agricultural engineering.eu 79(4), 2024, 261–274

DOI: 10.15150/ae.2024.3326

agricultural engineering.eu

Application of a direct georeferencing method of drone images for smart farming

Ralph L. Stoop, Markus Sax, Dejan Seatovic, Thomas Anken

In smart farming, information collected by autonomous devices needs to be related to its exact field location, that is, georeferenced. In this work, we study the applicability and accuracy of a simplified direct georeferencing method of drone images for typical smart farming applications, such as weed detection. Based solely on an affine homography, the method results in accuracies < 0.82 m measured at 30 m above ground level even in the presence of pronounced steepness when applied to a real-time kinematic (RTK) enterprise drone (DJI Matrice 300 RTK). Our method uses only single images and does not rely on feature matching; therefore, it is inexpensive to compute. Depending on the targeted use case, the proposed georeferencing method yields errors in the order of the object of interest's dimensions, which we demonstrate for our envisioned use case of dock plant ('Rumex') detection on meadows. The method may be seen as an upper bound for georeferencing errors that can be applied easily to other drone systems.

Keywords

Smart farming, direct georeferencing, unmanned aerial vehicles, weed localization

Feeding the world's population has always been a major challenge for humankind and is now compounded by climate change and the loss of biodiversity, which are strongly complicating efforts to maintain agricultural productivity in the 21st century. To face these challenges, new concepts and methods are needed to achieve the sustainable intensification of agriculture. Smart farming aims to apply digital technologies on a large scale to agriculture and is considered one important ingredient toward this goal (Walter et al. 2017, Storm et al. 2024). It includes the use of intelligent systems, for example, autonomous vehicles, equipped with different sensors to provide spatio-temporal information about the status of a crop field. Georeferencing is the process of associating a geographic location with a particular object or feature and is a crucial step in obtaining such spatio-temporal maps. It involves aligning data from various sources, such as maps, satellite and terrestrial imagery, and GPS coordinates, to create an accurate spatial reference of a feature. Therefore, georeferencing is essential for smart farming applications based on unmanned vehicles and intelligent systems.

Applied to standard RGB cameras, georeferencing associates a geographic location in the 3D world with each pixel in an image. However, this process is inherently limited by the fact that images are produced through the lossy mapping of the 3D world's structural data to the 2D image plane, which is performed by the camera. This mapping, or transformation, is typically represented by a perspective projection, a specific type of 3D to 2D projection that also underlies the functioning of the human eye as well as many cameras. For example, perspective projection gives rise to the well-known phenomenon in which railway lines observed by the engineer, despite being parallel in reality, appear to converge until they ultimately meet at a single point toward the horizon. The loss of information

received 15 July 2024 | accepted 16 December 2024 | published 30 December 2024

that underlies such optical effects significantly complicates the reconstruction of a scene's structure from a single image. Consequently, more information is required for this reconstruction task. In nature, most animals possess at least two eyes, and information from each can be combined to more accurately estimate the dimensions of objects and distances. Similarly, georeferencing methods seek to reconstruct and localize the scene of an image by integrating it with additional information, either derived from other sensor data or from additional images of the same scene captured from a slightly different position. Although there are many different georeferencing approaches, it is instructive to differentiate between direct georeferencing methods that are based solely on Global Navigation Satellite Systems (GNSS) data (direct georeferencing) and indirect georeferencing methods that require ground information collected in the field, also called aerial triangulation (Mian et al. 2016).

In aerial triangulation, the required geospatial information stems from visual objects with accurately known locations, usually provided by ground control points (GCPs). Based on this information, the scene can be reconstructed using standard aerial triangulation techniques, particularly bundle adjustments. Although these methods are accurate, tested, and established, the need for GCPs hampers their applicability in cases of emergency scenarios, inaccessible terrain, and low-cost surveillance (Syetiawan 2020). The latter typically applies to smart farming applications, where data collection must be as cost-effective as possible (Hemerly 2014, Xiang and Tian 2011).

Recently, direct georeferencing has emerged as an alternative approach that does not require GCPs. These methods combine the visual system with sensor readings from GNSS and inertial systems (Cucci et al. 2017, Gabrlik et al. 2018, Syetiawan et al. 2020). Although many different implementations and methods for direct georeferencing exist, the usual pipeline consists of several steps. First, the camera position and orientation (camera pose) at each image snapshot are initialized based on the GNSS locations. Second, distinctive and comprehensive image regions are extracted for each image. These visual features offer a robust way to describe an image and relate it geometrically to other images depicting the same scene. This is achieved by comparing and matching the visual features of all images, which results in a comprehensive correspondence map. Third, from the initial camera poses and feature correspondences, a sparse point cloud containing the 3D coordinates of all feature points is calculated, and the camera poses are refined, typically simultaneously using bundle adjustment (Hartley and Zissermann 2004). In this step, the refinement of camera poses and scene feature points is typically formulated as a joint optimization problem, wherein bundle adjustment seeks to find a globally consistent configuration. Lastly, a densification step is applied to the point cloud to recover the dense structure of the scene. In the case of UAVs, a so-called orthomosaic is obtained by reprojecting the dense model onto a common plane, thereby providing a highly accurate, rectified planar view of the area of interest typically used in arial surveillance. The high accuracy and precision of this process have made direct georeferencing an important technique that is widely implemented in commercial tools, such as Agisoft's Metashape (Agisoft LLC, St. Petersburg, Russia) or Pix4D (Pix4D, Lausanne, Switzerland). Note that direct georeferencing can still be combined and refined using ground truth data from GCPs (Liu et al. 2022).

Despite the high accuracy of direct georeferencing, its practical application in agricultural applications is still hampered by their demand for high-quality data. State-of-the-art direct georeferencing methods require an extensive set of images with distinct and unambiguous visual features that remain constant throughout the entire flight of the UAV. Furthermore, high image overlaps of around 70 % are typically required to allow feature matching during bundle adjustment (Seifert et

al. 2019, Pix4D 2024). Another crucial aspect is the high computational complexity of 3D scene reconstruction. Given the high resolution of modern UAV images (e.g., 12–20 MP for consumer drones and 45 MP for enterprise drones with full-frame camera sensors), creating an orthomosaic may take hours or even days and has yet to meet real-time requirements. This is particularly unfortunate, as, for many applications, the provided accuracy of these methods might not even be needed, and less precise alternative approaches might be better suited.

Recently, different mapping methods have been suggested to further achieve orthomosaics in real time. Bu et al. (2016) used simultaneous location and mapping (SLAM) to estimate the camera position and attitude and combined it with an incremental stitching procedure to obtain a real-time map fusion. However, stitching is still done using only visual features. Hinzmann et al. (2017) further included a reconstruction of the surface structure to obtain a dense reconstruction of the scene. Kern et al. (2020) introduced their automated pipeline, 'OpenREALM,' which represents the first system capable of generating real-time orthomosaics. The pipeline incorporates state-of-the-art SLAM methods with dense reconstruction. Even though the proposed methods are capable of generating high-quality orthomosaics in real time, their performance still strongly depends on the quality of the feature matching, and they therefore require images with distinct visual features. Moreover, the quality of the orthomosaic is only a pertinent consideration in the context of applications that necessitate its utilization. However, in many agricultural applications of UAVs, such as weed detection, only the coordinates of the specific object of interest are required, without the need for a full 3D reconstruction of the scene or the orthomosaic. Furthermore, the proposed methods for real-time orthomosaics still depend on visual feature matching, for which images with high overlaps are needed. Unfortunately, this increases the time the UAV requires to fully scan the area of interest. Since a human operator ("pilot") is still usually required on the site to control the UAV, the flying time crucially determines the total cost of the service.

Popović et al. (2017a) proposed a sophisticated informative path planning method that fuses images acquired at different flying altitudes to reduce flying time while maintaining acceptable classification rates in the typical setting of UAV-based weed detection. Their multiresolution approach could be further extended to other sensory systems based on UAVs, such as temperature maps (Popović et al. 2017b). However, the 3D reconstruction and georeferencing are still done using standard structure-from-motion approaches using a commercial tool (Pix4D), with sufficient overlap of the images needed to reconstruct the scene (Pretto et al. 2021).

This work is based on practical considerations gained in the context of the "Rumex project," which has been co-financed by Innosuisse (Innosuisse 2024). The objective is to develop an automated system for detecting Rumex plants on meadows using UAV images and estimating their position in the 3D world (Sax et al. 2023). The Rumex's coordinates can then be sent to an autonomous vehicle for treatment in the future. For the approach to be useful in practice, it is necessary to maximize the quality of the binary classification (Rumex/no Rumex) of a unit cell, thereby maximizing the F1 score. In other words, the objective is to minimize false positives while maximizing the true positives. Our preliminary findings indicate that this requires acquiring images at a minimum of 12 m and maintaining a constant flying altitude. This maximizes the resolution of the images, which is needed to robustly detect the fine visual features of the plant. In light of these constraints, it is not feasible to alter the flying altitude as proposed by Popović et al. (2017a). Consequently, the only viable option is to modify the side-overlap to further reduce the overall flying time. The dependency of

the flying time on the side-overlap can be approximated via hyperbolic dependence *flying time* ~ 1/(1 – *overlap*), where ~ denotes equality up to a constant, as illustrated in Figure 1a.

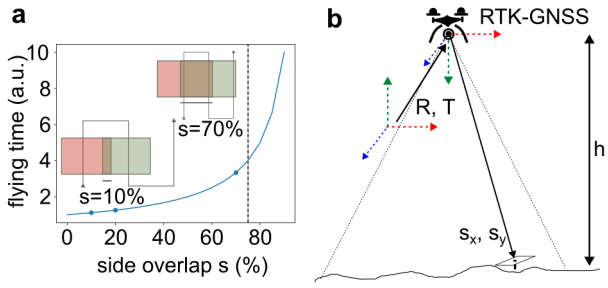

Figure 1: a) Illustration of the flying time vs. side overlap, b) Schematic of the georeferencing method based on affine transformation, including anisotropic scaling s_x, s_y, rotation R, and translation T; the altitude above ground level (AGL) is either assumed to be fixed $h = h_{init}$ ("Original") or refined $h = h_{refined}$ ("AGL-adjusted") to account for the structure of the surface

This is in accordance with experimental findings (Seifert et al. 2019). Typically, a side-lap of 70 % is recommended, which strongly increases the flying time (Pix4D 2024, Seifert et al. 2019). To address this issue, the proposed direct georeferencing approach does not depend on overlapping images. The method estimates the location of objects from their position in images using only the sensory data of the UAV in conjunction with an affine transformation. Therefore, only the location of the detected objects (in this case, the Rumex plants) is of interest, and no orthomap is provided. Our approach is largely based on previous work on single image-based, direct georeferencing that does not depend on feature matching (Helgesen et al. 2019, Hemerly 2014, Leira et al. 2015, Pfeifer et al. 2012, Xiang and Tian 2011). Although the problem is greatly simplified by treating it as 2-dimensional, the resulting accuracy is still satisfactory for many applications in precision farming, particularly in the detection of weed. Moreover, our approach does not assume precise camera calibration and minimizes the number of pre-processing steps typically employed to enhance predictions, such as sensor data fusion. Our approach is motivated by its straightforward implementation and simplicity for practical applications, offering a convenient estimate of the upper bound for georeferencing errors in typical smart farming applications. Such estimates may facilitate a more comprehensive understanding of the requirements for UAV georeferencing, which are necessary for the selection of parameters for UAV flights, including altitude, speed, and overlaps. To extend the applicability of the method to regions with changing slopes, we combine the direct georeferencing approach with the region's digital elevation map to correct for altitude errors.

Materials and Methods

Direct Georeferencing Model

To calculate the real-world coordinates of the target GCPs, we used the simplified camera model depicted in Figure 1b. The model assumes that the ground surface scene is a perfect plane parallel to the camera's image plane. At any given moment of time, the model assumes that the optical axis of the camera is aligned with the vertical direction (nadir) and that the distance between the two planes is always constant (constant flying height or altitude above ground level, AGL). Under these assump-

tions, the transformation of the two-dimensional (2D) image (i.e., pixel) coordinates $p_{im} = (x_{im}, y_{im})^T$ to the world coordinates $p_w = (x_w, y_w)^T$ (in longitude, latitude) is given by a homography $H: P^2 \rightarrow P^2$; $p_{im} \rightarrow p_w = Hp_{im}$, with P^2 being the projective space. Using homogenous coordinates, H can be described by a 3 × 3 matrix and is defined only up to scale (8 degrees of freedom, DOF). Assuming an affine transformation consisting of a single translation ($t_{x,y}$, in the x and y directions, respectively), rotation about the z axis (yaw angle θ) and anistropic scaling ($s_{x,y}$ in the x and y directions) finally leads to the following homography with 5 DOF:

$$\begin{pmatrix} x_w \\ y_w \\ 1 \end{pmatrix} = \begin{pmatrix} s_x \cos\theta & -s_y \sin\theta & t_x \\ s_x \sin\theta & s_y \cos\theta & t_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_{im} \\ y_{im} \\ 1 \end{pmatrix} (1) \qquad \text{(Eq. 1)}$$

Given our constant AGL assumption, $s_{x,y}$ remain constant during a flight and are derived from the ground sampling distance (GSD) in the x and y directions. The assumption of the perfect nadir condition imposes roll and pitch angles to be zero. Therefore, the only IMU data considered is that of the yaw angle (θ) at each image taken. Lastly, $t_{x,y}$ are given by the GNSS RTK position of the UAV at each image point. Note that our model assumes that the GNSS RTK signal corresponds to the projection center of the camera and therefore neglects any offset that is typically present (Correia et al. 2022). However, since all flights are carried out in RTK mode, the coordinates are already corrected by the offsets in the longitudinal, latitudinal, and vertical directions by the DJI's internal processing. Furthermore, perfect time synchronization of all sensor readings is assumed.

Although the assumption of a perfectly flat scene might be sufficient in some applications, it seems too restrictive for many crop fields in practice, particularly for countries like Switzerland, where considerable elevation differences and slopes are often present. To handle these situations, the model can be applied iteratively to correct for some errors due to non-flat surfaces: initially, the pre-assumed flying height (here h_{init} = 30 m) is used with Equation 1 to estimate the position of the object of interest. Then, the actual vertical distance between the UAV and the object is refined using a publicly available digital terrain model (DTM) with 1 m resolution (Swisstopo, Wabern, Switzerland). The refined height is given as $h_{refined} = z_{UAV} - z_{pred}$, where z_{UAV} is the UAV's altitude measured by the drone's GNSS, and z_{pred} is the altitude value of the DTM at the location of the initial prediction. Finally, the refined position of the object of interest is calculated again using Equation 1, where h_{init} is updated by $h_{refined}$. Although this procedure can be applied multiple times, our findings indicate that one iteration is sufficient for our dataset, with convergence of the error occurring after one iteration.

Measurement Site and Flight Parameters

A DJI's Matrice 300 RTK (Shenzhen, China) UAV was utilized in this study. This drone had a positional accuracy of 1 cm (horizontal) and 1.5 cm (vertical) when the RTK was fixed. A full-frame 45 MP camera (Zenmuse P1) with a 50 mm objective lens was connected to the drone via a corresponding gimbal (all components from DJI, Shenzhen, China) and set to nadir configuration. For flight planning, we used the commercial software UGCS (SPH Engineering, SIA, Riga, LV-1050, Lativa) in photogrammetry mode, which ensured an approximately constant AGL that could be as small as 12 m.

To evaluate the performance of the georeferencing method for different flight configurations, 28 GCPs were distributed over a total area of approximately 2.3 ha in location "Waldegg" (Lat/Long: 47.49290,

8.92094) near Aadorf, Switzerland (Figure 2a). The area's topography consists of a flat region in the south-east and a steeper area toward the north-west (Figure 2b). The total height difference is roughly 20 m. The ground-truth position of each GCP was measured using Trimble R780 with TDC600 (Trimble Inc., Westminster, Colorado, USA).

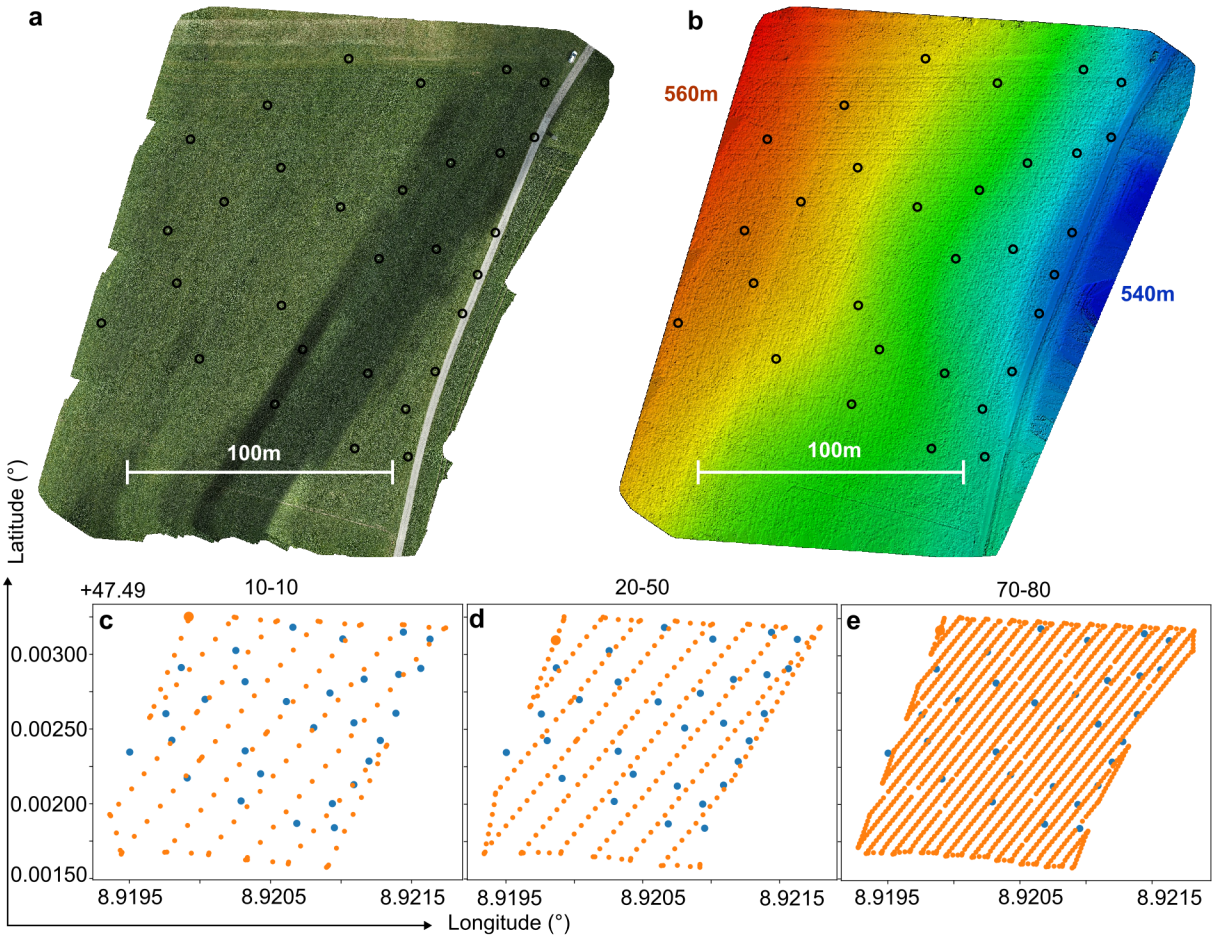

Figure 2: Experimental site 'Waldegg' reconstructed from the high overlap (70–80) flight via standard photogrammetry (Pix4D), a) The 2D reconstruction of the scene (orthomosaic), b) The obtained surface model (digital surface model, DSM), black open circles indicate GCP positions, c)–e) UAV position at each image acquisition for three flights with different sides and front overlaps: 10–10 (c), 20–50 (d), and 70–80 (e), blue-filled circles indicate the GCP positions used to determine the accuracy of georeferencing

Three subsequent flights at AGL = 30 m with different side and front overlaps (10–10, 20–50, 70–80) were carried out, as shown in Figure 2c–e. For the selected flight speed of v = 2.5 m/s, the flight time was approximately 10 min for the 10–10 flight, 12 min for the 20–50 flight, and 27 min for the 70–80 flight. The camera trigger was set to automatic based on equal distances. For all flights at AGL = 30 m, the ground sampling distance was 0.26 cm. Given the typical noise present in the UAV's trajectory, the 10–10 flight represents the non-overlapping case, in which each GCP was ideally seen only in a single image. The 70–80 flight represents the standard flight needed for photogrammetric processing, and each GCP was seen in approximately 4 (side overlap) × 5 (front overlap) images. The standard rectangular flying pattern, which was also employed in this work, necessitates that the UAV visit a GCP typically in four distinct flight lines. In these lines, groups of two are shifted, but they oth-

erwise maintain the same horizontal orientation. In our case, the lines had an approximate angle of $\theta_1 \approx 30°$ and $\theta_2 \approx -150°$, hence rotated by 180°. Note that GCPs located at the border of the surveyed area were usually seen in a lower number of images, since they were often only visited three times. Throughout all flights, the UAV translated at a constant speed of 2.5 meters per second (m/s).
To further validate our approach, we conducted additional measurements at a flat parking space in Tänikon (Aadorf, Switzerland) at an AGL = 12 m, GSD = 0.11 cm and a constant flying speed of 1 m/s. The camera triggering mode was an equal distance. Using this configuration, the flight time was roughly 25 min/ha. Fifteen GCPs were distributed along two straight lines over a small area of roughly 300 m^2 and measured using a Trimble R8. The selected low AGL was identical to the one used in the Rumex project. Aligning all GCPs along two lines allowed for a better interpretation of the obtained accuracies.

Accuracy Analysis

GCP image coordinates p_{im} (in pixel) were manually determined for each image containing at least one GCP. Due to their sparse distribution, only very few images contain more than one GCP, and most contain none. However, the opposite was not true, as the same GCP was typically seen in different images in the high overlap case. The resulting collection of images containing GCPs and their locations in the image were then used to assess the accuracy of georeferencing: for each GCP image coordinate p_{im}, we obtained metadata, such as longitude/latitude/altitude measured by the RTK GNSS and yaw angle of the corresponding image, which, together with the constant values for the sensor size and focal length, were used as input for the georeferencing method to return the estimated longitude and latitude. Given the ground truth position of the GCPs, prediction errors were either determined by component-wise comparison, yielding the longitudinal and latitudinal errors $\Delta Long = Long_{target} - Long_{pred}$ and $\Delta Lat = Lat_{target} - Lat_{pred}$, or by the total error, determined via the Euclidian distance between error and prediction, $E=\sqrt{(\Delta Long^2 + \Delta Lat^2}$). All differences in longitude and latitude were approximated by meters via the conversion $c_{lat \rightarrow m} = 111000.0\ m/°Lat$ and $c_{long \rightarrow \mathrm{m}} = 111319.5 \cdot cos\ (Lat) m/°Long$. Uncertainty in these scaling factors was approximately ± 0.02 m.

Results and Discussion

For all three flights, the average and standard deviation of the total error for the original and AGL-adjusted georeferencing methods are summarized in Table 1.

Table 1: Average and standard deviation of the total georeferencing error E for the original and AGL-adjusted methods

Overlap (side-front)	Average of total error E: original method in m	Standard deviation of total error E: original method in m	Average of total error E: AGL-adjusted method in m	Standard deviation of total error E: AGL-adjusted method in m	Flight time in min
10–10	1.05	0.72	0.82	0.53	10
20–50	0.99	0.77	0.80	0.75	12
70–80	0.92	0.68	0.63	0.54	27

Values have an uncertainty ± 0.02 m for distances and < 1 min for time.

The application of the AGL adjustment yielded a mean total error of E = 0.82, 0.80, and 0.63 m for the 10–10, 20–50, and 70–80 flights, respectively. Thus, mean accuracies ≤ 0.82 m were reached, despite the strong assumptions imposed by our affine homography approach. Surprisingly, the average error was minimal for the highest overlap flight, which was unexpected, as our direct georeferencing method was based on only single images. However, this observation can be attributed to the very different sample sizes for the different flights, as high overlaps resulted in an increased number of samples. To better interpret the obtained accuracies, the histograms of *ΔLat* and *ΔLong* for all three flights using only the predefined AGL of 30 m (blue) and the AGL-adjusted method (orange) are shown in Figure 3.

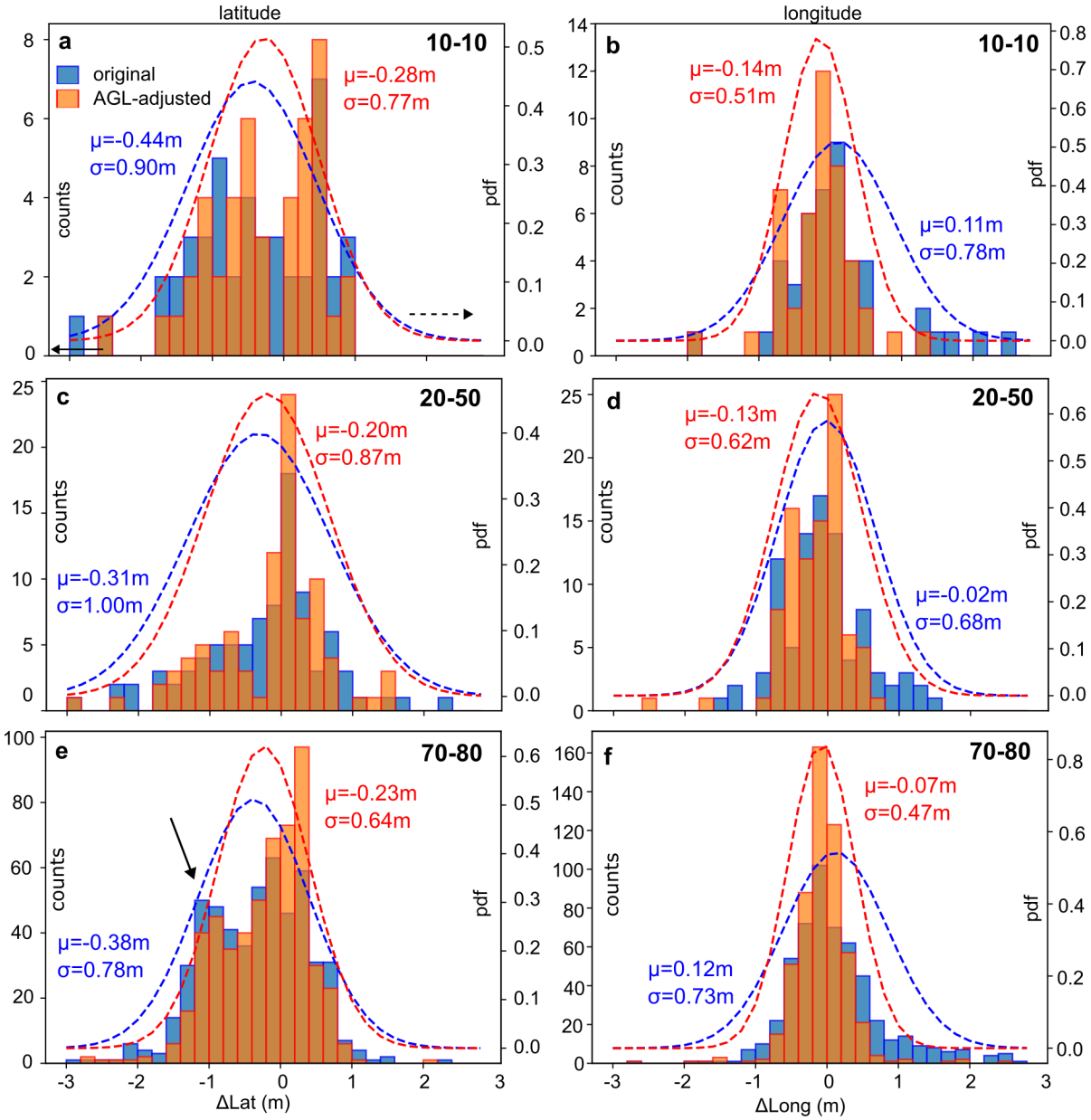

Figure 3: Horizontal georeferencing errors for latitude (*ΔLat*, left) and longitude (*ΔLong*, right) and for the initial (blue) and AGL-adjusted method (orange) for the 10–10 (a,b), 20–50 (c,d), and 70–80 (e,f) flights

For each histogram, we estimated the parameters of a Gaussian distribution, its mean μ, and standard deviation σ (crossed lines in Figure 3). Assuming a Gaussian distribution of the errors is reasonable when full 3D reconstruction of the scene is conducted (Hartley et al. 2004). As shown in Figure 3, this assumption seems too restrictive for our georeferencing model, especially since symmetry around the distribution's mode is generally not supported by our data, particularly for ΔLat. However, for the sake of comparison with the standard photogrammetry literature, we nevertheless utilize the first two moments as the primary figure of merit in the remainder of this work.

As with the total error, the AGL correction significantly enhanced the predictions for both latitude and longitude, as evidenced by a reduction in the absolute values of both moments, particularly the standard deviation. This finding suggests that a fixed value of AGL = 30 m does not always accurately represent the actual altitude of the UAV with respect to the GCP. For example, if the terrain-following mode does not maintain a fixed AGL, our initial assumption of AGL = 30 m will obviously be wrong. Even if the UAV has the correct AGL with respect to the surface point perpendicular to the UAV, the GCP of interest might still be shifted horizontally and therefore have a different altitude. A single iteration of the adaptive height estimation using the actual altitude from the DTM of the GCP's estimated position decreases the error significantly and underscores the importance of using terrain models to properly estimate the vertical height difference between the object of interest and the UAV.

Interestingly, a clear difference between latitudinal and longitudinal errors was observed across all three flights. Specifically, a reduced standard deviation and sharper distribution for $\Delta Long$, was evident. This observation can be primarily attributed to the emergence of a secondary peak around ΔLat = -1 m, most pronounced in the 70–80 flight (black arrow in Figure 3e). To better understand the origin of this peak, a revised histogram of the latitudinal (violet histogram bars) and longitudinal (green histogram bars) errors for the 70–80 overlap flight with the results for the two flying orientations $\theta_1 \approx 30°$ and $\theta_2 \approx -150°$ separated, is shown in Figure 4a and Figure 4b, respectively.

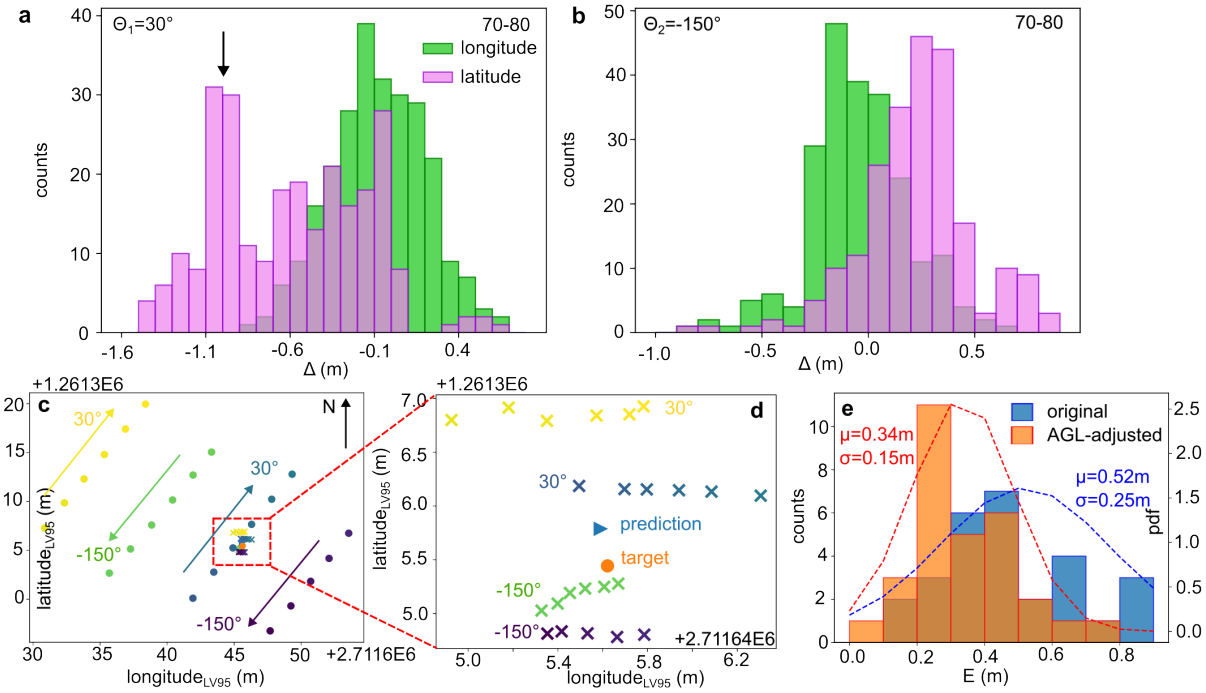

Figure 4: Flight angle-dependent prediction error for the 70–80 overlap flight for a) $\theta \approx 30°$ and b) $\theta \approx -150°$, c) Example of UAV positions (filled circles) and predictions ('x') for GCP7 show the flying pattern of the UAV with two yaw angles, d) Zoomed region of the red rectangle from c shows predictions from the georeferencing method together with the actual GCP position ('target'), the triangle shows the position of the average overall predictions ('prediction'), and the filled circle indicates the target's actual position, e) Total errors resulting from prediction averaging within clusters from individual GCPs

We can fully attribute the emergence of the second peak in the latitudinal error histogram to the predictions from images acquired at a yaw angle of $\theta_1 \approx 30°$, indicated by the black arrow in Figure 4a. At θ_2, no clear second peak was observed for the latitudinal error. Note that the longitude seems to follow a Gaussian or, at least, a monomodal distribution. To better understand this behavior, we show examples of UAV positions at image acquisition (filled circles) and corresponding predictions (crosses) for a specific GCP ('GCP7') in Figure 4c. The four different colors indicate the four different flying lines, where images are acquired consequentially and at a fixed θ. To facilitate comparison, the longitude and latitude values are approximated by LV95 coordinates, where the unit is expressed in meters. This allows for relating the distances from Figure 4c and 4d to the errors in the histograms of Figure 3 and 4. Depending on its position, each GCP was covered by four or three flying lines, leading to a varying number of resulting predictions, as shown in the red crossed-line rectangle of Figure 4c.

To better investigate how the UAV position determines the prediction, a zoomed version of the rectangle's area is shown in Figure 4d. As with many other GCPs, we find that the predictions from individual flying lines often show more variance along the longitudinal direction and less variation in their latitude. A pronounced example is given by the yellow predictions with θ = 30° in Figure 4e, where the total spreading in the longitudinal direction is about 0.9 m compared to only 0.3 m in latitude. The opposite is observed when comparing the predictions from different flying lines. In this case, most of the variance is captured by the latitudinal direction. Both effects point toward contributions from drift: for each flying line, the yaw angle remains approximately constant and only $t_{x,y}$ (i.e. the UAV's position) and $p_{x,y}$ (i.e. the position of the GCP in the image) are changing. Note that most

of the change in pixel coordinates occurs in the y-component. The x-component only changes due to non-zero pitch and roll angles. Due to the linearity of our georeferencing model and the fact that $t_{x,y}$ only contributes as a translation, the resulting predictions were typically also correlated along a line with a yaw-dependent orientation in the xy plane. Overall, for a single GCP, the error variations in the latitudinal direction are often larger than in the longitude, which can be seen directly from the fact that the values for the latitude do not overlap in Figure 4e, while the ones for longitude do. This effect leads to a broader distribution of the latitudinal compared to the longitudinal error in Figure 3e and 3f.

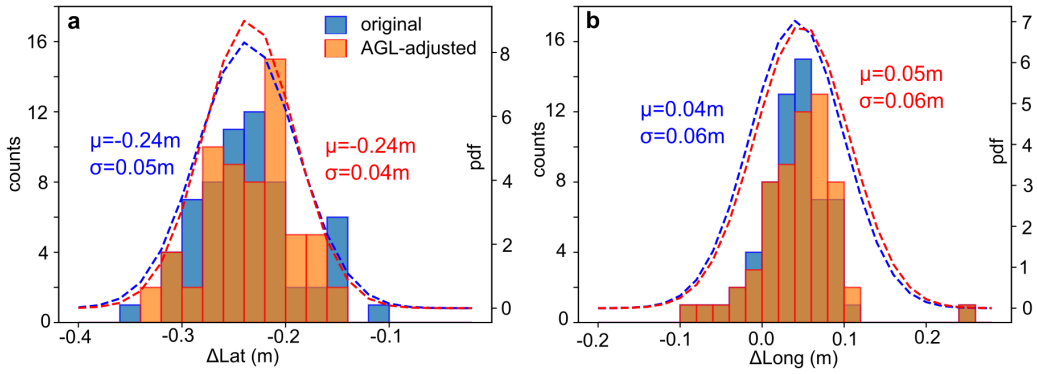

Figure 5: Histograms of the a) latitudinal and b) longitudinal error at AGL = 12 m for the original (blue) and AGL-adjusted (orange) georeferencing method

When flying with overlap, it is possible to utilize multiple predictions derived from disparate images showing the same object, provided that the object can be identified unambiguously. In the simplest case, the numerical average of all predictions of the object can be used. However, in a typical scenario, object identification is challenging and requires matching feature points between overlapping images, which our method aims to avoid. If the correspondence of predictions and objects is unknown, unsupervised clustering is the sole option for collecting predictions that are spatially close to each other. In the context of sparse object distribution, the present study employs an expectation-maximation algorithm („mean-shift“, Comaniciu and Meer 2002) with a bandwidth of approximately 5 m to cluster predictions of individual GCPs' locations. Following our previous discussion on the spatial distribution of the predictions, it should be clear that the samples are not independently identically distributed (iid), and therefore their average will generally be biased, particularly in terms of their latitude for uneven numbers of flying lines. The resulting histogram for the total error of the 70–80 flight is shown in Figure 4e. Averaging reduces the mean of the total error from 0.92 to 0.52 m for the constant AGL method, and from 0.63 to 0.34 m for the AGL-adjusted method.

Finally, our homography-based approach was applied to the typical setting of weed detection, in which we chose a very low AGL = 12 m. The low AGL ensures the robust detection of the rumex plants using convolutional neural networks. Given the systematic distribution of GCPs along two parallel lines with a distance of 2 m, a single flight line was sufficient, and no side overlap was used. The resulting histograms of *ΔLat* and *ΔLong* are shown in Figure 5. We observed no clear difference between the original and AGL-adjusted methods, which is expected due to the flatness of the investi-

gated area. Note that for Figure 5, we did not average the predictions of the same GCP from different images. The corresponding average total error E = 0.25 m is roughly a factor of three smaller than the value obtained at AGL = 30 m, in agreement with the linear dependence of the error on the AGL.

Conclusions

In this work, we demonstrate that ready-to-use UAV-based RTK GNSS are sufficient to provide < 0.82 m accuracy from 30 m height, even in the presence of noticeably variable terrain, even when using the simplest possible affine mapping in a purely direct georeferencing setting. Although the achieved accuracies at AGL = 30 m are not sufficient for direct spot spraying of herbicides, the rumex locations obtained from our simple method could be delivered directly to the farmer for surveillance or for more targeted manual removal in the context of organic farming. Lower flying heights result in georeferencing errors of 0.25 m at AGL = 12 m, which may be sufficient for automated treatment. However, any treatment device, particularly for hot water treatment of the rumex' roots in organic farming, will likely require its own visual system. In this case, the approximate positions of the rumex plants computed by the proposed method could serve as the waypoints of the autonomous device and could be used for optimized path planning. Given the great simplicity of the approach with respect to flight operation and post-processing, it seems sufficient for many agricultural applications. Even though highly accurate direct georeferencing methods exist, their applicability to practical applications is often hampered by flight time or computation. As shown in this work, affine homographies are a good initial starting point, and more advanced techniques might only be considered when higher accuracies are needed, or sensory data are noisier. At AGL = 30 m and flight speed v = 2.5 m/s, the proposed georeferencing method allows for a reduction in flight time by almost a factor of three in comparison to flights with high side overlaps, given that only minimal side overlap (10 %) is required.

References

Bu, S.; Zhao, Y.; Wan, G.; Liu, Z. (2016): Map2DFusion: Real-time incremental UAV image mosaicing based on monocular SLAM. In 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4564–4571, https://doi.org/10.1109/IROS.2016.7759672

Comaniciu, D.; Meer, P. (2002): Mean shift: a robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(5), pp. 603–619, https://doi.org/10.1109/34.1000236

Correia, C.A.M.; Andrade, F.A.A.; Sivertsen, A.; Guedes, I.P.; Pinto, M.F.; Manhães, A.G.; Haddad, D.B. (2022): Comprehensive direct georeferencing of aerial images for unmanned aerial systems applications. Sensors 22(2), pp. 604-629, https://doi.org/10.3390/s22020604

Cucci, D.A.; Rehak, M.; Skaloud, J. (2017): Bundle adjustment with raw inertial observations in UAV applications. ISPRS Journal of Photogrammetry and Remote Sensing 130, pp. 1–12, https://doi.org/10.1016/j.isprsjprs.2017.05.008

Gabrlik, P.; la Cour-Harbo, A.; Kalvodova, P.; Zalud, L.; Janata, P. (2018): Calibration and accuracy assessment in a direct georeferencing system for UAS photogrammetry. International Journal of Remote Sensing 39(15–16), pp. 4931–4959, https://doi.org/10.1080/01431161.2018.1434331

Hartley, R.I.; Zisserman, A. (2004): Multiple View Geometry in Computer Vision. Second ed. Cambridge University Press

Helgesen, H.H.; Leira, F.S.; Bryne, T. H.; Albrektsen, S.M.; Johansen, T.A. (2019): Real-time georeferencing of thermal images using small fixed-wing UAVs in maritime environments. ISPRS Journal of Photogrammetry and Remote Sensing 154, pp. 84–97, https://doi.org/10.1016/j.isprsjprs.2019.05.009

Hemerly, E.M. (2014): Automatic georeferencing of images acquired by UAV's". International Journal of Automation and Computing 11(4), pp. 347–352, https://doi.org/10.1007/s11633-014-0799-0

Hinzmann, T., Schönberger, J.L.; Pollefeys, M.; Siegwart, R. (2018): Mapping on the fly: Real-time 3D dense reconstruction, digital surface map and incremental orthomosaic generation for unmanned aerial vehicles. In Field and Service Robotics, 383–396. Cham: Springer International Publishing, https://doi.org/10.1007/978-3-319-67361-5_25

Innosuisse (2024): Schweizerische Agentur für Innovationsförderung. "Verbesserung der Schweizer Agrarproduktion durch Digitalisierung," published 22 September 2022. https://www.innosuisse.admin.ch/de/verbesserung-der-schweizer-agrarproduktion-durch-digitalisierung, accessed on 15 October 2024

Kern, A.; Bobbe, M.; Khedar, Y.; Bestmann U. (2020): OpenREALM: Real-time mapping for unmanned aerial vehicles. In: 2020 International Conference on Unmanned Aircraft Systems (ICUAS), pp. 902–911, https://doi.org/10.1109/ICUAS48674.2020.9213960

Leira, F.S.; Trnka, K.; Fossen, T.I.; Johansen, T.A. (2015): A light-weight thermal camera payload with georeferencing capabilities for small fixed-wing UAVs. In: 2015 International Conference on Unmanned Aircraft Systems (ICUAS), pp. 485–494, https://doi.org/10.1109/ICUAS.2015.7152327

Liu, X.; Lian, X.; Yang, W.; Wang, F.; Han, Y.; Zhang, Y. (2022): Accuracy assessment of a UAV direct georeferencing method and impact of the configuration of ground control points. Drones 6(2), https://doi.org/10.3390/drones6020030

Mian, O.; Lutes, J.; Lipa, G.; Hutton, J. J.; Gavelle, E.; Borghini S. (2016): Accuracy assessment of direct georeferencing for photogrammetric applications on small unmanned aerial platforms. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-3/W4, pp. 77–83, https://doi.org/10.5194/isprs-archives-XL-3-W4-77-2016

Pix4D (2024): Pix4D documentation. https://support.pix4d.com/hc/en-us/articles/115002471546, accessed on 9 October 2024

Pfeifer, N.; Glira, P.; Briese, C. (2012): Direct georeferencing with on board navigation components of light weight UAV platforms. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXIX-B7, pp. 487–492. https://doi.org/10.5194/isprsarchives-XXXIX-B7-487-2012

Popović, M.; Hitz, G.; Nieto, J.; Sa, I.; Siegwart, R.; Galceran, E. (2017a): Online informative path planning for active classification using UAVs. In: 2017 IEEE International Conference on Robotics and Automation (ICRA), pp. 5753–5758, https://doi.org/10.1109/ICRA.2017.7989676

Popović, M.; Vidal-Calleja, T.; Hitz, G.; Sa, I.; Siegwart, R.; Nieto, J. (2017b): Multiresolution mapping and informative path planning for UAV-based terrain monitoring. In: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1382–1388, https://doi.org/10.1109/IROS.2017.8202317

Pretto, A.; Aravecchia, S.; Burgard, W.; Chebrolu, N.; Dornhege, C.; Falck, T.; Fleckenstein, F. et al. (2021): Building an aerial–ground robotics system for precision farming: an adaptable solution. IEEE Robotics & Automation Magazine 28(3), pp. 29–49. https://doi.org/10.1109/MRA.2020.3012492

Sax, M.; Stoop, R.L.; Nasser, H.-R-; Seatovic, D.; Keel, S.; Lehrmann, A.; Höfer,T.; Anken, T. (2023): Broad-leaved dock control by unmanned aerial vehicles: What image quality is needed for a plant detection? In Land. Technik AgEng 2023, VDI-Berichte 2427, pp. 505–512,Düsseldorf, VDI Verlag

Seifert, E.; Seifert, S.; Vogt, H.; Drew, D.; van Aardt, J.; Kunneke, A.; Seifert, T. (2019): Influence of drone altitude, image overlap, and optical sensor resolution on multi-view reconstruction of forest images. Remote Sensing 11(10), pp. 1252-1271. https://doi.org/10.3390/rs11101252

Storm, H.; Seidel, S. J.; Klingbeil, L.; Ewert, F.; Vereecken, H.; Amelung, W.; Behnke, S., et al. (2024): Research priorities to leverage smart digital technologies for sustainable crop production. European Journal of Agronomy 156(1), pp. 127178-127185. https://doi.org/10.1016/j.eja.2024.127178

Syetiawan, A., Gularso, H.; Kusnadi, G. I.; Pramudita, G. N. (2020): Precise topographic mapping using direct georeferencing in UAV. IOP Conference Series: Earth and Environmental Science 500(1), pp. 012029-012039, https://doi.org/10.1088/1755-1315/500/1/012029

Walter, A., Finger, R.; Huber, R.; Buchmann, N. (2017): Smart farming is key to developing sustainable agriculture. Proceedings of the National Academy of Sciences 114(24), pp. 6148–6150. https://doi.org/10.1073/pnas.1707462114

Xiang, H.; Tian, L. (2011): Method for automatic georeferencing aerial remote sensing (RS) images from an unmanned aerial vehicle (UAV) platform. Biosystems Engineering 108(2), pp. 104–113. https://doi.org/10.1016/j.biosystemseng.2010.11.003

Authors

Dr. Ralph Lukas Stoop and **Markus Sax** are research assistants and **Dr. Thomas Anken** is the group leader of the Digital Production Group at Agroscope, Tänikon 1, CH-8356 Ettenhausen, Switzerland. E-mail: ralph.stoop@agroscope.admin.ch

Prof. Dr. Dejan Seatovic is a full professor at ILT Institut für Laborautomation und Mechatronik of the Eastern Switzerland University of Applied Sciences (OST), Oberseestrasse 10, CH-8640 Rapperswil, Switzerland

Acknowledgments

We would like to thank Roland-Hassan Nasser, Patrick Meyer, Sebastian Kehl, Alexander Lehrmann, and Timo Höfer for fruitful discussions within the Innosuisse project “Automated broad-leaved dock control” and Mickey Shavit-Zepeda for carefully reading the manuscript. We would like to thank the Innosuisse Swiss Innovation Agency (project 47541.1 IP-EE) for funding.