots, l, \dots, L\}$, $c \in \mathcal{C} = \{1, \dots, c, \dots, C\}$, $i \in \mathcal{I} = \{1, \dots, i, \dots, I\}$, $t \in \mathcal{T} = \{1, \dots, t, \dots, T\}$

Table 1: Overview of the parameters estimated by adjustment. S = no. of strips, L = no. of laser scanners, C = no. of cameras, I = no. of images, n_r = no. of radial distortion coefficients, n_t = no. of tangential distortion coefficients, T = no. of image tie points. The number of trajectory correction parameters is given for the *Bias Trajectory Correction Model*.

- **ground control points (GCP):** datum-defining single points with known coordinates in object space (e -system), e.g. measured by GNSS or total stations, and image space (c -system). These points are matched with the images only, i.e. it is not tried to identify them in the lidar point cloud, e.g. on the basis of intensity values.

Five different correspondence types can be established between these data inputs; a summary is given in Figure 2 and Table 2. As can be seen, two correspondence types can be associated to the strip adjustment of lidar strips and the bundle adjustment of aerial images respectively, whereas the fifth type is newly introduced in this work to establish a link between the laser scans and the aerial images. The various correspondences serve to define the observations used for parameter estimation in the hybrid adjustment.

Correspondences		
type	correspondences between	where
STR-to-STR	lidar strip & lidar strip	object space
CPC-to-STR	control point cloud & lidar strip	object space
IMG-to-IMG	image points & image points	image space
IMG-to-GCP	image tie points & GCPs	object space
IMG-to-STR	image points & lidar strip	object space

Table 2: Correspondences established in the hybrid adjustment.

As in the ICP algorithm (Besl and McKay, 1992), correspondences in object space are established on a point basis. This is advantageous as (i) the highest possible resolution level of the data is exploited, (ii) no time-consuming pre-processing of the

data is required (in contrast to correspondences which are found by segmentation and/or interpolation), and (iii) no restrictions are imposed on the object space (e.g. the presence of rooftops or horizontal fields). As described in the previous section, each point is calculated by applying the direct georeferencing equations and the collinearity equations extended by additional parameters. The STR-to-STR, CPC-to-STR, and IMG-to-STR correspondences are established in three distinct steps (cf. Rusinkiewicz and Levoy (2001)): the *selection*, the *matching*, and the *rejection* step.

Selection of correspondences Due to the high number of points in lidar, it is simply not possible to use *each* point as correspondence within the overlap volume of two point clouds. Thus, a subset of points needs to be selected. For this we introduced four different selection strategies in Glira et al. (2015). Sorted by increasing computational complexity, these are: (i) *Random Sampling (RS)*, (ii) *Uniform Sampling (US)*, (iii) *Normal Space Sampling (NSS)*, and (iv) *Maximum Leverage Sampling (MLS)*. While RS constitutes the simplest method, US provides a homogeneous distribution of correspondences. NSS, in turn, considers the fact that some parameters of the sensor model can only be estimated based if sufficient tilted surface with arbitrary orientation are available. The correspondences are, thus, selected in angular space rather than in the spatial domain. The MLS model, finally, constitutes the most complex strategy selecting those points, which are best suited for the estimation of the parameters. For this, the effect of each point on the parameter estimation, i.e. its leverage, is considered. The points with the maximum leverage (= the lowest redundancy) are selected. More details on the correspondence

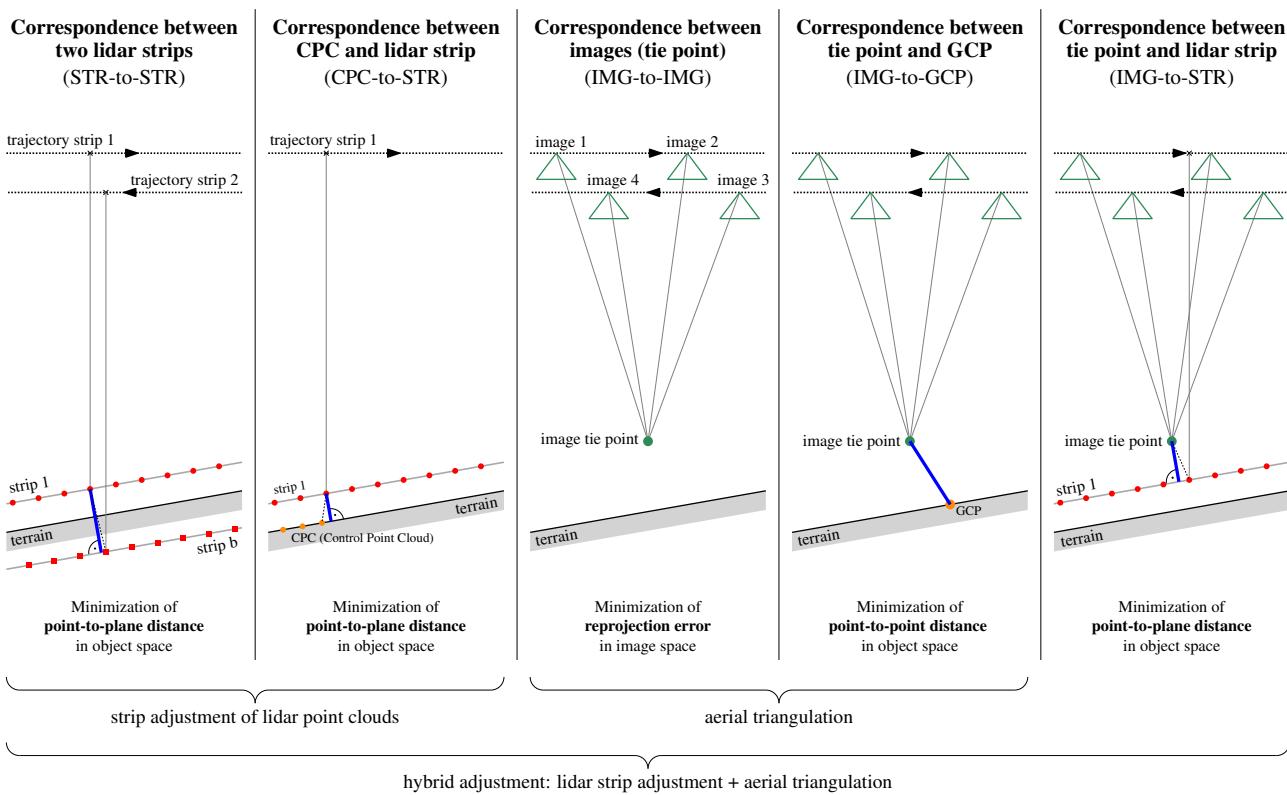


Figure 2: Types of correspondences used in the hybrid adjustment.

selection are provide in Glira et al. (2015).

Matching of correspondences In this step the correspondences are established, i.e. each point previously selected by one (or a combination) of the selection strategies is paired to one point in the overlapping point cloud. The simplest strategy is to match the selected points to their closest points (*nearest neighbours*) as proposed by Besl and McKay (1992). We found that for lidar and image data this is an adequate choice, mainly due to the good initial relative orientation and the high point density of lidar strips. The search for closest points can be realized efficiently using k-d trees.

Rejection of correspondences The aim of this step is the a priori detection and rejection of false correspondences (outliers), as they may have a large effect on the result of the adjustment. The proposed correspondence rejection criteria are: (i) Rejection based on the reliability of the normal vectors of corresponding points, (ii) Rejection based on the angle α between the normal vectors of corresponding points, (iii) Rejection based on the distance between corresponding points, and (iv) Rejection of correspondences in non-stable areas. Details can be found in Glira (2018). As it is not guaranteed that all outliers in the observation data are rejected a priori with these strategies, a robust adjustment method is highly recommended for the detection and removal of the remaining ones.

Error metric The error metric defines which type of distance is minimized between two corresponding points. In the hybrid adjustment two types of error metrics are used:

- **Point-to-point error metric** This error metric minimizes the Euclidean (unsigned) distance between corresponding points. It is defined as

$$d_{s[p]} = \|\mathbf{p}_{[p]} - \mathbf{q}_{[p]}\| \quad (18)$$

where $\mathbf{p}_{[p]}$ and $\mathbf{q}_{[p]}$ are the corresponding points of the $[p]$ -th correspondence. This error metric is used for the IMG-to-GCP

correspondences only, as only these are real point-to-point correspondences in object space.

- **Point-to-plane error metric** This error metric minimizes the perpendicular (signed) distance of one point to the tangent plane of the other point. It is defined as

$$\mathbf{d}_{p[p]} = (\mathbf{p}_{[p]} - \mathbf{q}_{[p]})^T \mathbf{n}_{[p]} \quad (19)$$

where $\mathbf{n}_{[p]}$ is the normal vector associated to the point $\mathbf{p}_{[p]}$ (with $\|\mathbf{n}_{[p]}\| = 1$). In contrast to the point-to-point error metric, it is not necessary that the corresponding points are identical in object space. Thus, it is suitable for matching point clouds with a different ground sampling. The only requirement is that the corresponding points belong to the same plane in object space (e.g. roof or street). This error metric is characterized by a high convergence speed, as flat regions can slide along each other without costs. It is used for the STR-to-STR, CPC-to-STR, and IMG-to-STR correspondences.

IMG-to-STR correspondences One of the most important components in the hybrid adjustment are the correspondences established between image tie points and lidar strips (IMG-to-STR), cf. Figure 2. When integrating lidar and image tie point measurements, the specific characteristics of both measurement techniques must be considered. Even though both techniques ultimately serve the mapping of the Earth's surface, they have a slightly different view on it. Lidar is an active measurement system operating at a single wavelength (mono-spectral), usually in the visible or near infrared range. It relies on the diffuse backscattering of the emitted laser pulse. In general, lidar provides a rather uniform sampling of the Earth's surface. In contrast, photogrammetry is a passive measurement system capturing the scattered solar radiation in the optical spectrum. The photogrammetric reconstruction process relies on sufficient texture. Consequently, corners and edges can be well reconstructed, whereas the accuracy decreases in areas with low texture or a bad signal-to-noise ratio, e.g. shadowed areas. The main geometrical differences between the two techniques stem from the ability of the lidar pulse

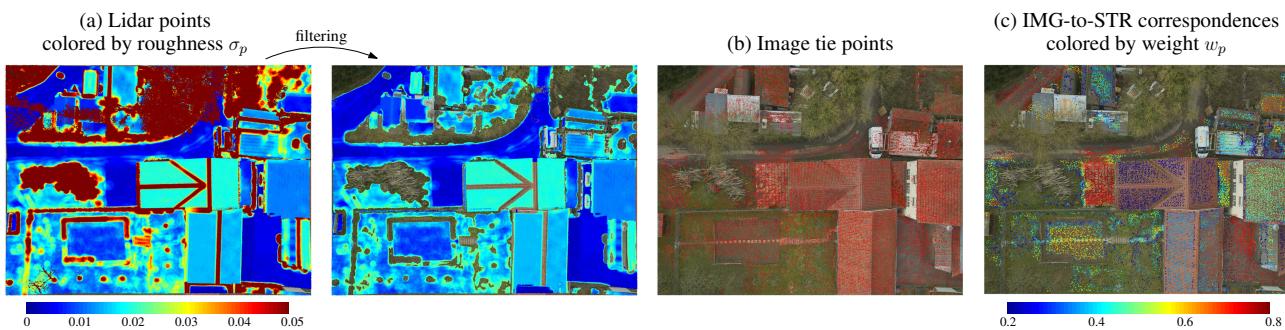


Figure 3: Correspondences between image tie points and lidar strips (IMG-to-STR) in image space of a single image.

to penetrate small-scale structures (e.g. vegetation), whereas the photogrammetric reconstruction leads mostly to points from the topmost surface (Mandlburger et al., 2017). For instance, in case of a grass field, the laser penetrates the grass layer to a certain extent, whereas the triangulated image points describe the top of the grass layer. Summarizing, correspondences between image tie points and lidar strips should only be established in areas in which the scanner and the cameras have the same view on the Earth's surface.

In this work, we have chosen a rather simple approach by limiting the correspondences to smooth and textured areas. In Figure 3 a set of IMG-to-STR correspondences is shown in image space of a single image. The depicted scene includes relatively flat areas (roads, roofs, facades, bare soil) and rough areas (low and high vegetation). In the first step a subset of image tie points can be selected by one of the selection strategies presented above. However, if feasible (e.g. in terms of computer memory), this step can be skipped so that *all* tie points are used for matching, as we did in this example. The correspondences are then established by matching the image tie points with the nearest neighbour in the lidar point clouds (*matching* step). Finally, potentially wrong correspondences (outliers) are rejected by the criteria described above (*rejection* step). We rejected in this example all correspondences with an estimated roughness $\sigma_p \geq 0.02$ m and with point-to-plane distances outside the range $d \pm 3\sigma_{\text{mad}}$. As can be seen, the resulting correspondences are predominantly in smooth areas, e.g. on streets, terraces, roofs, and bare soil.

Stochastic model All correspondences (i.e. observations) have a certain accuracy, which is considered in the *stochastic model*. With the nominal lidar point measurement precision σ_t , the STR-to-STR correspondences have weight $w_{dp} = 1/(2\sigma_t^2)$, whereas CPC-to-STR have $1/\sigma_t^2$, i.e. a control point cloud is assumed to be free of errors. IMG-to-IMG correspondences are built in image space and consequently their weight depends on the image measurement precision σ_c and is $1/\sigma_c^2$. The ground control points are introduced with their accuracy, which can be different in the coordinate directions. Finally, the IMG-to-STR correspondences discussed previously are introduced with weight $w_{dp} = 1/(\sigma_t^2 + \sigma_c^2) \cdot w_p$, where σ_t is an estimate for the tie point precision in object space. Here, an additional, empirically motivated weight ($0 \leq w_p \leq 1$) based on the roughness measure of each correspondence, is used - it is defined by $w_p = 1 - \sigma_p/\sigma_{p,\max}$, whereby $\sigma_{p,\max}$ is the maximum allowed roughness measure.

2.3 Workflow

A simplified flowchart of the hybrid adjustment is shown in Figure 4. For the sake of simplicity, we thereby omitted all I/O-steps, i.e. the import and export of the data. The workflow can be divided into three stages: the *pre-processing stage*, the *main iteration loop*, and the *post-processing stage*.

The *pre-processing stage* includes three image-based steps. The two objectives of this stage are to establish the IMG-to-IMG correspondences and to get a first estimate of the 3D coordinates of the image tie points. These coordinates are estimated in a pure aerial triangulation (i.e. without consideration of lidar strips and CPCs and are used for the IMG-to-STR matching in object space later on. The *main iteration loop* starts with the direct georeferencing of the lidar strips. Thereby the *current* parameters are used, i.e. the priors of the parameters in the first iteration and the parameters estimated by the hybrid adjustment for all subsequent iterations. Then, as in the ICP algorithm, the correspondences in object space (STR-to-STR, CPC-to-STR, IMG-to-STR) are newly established in each iteration. After the hybrid adjustment a convergence criteria is tested, e.g. the relative change of the weighted sum of squared errors. If it is not met, a new iteration starts. Usually, due to the high convergence speed of the point-to-plane error metric only 3 to 5 iterations are needed until convergence is reached. Finally, the lidar strips are georeferenced with the last parameter estimates in the *post-processing stage*.

2.4 Experimental results

The potential of the hybrid adjustment is demonstrated on the basis of high-resolution data captured by UAVs in the course of a research and development project initiated by the German Federal Institute of Hydrology (BfG) in Koblenz in partnership with the Office of Development of Neckar River Heidelberg (ANH). The study area is located in Hessigheim, Germany, (48°59'67"N, 9°11'20"E; WGS 84), cf. Figure 5. The lidar data was acquired from a RIEGL RiCopter UAV with the RIEGL VUX-1LR scanner. As GNSS-inertial solution an Applanix APX-20 board was used. The accuracy of the post-processed flight trajectory is 2–5 cm for the 3D positions, 0.015° for the roll and pitch angles, and 0.035° for the yaw angle. The data was collected in a flying height of about 40 m above ground and a flying speed of about 8 m/s. Oblique imagery with an average GSD of 20 mm was concurrently captured with two Sony Alpha 6000 cameras (*cam1/cam2*) mounted on the same UAV octocopter UAV platform. As precise time stamps are available for these images, they are handled as *coupled images* in the adjustment.

High-resolution nadir images with an average GSD of 4 mm and 80/60 overlap were acquired in a second flight campaign with the CopterSystems CS-SQ8 copter and a PhaseOne iXU-RS 1000 camera (*cam3*). Due to inaccurate time stamps, these images are introduced as *loose images* into the adjustment. Ground truth data was measured by a combination of GNSS static baseline, tacheometry measurements and precise levelling. The accuracy of the thereby measured points is between 2 and 4 mm. As control point clouds (CPC) gable roof shaped structures, fixed on tripods, are used. A dense point cloud was derived from the observed corner points of these structures. Checkerboard targets with a diameter of 27 cm are used as ground truth data for the aerial images. These points serve as ground control points (GCP)

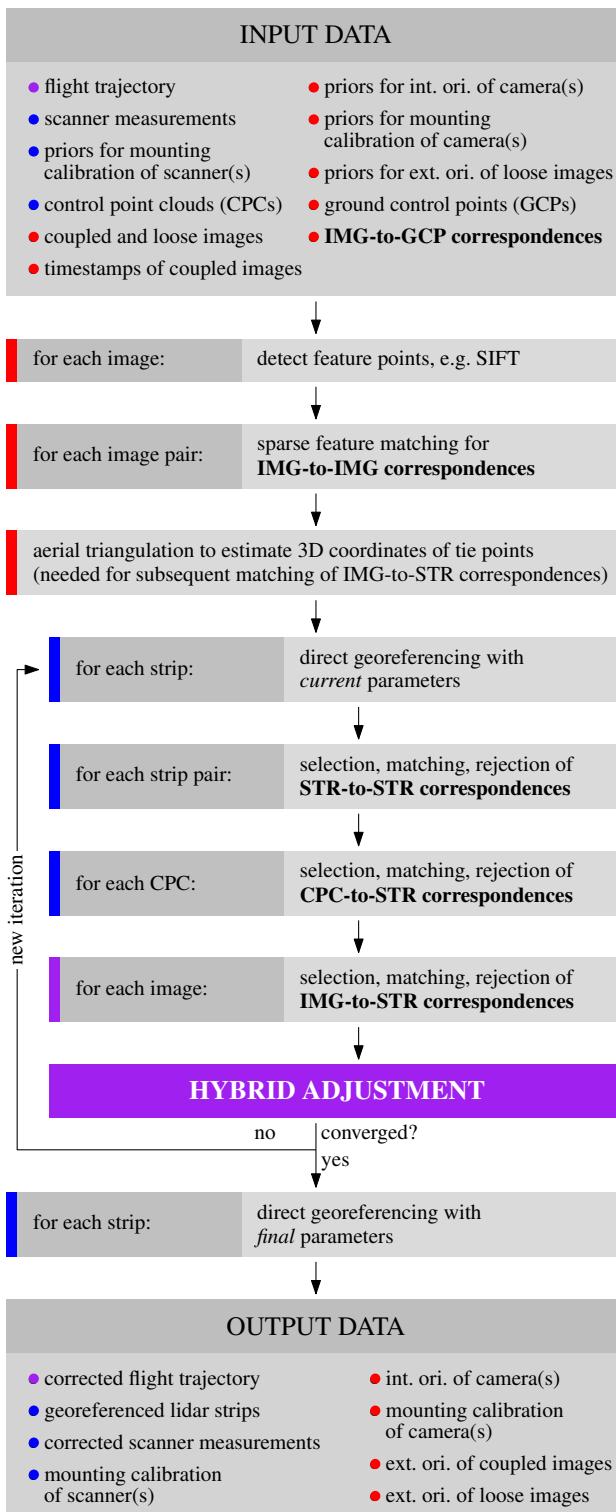


Figure 4: Flowchart of the hybrid adjustment method. **Blue:** lidar related steps/data. **Red:** image related steps/data. **Purple:** lidar and image related steps/data.

and as check points (CP). In sum, 4 longitudinal lidar strips, 506 oblique images (coupled images), and 76 nadir images (loose images) have been chosen. It is noted, that preliminary results from the same flight campaign have been published by Cramer et al. (2018).

The results of the adjustment are summarized in Table 3. They confirm the appropriateness of the a priori stochastic model and the functional model of the adjustment. A comprehensive discus-

correspondences	unit	n	mean	σ
STR-to-STR	[m]	35765	0.000	0.009
CPC-to-STR	[m]	769	0.000	0.004
IMG-to-IMG x/y (cam1)	[px]	93456	0.00/0.00	0.35/0.35
IMG-to-IMG x/y (cam2)	[px]	71549	0.00/0.00	0.36/0.34
IMG-to-IMG x/y (cam3)	[px]	18689	0.00/0.00	0.43/0.35
IMG-to-STR	[m]	15155	0.000	0.006
GCPs and CPs			mean	min max
GCP x/y	[m]	3	0.003/0.001	-0.002/-0.003 0.008/0.004
GCP z	[m]	3	-0.013	-0.021 0.007
CP x/y	[m]	2	-0.002/0.004	-0.006/0.002 0.003/0.006
CP z	[m]	2	0.018	0.005 0.031

Table 3: Residuals of the hybrid adjustment.

sion of the results can be found in Glira (2018).

Finally, we would like to point to one of the most important benefits of the hybrid adjustment. In Figure 6 the estimated roll angle correction of a single strip is shown (i) using the lidar data only (red), and (ii) using lidar *and* image data simultaneously in the hybrid adjustment (blue). Using the lidar data only, the spline correction function oscillates to a relatively high degree. These oscillations typically occur in case of overfitting, i.e. when the spline function is too flexible and consequently tends to overcompensate the real trajectory errors. However, by additionally considering the (coupled) images in the adjustment, the spline function needs to satisfy additional geometric constraints and in consequence unmotivated oscillations are strongly mitigated. This leads to a highly increased overall block stability, making local and global deformations (e.g. a bending of the whole block) more unlikely.

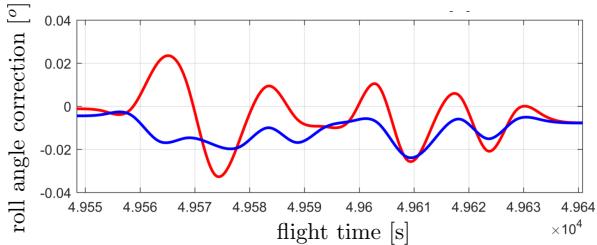


Figure 6: Comparison of estimated spline functions for the correction of the roll angle of a single flight strip between lidar strip adjustment (red) and hybrid adjustment (blue).

3. SUMMARY AND CONCLUSIONS

This paper gave an overview on the hybrid adjustment of airborne lidar strips and images. It is rigorous in the sense that it ties the two sensors together at the trajectory and at the ground. It is also rigorous in the sense, that the functional model is, according to our experiences, complete and for most parts very close to the physical realization of the measurement system. Concerning the strip adjustment part, the unknowns are restricted to the calibration parameters of the laser scanner and the laser scanner's mounting. The trajectory correction models are driven by the deficiencies of the provided trajectories and more empirical.

The strip to strip correspondences are restricted to pairwise correspondences. This is justified by the typically low number of overlaps between laser strips, especially in comparison to images, and, more importantly, by the lack of exact point correspondences, as they are found between homologous image points (or rather features). The tie points of bundle block adjustment are the dominating number of unknowns, augmented by the interior orientation and mounting parameters of the cameras (or interior and exterior orientation in the case of “loose” images).

The stochastic model is, in comparison, quite simple. A special challenge are the strip to image tie point correspondences. Given

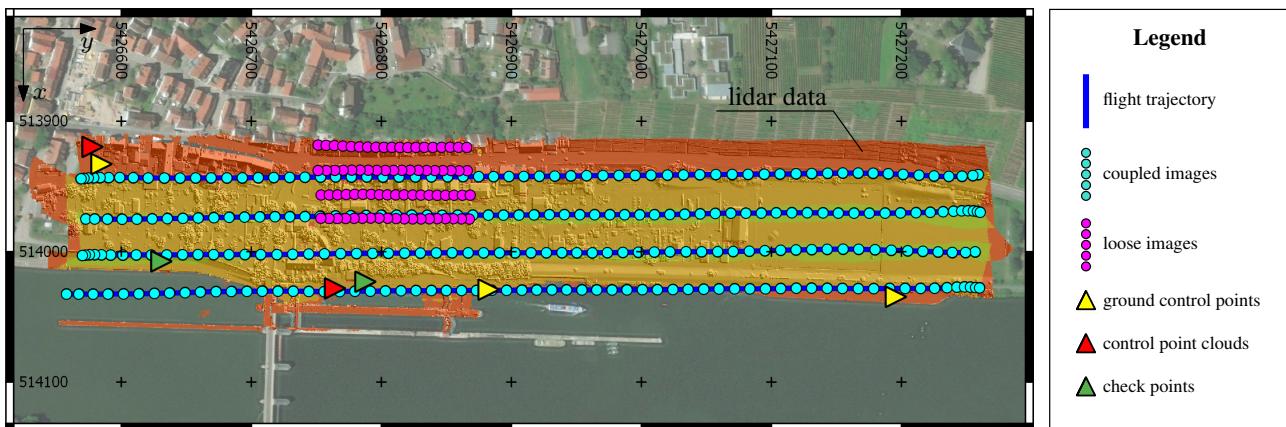


Figure 5: Data overview. The area covered by the lidar data is colored according to the strip overlap.

the very high precision of a few centimeter, for both, laser scanning and reconstruction from photos, it is not yet obvious, how to select only appropriate correspondences and which precision to assign to them. The suggested approach, including selection and weighting of correspondences, proofed feasible due to the results of the hybrid adjustment. Improving the overall handling of strip to tie point correspondence and the associated stochastic model in particular, appears to be the most important research in the domain of hybrid orientation of airborne lidar point clouds and aerial images.

References

- Bäumker, M. and Heimes, F., 2001. New calibration and computing method for direct georeferencing of image and scanner data using the position and angular data of an hybrid inertial navigation system. In: OEEPE Workshop, "Integrated Sensor Orientation", Hannover, Germany.
- Besl, P. J. and McKay, N. D., 1992. Method for registration of 3-d shapes. In: Robotics-DL tentative, International Society for Optics and Photonics, pp. 586–606.
- Brown, D. C., 1971. Close-range camera calibration. Photogrammetric Engineering 37(8), pp. 855–866.
- Cramer, M., Haala, N., Laupheimer, D., Mandlburger, G. and Havel, P., 2018. Ultra-high precision uav-based lidar and dense image matching. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-1, pp. 115–120.
- Glira, P., 2018. Hybrid Orientation of LiDAR Point Clouds and Aerial Images. PhD thesis, TU Wien.
- Glira, P., Pfeifer, N. and Mandlburger, G., 2016. Rigorous strip adjustment of UAV-based laserscanning data including time-dependent correction of trajectory errors. Photogrammetric Engineering & Remote Sensing 82(12), pp. 945–954.
- Glira, P., Pfeifer, N., Briese, C. and Ressl, C., 2015. A correspondence framework for ALS strip adjustments based on variants of the ICP algorithm. PFG Photogrammetrie, Fernerkundung, Geoinformation 2015(4), pp. 275–289.
- Hebel, M. and Stilla, U., 2012. Simultaneous calibration of ALS systems and alignment of multiview LiDAR scans of urban areas. Geoscience and Remote Sensing, IEEE Transactions on 50(6), pp. 2364–2379.
- Holopainen, M., Vastaranta, M. and Hyppäpä, J., 2014. Outlook for the next generations precision forestry in finland. Forests 5(7), pp. 1682–1694.
- Jaboyedoff, M., Oppikofer, T., Abellán, A., Derron, M.-H., Loyer, A., Metzger, R. and Pedrazzini, A., 2010. Use of lidar in landslide investigations: a review. Natural Hazards 61(1), pp. 528.
- Kalman, R. E., 1960. A new approach to linear filtering and prediction problems. Journal of Fluids Engineering.
- Kraus, K., 1997. Photogrammetry, Vol.2, Advanced Methods and Applications. Duemmler / Bonn.
- Mandlburger, G., Hollaus, M., Glira, P., Wieser, M., Milenkovic, M., Riegl, U. and Pfennigbauer, M., 2015. First examples from the RIEGL VUX-SYS for forestry applications. In: Proc. SilviLaser 2015; 28–30 Sep. 2015, La Grande Motte, France.
- Mandlburger, G., Wenzel, K., Spitzer, A., Haala, N., Glira, P. and Pfeifer, N., 2017. Improved topographic models via concurrent airborne lidar and dense image matching. ISPRS Annals of Photogr., Remote Sensing & Spatial Information Sciences.
- Mulla, D., 2013. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. Biosystems Engineering 114(4), pp. 358–371.
- Rusinkiewicz, S. and Levoy, M., 2001. Efficient variants of the ICP algorithm. In: 3-D Digital Imaging and Modeling, 2001. Proceedings. Third International Conference on, IEEE, Quebec City, Canada, pp. 145–152.
- Skaloud, J. and Lichten, D., 2006. Rigorous approach to bore-sight self-calibration in airborne laser scanning. ISPRS Journal of Photogrammetry and Remote Sensing 61(1), pp. 47–59.
- Skaloud, J., Schaer, P., Stebler, Y. and Tomé, P., 2010. Real-time registration of airborne laser data with sub-decimeter accuracy. ISPRS Journal of Photogrammetry and Remote Sensing 65(2), pp. 208–217.
- Toschi, I., Remondino, F., Rothe, R. and Klimek, K., 2018. Combining Airborne Oblique Camera and Lidar Sensors : Investigation and New Perspectives. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XLII-1, October, Karlsruhe, pp. 437–444.
- Xiao, Y. and Zhan, Q., 2009. A review of remote sensing applications in urban planning and management in china. In: IEEE Joint Urban Remote Sensing Event.