ISSN 1648-7788 / eISSN 1822-4180

VILNIUS TECH
Vilnius Gediminas Technical University

AVIATION

2026
Volume 30
Issue 4
Pages 251–262
https://doi.org/10.3846/aviation.2026.28129

RGB AND NEAR-INFRARED AERIAL IMAGERY DATASET OF A FORESTED AREA WITH RTK-REFERENCED UAV TRAJECTORIES FOR VISUAL SLAM AND MULTISPECTRAL NAVIGATION DEVELOPMENT AND BENCHMARKING

Dominykas PETRULAITIS ✉, Laurynas ARŠTIKYS, Arnoldas NOVIKEVIČIUS

Department of Aeronautical Engineering, Antanas Gustaitis' Aviation Institute, Vilnius Gediminas Technical University, LT-10223 Vilnius, Lithuania

Article History:
- received 5 June 2026
- accepted 17 June 2026

Abstract. Visual simultaneous localization and mapping (SLAM) is increasingly utilized for autonomous unmanned aerial vehicle operations in GNSS-degraded environments. However, there are limited public datasets that provide multispectral aerial imagery alongside the centimeter-level trajectory ground truth which is necessary for such algorithms, particularly in forested areas. This paper presents a dataset that addresses this deficiency. The study validates both the acquisition methodology and the benchmark quality of the resulting data. A custom RTK-referenced UAV was employed to capture time-synchronized RGB and near-infrared imagery over a 2.51 ha mixed coniferous-deciduous forest polygon in the Vilnius district of Lithuania, conducted across seven sessions in spring 2026 to reveal inter-day variations in illumination and weather conditions. Two intentionally distinct sets of cameras were utilized: an economical rolling-shutter camera pair and a global-shutter camera pair facilitate the evaluation of the same algorithm across varying shutter regimes and radiometric setups. Each session combines a photogrammetric grid mission with linear and diagonal localization transects at 80 meters above ground level. The release includes 629 raw and rectified RGB-NIR image pairs, intrinsic and stereo calibration, comprehensive per-frame telemetry, per-session METAR records, and mission waypoints, facilitating reproducible benchmarking of visual SLAM, visual odometry, and multispectral navigation in forested environments.

Keywords: multispectral imagery, near-infrared, unmanned aerial vehicle, visual simultaneous localization and mapping, real-time kinematic positioning, dataset, forest remote sensing, normalised difference vegetation index, navigation algorithm benchmark.

✉Corresponding author. E-mail: *dominykas.petrulaitis@vilniustech.lt*

Notations

Variables and functions

B – symmetric matrix $B = K^{-T} K^{-1}$ used to recover K;
c_x, c_y – principal-point coordinates in pixels;
e_{repr} – mean reprojection error (pixels);
f – lens focal length (mm);
f_x, f_y – focal lengths in pixels along the two image axes;
H_i – homography mapping the planar target to the i-th image;
K – camera intrinsic matrix;
k_1, k_2, k_3 – radial lens-distortion coefficients;
M – number of detected corners per target view;
N – number of calibration images;
$P = (X, Y, Z)$ – a 3-D scene point in the camera reference frame (Z = depth along the optical axis);
p_1, p_2 – tangential lens-distortion coefficients;
R, T – stereo rotation and translation between the two cameras; $|T|$ is the inter-lens baseline.
R_i, t_i – extrinsic rotation matrix and translation vector of view i;
r – radial distance from the principal point, $r^2 = x^2 + y^2$;
(u, v) – projected image-plane coordinates, in pixels;
(x, y) – normalised undistorted image coordinates, $(x, y) = (X/Z, Y/Z)$;
(x_d, y_d) – distorted normalised image coordinates.

Abbreviations

AGL – above ground level;
ATE – absolute trajectory error;
CRediT – contributor roles taxonomy;
CSV – comma-separated values;
DOI – digital object identifier;
DSO – direct sparse odometry;
EXIF – exchangeable image file format;
EYVI – ICAO location indicator for Vilnius International Airport;
GNSS – global navigation satellite system;

Copyright © 2026 The Author(s). Published by Vilnius Gediminas Technical University

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (*http://creativecommons.org/licenses/by/4.0/*), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

GPS – global positioning system;
GSD – ground sample distance;
ICAO – International Civil Aviation Organization;
IMU – inertial measurement unit;
IRCUT – infrared-cut (visible-band) camera variant;
LiDAR – light detection and ranging;
LitPOS – Lithuanian permanent GNSS reference-station network;
MAVLink – micro air vehicle link protocol;
METAR – meteorological aerodrome report;
NDVI – normalised difference vegetation index;
NIR – near-infrared;
NIR1 – near-infrared-pass camera variant;
NoIR – 'no IR-cut filter' camera variant;
ORB-SLAM3 – oriented FAST and rotated BRIEF SLAM, version 3;
PNG – portable network graphics;
PPK – post-processed kinematic;
QE – quantum efficiency;
RAM – random-access memory;
RGB – red, green, blue;
RH – relative humidity;
RMS – root-mean-square (error);
RTCM – radio technical commission for maritime services (correction format);
RTK – real-time kinematic;
SLAM – simultaneous localization and mapping;
SRAM – static random-access memory;
STM32H7 – STMicroelectronics 32-bit ARM Cortex-M7 microcontroller series;
UAV – unmanned aerial vehicle;
UTC – coordinated universal time;
VINS-Fusion – visual-inertial system (Fusion) estimator;
WB – white balance.

1. Introduction

Visual simultaneous localization and mapping (SLAM) and visual odometry are fundamental technologies facilitating autonomous operation of unmanned aerial vehicles (UAVs) in areas lacking or experiencing degradation of GNSS signals (Campos et al., 2021). Forest canopies represent one of the most challenging operating regimes for these algorithms because feature-based pipelines rely on local descriptors whose discriminative power collapses under dense, self-similar canopy texture – the descriptors of different scene patches become near-indistinguishable, and the matcher's correspondence ambiguity grows accordingly (Liu et al., 2025).

Despite the availability that includes several multispectral UAV datasets in recent years, nearly all focus on plant phenotyping, precision agriculture, or vegetation disease analysis. Representative examples include the large-scale semantic weed-mapping framework released by (Sa et al., 2018), which combined aerial multispectral imagery with deep neural networks over sugar-beet fields, the paddy-field multispectral collection of nitrogen-requirement estimation released by Fonseka et al. (2024), and the vineyard dataset discussed by Vélez et al. (2023), which documents Botrytis cinerea infestation under several flight geometries. To the authors' knowledge, there are little to none public datasets that integrate multispectral aerial images with centimeter-level trajectory ground truth necessary for comparing SLAM and visual-odometry methods.

This paper presents a dataset containing time-synchronized RGB and NIR aerial imagery, calibration files, telemetry logs, RTK trajectories, and METAR weather records, all gathered using a custom RTK-equipped UAV over a 2.51 ha forested polygon in Lithuania during seven flight sessions in spring 2026. The dataset aims to facilitate the development, benchmarking, and reproducibility of SLAM and visual-odometry algorithms in forest environments. It is intentionally released with both raw and aligned image streams to equally support downstream applications, such as structure-from-motion canopy reconstruction, feature-detector comparisons across spectral bands, and vegetation-index studies (Kazemi & Parmehr, 2023).

The research objective of this study is therefore two-fold. The first aim is to develop and validate an acquisition methodology that couples time-synchronised dual-band (RGB and NIR) aerial imagery with centimetre-level RTK trajectory ground truth under the GNSS-degraded conditions of a closed forest canopy. The second aim is to verify, through calibration analysis, structure-from-motion reconstruction and per-frame fix-type screening, that the resulting data are of sufficient geometric and radiometric quality to serve as a reproducible benchmark for visual SLAM, visual-odometry and multispectral-navigation algorithms.

2. Data description

The dataset is intended for researchers in computer vision, robotics, remote sensing, and aerospace engineering who are developing visual simultaneous localization and mapping, visual odometry, photogrammetry, and multispectral navigation techniques for UAV operations in GNSS-compromised forest environments. The centimetre-level RTK trajectories offer measurable ground truth for assessing localization accuracy and comparing algorithm performance under consistent flight conditions.

To the authors' knowledge, this dataset is one of the few openly licensed airborne RGB and NIR collections obtained over a Northern European forest, accompanied by centimetre-level RTK trajectory ground truth. It thus facilitates the connection between primarily agriculture-oriented multispectral UAV datasets and traditional urban or indoor SLAM standards. The released camera-intrinsic and stereo-extrinsic calibration files allow the imagery to be utilized immediately in open-source SLAM, visual odometry, and photogrammetric frameworks without necessitating further camera calibration.

Two other pairs of cameras were utilized in the same study area: a rolling-shutter OV5647 *NoIR* pair and a global-shutter AR0234 pair. This setup allows researchers

Table 1. Dataset specifications

Item	Description
Subject	Aerospace engineering; Computer vision; Remote sensing.
Specific subject area	RGB and near-infrared (NIR) aerial imagery with centimetre-level RTK trajectory ground truth for visual SLAM, visual-odometry and multispectral navigation-algorithm development over forested terrain.
Type of data	Image (PNG: RGB and NIR; raw and rectified-aligned) Table (CSV telemetry with per-frame camera exposure / gain / WB settings) Calibration imagery (chessboard PNG pairs) Text (per-session METAR records; mission waypoint files) Derived orthophoto
Methods of obtaining the data	UAV RGB + multispectral imaging with onboard RTK ground truth. Aerial platform: custom *Holybro X500* quadcopter (10-inch propellers) Flight controller / companion: *Pixhawk 6X* (dual STM32H7) / Raspberry Pi 5 RTK GNSS: *Mateksys M9Z*, *LitPOS* network RTCM3 corrections Flight speed: 1.5 m/s (grid) / 2.0 m/s (line) Overlap: Side ≥ 80 % Flight altitude: 80 m AGL Main camera: *Kurokesu* AR0234 global-shutter pair; 1920 x 1200, 3.45 µm pixel; *Kurokesu* L135-3.5-M12-NF lens (*f* = 3.5 mm, f/1.8) IRCUT (RGB) band: 370–700 nm NIR1 band: 830–1200 nm Preliminary camera: *OmniVision* OV5647 *NoIR* pair; 2592 x 1944, 1.4 µm pixel; integrated M12 lens (*f* = 2.8 mm) Sensor characteristics: manual exposure, white balance locked at 5600 K; per-frame exposure / gain / WB logged in *telemetry.csv*
Data format	Raw and rectified (aligned)
Description of data collection	Seven flight sessions in 2026 (04-25, 04-28, 05-06, 05-08, 05-11, 05-17 and 05-19) over the same 2.51 ha forest polygon. Each session combines a raster grid mission (orthophoto generation and SLAM-map initialisation) with one or more linear missions (straight and diagonal). All flights at 80 m AGL with ≥ 80% side overlap; daylight only; RTK corrections from *LitPOS*.
Data source location	Institution: Antanas Gustaitis' Aviation Institute, VILNIUS TECH Region: Vilnius district Country: Lithuania Latitude and longitude (site centroid): 54.7704 N, 25.3039 E
Data accessibility	The dataset described in this paper is openly available on request to corresponding author. Licence: CC-BY-4.0

to examine the impact of shutter type and radiometric configuration on the algorithm's performance under similar environmental and flying conditions. The raster-grid missions facilitate photogrammetric orthophoto generation and two-dimensional canopy reconstruction, while the integration of RGB and NIR channels allows for NDVI calculation and other multispectral index analyses, thereby enhancing the dataset's relevance to forestry and remote sensing research. The principal characteristics, acquisition parameters, data types, and accessibility conditions of the dataset are summarised in Table 1.

2.1. Study area

All flights were conducted over a single mixed coniferous-deciduous forest polygon in the Vilnius district, Lithuania, centred at 54.7704 N, 25.3039 E. The surveyed polygon covers 2.51 ha and was selected as a controlled outdoor test site in which the same canopy structure could be repeatedly observed under varying natural illumination. The dominant land cover consists of a mixed stand of Scots pine, Norway spruce and silver birch (Lietuvos Respublikos aplinkos ministerija, 2016), with no major elevation changes across the acquisition area, therefore, changes between sessions are mainly attributable to illumination, viewing geometry and atmospheric conditions rather than terrain itself.

The site is representative of a visually challenging forest environment for UAV-based perception. The repeated canopy texture, partial self-similarity of tree crowns and limited number of persistent artificial landmarks make feature correspondence and trajectory estimation more difficult than in urban scenes. At the same time, the coexistence of coniferous and deciduous vegetation provides spectrally distinct targets in the multispectral bands. This makes the polygon suitable for evaluating whether multispectral information, including NIR-derived vegetation contrast and NDVI-based representations, can provide additional environmental cues for visual SLAM, visual odometry and multispectral navigation studies.

Data were collected from late April to May 2026, with the same polygon re-flown in seven sessions. This acquisition strategy preserves the spatial layout of the scene while introducing inter-day variation in sun elevation, solar azimuth, cloud cover and atmospheric scattering. Consequently, the dataset enables users to distinguish between geometric repeatability and appearance variability when testing localization or mapping algorithms. The fixed study area also supports direct comparison between the preliminary OV5647 rolling-shutter camera pair and the later AR0234 global-shutter RGB-NIR camera pair, because both camera systems observed the same forest polygon rather than different test sites.

2.2. Directory layout, file formats and naming

The release is structured within a singular top-level dataset/folder with four simultaneous sub-trees. calib/ contains the intrinsic and stereo calibration chessboard images, organized by camera pair (*Kurokesu* camera global-shutter pair and OV5647 camera pair) and then by channel (rgb/, nir/). The "raw/" directory contains the unprocessed RGB and NIR frames along with the per-mission *telemetry.csv*, while the "aligned/" directory offers the stereo-rectified RGB–NIR pairs. Both directories adhere to the same per-session/per-mission structure, with session folders divided into gridv2, straight, and diagonal mission profiles. The additional sub-tree contains the Mission Planner waypoint files used to execute the mission profiles.

Image file names directly encode acquisition metadata: the *Kurokesu* camera pair employs the format <CAM>_<index>_<YYYY>_<MM>_<DD>_<HH MMSS>.png, whereas the OV5647 camera pair utilizes <CAM>_<index>_<YYYY>_<MM>_<DD>.jpg. In any mission folder, the RGB and NIR photos with identical indices constitute a synchronized pair, allowing for direct matching of consecutive frames between the two channels without external reference.

All imagery is distributed as standard 8-bit raster files whose container format reflects the acquisition hardware. The five AR0234 sessions are stored losslessly as 1920 × 1200 PNG files with an average size of approximately 3.7 MB per raw frame and 3.9 MB per rectified frame. The two preliminary OV5647 sessions were recorded at 1920 × 1080 pixels: session 04-25 and the 04-28 grid mission are stored as JPEG files of approximately 0.1–0.5 MB per frame, while the straight and diagonal missions of 04-28 are archived losslessly as PNG files of approximately 2.4 MB each. Each per-mission *telemetry.csv* table contains one row per synchronised image pair and 43 columns, combining the image file names, an ISO-8601 trigger timestamp, dual-GNSS position, fix type and reported accuracy, RTK baseline components, vehicle attitude, battery voltage, barometric altitude and trigger metadata. The three mission-plan files in additional/ follow the plain-text QGC WPL 110 waypoint format used by Mission Planner.

In total, the release occupies approximately 7.24 GB of storage. The raw/ and aligned/ sub-trees account for approximately 3.5 GB and 3.4 GB, respectively, calib/ adds approximately 0.3 GB containing the 164 chessboard calibration images (82 synchronised RGB–NIR pairs across the two camera systems), and the waypoint files occupy less than 1 MB in total. These figures, together with the total flight time and flight distance of the campaign, delimit the overall scope of the dataset.

2.3. Atmospheric conditions (METAR)

The EYVI records indicate tranquil yet visually diverse acquisition settings across the seven sessions. Surface winds remained mild, ranging from 9 knots (about 4.6 m/s) on May 17 to 12 knots (approximately 6.2 m/s) on May 8, with no gusts recorded; these conditions are evident in the smoothness of the RTK trajectories and minimize wind-induced motion blur, which is beneficial for the assessment of visual SLAM. The air temperature varied between 5 °C (05-08) and 20 °C (05-06), while station pressure fluctuated from 999 hPa (04-25) to 1018 hPa (05-08), suggesting the presence of multiple distinct synoptic conditions. Visibility exceeded 10 km in six sessions and reached 9000 m on

Table 2. Per-session METAR observations from EYVI (Vilnius International) (AllMetSat, 2026)

Session	Report time (UTC)	Wind, kt	Temperature, °C	Pressure, hPa	Visibility, km	Cloud / sky	RH, %
04-25	13:20	11 W/SW	13	999	≥ 10	SCT 4000 ft (1200 m)	54
04-28	03:50	11 NE	7	1015	≤ 9	OVC 400 ft (120 m)	93
05-06	11:20	10 N (var NW-NNE)	20	1012	≥ 10	SCT 2600 ft (790 m)	68
05-08	07:20	12 NW (var W-N)	5	1018	≥ 10	SCT 2700 ft (820 m)	56
05-11	11:50	10 SSE (var E-S)	15	1009	≥ 10	SCT 3100 ft (940 m)	59
05-17	10:20	9 NE (var N-ENE)	17	1008	≥ 10	NSC below 1500 m	55
05-19	14:00	10 W/NW (var W-NW)	16	1016	≥ 10	SCT 1700 ft (520 m)	72

Note: encoded METAR example: EYVI 251320Z 25011KT 9999 SCT040 13/04 Q0999 NOSIG. Cloud-cover abbreviations: SCT – scattered, OVC – overcast, NSC – no significant cloud, NOSIG – no significant change expected.

April 28, indicating low air attenuation for both visible and near-infrared channels.

Cloud cover exhibited significant variability and is the most critical factor for appearance-based navigation. Six sessions were conducted under scattered clouds (SCT, base 520–1200 m) or in the absence of significant cloud cover (05-17), resulting in direct solar illumination with distinct canopy shadows. In contrast, session 04-28 occurred under overcast conditions (400 ft, 93% RH), producing flat, diffuse lighting that provides a valuable low-contrast reference for assessing robustness to variations in illumination. Reporting periods range from early morning (06:50 local on April 28) to late afternoon (17:00 local on May 19), resulting in significant variations in sun elevation and shadow orientation between sessions. Since EYVI is situated approximately 12 km northwest of the test polygon, the numbers in Table 2 reflect regional conditions during each flight rather than precise on-site measurements, and are used as environmental context for the interpretation of the imagery.

2.4. Per-session camera settings

The manual exposure mode is used for all seven sessions. The exposure period is customized for each session to correspond with the current illumination, and the white balance is locked to a 5600 K daylight illuminant. In order to directly compare the OV5647 and AR0234 camera pairs, sensitivity is presented as the ISO equivalent. The final image-formation settings for each session are listed in Table 3.

Any color variation across sessions is a reflection of scene illumination rather than camera processing because the white balance is the same throughout. Users comparing imagery across sessions should take these radiometric changes into consideration because the exposure time and ISO equivalent fluctuate depending on the light of each flight, ranging from ISO 100 under bright clouds to ISO 800 on the cloudy morning of session 05-08.

2.5. RTK ground-truth quality and reference orthophoto

The acquisition campaign was designed around a singular fixed study polygon, re-flown on many occasions, ensuring that the released imagery depicts the same scene geometry with varying light, sun azimuth, and air circumstances, but the underlying canopy structure remains practically unchanged. Each session provides two complementary capture geometries over the same terrain: a dense raster grid and one or more linear transects. This enables data users to analyze daily appearance variability and cross-session repeatability without interference from alterations in the scene itself. The grid mission of each session delivers the necessary high-overlap coverage for photogrammetric reconstruction; specifically, the NIR grid frames from session 05-06 were integrated through structure-from-motion stitching into a metric orthophoto of the polygon (Figure 1). The orthophoto serves as the reference frame against which the visual SLAM trajectories of the subsequent line missions can be evaluated, and the successful reconstruction verifies that the released grid imagery possesses adequate overlap and geometric consistency for multi-view processing.

Beyond the reconstruction integrity demonstrated by the orthophoto, the second quality dimension of the dataset is the stability of the state per-frame ground-truth poses, which directly governs the strength of any trajectory-evaluation benchmark derived from the released data. RTK GNSS positioning achieves centimetre-level horizontal accuracy when the receiver maintains a fixed integer-ambiguity solution (RTK_FIXED). When the solution falls back to a float-ambiguity state (RTK_FLOAT), accuracy typically degrades to the decimetre level, and a standalone three-dimensional fix yields metre-level positioning. The dependence of UAV photogrammetric and trajectory

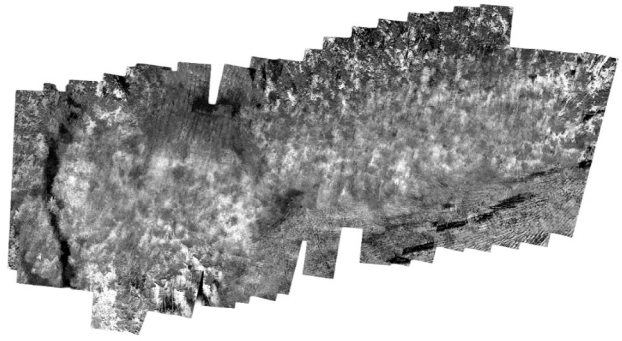


Figure 1. Orthophoto of the 2.51 ha imaged test polygon (NIR channel, grid mission of session 05-06)

Table 3. Per-session camera image-formation settings

Session	Exposure, μs	Shutter, s	ISO (equivalent)	White balance, K	Notes
04-25	2000	1/500	200	5600 locked	OV5647 camera pair
04-28	500	1/2000	200	5600 locked	OV5647 camera pair
05-06	489	1/2045	172	5600 locked	Bright midday
05-08	7996	1/125	800	5600 locked	Overcast morning, exposure boosted
05-11	244	1/4098	172	5600 locked	Bright afternoon
05-17	244	1/4098	172	5600 locked	Bright afternoon
05-19	1000	1/1000	100	5600 locked	Cloudy afternoon

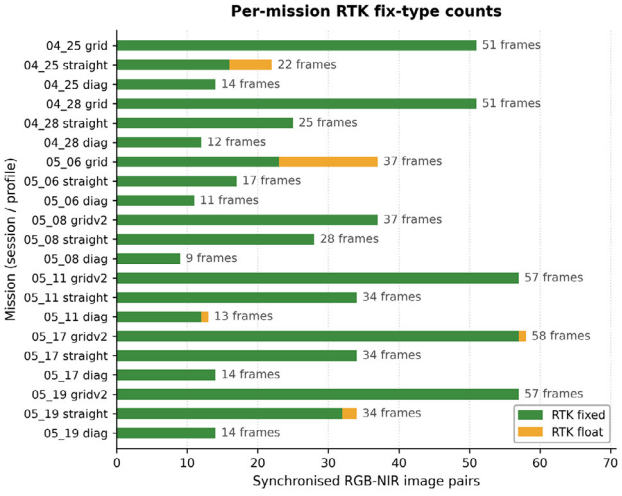


Figure 2. Per-mission RTK fix-type breakdown across the seven campaign sessions, computed from the released per-mission telemetry

accuracy on the maintained RTK fix state has been demonstrated for low-altitude flights over challenging terrain, where switching from RTK to standalone GNSS more than doubled the residual mapping error (Kim et al., 2023). Per-frame logging of the fix type is therefore essential to retain only those epochs that genuinely deliver centimetre-level ground truth.

Figure 2 reports the per-mission RTK fix-type distribution extracted from gps1_fix_type, with RTK_FIXED epochs shown in green and RTK_FLOAT epochs in amber. Sixteen of the 21 released missions maintain a 100% RTK_FIXED rate. The largest deviations are 04-25/straight (6 of 22 frames in RTK_FLOAT) and 05-06/grid (14 of 37 frames in RTK_FLOAT), and isolated RTK_FLOAT epochs occur in 05-11/diag, 05-17/gridv2 and 05-19/straight (one, one and two frames, respectively). The exact cause of these degradations could not be established with certainty. Adverse weather during the affected flights may have contributed, and individual RTCM correction messages may have been lost when the line of sight between the ground-station telemetry radio and the aircraft was temporarily obstructed over the extent of the surveyed area. In total, 605 of the 629 released frames (96.2%) carry an RTK_FIXED solution and the remaining 24 frames (3.8%) an RTK_FLOAT solution. The per-frame accuracy values logged alongside each fix allow these epochs to be down-weighted or excluded in any benchmark, so that algorithm evaluation is not biased by transient ground-truth degradation.

3. Experimental design, materials and methods

Figure 3 illustrates the data production workflow and the origin of each released artefact: each stage is shown together with the artefacts it produces, all of which are organised into the final dataset directory tree. Mission planning produced the waypoint files included in the release. The flight campaign generated the raw RGB and NIR frames together with the per-mission telemetry tables. Laboratory calibration of both camera pairs yielded the intrinsic and stereo calibration files as well as the chessboard imagery, and the processing stage transformed the raw frames into the stereo-rectified pairs and the reference orthophoto. The final stage organised all artefacts into the released directory tree.

3.1. UAV platform and avionics

The acquisition platform comprises a *Holybro X500* quadcopter equipped with 10-inch propellers, a *Pixhawk 6X* flight controller, and a *Raspberry Pi 5* companion computer featuring 16 GB of RAM. The *Pixhawk 6X* is equipped with dual STM32H7 cores, a redundant IMU, and 4 MB of SRAM. Its computational budget was chosen especially to enable telemetry logging and continuous RTCM stream integration. The companion computer and the

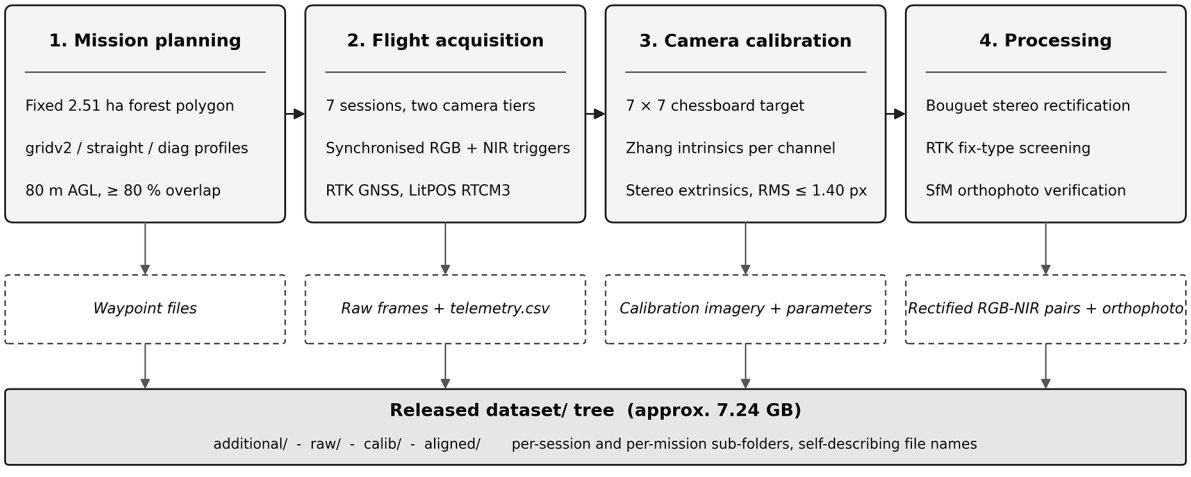


Figure 3. Schematic workflow of the dataset production

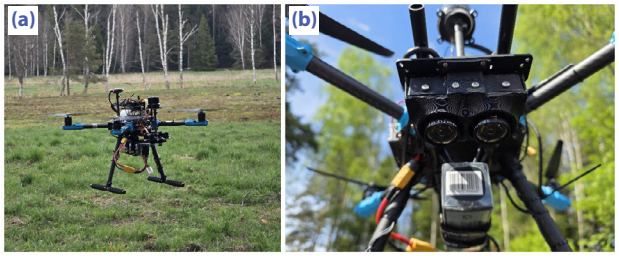


Figure 4. The *Holybro X500* platform with the synchronised dual-camera payload and RTK antenna: (a) – Front view, (b) – sensor-payload close-up showing the camera pair

flight controller communicate data using *MAVLink*. All telemetry signals (attitude, IMU, RTK position, mission state), along with the camera-side image-formation parameters, are synchronized with the camera shutter triggers and recorded in the per-session *telemetry.csv* file. The assembled platform, with the synchronised dual-camera payload and the RTK antenna, is shown in Figure 4.

An onboard *Mateksys* M9Z RTK GNSS receiver delivers centimetre-level positional accuracy. The Lithuanian permanent GNSS reference-station network *LitPOS* transmits RTCM3 corrections in real time under open-sky conditions (*LitPOS*, 2026). *LitPOS* has been documented to provide 1–2 cm planimetric accuracy based on carrier-phase measurements (Baniulis et al., 2017). For each shutter activation, the receiver-reported fix quality (RTK_FIXED, RTK_FLOAT, or 3D_FIX) along with the associated horizontal and vertical accuracy values are recorded in *telemetry.csv* (column *gps1_fix_type*), enabling subsequent algorithm developers to weight or filter individual ground-truth samples on a per-frame basis.

3.2. Camera-payload evolution: from a *NoIR* rolling-shutter camera pair to a laboratory-characterised global-shutter camera pair

Throughout the campaign, two camera pairs were utilized. A rolling-shutter *Raspberry Pi* class camera pair based on the *OmniVision* OV5647 sensor in a *NoIR* configuration was employed for the two initial sessions (2026-04-25 and 2026-04-28). A global-shutter industrial *Kurokesu*

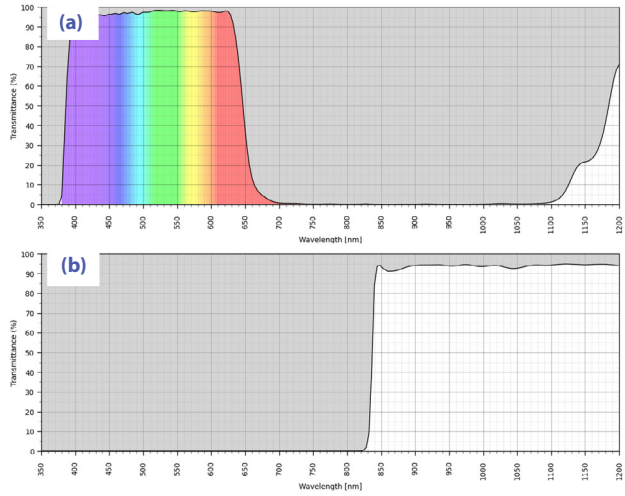


Figure 5. Quantum-efficiency curves of the global-shutter camera pair used from session 05-06 onwards: (a) – the visible (RGB / IRCUT) channel and (b) – the near-infrared (NIR1) channel (Kurokesu, 2026)

AR0234 pair (IRCUT variant for the visible RGB channel + NIR1 variant with a dedicated NIR pass-band filter) that shared a low-distortion *Kurokesu* L135-3.5-M12-NF lens and a 33 mm shared mounting plate baseline replaced the OV5647 pair from session 2026-05-06 onward (Kurokesu, 2026).

The update was motivated by two causes. First, the OV5647's rolling-shutter produces the distinctive jello-effect shear under UAV motion (Cai et al., 2020). This distortion is harmless in bright daylight when short exposure times keep the per-row readout smear small, but it becomes severe in cloudy conditions when the sensor needs longer exposures. The resulting row-by-row image deformation, along with the residual barrel distortion of the integrated M12 lens, propagates directly into feature-descriptor positions and biases feature-based visual SLAM.

The *Kurokesu* L135-3.5-M12-NF lens has a carefully characterized radial-tangential distortion model, eliminating both causes of geometric error, while the global shutter of the AR0234 completely removes jello shear. Second, NDVI and other indices can be directly calculated

Table 4. Specifications of the two camera pairs used across the campaign

Parameter	OV5647 *NoIR* pair (*OmniVision* Technologies, 2009)	AR0234 IRCUT + NIR1 pair (Kurokesu, 2026)
Sensor	*OmniVision* OV5647 (*NoIR*), rolling shutter	ON Semi AR0234, global shutter
Native resolution	2592 × 1944	1920 × 1200
Pixel size	1.4 µm × 1.4 µm	3.45 µm × 3.45 µm
Lens / focal length	Module-integrated M12, f = 2.8 mm (fixed)	*Kurokesu* L135-3.5-M12-NF, $f_{nominal}$ = 3.5 mm, f/1.8
Spectral configuration	Single *NoIR* sensor (IR-cut filter removed)	Two filtered channels with no spectral overlap: 370-700 nm (IRCUT) and 830-1200 nm (NIR1)
Effective GSD at 80 m AGL	≈ 4 cm/px	6.7–6.9 cm/px (before calibration)
Frame footprint at 80 m AGL	≈104 x 78 m	131.7 x 82.3 m
Sessions	04-25, 04-28	05-06, 05-08, 05-11, 05-17, 05-19

from two non-overlapping reflectance channels without the need for spectral-unmixing approximations thanks to the laboratory-characterized AR0234 pair, which has a flat passband spanning 830-1200 nm in the NIR1 channel and a sharp cut-off at ≈680 nm in the IRCUT channel (Park & Jeon, 2020). The laboratory-measured quantum-efficiency curves of the two channels, which substantiate these non-overlapping passbands, are shown in Figure 5. The release still includes the two OV5647 camera pair sessions, which allow users to compare the performance of the same SLAM and visual-odometry pipelines on rolling-shutter versus global-shutter imagery taken over the same polygon and investigate whether localization in the canopy is even possible with a low-cost *Raspberry Pi* class camera pair. Table 4 summarises the technical specifications of both camera pairs.

3.3. Image alignment and stereo calibration

Each camera pair is represented as a pinhole projector with a five-parameter Brown-Conrady radial-tangential distortion model. A three-dimensional point $P = (X, Y, Z)$ in the camera reference frame projects to homogeneous image-plane coordinates $(u, v, 1)$ according to the pinhole equation (Hartley & Zisserman, 2004):

$$\begin{bmatrix} u & v & 1 \end{bmatrix}^{\mathsf{T}} \sim K \cdot \begin{bmatrix} \frac{X}{Z} & \frac{Y}{Z} & 1 \end{bmatrix}^{T}, \quad K = \begin{bmatrix} \begin{bmatrix} f_x, 0, c_x \end{bmatrix}, \begin{bmatrix} 0, f_y, c_y \end{bmatrix}, \begin{bmatrix} 0, 0, 1 \end{bmatrix} \end{bmatrix}, \tag{1}$$

where (f_x, f_y) are the focal lengths in pixels along the two image axes, (c_x, c_y) is the principal-point pixel coordinate, and the symbol "~" denotes projective equivalence (equality up to a non-zero scalar). Residual lens distortion is modelled by the Brown-Conrady polynomial: let $(x, y) = \left(\frac{X}{Z}, \frac{Y}{Z}\right)$ be the normalized undistorted image coordinates and $r^2 = x^2 + y^2$ the squared radial distance from the principal point. The distorted normalized coordinates (x_d, y_d) are then (Bradski, 2000):

$$x_d = x \cdot \left(1 + k^1 \cdot r^2 + k^2 \cdot r^4 + k^3 \cdot r^6\right) + 2 \cdot p^1 \cdot x \cdot y + p^2 \cdot \left(r^2 + 2 \cdot x^2\right); \tag{2}$$

$$y_d = y \cdot \left(1 + k_1 \cdot r^2 + k_2 \cdot r^4 + k_3 \cdot r^6\right) + p_1 \cdot \left(r^2 + 2\,y^2\right) + 2 \cdot p_2 \cdot x \cdot y, \tag{3}$$

where (k_1, k_2, k_3) are the radial-distortion coefficients (k_1 dominating the smooth radial term, k_3 becoming significant only for very wide-angle optics) and (p_1, p_2) are the tangential-distortion coefficients, which model misalignment between the optical-axis elements of the lens and the sensor plane.

The nine parameters per camera (f_x, f_y, c_x, c_y, k_1, k_2, p_1, p_2, k_3) are determined utilizing the planar-target methodology (Zhang, 2000). The method utilizes the closed-form homography H_i that transforms the physical checkerboard plane to the i-th captured image. With a sufficient number of images N (generally $N \geq 10$) obtained from various target orientations, the orthonormality of the columns of the rotation matrix derived from each H_i yields two linear constraints per view on the symmetric matrix $B = K^{-T}K^{-1}$. The intrinsic matrix K is derived via Cholesky factorization of B, and the distortion coefficients, together with a non-linear refinement of $(K, \{R_i, t_i\})$, are obtained using Levenberg-Marquardt minimization of the global reprojection error (Bouguet, 2003):

$$\sqrt{\left(\left(\frac{1}{(N \cdot M)}\right) \cdot \sum_i^N \sum_j^M \left\| \pi\left(K, k, \left[R_i \;\; t_i\right]; P_j\right) - p_{ij} \right\|^2\right)} \tag{4}$$

where M is the number of corners per target view ($M = 49$ for a 7×7 inner-corner target), $\pi(\cdot)$ is the distortion-aware projection combining Equations (1)–(3), $k = (k_1, k_2, k_3, p_1, p_2)$ is the lens-distortion coefficient vector, P_j are the known 3-D coordinates of the j-th corner on the target plane ($z = 0$), and $p_{\{ij\}}$ is the measured 2-D pixel position of that corner in image i. Values of $e_{repr} \leq 0.5$ px are regarded as accurate for general photogrammetric applications, whereas $e_{repr} \geq 3$ px indicate an insufficient number of views, near-coplanar target orientations or motion blur.

This work utilizes the *OpenCV* pipeline *cv2.findChessboardCorners -> cv2.cornerSubPix -> cv2.calibrateCamera* function was employed as a straight numerical implementation of the planar-target methodology (Zhang, 2000). A planar 7×7 inner-corner chessboard with 31.8 mm square edges was captured in 40 synchronized stereo pairs at various orientations. Figure 6 illustrates the synchronized RGB/NIR view of one such pair (*pair_000*). All 49 inner corners of the target were detected and sub-pixel-refined in each channel, and the same imaging geometry was repeated across all 40 stereo pairs. Stereo extrinsics (R, T) were subsequently estimated using *cv2.stereoCalibrate* with the CALIB_FIX_INTRINSIC flag, based on the 40 pairings that passed corner verification, and the pair was rectified for spectral co-registration employing the Jean-Yves Bouguet algorithm. The obtained values are: stereo reprojection *RMS* = 1.40 px, baseline $|T|$ = 33.03 mm (corresponding to the hardware mounting plate within 0.1 mm), inter-camera rotation < 0.5 degrees. The residual parallax at 80 m AGL is $f \times \frac{|T|}{Z} \approx 0.49$ px, significantly lower than the corner-localization accuracy, hence validating pixel-wise spectral computation on the rectified pair. Table 5 summarises the observations.



Figure 6. Synchronised calibration pair (pair_000) used for joint intrinsic estimation of the camera pair: (a) – visible (IRCUT) channel; (b) – near-infrared (NIR1) channel

Table 5. Calibration of the *Kurokesu* camera pair (40 chessboard pairs)

Parameter	RGB (IRCUT)	NIR (NIR1)
Image size (pixels)	1920 × 1200	1920 × 1200
Focal length f_x (px)	1175.26	1193.56
Focal length f_y (px)	1175.80	1193.17
Principal point c_x (px)	936.50	971.04
Principal point c_y (px)	572.75	510.01
Radial k_1	–0.12489	–0.13261
Radial k_2	+0.11281	+0.12765
Tangential p_1	–0.00477	–0.00325
Tangential p_2	+0.00249	+0.00063
Radial k_3	–0.04493	–0.06217
Reprojection RMS (px)	1.336	1.378

3.4. Mission protocol and flight schedule

Within the provided dataset, each session generates two complementary image geometries using a two-phase technique. The first stage is a raster grid (lawnmower pattern, defined in *gridv2.waypoints*), which offers the dense, high-overlap coverage needed to facilitate photogrammetric or multi-view processing of the released frames and to reconstruct an orthomosaic of the polygon. One or more straight-defined line missions make up the second phase. Longer single-pass image sequences with consistent heading and viewpoint are contributed by waypoints (unidirectional linear transects) and *diag.waypoints* (diagonal traversals crossing the polygon center). To provide a single-pass image series with consistent heading and viewing geometry along the flight track, the *straight.waypoints* mission consists of a unidirectional transect of about 150 m along a fixed heading of around 71° from north. For grid missions, the cruising speed is 1.5 m s^{-1}, while

Table 6. Flight schedule, camera pair and per-mission image-pair counts (one pair = one synchronised RGB + NIR frame)

Session	Date / take-off (UTC)	Camera pair	Grid image pairs	Straight image pairs	Diagonal image pairs
04-25	2026-04-25 04:40	OV5647 *NoIR*	51	22	14
04-28	2026-04-28 08:12	OV5647 *NoIR*	51	25	12
05-06	2026-05-06 12:59	AR0234 IRCUT+NIR1	37	17	11
05-08	2026-05-08 07:18	AR0234 IRCUT+NIR1	37	28	9
05-11	2026-05-11 15:31	AR0234 IRCUT+NIR1	57	34	13
05-17	2026-05-17 14:20	AR0234 IRCUT+NIR1	58	34	14
05-19	2026-05-19 14:00	AR0234 IRCUT+NIR1	57	34	14

Note: Total released image pairs across the campaign: 629 (175 from the two preliminary OV5647 sessions and 454 from the five main AR0234 sessions). The reported take-off times correspond to the gridv2 mission, which is flown first in every session.



Figure 7. Flight patterns of the three UAV mission types flown over the same 2.51 ha test polygon, overlaid on satellite imagery (Esri, 2026): (a) – angle / diagonal localisation mission, traversing the polygon centre; (b) – grid acquisition mission used for SLAM-map initialisation or orthophoto generation; (c) – straight-line localisation mission

for line missions, it is 2.0 m s^{-1}. Every flight has at least 80% side image overlap and is conducted at 80 m AGL. To reduce exposure fluctuation and sensor noise, acquisitions are made throughout the day under steady lighting. To enable future users to replicate the exact flight geometry, the accompanying waypoint files are made available in the dataset's additional/subfolder. The flight schedule, the camera pair used in each session and the per-mission image-pair counts are summarized in Table 6.

The grid, straight and diagonal geometries depicted in Figure 7 were flown in each of the seven sessions, producing a total of 629 synchronised RGB + NIR image pairs over a single 2.51 ha polygon; in each panel of Figure 7, the dashed line links the take-off point to the first survey waypoint, and the arrows indicate the direction of flight. Across the campaign, these missions correspond to approximately 118 minutes of in-mission acquisition time and approximately 8.8 km of surveyed flight path, as computed from the first and last camera triggers and the logged RTK positions of each per-mission telemetry file. The dataset is appropriate for comparing rolling-shutter and global-shutter imagery taken over the same terrain, as well as for comparing the three flight geometries within a session, since it surveys the same polygon under two camera configurations and a variety of illumination and weather conditions.

4. Discussion

The central design choice behind this dataset – repeatedly surveying a single fixed forest polygon rather than many different sites – is what gives it value for benchmarking. Because the study area and the nominal flight geometries were held constant across all seven sessions, the differences between sessions are attributable almost entirely to illumination, cloud cover, camera configuration, exposure settings and aircraft motion rather than to changes in the underlying scene. This lets users distinguish geometric repeatability from appearance variability when evaluating visual-localisation and multispectral-processing methods over visually repetitive canopy.

Two quality dimensions support this use. The successful structure-from-motion reconstruction of the 05-06 grid mission confirms that the released frames have sufficient overlap and geometric consistency for multi-view processing, and the resulting orthophoto provides a common spatial reference for the trajectories estimated from the straight and diagonal missions. The per-frame RTK records confirm that these trajectories are usable as quantitative ground truth, because most epochs carry an RTK_FIXED solution and the few RTK_FLOAT or unclassified epochs are flagged in the telemetry, so unreliable poses can be excluded rather than mistaken for algorithm error. The calibration further supports the rectified AR0234 pairs for cross-spectral work, with a recovered stereo baseline matching the mounting plate to within 0.1 mm and sub-pixel residual parallax at 80 m AGL that makes the aligned pairs suitable for pixel-level NDVI.

The most deliberate aspect of the release is the retention of the preliminary OV5647 rolling-shutter sessions alongside the later AR0234 global-shutter data. The OV5647 pair is the lower-quality system, with rolling-shutter jello shear, greater residual distortion from its 1.4 µm pixels and module-integrated M12 lens, and a single *NoIR* sensor lacking the separated visible and NIR channels of the AR0234. It was kept rather than discarded because its value lies precisely in being a realistic low-cost reference acquired over identical ground. Both camera tiers observed the same polygon, flew the same grid, straight and diagonal geometries, and shared the same centimetre-level RTK ground truth, so they form a controlled comparison rather than two unrelated collections. Running the same SLAM, visual-odometry or multispectral pipeline on an accessible *Raspberry Pi* class rig and on an industrial global-shutter configuration lets users separate algorithms that are genuinely robust from those that implicitly depend on a particular sensor class, quantify the performance gap between consumer-grade and industrial hardware on the same terrain, and study a method before and after a representative camera upgrade. Discarding these sessions would have removed that lower bound and confined the dataset to best-case hardware alone.

The remaining considerations are environmental variability and scope. The sessions span diffuse overcast illumination, direct sunlight, changing shadow orientation and a wide range of exposure settings, which is valuable for probing robustness to appearance change. Any direct radiometric comparison must nevertheless account for these differences, and the supplied METAR observations and per-frame camera records aid interpretation without replacing on-site illumination measurements. The scope of the release is deliberately bounded. All data originate from a single 2.51 ha mixed coniferous–deciduous study site, so transferability to other forest structures and stand densities remains untested. All missions were flown at a single nominal altitude of 80 m AGL, which fixes the ground sample distance and limits scale variation, and the acquisitions cover only one late-April-to-May spring window, leaving summer, autumn and leaf-off canopy states unrepresented. The cameras are rigidly mounted without mechanical gimbal stabilisation, and no *LiDAR*, thermal or on-site illumination measurements are provided. Even so, the controlled repeated-acquisition design, the centimetre-level trajectory reference and the calibrated multispectral imagery together provide a reproducible basis for benchmarking navigation-relevant algorithms under GNSS-degraded forest conditions, and future extensions toward additional seasons, forest types, altitudes and modalities would broaden its applicability.

5. Conclusions

An acquisition methodology that couples time-synchronised RGB and NIR aerial imagery with centimetre-level RTK trajectory ground truth in a closed-canopy forest was developed and validated, and the validated result

was released as a public dataset. The resulting dataset is among the first publicly available collections in which multispectral imagery of a forest environment was acquired specifically for visual SLAM and navigation research and is paired with per-frame RTK-grade reference trajectories. The campaign comprises seven flight sessions over the same 2.51 ha polygon and provides 629 synchronised RGB and NIR image pairs (approximately 7.24 GB in raw and rectified form) acquired over approximately two hours of in-mission flight time and approximately nine kilometres of surveyed flight path.

The camera intrinsic and stereo calibration results demonstrate that the released data can serve as a quantitative benchmark. The calibration converged to per-channel reprojection residuals of 1.34–1.38 px and recovered the inter-camera baseline to within 0.1 mm of the mechanical design, a result that follows directly from pairing a global-shutter sensor with a laboratory-characterised low-distortion lens. The corresponding residual parallax of approximately 0.49 px at the operating altitude implies that pixel-level NDVI is computed on the rectified pairs without further geometric correction. The structure-from-motion reconstruction of the grid imagery confirmed sufficient overlap and geometric consistency for multi-view processing, and 96.2% of the released frames carry a centimetre-level RTK_FIXED solution, with every degraded epoch flagged in the per-frame telemetry so that trajectory evaluation remains unbiased.

The principal novelty of the release is therefore the combination of multispectral aerial imagery of a forested area, acquired specifically for the development and benchmarking of SLAM and navigation algorithms, with centimetre-level trajectory ground truth for every frame. The use of two different camera pairs strengthens the dataset further. Both pairs observed the same polygon, mission geometries and ground truth, so researchers can isolate how shutter regime, optical quality and spectral configuration influence visual SLAM, visual odometry and multispectral navigation. In practical terms, the dataset directly supports the benchmarking of localisation algorithms in GNSS-degraded forest environments, the development of NDVI-assisted navigation and cross-spectral feature matching, photogrammetric canopy reconstruction, and algorithm prototyping with low-cost hardware. The main limitations are the single study site, the single 80 m flight altitude and the spring-only acquisition window. Planned extensions therefore include additional seasons, forest types and flight altitudes, as well as gimbal-stabilised acquisition.

Data availability

The dataset described in this paper is openly available on request to corresponding author.

Ethics statement

This work did not involve any human subjects, animal experiments or data collected from social-media platforms. All UAV flights were conducted in compliance with the national civil-aviation regulations of the Republic of Lithuania and with the consent of the relevant land managers.

Funding

This research received no external funding.

Author contributions

Following the CRediT taxonomy: D. Petrulaitis – Conceptualization, Methodology, Hardware, Software, Investigation, Data curation, Writing – original draft. L. Arštikys – Conceptualization, Methodology, Hardware, Investigation, Writing – review & editing. A. Novikevičius – Conceptualization, Methodology, Hardware, Investigation, Writing – review & editing.

Disclosure statement

The authors declare that they have no competing financial, professional or personal interests from other parties.

References

AllMetSat. (2026). *METAR and TAF reports for Vilnius Airport (EYVI), Lithuania*. https://en.allmetsat.com/metar-taf/estonia-latvia-lithuania.php?icao=EYVI

Baniulis, R., Galinauskas, K., Marozas, L., Paršeliūnas, E., & Petniūnas, M. (2017). An analysis of RTK network LitPOS performance. In *10th International Conference on Environmental Engineering* (*ICEE 2017*) (Vol. 10, pp. 1–9). Vilnius TECH Conferences. https://doi.org/10.3846/enviro.2017.161

Bouguet, J.-Y. (2003). *Camera calibration toolbox for MATLAB*. https://robots.stanford.edu/cs223b04/JeanYvesCalib/

Bradski, G. (2000). The OpenCV library. *Dr. Dobb's Journal of Software Tools*, *25*(11), 120–125.

Cai, Y., Lam, E., Howlett, T., & Cai, A. (2020). Spatiotemporal analysis of "jello effect" in drone videos. In J. Chen (Rds), *Advances in Human Factors in Robots and Unmaned Systems. AHFE 2019. Advances in Intelligent Systems and Computing* (Vol. 962, pp. 197–207). Springer. https://doi.org/10.1007/978-3-030-20467-9_18

Campos, C., Elvira, R., Rodríguez, J. J. G., Montiel, J. M. M., & Tardós, J. D. (2021). ORB-SLAM3: An accurate open-source library for visual, visual-inertial, and multimap SLAM. *IEEE Transactions on Robotics*, *37*(6), 1874–1890. https://doi.org/10.1109/TRO.2021.3075644

Esri. (2026). *World imagery* [Map]. https://developers.arcgis.com/documentation/security-and-authentication/

Fonseka, C. L. I. S., Halloluwa, T., Hewagamage, K. P., Rathnayake, U., & Bandara, R. M. U. S. (2024). A dataset of unmanned aerial vehicle multispectral images acquired over a field to identify nitrogen requirements. *Data in Brief*, *54*, Article 110479. https://doi.org/10.1016/j.dib.2024.110479

Hartley, R., & Zisserman, A. (2004). *Multiple view geometry in computer vision* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511811685

Kazemi, F., & Parmehr, E. G. (2023). Evaluation of RGB vegetation indices derived from UAV images for rice crop growth monitoring. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *X-4/W1-2022*, 385–390. https://doi.org/10.5194/isprs-annals-X-4-W1-2022-385-2023

Kim, H., Hyun, C.-U., Park, H.-D., & Cha, J. (2023). Image mapping accuracy evaluation using UAV with standalone, differential (RTK), and PPP GNSS positioning techniques in an abandoned mine site. *Sensors*, *23*(13), Article 5858. https://doi.org/10.3390/s23135858

Kurokesu. (2026). *Optical filters for cameras*: *IR-cut and NIR longpass*. https://www.kurokesu.com/item/FILT

Lietuvos Respublikos aplinkos ministerija. (2016). *Miško savininko elementorius* [Forest owner's primer] (3rd ed.). Baltijos kopija.

LitPOS. (2026). *LitPOS service*: *Lithuanian permanent GNSS reference station network*. https://www.geoportal.lt/geoportal/web/litpos-paslauga

Liu, Y., Feng, Z., Zhang, H., & Dong, W. (2025). Post-integration based point-line feature visual SLAM in low-texture environments. *Scientific Reports*, *15*, Article 14606. https://doi.org/10.1038/s41598-025-97250-6

OmniVision Technologies. (2009). *OV5647*: *5-megapixel CMOS image sensor* [Data sheet]. https://cdn.sparkfun.com/datasheets/Dev/RaspberryPi/ov5647_full.pdf

Park, Y., & Jeon, B. (2020). An acquisition method for visible and near infrared images from single CMYG color filter array-based sensor. *Sensors*, *20*(19), Article 5578. https://doi.org/10.3390/s20195578

Sa, I., Popović, M., Khanna, R., Chen, Z., Lottes, P., Liebisch, F., Nieto, J., Stachniss, C., Walter, A., & Siegwart, R. (2018). WeedMap: A large-scale semantic weed mapping framework using aerial multispectral imaging and deep neural network for precision farming. *Remote Sensing*, *10*(9), Article 1423. https://doi.org/10.3390/rs10091423

Vélez, S., Ariza-Sentís, M., & Valente, J. (2023). Dataset on unmanned aerial vehicle multispectral images acquired over a vineyard affected by *Botrytis cinerea* in northern Spain. *Data in Brief*, *46*, Article 108876. https://doi.org/10.1016/j.dib.2022.108876

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(11), 1330–1334. https://doi.org/10.1109/34.888718