

A Robust Integrated Approach for Near Real-Time Seamless Orthomosaic Generation Using Off-the-Shelf UAVs for Ultra-High Resolution Mapping Applications

Muhammad Farhan Humayun, Ali Salmasi, Jiaqiang Zhang, Tomi Westerlund, Jukka Heikkonen

Department of Computing, University of Turku, FI-20014, Finland
(farhan.humayun, ali.a.salmasi, jiaqiang.zhang, toveuwe, jukhei)@utu.fi

Keywords: Orthomosaicing, Georeferencing, Small Scale UAVs, Aerial Imaging, UAV Mapping

Abstract

The widespread availability of small-scale, off-the-shelf Unmanned Aerial Vehicles (UAVs) has opened various avenues for drone-based photography and videography, where limited human expertise is needed to carry out drone flying and post-processing of the acquired media. However, these off-the-shelf, closed-loop systems lack the flexibility and range of sensors/metadata to perform specialized operations, such as accurate mapping for Geographical Information Systems (GIS) applications. To achieve survey-grade precision, often larger UAV platforms are required, which can carry more sophisticated payloads leading to an increase in the overall cost of the systems while introducing sensor calibration complexities. To address these issues, we present an efficient approach using progressive, sequence-based feature matching and image alignment to generate high-quality, seamless, orthorectified mosaics using low-cost UAVs equipped with general-purpose imaging and positioning sensors. Our proposed approach does not require any prior knowledge of the environment or specific flight planning protocols. We validate the proposed approach through real-world, outdoor experiments by collecting and processing flight data and generating geo-referenced orthomosaics. The results showcase that our approach delivers operationally acceptable accuracy while achieving a mosaicking speed of 2.3 frames per second using consumer-grade UAV platforms with minimal sensor data. The proposed solution aims to enable rapid and cost-effective, UAV-based mapping for applications in forestry, hydrology, urban planning and disaster management.

1. Introduction and Related Works

In the past couple of decades, advancements in hardware manufacturing technology have paved the way for miniaturized Unmanned Aerial Vehicles (UAVs) commonly known as off-the-shelf drones. These low-cost imaging platforms offer unprecedented flexibility, spatial resolution, and temporal frequency for capturing geospatial information. The integration of UAVs into photogrammetry and remote sensing has facilitated the generation of high-resolution orthomosaics and geo-referenced datasets, vital for a variety of applications including precision agriculture, forestry (Remondino et al., 2011), infrastructure monitoring, urban planning (Zhang and Zhu, 2023) and disaster assessment (Mohd Daud et al., 2022).

Typically, UAVs fly at low altitudes compared to the Earth observation (EO) satellites which typically operate in the Low Earth Orbit (LEO). Therefore, they have narrower swath widths compared to their satellite counterparts. Due to less area coverage, a basic prerequisite for the utilization of drone-based imagery in mapping applications is the accurate orthomosaicing of the image or video frames captured from the drone. Orthomosaicing refers to the process of generating geometrically corrected, seamless composites that represent the Earth's surface without perspective distortions.

Another process that is commonly carried out before the images can be used in Geographical Information Systems (GIS) applications is geo-referencing. This involves accurate knowledge of the geographical positioning of the UAV along with the calibration of the imaging sensors. Various studies have emphasized the necessity of careful calibration of imaging and navigational sensors on UAV platforms for generating accurate results. Some of the widely used techniques include Bundle Block Adjustment (BBA) and lever-arm calibration (Turner et

al., 2014; Daakir et al., 2015). Camera calibration involves the precise determination of intrinsic and extrinsic orientation parameters. Intrinsic parameters involve the internal characteristics of the camera including focal length and lens distortions, whereas extrinsic parameters are related to the external positioning and orientation of the camera in the 3D space (Abeho et al., 2024; Pérez et al., 2011).

While commercial UAVs are commonly used these days for regular photography, they are generally equipped with low-grade, closed-loop sensors to keep the cost and weight of the UAV at relatively lower levels. Closed-loop UAV systems, in the context of this work, refer to commercial, off-the-shelf drone platforms in which the onboard sensors (e.g., GPS, IMU, and camera) are tightly integrated and managed internally by the flight controller, without providing direct access to raw sensor data streams. As a result, users are restricted from independently accessing and processing the data and from possibly integrating other sensors with the platform, limiting the system's flexibility for advanced photogrammetric or georeferencing tasks. As a result, it is often difficult to generate reasonable results with such systems. For the majority of the mapping applications, UAV platforms with sophisticated payloads such as Real Time Kinematics (RTK) based GNSS sensors, highly sensitive Inertial Measurement Units (IMUs) and imaging systems with minimal distortions are used (Hutton et al., 2020; Oliveira et al., 2018). This leads to an increase in the overall cost of the UAV platforms as well as an addition to their size and weight, making it difficult to operate them under varying environmental conditions.

In this paper, we present a streamlined workflow for generating geo-referenced orthomosaics using sensor data from off-the-shelf, low-cost UAVs. Our proposed approach is based on a progressive image alignment pipeline, where we initially de-

signed an in-door experimental setup to precisely determine the angular field of view (AFoV) of the UAV camera. This is used to determine and validate the intrinsic parameters of the camera and for calculating the image swaths in the direct geo-referencing approach. Once the parameters are established, we integrate them with robust feature matching and image alignment modules to generate the desired outputs. The main contribution of the paper is the near real-time generation of seamless orthomosaics using progressive sequence-based image matching while without utilizing specialized sensors such as RTK and IMU data. Unlike traditional commercial photogrammetry software (e.g., Pix4D, and Agisoft Metashape) that rely on dense Structure-from-Motion (SfM), Multi-View Stereo (MVS) reconstructions and Digital Surface Models (DSMs), our approach avoids heavy computation by using a progressive, sequence-aware matching and transformation pipeline. The methodology is specifically optimized for off-the-shelf UAVs without requiring RTK-GNSS or detailed flight planning, making it more applicable to low-cost deployments.

1.1 Types of Geo-referencing

Geo-referencing techniques can be broadly categorized into two types; direct and indirect methods. Direct geo-referencing involves real-time onboard GNSS and IMU data to determine image orientation and position (Mostafa and Schwarz, 2001). While this method is efficient and scalable, its accuracy is often limited by sensor calibration, synchronization issues, and environmental noise (Turner et al., 2014). Pfeifer et al. (2012) introduced an integrated approach using a range of onboard sensors and photogrammetric test field calibration. Hemerly (2014) presented an automatic geo-referencing approach through precise estimation of the pixel positions of image corners in the model plane. Khoramshahi et al. (2019) proposed a framework for overcoming the synchronization and calibration issues through the estimation of interior camera parameters and precise alignment of the imaging and navigational sensors for a multi-projective camera system. Similarly, Zeybek et al. (2023) proposed a methodology based on post-processing kinematic (PPK) to improve the overall positional accuracy of the UAV sensors contributing to improved geo-referencing results.

Indirect geo-referencing relies on the establishment of Ground Control Points (GCPs), which serve as known reference locations in the imagery and ground space. Although typically more accurate, it requires time-intensive fieldwork and becomes impractical in many operational contexts. Jóźków and Toth (2014) compared the results of direct and in-direct geo-referencing methods on UAV images and concluded that the in-direct methods provide better accuracy. To address these challenges, hybrid approaches combining direct measurements with image-based tie-point matching using computer vision algorithms and camera self-calibration techniques have emerged as promising alternatives (Faraji et al., 2016; Haala et al., 2022).

1.2 Orthomosaicing Techniques

Traditional orthomosaicing involves correcting each image for lens distortion, terrain relief, and sensor orientation, then stitching them based on overlapping areas and matching features. To enhance automation and robustness of the process, Faraji et al. (2016) proposed a computer vision-based orthorectification pipeline that uses feature point detection and robust matching. Similarly, Angel et al. (2020) proposed a computationally efficient methodology to geo-rectify and mosaic

UAV-based hyper-spectral images using Speeded-Up Robust Features (SURF) and Maximum Likelihood Estimator Sample Consensus (MLESC) algorithms.

Other Modern workflows integrate techniques such as SfM, MVS and DSMs to automate 3D scene reconstruction and image alignment. However, these workflows are sensitive to camera calibration, image overlap, lighting conditions, and surface complexity. Ludwig et al. (2020) introduced a reproducible orthomosaic generation workflow incorporating automated point cloud filtering to minimize checkpoint errors. Kern et al. (2020) developed an open source framework for automatic orthomosaicing and geo-referencing of a continuous stream of UAV images with multiple options including 2D mosaicing and 3D reconstruction.

1.3 Machine Learning based Advancements

Recent studies have focused on multi-sensor fusion and deep learning approaches to improve the results of orthomosaicing. For instance, Park et al. (2022) introduced a deep learning framework for image segmentation and inpainting to remove transient objects (e.g., vehicles) to improve the results of orthorectification. Similarly, Kern et al. (2020); Fanta-Jende et al. (2023) proposed techniques for integrating visual-inertial odometry, GNSS data, and semantic segmentation to generate spatially accurate, context-rich maps. Recent studies have demonstrated the generation of highly precise digital orthophotos using Neural Radiance Fields (NeRF) (Chen et al., 2024; Yue et al., 2025). NeRF is a fully connected neural network model for 3D reconstruction of scenes based on the estimation of volume density and emitted radiance at specific spatial positions and viewing angles. The downside of these deep learning approaches is the requirement of training data and large computational resources.

2. Methodology

The proposed seamless orthomosaic generation pipeline is described in this section. It consists of various steps including experimental setup, robust feature detection and matching, calculation of intrinsic and extrinsic camera parameters, image transformations, seamless mosaicing and geo-referencing. The following subsections provide detailed descriptions of each module involved in the approach.

2.1 Determining the Camera AFoV

Since the UAV that is part of the study is a closed loop system and the Angular Field of View (AFoV) of the camera is undocumented, we initially design an indoor setup to precisely determine the AFoV. Figure 1(a) shows the imaging geometry of the drone. Here the UAV camera is pointing straight down at the nadir angle which corresponds to a gimbal tilt of -90° and a gimbal pan equal to 0° . Figure 1(b) shows the dimensions of the corresponding image footprint on the ground. In this case, the swath width of the frame is given by the following equation:

$$x = 2 \times h \times \tan(\theta/2) \quad (1)$$

where x is the swath width of the imaging frame at nadir viewing angle, h is the height of the UAV and θ is the AFoV of the camera. By fixing the height of the UAV and measuring the

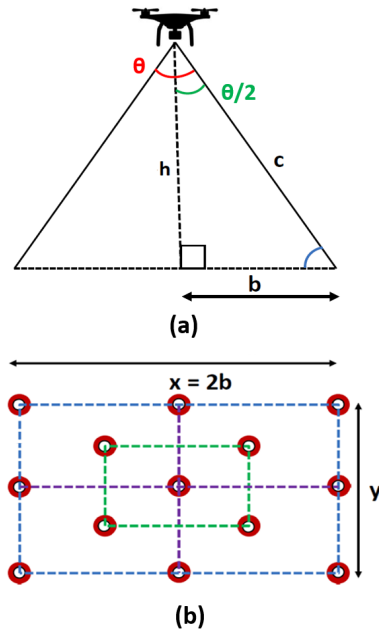


Figure 1. (a) UAV imaging geometry at Nadir view point, (b) image footprint formed on the ground.

horizontal distance on the ground, we can calculate the AFoV of the camera precisely. This is useful for the calculation of intrinsic camera parameters as well as computing swath widths and ground spatial distance (GSD) of the images. GSD is the distance between the center of two consecutive image pixels as measured on the ground and is an essential metric for quantifying the spatial resolution of an image.

2.2 Robust Feature Detection

The next step in the proposed workflow is the detection of useful features in each image frame. Different algorithms exist in the literature for the detection of key feature points in an image including Scale Invariant Feature Transform (SIFT), Oriented FAST and rBRIEF (ORB), Binary Robust Invariant Scalable Keypoints (BRISK) and accelerated KAZE (AKAZE) (Tareen and Saleem, 2018). For this particular study, we choose the SIFT algorithm because it provides sufficient features in terms of detection density as compared to ORB and AKAZE detectors. The feature detections produced by BRISK algorithm are too dense and thus have a relatively large computational overhead for outlier removal. The SIFT algorithm basically constructs a scale space by convolving the given image with a Gaussian blurring mask and then computing the difference of Gaussians as given by the following equations, respectively:

$$S(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (2)$$

$$D(x, y, \sigma) = S(x, y, k\sigma) - S(x, y, \sigma) \quad (3)$$

where S is the scale-space image, I is the original image, G is the Gaussian kernel, and k is a multiplicative constant between scales. Finally, key-point localization is performed using Taylor series expansion and orientation is assigned to each key point using gradient evaluation. The final descriptor of each key point includes its (x,y) coordinates with respect to the image frame, its scale and orientation.

2.3 Precise Feature Matching

Next, we perform feature matching using the Best of Two Nearest Neighbours (Bo2NN) approach. First, the L2 norm of each descriptor is calculated using the following equation:

$$\text{distance}(\mathbf{d}_i, \mathbf{d}_j) = \|\mathbf{d}_i - \mathbf{d}_j\|_2 \quad (4)$$

This is computed for every descriptor d_i in image A against every d_j in a set of neighboring images Z , defined as:

$$Z = \{A \pm 1, A \pm 2, \dots, A \pm n\} \quad (5)$$

where A is the index of the current image in the acquisition sequence, and n determines how many adjacent images (before and after A) are considered for matching. The value of n depends on the expected overlap between consecutive images, which is influenced by the UAV's speed and the imaging rate. Here, we empirically set $n = 7$, which provides a clear computational advantage while retaining sufficient pairs of images to generate high quality mosaicing results. This optimization leverages sequential image capture to avoid matching distant, non-overlapping frames thus reducing the computational costs.

For each descriptor in the frame A , we find the two closest descriptors in the set of comparing images and perform Lowe's ratio test on them which is given by the following equation:

$$\frac{d_1}{d_2} < r \quad \text{where } r \approx 0.75 \quad (6)$$

If the above ratio is greater than or equal to 0.75, we keep the closest matched descriptor. The 0.75 threshold balances sensitivity and specificity in Lowe's ratio test, empirically maximizing accurate matches while suppressing outliers (Lowe, 2004). The Bo2NN method provides high-quality stable matches while reducing false positives in repetitive areas of the images. Next, we estimate the homography matrix H to relate the coordinates of two matching points between the images as given below:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} \sim H \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (7)$$

where H is a 3×3 matrix. To compute the outliers, we use the RANSAC algorithm which computes the Euclidean distance of the reprojection errors using the homography matrix and if this distance is greater than a certain threshold, we count the feature as an outlier. A pairwise confidence matrix is also computed to determine the confidence score corresponding to each pair of images.

2.4 Camera Parameter Estimation

Two sets of camera parameters are computed, i.e., intrinsic and extrinsic parameters. The intrinsic parameters are defined by the following matrix:

$$\mathbf{K} = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (8)$$

here, f_x and f_y are the focal lengths on the x and y axes, s is the skew coefficient and c_x, c_y are coordinates of the principal point. The intrinsic parameters were estimated using a combination of UAV camera specifications and the empirically calculated AFoV from the indoor setup. The extrinsic parameters accommodate for the external orientation of the camera including scaling, translation and rotation. These are defined by the following equation:

$$\mathbf{X}_{\text{camera}} = \lambda \cdot (\mathbf{R} \cdot \mathbf{X}_{\text{world}} + \mathbf{T}) \quad (9)$$

where $\mathbf{R} \in \mathbb{R}^{3 \times 3}$ is the rotation matrix, $\mathbf{T} \in \mathbb{R}^{3 \times 1}$ is the translation vector, $\mathbf{X}_{\text{world}} \in \mathbb{R}^{3 \times 1}$ are the real World coordinates and λ is the scale factor. Extrinsic parameters were derived by performing sequential image alignment, in conjunction with GPS-derived UAV positions.

2.5 Image Mosaicing

Once the transformation matrix of each image is computed, we apply the transformations to bring the images to a common plane before proceeding with the generation of mosaiced masks along with the corresponding placement positions of each image within the mosaiced mask. For seamless mosaicing of images, we precisely identify the areas of the images which relate to their corresponding positions in the final mosaic.

3. Experimental Setup

The following subsections describe the hardware specifications of the UAV along with the details of the indoor setup for camera AFoV estimation and the outdoor experiments for the collection of real-world imaging and positional data.

3.1 Drone and Camera Specifications

For the purpose of this study, we used the Autel Evo II Pro V2 drone which is a light-weight, off-the-shelf, commercial quadcopter equipped with an RGB camera. Figure 2 shows the snapshot of the drone along with the laser rangefinder which is used to determine the precise dimensions of the UAV footprint during the indoor experiment.



Figure 2. Autel EVO II Pro V2 Imaging Drone along with Bosch GLM 30 laser rangefinder

The specifications of the drone platform are summarized in Table 1. Due to its small size, the UAV has limited take-off weight capacity and flying time. It is equipped with a standard GNSS receiver which can receive GPS signals for determining the horizontal coordinates and the flight altitude. The GLM 30 laser rangefinder has a reported accuracy of $\pm 2\text{mm}$. The camera specifications are summarized in Table 2.

Table 1. Specifications of Autel EVO II Pro V2 Imaging Drone

Parameter	Value
Dimensions	457×558×110 mm
Max take-off Weight	2000 g
Max Flying Altitude	7000 m
Max Flying Time	40 mins
Max Hovering Time	35 mins
Gimbal Tilt Range	-90° to +30°
Gimbal Pan Range	-90° to +90°

Table 2. Specifications of the builtin RGB Camera

Parameter	Value
Sensor	1 inch CMOS
Effective Megapixels	20MP
Lens Focal Length	28.6 mm
Aperture	f/2.8 - f/11
Focus Range	1 m to infinity
Photo Resolution	5472 x 3076 (16:9)

3.2 Indoor Experiment

Figure 3 shows the indoor lab setup for calculation of the AFoV of the UAV camera. We prepare the ground area by measuring the horizontal distance ($x = 2\text{m}$) and vertical distance ($y = 1.125\text{m}$) using a laser distance meter. y is calculated using the photo dimensions as mentioned in Table 2. The UAV was flown on top of the ground setup with its camera tilted at nadir viewing angle, pointing at the marked center point in the figure 3. After that we measured the precise height of the UAV using the laser distance meter. The vertical height (h) in Figure 1 (a) is measured to be 1.56 meters. The AFoV is calculated using the equation 1 which comes out to be 65.2° .

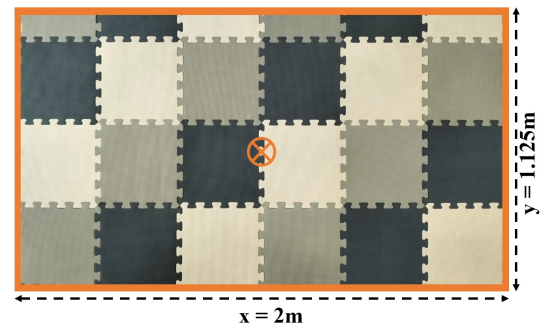


Figure 3. Visualization of ground area setup for calculating AFoV of the UAV Camera.

3.3 Outdoor Study Area

The outdoor area selected for this study is located in the city of Turku, in South-Western Finland. In Figure 4(a), the map shows the exact region of interest used for data collection. The primary motivation for choosing this site was its diverse range of surface features, including paved surfaces, a well-defined river channel, patches of soft ground covered with grass, and a pedestrian bridge. Additionally, the area's ease of access and relative distance from the densely populated city center made it a practical and convenient choice for conducting UAV-based experiments. Figure 4(b) shows the flight trajectories of the drone along with the corresponding imaging points. We performed two sweeps of the study area; one from East to West

and the second from West to East. The observed irregularities in the flight trajectories are primarily attributed to manual piloting and wind-induced drift during flight. The experiment was intentionally conducted without the use of advanced autonomous flight planning tools to simulate real-world, low-cost UAV operations under minimal supervision or technical expertise. The height of the UAV was fixed at a vertical altitude of 55 m above ground level during the entire duration of the flight in which the images were captured. Also, the camera was always pointed at the nadir angle and hence looking vertically down at all times.

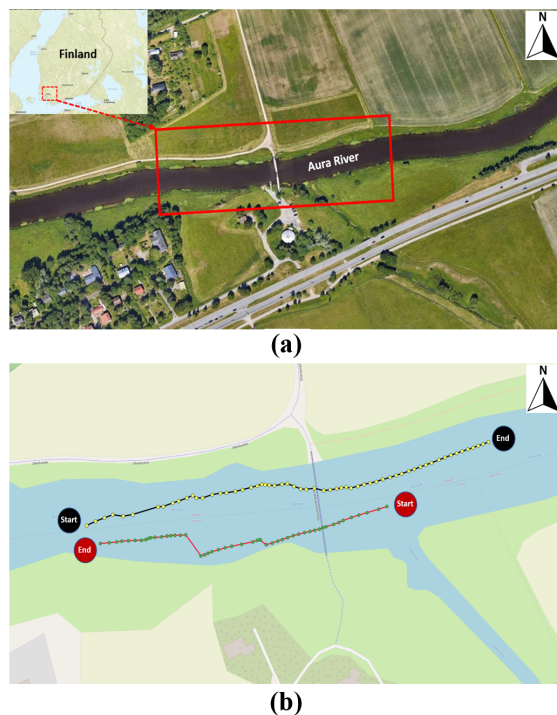


Figure 4. (a) Map of the outdoor study area on the Aura river Finland (Coordinates: 60°27'56.59"N, 22°20'6.28"E), (b) Flight trajectories of two passes: West to East and East to West.

4. Results and Analysis

To showcase the results and validate the outputs and effects of each module in the pipeline, we take the working example of the three images as depicted in Figure 5. As shown, there is a certain degree of overlap between the three images which is essential for the image matching and mosaicing process. It can also be observed that apart from the pedestrian bridge, most parts in the images do not have well-defined objects or features and this could be challenging for the feature detection and matching algorithms. These parts contain grass, snow and water.

The initial step involves the detection of good-quality features in each image for mapping to subsequent images down the processing pipeline. Figure 6 shows the output of the detected SIFT features along with the exact number of features detected for each input image. It can be seen that the SIFT algorithm effectively identifies useful features across the images. Most of the detected features lie over the bridge and towards the left side of the images where there is more ground surface to present better features as compared to the central parts of the images which contain water.

Table 3 shows the confidence matrix generated as a result of the feature matching process on the three input images. Based on

the information in the table, it can be seen that image 6 (a) has a high matching confidence score with image 6 (b). Also, image 6 (b) has a high matching score with image 6 (c). However, the corresponding confidence scores of images 6 (a) and 6 (c) are quite low. This means that image 6 (b) has distinct overlapping regions which are present separately in images 6 (a) and 6 (c).

Table 3. Pairwise, matching confidence matrix of Images as shown in Figure 6.

	Image 6 (a)	Image 6 (b)	Image 6 (c)
Image 6 (a)	0.0	2.047	0.48
Image 6 (b)	2.047	0.0	2.086
Image 6 (c)	0.48	2.086	0.0

Figure 7 illustrates the output of the feature matching process along with the number of inliers. Figure 7 (a) shows the feature matching results of images 6 (a) and 6 (b). Similarly, figure 7 (b) shows the feature matching of images 6 (b) and 6 (c), respectively. Again, it can be seen that most of the feature matches correspond to the left side of the images because it has a high density of detected features. The algorithm was able to successfully identify good feature matches across all three images, despite visual variability.

Once the transformation process is carried out using the matched features and the estimated camera parameters, the next step generates the mosaiced masks of the three images on a unified plane. This means that all the transformed images are projected to a single frame while also retaining the initial spatial positions such as corner coordinates as well as the width and height of all individual frames involved in the final mosaic. Figure 8 (a) shows the mosaiced mask of the images and Figure 8 (b) shows the corresponding positions of initial image frames within the final mosaiced mask. Based on the positions of the masks, it can be observed that there is a significant overlap (approximately 50%) in the consecutive image frames.

After that, the exact seam lines are extracted where the images are to be stitched together to form the final mosaic. This determines the actual areas of the individual images which will form part of the final product. Figure 9 illustrates the output of this step. It can be seen that the images were captured sequentially in the relative direction of the UAV flight. Therefore, image 9 (a) has regions stripped off from the top whereas, 9 (c) has regions cropped from the bottom. On the other hand, image 9 (b) has stripped areas from both the top and bottom because it forms the center part of the final output.

To enhance the visual appearance of the mosaicing results, figure 10 (a) shows the cropped regions of the individual masks within the final mosaic mask in color-coded form with three colors pertaining to the three images. Figure 10 (b) shows the final mosaic product of the three images under consideration with the seam lines drawn in red color to demarcate the boundaries of each image frame.

Figure 11 showcases the result of seamless orthomosaic generated from a set of sixty consecutive image frames captured from the UAV flight trajectories as shown in Figure 4 (b). The average number of SIFT features detected across all the images is 3176.26. Similarly, the average number of inliers retained after progressive matching and RANSAC filtering across different image pairs is 214.59. The total time taken to process sixty image frames was 26.3 seconds which provided an average mosaicing speed of 2.3 frames per second (fps). Careful

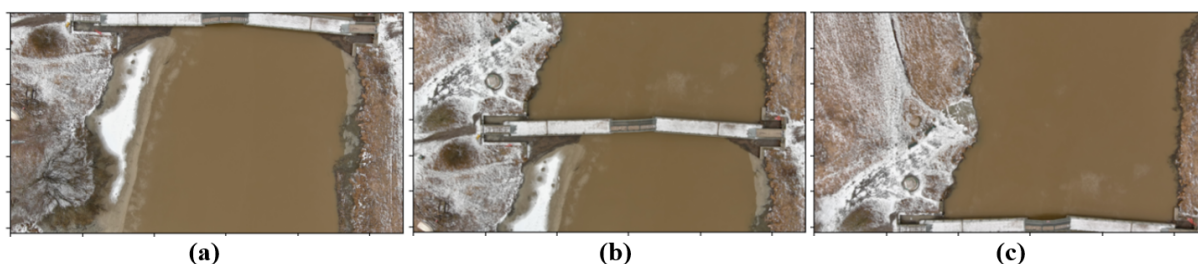


Figure 5. Three selected images to illustrate the intermediate results of image mosaicing.

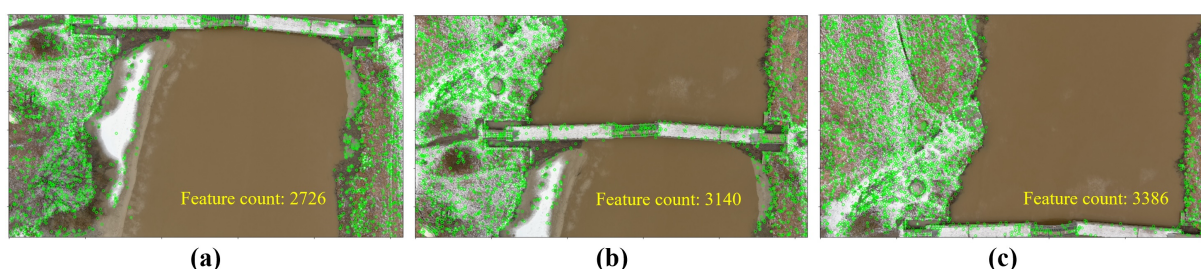


Figure 6. Features detected through SIFT algorithm on three selected images.

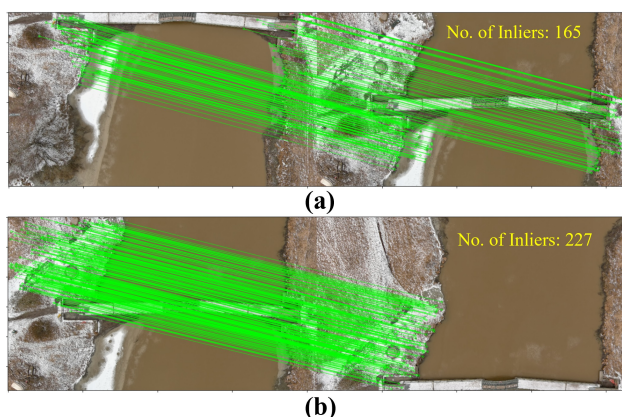


Figure 7. (a) Feature matching (top row) images 6 (a) and 6 (b), (bottom row) images 6 (b) and 6 (c)

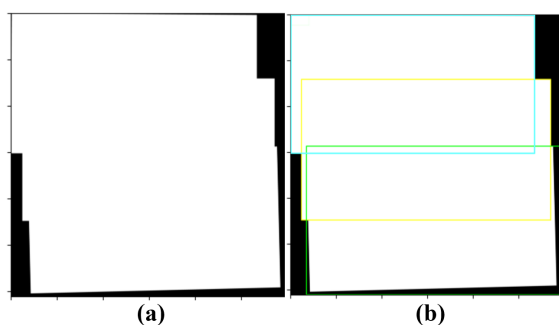


Figure 8. (a) Mosaiced mask in unified plane (b) Corresponding positions of individual image frames.

inspection of the final orthomosaic leads to the observation that the central and eastern parts of the product are very well aligned and almost perfectly mosaiced. However, the western parts of the mosaic contain some inconsistencies around the bank of the river as well as the portion that comprises of very thin power lines. This is because there are not sufficient distinct features present in these areas and the corresponding image-matching

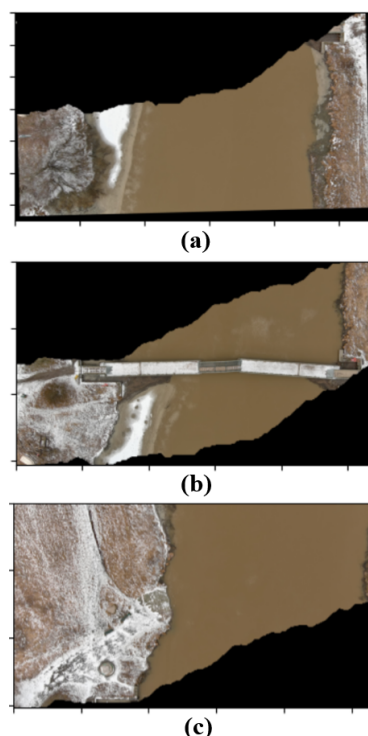


Figure 9. Extraction of exact seam positions to perform final stitching of images.

pairs have relatively low confidence scores.

Figure 12 illustrates the final result of the geo-referenced orthomosaic. For comparison of geo-spatial positioning and orientation, the orthomosaics are overlaid on the Open Steet Maps (OSM) base layer. Figure 12 (a) shows the geo-referencing result of the orthomosaic generated using the proposed methodology. Four visually identified GCPs were used to carry out the process. Figure 12 (b) shows the result of direct geo-referencing with no orthorectification of each image frame using the GPS coordinates and calculation of the image footprints through equation 1. By comparing the two figures, it is evid-

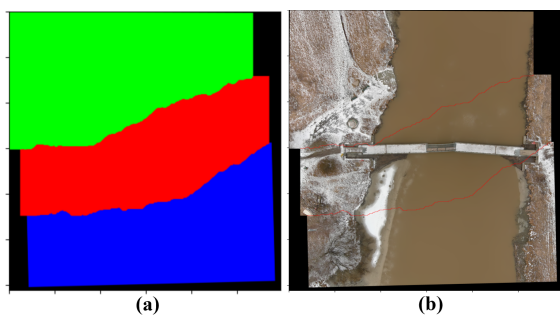


Figure 10. Results of mosaicing: (a) Color coded cropped masks. (b) Final mosaic with seam lines.

ent that the accuracy of the final product generated through the proposed methodology, shown in Figure 12 (a) is notably better.

The entire pipeline was implemented in Python within an Anaconda environment and executed on a local machine equipped with an Intel Core Ultra 9 185H 2.50 GHz CPU, 32GB RAM, and a 1TB SSD, without any discrete GPU acceleration. The reported mosaicking speed of 2.3 frames per second was achieved under these conditions using single-threaded processing.

5. Conclusion

Developing highly accurate photogrammetric products using off-the-shelf, small-scale UAVs remains a significant challenge due to the reliance on general-purpose, low-grade imaging and positioning sensors aimed at minimizing cost and size. In this paper, we proposed a streamlined approach for generation of seamless geo-referenced orthomosaics with images obtained with low-cost UAVs and limited sensor data. We empirically calculate the AFoV of the UAV camera and perform image mosaicing through progressive feature matching and image transformations while reducing the computational overhead by restricting matches to relevant consecutive images only. Using the proposed approach, we achieved seamless, near real-time mosaicing at a speed of 2.3 fps with minimal sensor data. The final orthomosaic is geo-referenced using a minimal set of visually identified GCPs, enabling its integration into GIS mapping workflows. While traditional photogrammetric workflows with RTK, GCPs, and SfM pipelines may still offer superior precision, our method provides a viable, low-cost alternative for rapid mapping where such infrastructure is unavailable. Certain limitations exist in our work which include some inconsistencies in the mosaicing process in areas where the input images do not contain sufficient quality descriptors and the use of GCPs for generating the final geo-referenced product. Future work will focus on addressing these challenges by developing more robust feature matching and image transformation techniques, as well as incorporating quantitative quality assessment methods. These may include Root Mean Square Error (RMSE) analysis against surveyed Ground Control Points (GCPs), and computation of reprojection errors from homography estimation to benchmark the geometric accuracy of the generated orthomosaics.

Acknowledgments

This research was a part of the Ministry of Education and Culture's Doctoral Education Pilot under Decision No.

VN/3137/2024-OKM-6 (Digital Waters (DIWA) Doctoral Education Pilot related to the DIWA Flagship (Decision no. 359247) funded by the Research Council of Finland's Flagship Programme.

References

- Abeho, D. R., Shoko, M., Odera, P. A., 2024. Effects of camera calibration on the accuracy of Unmanned Aerial Vehicle sensor products. *International Journal of Engineering and Geosciences*, 9(3), 314–323.
- Angel, Y., Turner, D., Parkes, S., Malbeteau, Y., Lucieer, A., McCabe, M. F., 2020. Automated Georectification and Mosaicking of UAV-Based Hyperspectral Imagery from Push-Broom Sensors. *Remote Sensing*, 12(1).
- Chen, S., Yan, Q., Qu, Y., Gao, W., Yang, J., and, F. D., 2024. Ortho-NeRF: generating a true digital orthophoto map using the neural radiance field from unmanned aerial vehicle images. *Geo-spatial Information Science*, 0(0), 1–20.
- Daakir, M., Pierrot-Deseilligny, M., Bosser, P., Pichard, F., Thom, C., 2015. UAV Onboard Photogrammetry and GPS Positioning For Earthworks. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-3/W3, 293–298.
- Fanta-Jende, P., Steininger, D., Kern, A., Widhalm, V., Apud Baca, J. G., Hofstätter, M., Simon, J., Bruckmüller, F., Sulzbachner, C., 2023. Semantic Real-Time Mapping with UAVs. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 91(3), 157–170.
- Faraji, M. R., Qi, X., Jensen, A., 2016. Computer vision-based orthorectification and georeferencing of aerial image sets. *Journal of Applied Remote Sensing*, 10(3), 036027.
- Haala, N., Kölle, M., Cramer, M., Laupheimer, D., Zimmermann, F., 2022. Hybrid georeferencing of images and LiDAR data for UAV-based point cloud collection at millimetre accuracy. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, 4, 100014.
- Hemerly, E. M., 2014. Automatic Georeferencing of Images Acquired by UAV's. *International Journal of Automation and Computing*, 11(4), 347–352.
- Hutton, J. J., Lipa, G., Baustian, D., Sulik, J., Bruce, R. W., 2020. High Accuracy Direct Georeferencing of the Altum Multi-Spectral UAV Camera and its Application to High Throughput Plant Phenotyping. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B1-2020, 451–456.
- Józków, G., Toth, C., 2014. Georeferencing experiments with UAS imagery. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-1, 25–29.
- Kern, A., Bobbe, M., Khedar, Y., Bestmann, U., 2020. Openrealm: Real-time mapping for unmanned aerial vehicles. *2020 International Conference on Unmanned Aircraft Systems (ICUAS)*, 902–911.
- Khoramshahi, E., Campos, M. B., Tommaselli, A. M. G., Vilijanen, N., Mielonen, T., Kaartinen, H., Kukko, A., Honkavaara, E., 2019. Accurate Calibration Scheme for a Multi-Camera Mobile Mapping System. *Remote Sensing*, 11(23).



Figure 11. Ortho-mosaic generated from Sixty frames captured during out-door UAV flight.

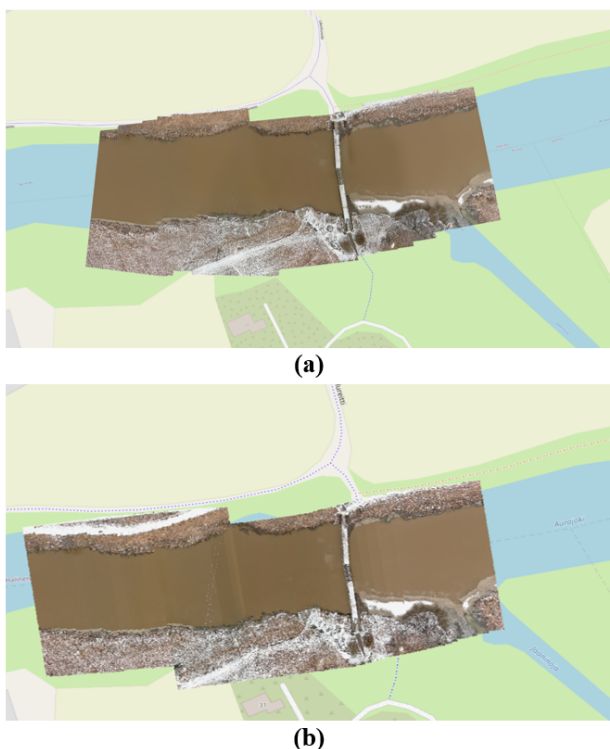


Figure 12. Results of geo-referencing overlaid on OSM base layer: (a) Orthomosaic generated through proposed methodology. (b) Direct geo-referencing approach.

Lowe, D. G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2), 91-110.

Ludwig, M., M. Runge, C., Friess, N., Koch, T. L., Richter, S., Seyfried, S., Wraase, L., Lobo, A., Sebastià, M.-T., Reudenbach, C., Nauss, T., 2020. Quality Assessment of Photogrammetric Methods—A Workflow for Reproducible UAS Orthomosaics. *Remote Sensing*, 12(22).

Mohd Daud, S. M. S., Mohd Yusof, M. Y. P., Heo, C. C., Khoo, L. S., Chainchel Singh, M. K., Mahmood, M. S., Nawawi, H., 2022. Applications of drone in disaster management: A scoping review. *Science Justice*, 62(1), 30-42.

Mostafa, M. M., Schwarz, K.-P., 2001. Digital image georeferencing from a multiple camera system by GPS/INS. *ISPRS Journal of Photogrammetry and Remote Sensing*, 56(1), 1-12.

Oliveira, R. A., Khoramshahi, E., Suomalainen, J., Hakala, T., Viljanen, N., Honkavaara, E., 2018. Improving the Spatial Accuracy of UAV-Processed Georeferencing For Hyperspectral Drone Remote Sensing. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2, 789–795.

Park, J., Cho, Y. K., Kim, S., 2022. Deep learning-based UAV image segmentation and inpainting for generating vehicle-free orthomosaic. *International Journal of Applied Earth Observation and Geoinformation*, 115, 103111.

Pérez, M., Agüera, F., Carvajal, F., 2011. Digital Camera Calibration using Images Taken from an Unmanned Aerial Vehicle. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVIII-1/C22, 167–171.

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, 487–492.

Remondino, F., Barazzetti, L., Nex, F., Scaioni, M., Sarazzi, D., 2011. UAV Photogrammetry for Mapping and 3D Modeling – Current Status And Future Perspectives. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVIII-1/C22, 25–31.

Tareen, S. A. K., Saleem, Z., 2018. A comparative analysis of sift, surf, kaze, akaze, orb, and brisk. *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, 1–10.

Turner, D., Lucieer, A., Wallace, L., 2014. Direct Georeferencing of Ultrahigh-Resolution UAV Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 52(5), 2738-2745.

Yue, D., Liu, X., Wan, Y., Zhang, Y., Zheng, M., Fan, W., Zhong, J., 2025. NeRFOrtho: Orthographic Projection Images Generation based on Neural Radiance Fields. *International Journal of Applied Earth Observation and Geoinformation*, 136, 104378.

Zeybek, M., Taşkaya, S., Elkhachy, I., Tarolli, P., 2023. Improving the Spatial Accuracy of UAV Platforms Using Direct Georeferencing Methods: An Application for Steep Slopes. *Remote Sensing*, 15(10).

Zhang, Z., Zhu, L., 2023. A Review on Unmanned Aerial Vehicle Remote Sensing: Platforms, Sensors, Data Processing Methods, and Applications. *Drones*, 7(6).