

## Optimizing high-resolution multi-view drone imaging for detecting foreign grains in gluten-free oat production fields

Roope Näsi<sup>1</sup>, Raquel A. Oliveira<sup>1</sup>, Stefan Rua<sup>1</sup>, Ehsan Khormashahi<sup>3</sup>, Axel Päivänsalo<sup>1</sup>, Oiva Niemeläinen<sup>2</sup>, Eija Honkavaara<sup>1</sup> and Markku Niskanen<sup>2</sup>

<sup>1</sup> Dept. of Remote Sensing and Photogrammetry, Finnish Geospatial Research Institute (FGI) in National Land Survey of Finland (NLS), Vuorimiehentie 5, 02150, Espoo, Finland - (raquel.alvesdeoliveira, roope.nasi, stefan.rua, axel.paivansalo, eija.honkavaara)@nls.fi

<sup>2</sup> Natural Resources Institute Finland (Luke), Finland – (markku.niskanen@luke.fi, oivaniemelainen@gmail.com)

<sup>3</sup> Dept. of Geographical and Historical Studies, Faculty of Social Sciences and Business Studies, University of Eastern Finland (ehsan.khoramshahi@uef.fi)

**Keywords:** UAV, drone, remote sensing, photogrammetry, agriculture, plant classification, deep learning, oat, clustering, multi-view

### Abstract

To reduce the high cost of manually detecting and removing gluten-containing grains from oat crops, drone imaging and deep learning can be used to automate the detection process. In a previous work, a multi-image object detection approach was proposed utilizing high-resolution RGB images captured by a drone using multi-view technology, including nadir and four oblique angles. The images were georeferenced using bundle block adjustment (BBA), and a semi-supervised object detection model (Faster R-CNN) was trained to identify foreign grains. The detector outputs were projected into ground coordinates using a photogrammetric technique. These coordinates were then analyzed using a clustering approach to generate a detection map of barley plant locations. In this study focused on three main objectives. First, it aimed to optimize parameters related to the clustering phase. Second, it evaluated drone data capture settings by assessing whether fewer images could maintain acceptable detection accuracy to reduce flight time. Third, it tested whether direct georeferencing could produce results comparable to those obtained using BBA-based georeferencing. The study showed that using fewer images—for example, two view angles and a side overlap of 80%—could maintain good detection accuracy (omission error of 0.14 and commission error of 0.27). This setup would reduce data collection time from 100 min/ha to 40 min/ha—a substantial improvement for practical field operations. Direct georeferencing showed promising practical results, even though error statistics increased slightly compared to BBA-based georeferencing. These improvements could significantly reduce data capture and processing time, representing a meaningful step toward a practical, cost-effective solution for end-users aiming to detect weedy foreign barley in gluten-free oat production fields.

### 1. Introduction

Oats are part of a healthy diet and are especially important for people with celiac disease, who cannot consume gluten-containing grains such as wheat and rye (Butt et al. 2008; Varma et al. 2016). Although oats are naturally gluten-free, the purity of the final product can be compromised by the presence of foreign grains in the crop. Contamination on the farm can occur through various means, such as contaminated seed material, seed transportation by birds and animals, or flooding. Currently, the detection and removal of foreign grains are done manually, with workers inspecting fields on foot to identify and remove them. This monotonous and labor-intensive phase significantly increases the cost of producing gluten-free oats. The process could be optimized by developing methods that use drone imaging and artificial intelligence to detect foreign grains.

Remote sensing technologies have the potential to provide tools for detecting foreign grains in the crop, enabling farmers to manage their fields more effectively. We previously developed a deep learning-based multi-image object detection method to identify individual foreign grains in oat fields (Khoramshahi et al., 2023). The approach utilizes high-resolution RGB images captured by a drone using multi-view technology, including nadir and oblique angles (front, back, left, and right), allowing each individual plant to theoretically appear in multiple images. This enhances the visibility of even those individuals growing beneath the main vegetation layer. Then images were georeferenced using bundle block adjustment (BBA) and a semi-supervised object detection architecture (Faster R-CNN),

was trained to detect foreign grains. The outputs of the object detector were transformed into ground coordinates using a photogrammetric technique. These coordinates were then analyzed using a clustering approach to generate a probabilistic map of the barley plant locations. The results were promising, with over 80% of the foreign barley plants successfully detected (Khoramshahi et al., 2023).

However, it is essential to optimize the entire process chain — from drone flight to calculation and data transfer — to ensure that the results are available to the user (human or robot) as quickly as possible. In this context, the objectives of this study were to 1) test whether the BBA, which requires a lot of computing time and power, could be omitted, i.e. whether we can achieve sufficient accuracy results with direct georeferencing, 2) test if less images could still provide acceptable accuracy in order to speed up the drone flight time, and 3) simplify and optimize hyperparameters related to clustering phase.

### 2. Materials and Methods

The simplified methodology pipeline, starting from multi-angle drone data capture and ending to detection map with ground coordinates to support farmer in plant removal, is presented in Figure 1. This study focuses to two phases: 1) to optimize drone data capture and processing and 2) to optimize the 3D localization and clustering phase.

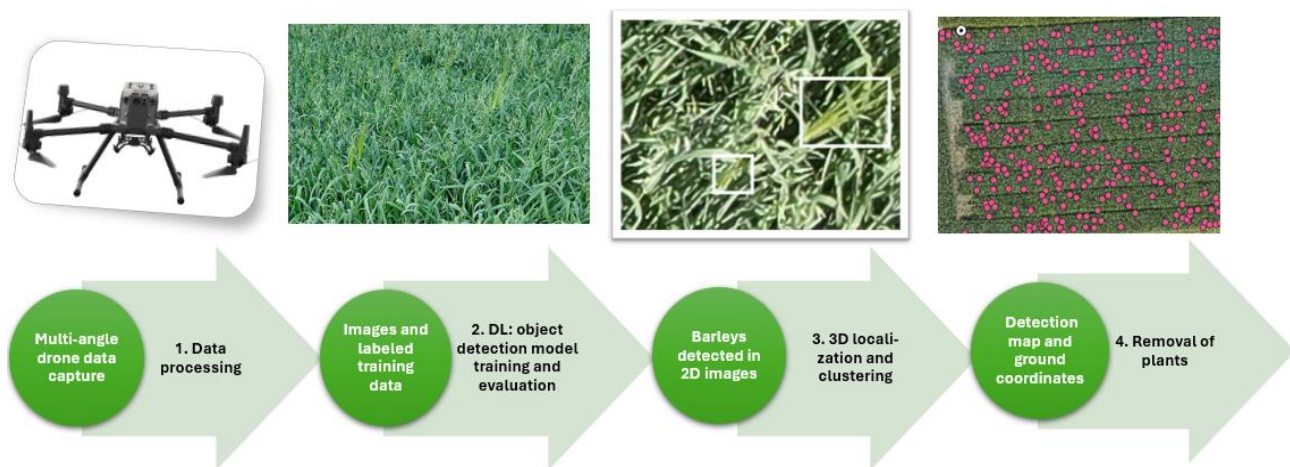


Figure 1. Methodological framework for detecting foreign grains in gluten-free oat production fields using high-resolution multi-view drone imaging.

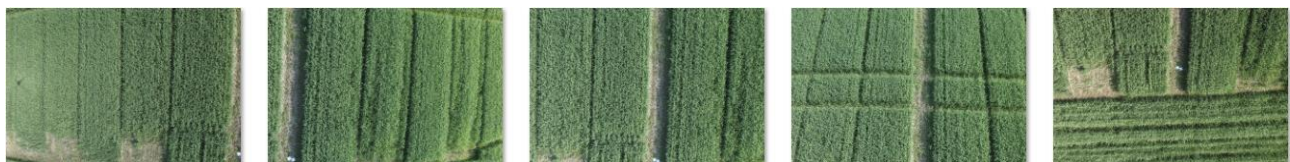


Figure 2. Five examples of multi-view images from study area. Starting from left: left, right, nadir front and back direction. Total number of images used in the analysis was 1495.

## 2.1 Datasets

A multi-angle drone dataset was captured using a DJI Matrice 300 RTK drone, equipped with a Zenmuse P1 RGB camera (8192 × 5460 pixels, 35 mm lens), a dual-frequency GNSS receiver, and an inertial measurement unit (IMU). The data collection took place on 9 July 2021 at 14:00 local time. The flight altitude was 12 meters, resulting in a ground sample distance (GSD) of approximately 1.6 mm for the images. The flight was conducted in “Smart Oblique Capture” mode, which acquires oblique images in four directions (left, right, back, and front) at a 66-degree tilt angle, in addition to nadir images (Figure 2). Image overlaps, based on nadir images, were set to 50% in the frontal direction and 80% in the side direction. When including all five view directions, each point in the study area was theoretically visible in 50 images. These settings resulted in a total drone flight time of 100 minutes to cover a 1-hectare field area.

The study trial was established on 29 May 2021 in Ilmajoki, Finland. The oat cultivar was sown at a standard sowing rate of 500 seeds/m<sup>2</sup>, with additional barley seeds mixed in at a rate of 0.5–1 seeds/m<sup>2</sup>. A detailed field survey was conducted to detect and record the coordinates of all foreign grain individuals within the oat stand. In total, 524 barley plants were identified. Ground control points were installed in the field, and their coordinates were measured during data collection.

A semi-supervised object detection framework, namely Unbiased Teacher v2 (Liu et al., 2022), using the Faster R-CNN architecture (Ren et al., 2016), was trained to detect foreign grains. At the image level, an average precision of 95% was achieved (see Khoramshahi et al., 2023, for more details). This study does not focus on the object detection phase itself but uses

the same model and image detections from Khoramshahi et al. (2023) to optimize subsequent processing phases using the dataset. The outputs of the object detector include the image coordinates of detected objects and a confidence score indicating the model’s certainty about each prediction. In this study, we filtered out all detections with a confidence score below 0.5. (Note: in Khoramshahi et al. (2023), the confidence threshold was set to 0.9 before the clustering phase.) The median confidence score was higher for nadir, right, and back images compared to left and front images (Figure 3).

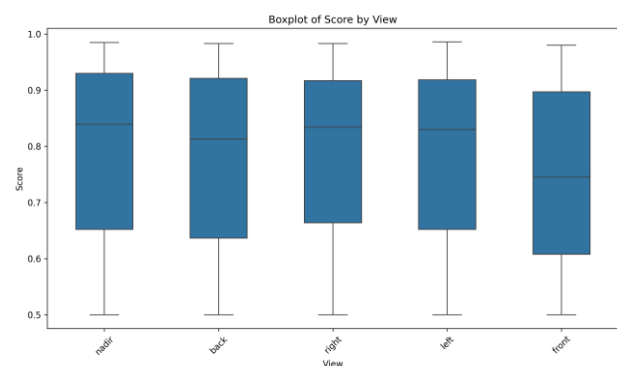


Figure 3. A boxplot visualizing the confidence score provided by object detector to all five view directions in the image dataset. Please note that detections where confidence score was lower than 0.5 was removed for the analysis.

## 2.2 Methods

In the 3D localization phase, the image coordinates of detected objects (i.e., the 2D positions of objects within the images) are transformed into real-world ground coordinates (i.e., 3D positions in a geographic coordinate system). This transformation is achieved using the classic photogrammetric collinearity equations. For simplification of ground coordinate estimation, a constant average altitude of the field was assumed for the study area. This was convenient as the area has very small variation in the terrain height.

Object detection using multi-view image datasets yielded multiple unlabelled points per foreign grain, and these points are slightly scattered instead of aligning perfectly at one ground location. Thus, DBSCAN (Density-Based Spatial Clustering of Applications with Noise) clustering method (Ester et al. 1996) was employed to analyse the large number of points from the object detection projected to ground coordinates. DBSCAN is a powerful density-based clustering algorithm well-suited for discovering clusters of various shapes and handling noise in data. It is widely used in various applications, from spatial data analysis to anomaly detection but it sensitive of choosing right parameters (Schubert et al. 2017). Thus, the algorithm's parameters—*epsilon* (the maximum distance between two samples to be accepted to add in same cluster) and *min\_samples* (the minimum number of samples in a neighbourhood for a point to be considered a core point)—were optimized using a simple grid search to enhance clustering accuracy. The range in grid search for *min\_samples* was from 3 to 15 and for *epsilon* 0.1 to 0.35 m. The final coordinate for each foreign grain was determined by calculating the mean position of all points within its corresponding cluster. An example of clustering in the test field is presented in Figure 4.

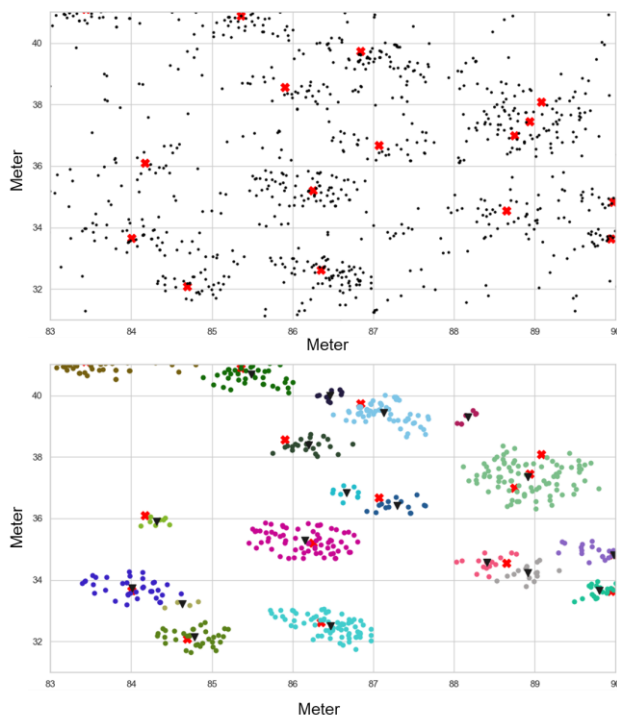


Figure 4. Up: An example of unlabelled points in the ground as a result of the multi-view object detection and 3D localization (red “x” indicates the locations of the ground truth). Down: Points clustered with DBSCAN method (black triangles indicate the centre coordinates of clusters). The points that are not considered to part of any cluster are not presented in this image.

Collecting images from five different view angles (four oblique + one nadir), as proposed in previous work, increases considerable the duration of drone-based data collection compared to setups utilizing fewer viewing angles. To investigate the potential for optimizing this process, various data collection scenarios were tested by altering the combinations of oblique and nadir images. In addition, a test of decreasing side overlap from original 80% to 60% by removing every second flight line was performed.

To test whether the BBA, which requires a lot of computing time and power, can be omitted in the data processing phase, i.e. whether we can get sufficiently accurate results with direct georeferencing, direct georeferencing results were firstly compared to those obtained through bundle block adjustment (BBA) using ground control points. Next, the accuracy of foreign grain detection was evaluated by comparing the estimated grain locations to ground-truth positions measured with a GNSS RTK receiver.

Two performance metrics were used to evaluate the detection on the ground level: Omission error =  $FN / (TP + FN)$  and commission error =  $FP / (TP + FP)$ , where FP is the false positive, TP is the true positive, FN is the false negative, TN is the true negative. Omission error indicates how many foreign grains were missed compared to the ground truth field data, while commission error reflects how many of the detected grains were not actually present in the ground truth data—i.e., a high commission error means the method is detecting too many false positives. As the false negatives should have higher cost than false positives in this application, i.e. it is worse to miss actual foreign grains than suspect too many of those, omission error was minimized. However, as omission and commission errors are often trade-offs, only tests where a commission error was less than 0.4 was considered when looking the minimum omission errors. In addition, root mean square error (RMSE) was used as performance metric, when comparing estimated coordinates by the method to coordinates measured directly in the field.

## 3. Results and Discussion

### 3.1 Optimizing parameters for BBA-based georeferencing

The grid search for optimizing DBSCAN clustering parameters showed that omission error was typically high when *epsilon* (the maximum distance between two samples to be considered part of the same cluster) was too small, while commission error tended to be high when *min\_samples* was too low (Figure 5). For example, the lowest omission error—under the condition that commission error does not exceed 0.4—was achieved with *epsilon* = 0.23 and *min\_samples* = 5 when using nadir and left-view images (Figure 5). The grid resulted in slightly different parameters to different scenarios (Table 1 and 2). Optimal value for *Min samples* varied from 3 to 8 and *epsilon* (*Max dist.*) from 0.23 to 0.35 m.

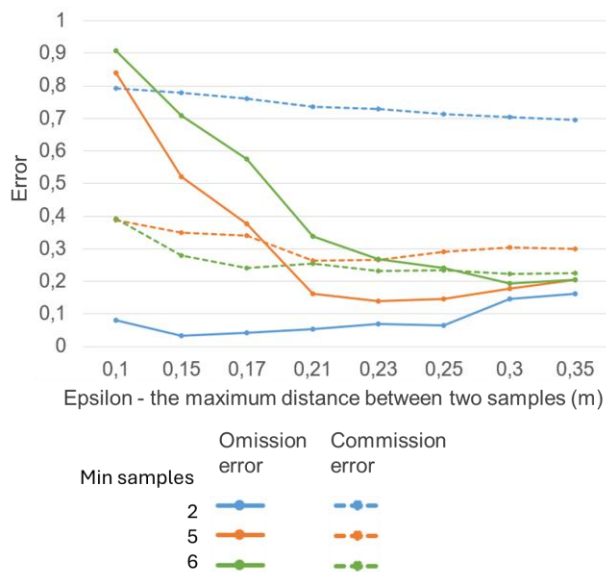


Figure 5: Optimization of DBSCAN parameters min\_samples and epsilon using nadir and left-view images. Note that min\_samples values of 4 and 5 were also analyzed but are omitted from the figure to improve readability.

| View angles      | Omission error | Commission error | RMSE (m) | Min samples | Max dist. (m) | Flight time / ha (min) |
|------------------|----------------|------------------|----------|-------------|---------------|------------------------|
| back             | 0.25           | 0.37             | 0.34     | 4           | 0.35          | 20                     |
| front            | 0.30           | 0.38             | 0.34     | 5           | 0.35          | 20                     |
| left             | 0.27           | 0.34             | 0.34     | 4           | 0.35          | 20                     |
| right            | 0.16           | 0.38             | 0.33     | 3           | 0.35          | 20                     |
| nadir            | 0.13           | 0.30             | 0.15     | 3           | 0.35          | 20                     |
| front+left       | 0.26           | 0.30             | 0.31     | 6           | 0.35          | 40                     |
| nadir+left       | 0.14           | 0.27             | 0.20     | 5           | 0.23          | 40                     |
| nadir+front+left | 0.19           | 0.34             | 0.23     | 6           | 0.25          | 60                     |
| all              | 0.19           | 0.34             | 0.24     | 8           | 0.23          | 100                    |

Table 1: The best results for each individual view angle and their selected combinations using the original **80% side overlap** in nadir images and the DBSCAN parameters (“Min samples”: minimum number of points required in a neighbourhood for a point to be classified as a core point, and “Max dist.”: Maximum distance between two samples for one to be considered within the other’s neighbourhood) at the time they were achieved. RMSE (Root Mean Square Error) was calculated by comparing estimated and measured coordinates. Flight time per hectare (min/ha) indicates the data collection efficiency in each scenario. BBA-based georeferencing.

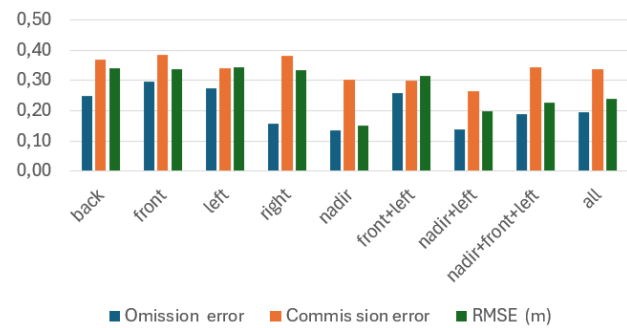


Figure 6: Omission and commission error and RMSE Root Mean Square Error) for each image views and combination.

The lowest omission errors were achieved when using nadir and nadir+ left datasets (Table 1, Figure 6). When comparing these two, nadir had smaller RMSE value but higher commission error than nadir+left.

| View angles      | Omission error | Commission error | RMSE (m) | Min samples | Max dist. (m) | Flight time / ha (min) |
|------------------|----------------|------------------|----------|-------------|---------------|------------------------|
| back             | 0.20           | 0.12             | 0.37     | 3           | 0.35          | 10                     |
| front            | 0.37           | 0.11             | 0.37     | 4           | 0.35          | 10                     |
| left             | 0.27           | 0.09             | 0.45     | 3           | 0.35          | 10                     |
| right            | 0.22           | 0.09             | 0.32     | 4           | 0.35          | 10                     |
| nadir            | 0.40           | 0.06             | 0.17     | 3           | 0.35          | 10                     |
| front+left       | 0.28           | 0.12             | 0.36     | 5           | 0.35          | 20                     |
| nadir+left       | 0.22           | 0.09             | 0.32     | 4           | 0.35          | 20                     |
| nadir+front+left | 0.22           | 0.16             | 0.30     | 5           | 0.3           | 30                     |
| all              | 0.24           | 0.12             | 0.24     | 6           | 0.3           | 50                     |

Table 2: The best results for each individual view angle and their selected combinations using the **60% side overlap** in nadir images and the DBSCAN parameters (“Min samples”: minimum number of points required in a neighborhood for a point to be classified as a core point, and “Max dist.”: Maximum distance between two samples for one to be considered within the other’s neighbourhood) at the time they were achieved. RMSE (Root Mean Square Error) was calculated by comparing estimated and measured coordinates. Flight time per hectare (min/ha) indicates the data collection efficiency in each scenario. BBA-based georeferencing.

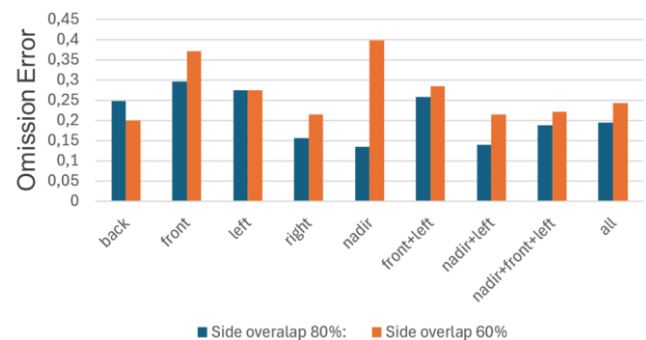


Figure 7: Omission error for each image views and combination in the case of 80 % and 60 % of side overlap.

As expected, omission error generally increased as side overlap decreased from 80% to 60% (Figure 7; Tables 1 and 2), with the most significant change observed when using only nadir images. The only exception was the back-view image dataset.

For example, in the nadir + left configuration, omission error rose from 0.14 to 0.22.

These findings suggest that using two view angles and a side overlap of 80%, or slightly less, would be an optimal approach for this application. This setup would reduce data collection time from 100 min/ha to 40 min/ha, which is a remarkable improvement for practical field operations.

### 3.2 Optimizing parameters for direct georeferencing

Considering the geometrical accuracies based on ground control points, the horizontal and vertical accuracies achieved using BBA were 0.035 m and 0.1 m, respectively. In contrast, direct georeferencing resulted in accuracies of 0.27 m horizontally and 0.13 m vertically. This would indicate that direct georeferencing would provide enough accurate coordinates for this application.

However, based on the ground truth barley locations and the estimated locations, the horizontal RMSE of foreign barley detection was 24 cm when using BBA and 35 cm when using direct georeferencing (Table 1, Table 3, Figure 8). In addition, with optimized parameters and using the full image dataset, the omission and commission errors were 0.19 and 0.34, respectively, when using BBA, and 0.39 and 0.35 when using direct georeferencing (Table 1, Table 3, Figure 8).

| View angles      | Omission error | Commission error | RMSE (m) | Min samples | Max dist. (m) | Flight time / ha (min) |
|------------------|----------------|------------------|----------|-------------|---------------|------------------------|
| back             | 0.37           | 0.35             | 0.40     | 5           | 0.35          | 20                     |
| front            | 0.61           | 0.39             | 0.59     | 5           | 0.3           | 20                     |
| left             | 0.47           | 0.40             | 0.51     | 7           | 0.35          | 20                     |
| right            | 0.77           | 0.36             | 0.54     | 4           | 0.3           | 20                     |
| nadir            | 0.33           | 0.38             | 0.34     | 3           | 0.35          | 20                     |
| front+left       | 0.47           | 0.40             | 0.51     | 7           | 0.35          | 40                     |
| nadir+left       | 0.31           | 0.38             | 0.31     | 5           | 0.35          | 40                     |
| nadir+front+left | 0.52           | 0.32             | 0.40     | 9           | 0.35          | 60                     |
| all              | 0.39           | 0.35             | 0.35     | 10          | 0.3           | 100                    |

Table 3: The best results for each individual view angle and their selected combinations using the **original 80% side overlap** in nadir images and the DBSCAN parameters (“Min samples”: minimum number of points required in a neighbourhood for a point to be classified as a core point, and “Max dist.”: Maximum distance between two samples for one to be considered within the other’s neighbourhood) at the time they were achieved. RMSE (Root Mean Square Error) was calculated by comparing estimated and measured coordinates. Flight time per hectare (min/ha) indicates the data collection efficiency in each scenario. Direct georeferencing.

When every second flight line was removed from the analysis—resulting in 60% side overlap in nadir images—direct georeferencing produced slightly larger errors compared to BBA-based georeferencing (Table 4, Figure 8). For both overlap configurations, RMSE was 25 cm using BBA and 35–37 cm with direct georeferencing (Figure 8).

There was more variation between the different view angles in the direct georeferencing results than in the BBA-based results.

Interestingly, using only the back-view images resulted in smaller errors than any other view in the case of 60% overlap (Table 4).

This particular dataset had a relatively high density of foreign grains due to its nature as a training trial, which may have caused more confusion with nearby plants and, therefore, more clustering errors. Thus, it would be of great interest to test how direct georeferencing would perform in a practical gluten-free oat production field with a lower density of foreign grains.

| View angles      | Omission error | Commission error | RMSE (m) | Min samples | Max dist. (m) | Flight time / ha (min) |
|------------------|----------------|------------------|----------|-------------|---------------|------------------------|
| back             | 0.23           | 0.25             | 0.41     | 3           | 0.35          | 10                     |
| front            | 0.63           | 0.32             | 0.62     | 5           | 0.35          | 10                     |
| left             | 0.76           | 0.39             | 0.58     | 3           | 0.23          | 10                     |
| right            | 0.79           | 0.36             | 0.51     | 3           | 0.3           | 10                     |
| nadir            | 0.70           | 0.34             | 0.34     | 3           | 0.35          | 10                     |
| front+left       | 0.65           | 0.36             | 0.57     | 6           | 0.35          | 20                     |
| nadir+left       | 0.44           | 0.37             | 0.32     | 4           | 0.35          | 20                     |
| nadir+front+left | 0.53           | 0.36             | 0.43     | 6           | 0.3           | 30                     |
| all              | 0.28           | 0.39             | 0.38     | 7           | 0.35          | 50                     |

Table 4: The best results for each individual view angle and their selected combinations using the **60% side overlap** in nadir images and the DBSCAN parameters (“Min samples”: minimum number of points required in a neighborhood for a point to be classified as a core point, and “Max dist.”: Maximum distance between two samples for one to be considered within the other’s neighborhood) at the time they were achieved. RMSE (Root Mean Square Error) was calculated by comparing estimated and measured coordinates. Flight time per hectare (min/ha) indicates the data collection efficiency in each scenario. Direct georeferencing.

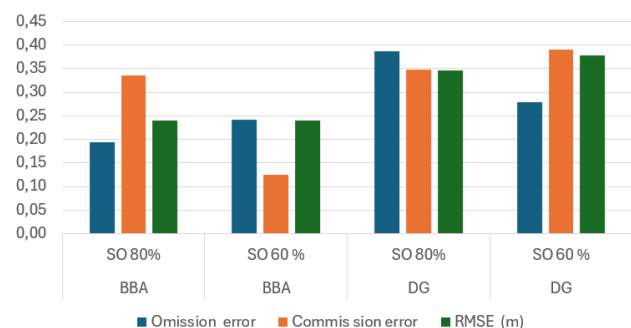


Figure 8. Omission and commission error and RMSE (Root Mean Square Error) when using 80 and 60 % overlap (SO) based on BBA (bundle block adjustment) and DG (direct georeferencing)

## 4. Conclusions and Future Work

This study showed that drone-based data collection and processing for foreign grains detection can be significantly speeded up achieving comparable accuracy than in previous study (Khoramshahi et al. 2023). For example, using only two view angles and a side overlap of 80% still maintained good detection accuracy. This setup alone would reduce data collection time from 100 minutes/ha to 40 minutes/ha—a substantial improvement for practical field operations. In

addition, the relatively good results achieved with only 60% side overlap suggest that the optimal overlap may be less than 80%, which could further shorten flight time. Direct georeferencing showed promising practical results, even if error statistics increased slightly compared to BBA-based georeferencing. This represents a significant improvement towards delivering a practical, cost-effective solution for the end-users aiming to detect weedy foreign barleys in gluten-free oat production fields. An important future work is to test and improve the full approach in practical gluten-free oat production field. In addition, we are developing tools to also assist foreign grain removal, not only detect and provide locations of them. Firstly, a mobile-based solution to guide the person on the optimal route to found foreign grains is planned to test. Secondly, automatic removing of foreign grains is studied using a cutter attached to the drone.

### Acknowledgements

This research was funded by European Union via regional (South Ostrobothnia) European Innovation Partnership (EIP) project “Using mobile tools to get rid of foreign grains - MOVI” (project id 452616). The dataset was collected during “Identification of unwanted cultivated species in seed multiplication and in gluten free oat cultivation - DrooniLuuppina” project number 144989 funded by the European Agricultural Fund for Rural Development (EAFRD). Arja Nykänen from SeAMK is thanked for her co-operation at the Ilmajoki trial site. Additionally, we would like to Niko Koivumäki and Teemu Hakala from the National Land Survey of Finland, for their valuable support during the project.

### References

- Butt, M. S., Tahir-Nadeem, M., Khan, M. K. I., Shabir, R., & Butt, M. S. (2008). Oat: unique among the cereals. *European journal of nutrition*, 47, 68-79.
- Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd* (Vol. 96, No. 34, pp. 226-231).
- Khoramshahi, E., Näsi, R., Rua, S., Oliveira, R. A., Päivänsalo, A., Niemeläinen, O., ... & Honkavaara, E. (2023). A novel deep multi-image object detection approach for detecting alien barleys in oat fields using RGB UAV images. *Remote Sensing*, 15(14), 3582.
- Liu, Y. C., Ma, C. Y., & Kira, Z. (2022). Unbiased teacher v2: Semi-supervised object detection for anchor-free and anchor-based detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9819-9828).
- Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*, 39(6), 1137-1149.
- Schubert, E., Sander, J., Ester, M., Kriegel, H. P., & Xu, X. (2017). DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems (TODS)*, 42(3), 1-21.
- Varma, P., Bhankharia, H., & Bhatia, S. (2016). Oats: A multi-functional grain. *Journal of Clinical and Preventive Cardiology*, 5(1), 9-17.