Rapid Georeferencing of Sensor-Limited Helicopter Imagery for Wildfire Response

Seongyu Kim[1], Jeonghyo Oh[1], Jangwoo Cheon[2], Impyeong Lee[1]

[1] Dept. of Geoinformatics, University of Seoul, Seoul, Republic of Korea - (ksg8048, ohrora98, iplee)@uos.ac.kr
[2] Geospatial Team, InnoPAM, Seoul, Republic of Korea - cjw@innopam.com

Keywords: Sensor-limited Georeferencing, Oblique Imagery, Image Retrieval, Virtual Ground Control Points, Wildfire Response

Abstract

In the initial response to wildfires, securing rapid and accurate geographic information is essential. However, helicopter imagery acquired on-site often lacks precise sensor metadata, such as camera pose and internal parameters, making the application of georeferencing difficult. In particular, obliquely captured wildfire imagery presents additional registration challenges due to severe viewpoint changes, scale variations, and low-texture environments. This study proposes an automated georeferencing pipeline capable of operating under these constraints. The proposed method consists of five stages: preprocessing, image retrieval, feature extraction and matching, Exterior Orientation Parameters (EOP) estimation, and orthomosaic generation. An initial Area of Interest (AOI) is defined using inaccurate initial position data, and the Region of Interest (ROI) within the reference map is obtained through a ResNet50-based image retrieval approach. Subsequently, virtual Ground Control Points (GCPs) are generated through deep learning-based feature matching. Elevation data is then assigned using a Digital Elevation Model (DEM), and EOP are estimated via Perspective-n-Point (PnP) and RANSAC algorithms. Intermediate frames are initialized via interpolation and refined through bundle adjustment to produce the final orthomosaic. Experimental results demonstrated that utilizing SuperGlue and LightGlue complementarily increased the number of successfully georeferenced intervals from 5 to 9. Furthermore, a minimum RMSE of 28.30 m was achieved in the most accurate interval. This method proves that by automating the feature-based georeferencing process, practical geographic information can be rapidly provided for initial disaster response, even in sensor-limited environments.

1. Introduction

Similar to other disasters, it is essential to suppress the spread of wildfires within the 'golden time.' The Korea Forest Service defines this golden time as the critical window to achieve initial suppression, stating that the nearest helicopter must arrive at the scene within 30 minutes (Korea Forest Service, 2026). If this golden time is missed, not only does the difficulty of fire suppression increase drastically, but the resulting damages also escalate (Calkin et al., 2014). To contain the fire spread and minimize damage, it is necessary to accurately locate the fire front and rapidly transmit this information to ground suppression crews and the command center. The fire front is the boundary between burning and unburned fuels; its movement drives the spread of the wildfire and serves as crucial data for predicting fire spread (Rothermel, 1972). Therefore, real-time data collection from the first-arriving platform at the scene is essential for the accurate prediction of the fire front and an effective initial response.

Commonly utilized resources for locating wildfires include fixed wildfire monitoring systems and aerial imagery acquired from early-dispatched helicopters. Since fixed monitoring systems collect data from stationary positions, they can easily identify the ignition point, where the fire starts, and the fire front, often referred to as the path of the fire. However, it is practically difficult for these monitoring systems to cover the entirety of mountainous terrains (Bao et al., 2015). Therefore, an alternative platform capable of covering broad areas is required.

Compared to other platforms, helicopters offer superior access to mountainous terrains and can cover extensive areas, making them a core platform in wildfire suppression operations. The first-arriving helicopter can provide real-time imagery and briefings, thereby supporting command-and-control decision-making. However, helicopters specifically designed for firefighting often lack precise sensors, such as gimbals or Inertial Measurement Units (IMUs). Furthermore, due to the operational nature of helicopters deployed to fire scenes, they typically fly at low altitudes and directly toward the target, resulting in the acquisition of oblique imagery. Such oblique helicopter imagery exhibits severe viewpoint and scale variations, drastically increasing the difficulty of image processing. Consequently, pinpointing the wildfire's location heavily relies on manual interpretation, which not only hinders the initial suppression response but also serves as a major factor in degrading the positional accuracy of the fire.

Currently, the National Institute of Forest Science in the Republic of Korea manually plots fire fronts onto situation maps using aerial imagery captured by helicopters (Park et al., 2025). Because this procedure depends on the visual inspection, experience, and subjectivity of a limited number of experts, it inevitably introduces inconsistencies in the results. Furthermore, since numerous helicopters are deployed to suppress even small wildfires, processing all the data generated by these helicopters through manual plotting is highly time-consuming.

These issues can be resolved by automating the plotting process. To automatically plot the fire front, the helicopter imagery must first be georeferenced. Georeferencing is the process of identifying points on a map that correspond to features in the helicopter imagery, and using these correspondences to map the imagery into real-world coordinates (Luft and Schiewe, 2021). The map image used for this purpose is referred to as a reference map. By establishing corresponding points between the reference map and the helicopter imagery, the locations of the fire front and ignition point can be accurately mapped to real-world coordinates and visualized. Georeferencing helicopter imagery requires the camera's position, orientation, and Interior Orientation Parameters (IOP). However, data acquired from helicopters designed specifically for firefighting purposes often lack the camera's orientation and IOP.

Previous studies have proposed various georeferencing methods to compensate for insufficient or inaccurate sensor data. When sensor information is absent or limited, approaches that reconstruct geometry directly from the imagery using feature matching and Structure-from-Motion (SfM) have predominantly focused on nadir or near-nadir Unmanned Aerial Vehicle (UAV) imagery (Zhuo et al., 2017, Turner et al., 2012). Some wildfire imagery studies have also attempted the geocorrection of data lacking Global Positioning System (GPS)/IMU (Ifimov et al., 2021). Conversely, many studies dealing with oblique aerial imagery have utilized external orientation data, such as GPS/Inertial Navigation System (INS) or GPS/IMU, to perform direct georeferencing or to enhance matching performance (Jiang and Jiang, 2017). Furthermore, although studies addressing the rough georeferencing of oblique imagery without metadata have been reported, they primarily rely on multi-view imagery of urban areas and geometric cues such as planes and straight lines (Verykokou and Ioannidis, 2016). Therefore, a rapidly applicable georeferencing solution for oblique helicopter wildfire imagery acquired from sensor-limited platforms, capable of functioning even under complex conditions like natural terrain and smoke, remains a significant research gap.

To address this gap, this study proposes an automated computational pipeline that generates georeferenced orthomosaics from sensor-limited oblique helicopter imagery. While conventional plotting tasks were conducted entirely manually, human intervention has been minimized by fully automating the feature-based georeferencing process. The outputs generated through this process provide the essential geospatial context required for practical situational awareness. Furthermore, this study aims to reduce processing time and improve georeferencing accuracy under these challenging conditions. Ultimately, by enabling efficient fire front mapping through an automated pipeline with minimal human intervention, this study seeks to enhance the efficiency of initial suppression decision-making. The main contributions of this study are as follows:

- Proposing an automated georeferencing pipeline capable of operating on sensor-limited oblique helicopter imagery.
- Introducing a method to generate virtual GCPs by combining reference map-based image retrieval with deep learning-based feature matching.
- Presenting a strategy to expand the successfully georeferenced intervals by leveraging the complementary characteristics of SuperGlue and LightGlue.

2. Methodology

This study proposes a method for generating orthomosaics using bundle adjustment, an indirect georeferencing technique, by extracting virtual GCPs from a reference map based solely on inaccurate initial positions. Indirect georeferencing is a method that determines the EOP and the 3D coordinates of ground points utilizing tie points between images and GCPs. However, because the locations of wildfires cannot be predicted, conducting advance GCP surveys in the affected areas is practically unfeasible. To address this issue, we propose a method that identifies the location being captured by the helicopter through a reference map and selects feature-based virtual GCPs. As shown in Figure 1, the proposed pipeline consists of five main stages: preprocessing, image retrieval, feature extraction and matching, EOP estimation, and orthomosaic generation.

2.1 Dataset

This study utilized imagery and helicopter trajectory data acquired from a firefighting helicopter, provided by the National Institute of Forest Science. The imagery was captured using a forward-facing IP camera installed inside the helicopter. The data was acquired during the 2025 Gyeongbuk wildfire, covering the regions of Uiseong and Yecheon. These areas consist primarily of forests and rural regions, and the imagery has a total duration of 45 minutes and 55 seconds. Frames were extracted from this imagery at 1-second intervals, resulting in a total of 2,725 image frames. Regarding the helicopter trajectory data, prior time synchronization with the imagery was not performed, resulting in a time offset between the two datasets. Furthermore, the trajectory data suffered from aperiodic acquisition intervals. Therefore, Kalman Smoothing was applied to the aperiodic trajectory data to mitigate noise, smooth the trajectory, and interpolate at 1-second intervals, enabling accurate positional assignment to each image frame. Subsequently,

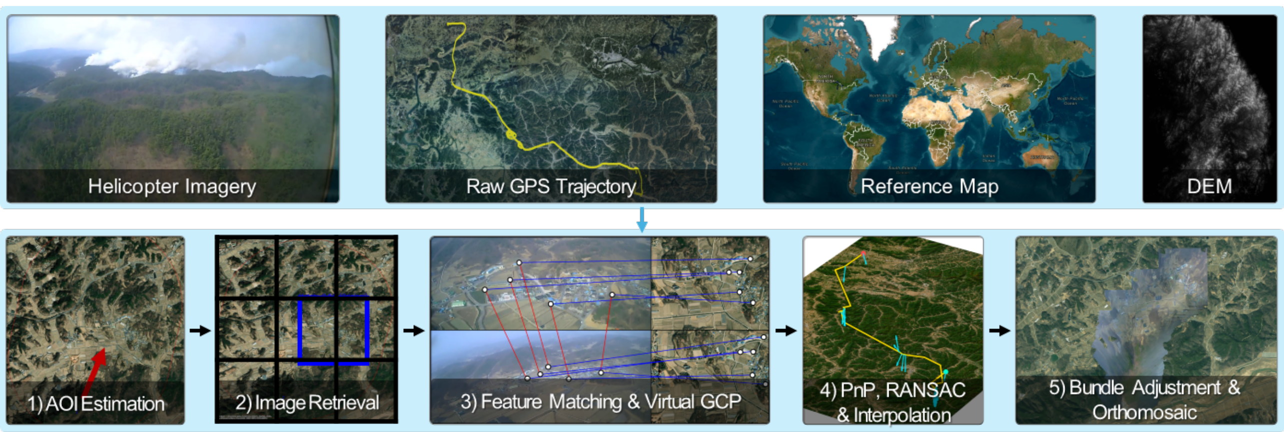

Figure 1. Overview of the proposed georeferencing pipeline from oblique helicopter imagery to orthomosaic output. Basemap: Esri World Imagery ©2026 Esri.

the temporal offset between the interpolated trajectory data and the image frames was roughly corrected to assign initial coordinates to each image frame. For the reference map data, Esri World Imagery was utilized. The corresponding imagery tiles for the study area were accessed and retrieved using the Python library contextily (Arribas-Bel et al., 2020). To incorporate terrain elevation data, a DEM with a 30 m spatial resolution (1 arc-second), developed by the National Aeronautics and Space Administration (NASA) Shuttle Radar Topography Mission (SRTM) project, was acquired via USGS EarthExplorer (NASA Jet Propulsion Laboratory, 2013).

2.2 AOI Estimation

Keyframes serve as the anchor points for the subsequent bundle adjustment process; therefore, they must possess distinct geometric structures and minimal noise or blur. Since the imagery was captured by a firefighting helicopter, the proportion of forested areas is dominant. Due to the lack of significant geometric variation in forests, there are limitations in extracting and matching feature vectors in such terrains. To ensure a stable estimation of the initial geometry, keyframes containing at least three man-made structures, where feature extraction is more reliable, were manually selected. This selection was made to ensure a stable geometric foundation during the initial stages of this study, and it can be extended to automated keyframe selection techniques in the future.

As previously mentioned, a time offset exists between the helicopter trajectory data and the imagery, causing the GPS coordinates of the trajectory to deviate from the actual captured locations. To compensate for the inaccurate positional information of a single viewpoint and to estimate the absent initial orientation data, trajectory data corresponding to two consecutive frames were utilized. A set was composed of a keyframe and a subsequent frame to calculate the heading vector. The helicopter trajectory data, originally acquired in latitude and longitude, were converted to the meter-based Universal Transverse Mercator (UTM) system to define the AOI. The heading and azimuth of the helicopter were derived by vectorizing the positional difference between the two frames. Based on this, the AOI was defined as a region extending 2 km forward and 1.5 km to the left and right.

2.3 Image Retrieval

When moving forward, a helicopter does not fly parallel to the ground but tilts in the direction of progression. Consequently, the imagery captured from the helicopter is acquired at an oblique angle. Since these oblique helicopter images exhibit significant viewpoint differences compared to the reference map, direct comparison is challenging. To enhance retrieval robustness against viewpoint variations, a trapezoidal homography reducing the bottom width was applied to the imagery frames (oblique-view augmentation). To further pre-compensate for tilt and perspective distortion caused by oblique capture, a perspective transformation-based homography was applied to the keyframes.

For retrieval, ResNet50 (He et al., 2016) was employed to match the augmented keyframes with the AOI of the reference map. Comprising 50 layers, ResNet50 can extract complex, high-dimensional features from images. Furthermore, its bottleneck structure allows for efficient operation with relatively few parameters despite its depth. By removing the final Fully Connected (FC) layer of the model, a 2048-dimensional feature vector was obtained. This vector was then normalized using L2-normalization, allowing the vector dot product to function as cosine similarity. Within the AOI, the area with the highest dot product value was selected as the ROI.

2.4 Feature Matching & Virtual GCP

To ensure robust feature matching, the keyframes were rotated and aligned to the north-up direction, consistent with the ROI, using the previously derived heading vector. After compensating for viewpoint differences, a deep learning-based feature extraction and matching network was applied to secure highly reliable correspondences between the two images. During this stage, multiple pipelines were evaluated and compared to analyze and compare the performance of state-of-the-art deep learning models. However, due to the nature of forested terrain, a limitation emerged where the extracted correspondences were heavily biased toward a few man-made structures with distinct geometric features. To suppress this clustering and ensure a uniform distribution of correspondences across the entire image, Adaptive Non-Maximum Suppression (ANMS) was applied to achieve spatial distribution. Based on these spatially distributed correspondences, it was verified whether at least four correspondence pairs were obtained to ensure the stable operation of the PnP algorithm for initial EOP estimation. Finally, the 2D planar coordinates (X, Y) on the reference map were extracted using the geometric relationships between the established correspondences. Since the reference map is a planar projection, complete 3D virtual GCPs were generated by integrating the elevation data corresponding to each coordinate from the previously constructed DEM.

2.5 PnP, RANSAC & Interpolation

To estimate the EOP of the keyframes from the virtual GCPs, the PnP and RANSAC algorithms were employed. Implementing the PnP algorithm requires the IOP of the camera and at least four GCPs. Pre-estimated IOP values were applied for the camera. The camera pose was calculated through PnP, while the RANSAC algorithm was used to exclude outliers among the virtual GCPs and estimate the pose based on inliers. After estimating the EOP for the keyframes, the initial EOP of the intermediate frames between keyframes were estimated and assigned using Spherical Linear Interpolation (SLERP) and linear interpolation. The orientation was estimated using quaternion-based SLERP, and the position was assigned through linear interpolation.

2.6 Bundle Adjustment & Orthomosaic

Subsequently, Agisoft Metashape (Agisoft LLC, 2026) was utilized to generate the final orthomosaic. The process followed a photogrammetric pipeline via the Metashape Python API, encompassing the Align Photos, Build Model, and Build Orthomosaic stages. During the Align Photos process, image matching was performed using the pre-estimated EOP as initial values. Specifically, during the optimization phase, different weights were assigned: a high weight was assigned to the EOP of keyframes where virtual GCPs were obtained, while a lower weight was assigned to the interpolated intermediate frames. By performing bundle adjustment under these constraints, the geometric stability of the entire image block was ensured.

3. Results and Analysis

To analyze the performance and practical applicability of each model from various perspectives, the evaluation metrics were categorized into two main aspects.

First, the performance of the deep learning-based matching pipelines was evaluated. In the complex wildfire imagery, characterized by a mix of forested and agricultural terrains, the robustness of each algorithm was assessed by analyzing the number and distribution of keyframes that successfully extracted virtual GCPs and maintained stable matching.

Second, the geospatial accuracy of the resulting orthomosaics was evaluated. Based on the keyframe segments successfully matched by each model, final merged orthomosaics were generated. By calculating positional errors against high-precision reference data, it was verified whether the geospatial accuracy reached a sufficient level for use in initial wildfire suppression operations.

3.1 Experimental Setup

To verify the effectiveness of the proposed georeferencing pipeline, a comparative evaluation was conducted by applying four state-of-the-art deep learning-based image matching pipelines: SuperPoint (DeTone et al., 2018)+SuperGlue, SuperPoint+LightGlue, OmniGlue, and GlueStick. These four models represent distinct evolutionary directions in deep learning-based matching technology.

The SuperPoint+SuperGlue combination is a standard model for point-based matching utilizing Graph Neural Networks (GNNs). SuperPoint+LightGlue is a recent lightweight pipeline that enhances computational efficiency and matching speed by improving the attention mechanism of the original SuperGlue. OmniGlue goes beyond simple local features by integrating foundation models to consider the semantic context of the entire image. Finally, GlueStick is a model that strengthens geometric constraints by simultaneously matching line-based structural information in addition to point features. For each of these feature-based deep learning models, pre-trained weights were utilized.

3.2 Performance Evaluation of a Deep Learning-Based Matching Pipeline

This evaluation was conducted on 48 manually selected keyframes, representative of the entire imagery. The average interval between keyframes is approximately 52 seconds. This interval serves as the basis for partitioning the entire trajectory into segments for bundle adjustment. Each keyframe encompasses diverse terrains and conditions, including forests, roads, agricultural lands, residential areas, and smoke from wildfires.

The matching performance results showed that SuperGlue succeeded in 17 keyframes, LightGlue in 9, OmniGlue in 2, and GlueStick in 0. Notably, as shown in Figure 2, cases where consecutive keyframes were all successfully matched, forming a stable segment occurred only three times for SuperGlue and twice for LightGlue. Theoretically, it is possible to generate orthomosaics over wider areas by skipping failed intermediate keyframes and connecting keyframes with larger intervals during bundle adjustment. However, given the nonlinear flight trajectories and extreme viewpoint variations characteristic of firefighting helicopters, interpolation over large intervals poses a risk of amplifying cumulative errors. Therefore, to strictly ensure geometric stability, this study limited the subsequent evaluation, including bundle adjustment and orthomosaic generation, to segments where adjacent keyframes were successfully matched in succession.

A notable aspect of this experiment is that despite using the same SuperPoint extractor, the matching success intervals of SuperGlue and LightGlue exhibited a highly complementary pattern. This pattern appears to result from the differences in the internal optimization mechanisms of each matching model interacting directly with environmental variations such as terrain features, altitude, and viewpoints in the wildfire helicopter imagery.

An analysis of the imagery in the segments where SuperGlue stably estimated virtual GCPs revealed that repetitive textures dominated across entirely forest-covered scenes and low-contrast characteristics caused by smoke. In forested areas lacking clearly distinguishable landmarks, attempting to match

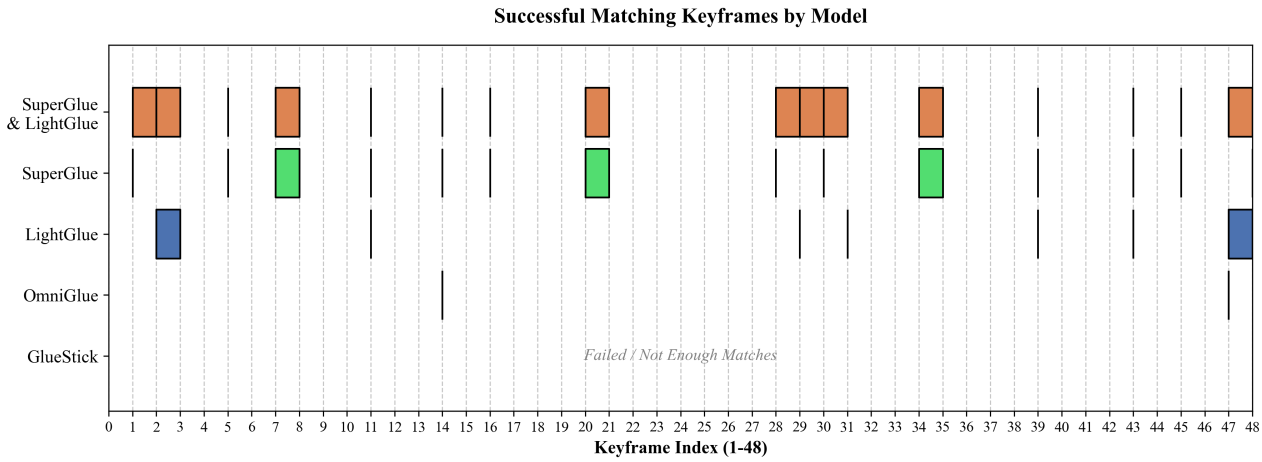

Figure 2. The keyframes that achieved successful geometric matching across the four pipelines were distributed throughout the 48 manually selected keyframes. SuperGlue (Sarlin et al., 2020) and LightGlue (Lindenberger et al., 2023) demonstrated complementary success intervals, whereas OmniGlue (Jiang et al., 2024) and GlueStick (Pautrat et al., 2023) failed to achieve a sufficient number of successful matches across the entire sequence.

based solely on the similarity of local feature points significantly increases ambiguity. However, SuperGlue utilizes an optimal transport layer based on the Sinkhorn algorithm to comprehensively consider not only the appearance of individual features but also the relative positions and geometric context of the features across the entire image. Consequently, it is determined that SuperGlue successfully achieved registration by enforcing strict 1:1 global matching even in extreme forest environments where individual features are visually similar.

On the other hand, the segments where LightGlue successfully achieved matching included numerous man-made features with clear geometric boundaries such as residential areas, distinct road networks, and agricultural plots in addition to forested areas, and they also exhibited clear contrast. This is the result of the positive effect of the adaptive point removal and selective attention mechanisms inherent to the LightGlue algorithm. LightGlue preemptively removes ambiguous feature points with low matching confidence, such as smoke or repetitive forests, in the early computational layers. Subsequently, it concentrates the computational resources of the network on distinct feature points with high confidence, such as the corners of buildings or roads. This implies that the strategy of LightGlue demonstrated robustness in environments where distinct landmarks exist, stably performing matching even amidst abrupt helicopter turns or oblique distortion.

A prominent point in the preceding matching performance evaluation results is the poor matching success rate of OmniGlue and GlueStick. Among the total 48 keyframes, OmniGlue derived meaningful virtual GCPs in only two segments (the 14th and 47th keyframes), and GlueStick failed to match in all keyframes. Examining OmniGlue first, this result appears to stem from a mismatch between the specific domain of wildfire helicopter imagery and the structural characteristics of the model. OmniGlue is a pipeline that utilizes the semantic context of the entire image for matching by internally integrating foundation models like DINO. However, the majority of the imagery captured from the helicopter consists of forested areas where the tree canopy repeats endlessly and smoke with irregular shapes. It is confirmed that in such visually and semantically monotonous and repetitive texture environments, the foundation model fails to perform meaningful object discrimination, leading to an exacerbation of matching ambiguity and ultimately causing the failure of the algorithm.

Conversely, the 14th keyframe where matching succeeded is a segment where this monotony is broken (Fig. 3(e)). This keyframe is an image that distinctly includes various man-made structures such as residential areas and roads in addition to forests. The appearance of these heterogeneous elements provided clear semantic boundaries within the monotonous natural terrain, and it appears to have helped OmniGlue effectively recognize the context of different objects, namely buildings and roads, to use them as cues for matching. The 47th keyframe is also a frame where the overall area of the image is composed of man-made structures such as residential areas (Fig. 3(f)). This segment is not only rich in distinct geometric features such as building corners and various forms of road networks but also has clear semantic information. Therefore, OmniGlue was able to operate optimally, which is the same reason other models could also easily estimate the initial positions in this segment. In conclusion, although OmniGlue could successfully perform matching by exploiting global context in environments containing distinct man-made structures, it exhibited clear limitations as a standalone georeferencing model for

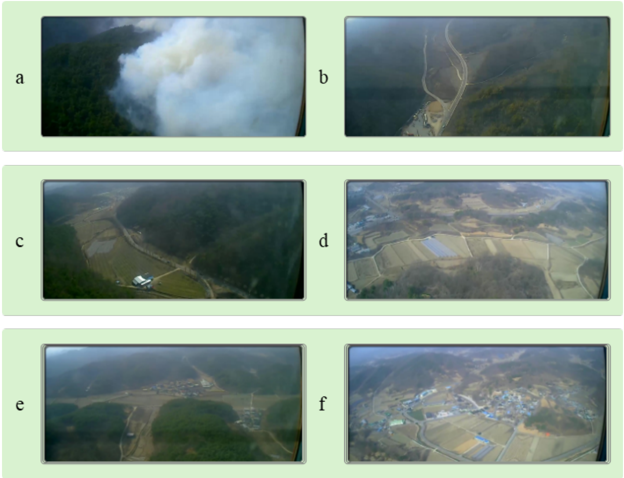

Figure 3. Keyframes with successfully estimated virtual GCPs. Top to bottom rows show the results using SuperGlue, LightGlue, and OmniGlue, respectively. (a) Keyframe 1, (b) Keyframe 20, (c) Keyframe 3, (d) Keyframe 48, (e) Keyframe 14, (f) Keyframe 47.

forest-dominated wildfire helicopter imagery, where such structural and semantic cues are limited.

In the case of the GlueStick model that performs matching by combining point and line features, it failed to derive meaningful matching results across all 48 keyframe segments. This can be analyzed from two aspects, namely the unique morphological characteristics of the forest domain and the overall quality degradation of the helicopter-captured imagery. Due to the domain characteristics of the wildfire helicopter imagery, there is an extreme lack of geometric structures from which to extract line segment features. GlueStick utilizes distinct straight-line structures such as building outlines or road networks as key cues for matching. However, the target area of this study is an environment where irregular forests and scattering smoke constitute the dominant elements of the screen, making the extraction of meaningful line segment features fundamentally impossible except for a few frames containing man-made structures.

The low visual quality of the imagery had a critical adverse impact on the extraction and matching of line segment features. An analysis of the image quality of the entire helicopter imagery (2,725 frames) used in this study using the Blind/Referenceless Image Spatial Quality Evaluator

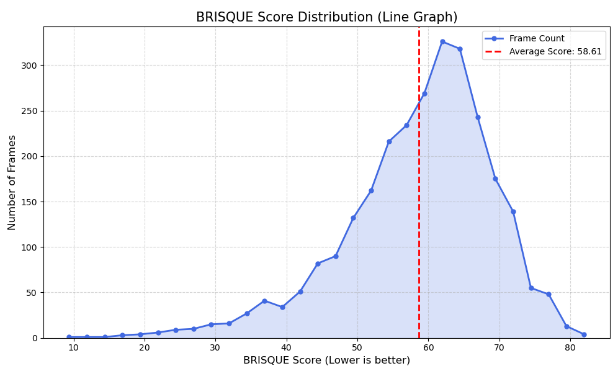

Figure 4. BRISQUE Score in Aerial Imagery.

(BRISQUE), a no-reference image quality assessment metric (Fig. 4), revealed an average score of 58.61. Typically, within the BRISQUE score range of 0 to 100, a score of 50 or higher indicates poor visual quality. Conversely, frames scoring 30 or below, which are evaluated as having clear image quality without noise or blur, accounted for only about 1.7% (48 frames) of the total.

While point-based features can be detected to some extent solely through local variations of surrounding pixels, line-based features can only be extracted as intact wireframes without interruption when the continuous gradient of pixels is clearly maintained. However, as evidenced by the BRISQUE scores, severe blur and outline distortion due to video compression existed throughout the imagery, and it is confirmed that this resulted in the blurring of object outlines, thereby creating an environment where the matching algorithm of GlueStick could not operate normally.

Consequently, even within the single domain of oblique wildfire helicopter imagery, SuperGlue, which performs global optimal transport, operated advantageously in forested areas dominated by repetitive patterns without distinct landmarks. Conversely, from viewpoints where distinct man-made structures or strong contrasts existed, LightGlue, a lightweight algorithm that excludes unnecessary noise and concentrates on high-confidence feature points, demonstrated outstanding performance. In other words, this indicates that the two models possess complementary strengths. Specifically, based on the dataset of this study, when each model was applied individually, the maximum number of valid estimation segments where bundle adjustment computation was possible was five. However, it was confirmed that when combining the complementary strengths of the two models in combination, stable geometric estimation was possible for a total of nine segments. This presents an important practical implication for overcoming the structural limitations of a single matching model and expanding the georeferencing coverage in complex disaster environments.

3.3 Geographic Accuracy Evaluation of Orthomosaic Images

The geospatial accuracy evaluation was conducted on three segments of the SuperGlue model and two segments of the LightGlue model that demonstrated meaningful continuous matching in the preceding matching performance evaluation. For analytical convenience, the successful segments of SuperGlue were defined chronologically as S1, S2, and S3, and the successful segments of LightGlue as L1 and L2. However, in the case of the S1 segment, although the estimation of the EOP of the keyframes at both ends was successful, the intermediate frames included the water dropping operation of the helicopter for wildfire suppression. Due to the resulting diffuse reflection on the water surface and extreme image distortion caused by the helicopter downwash, the S1 segment was excluded from the target for final orthomosaic generation considering geometric stability. Consequently, orthomosaics were generated for the remaining segments excluding S1, and as can be seen in Figure 5, it was confirmed that georeferencing was successfully performed even in environments including forests.

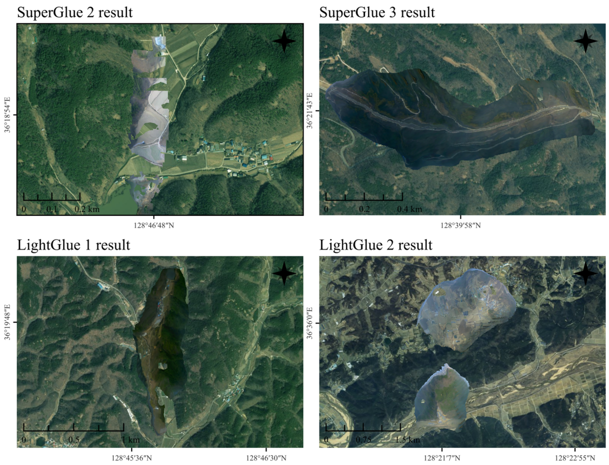

Figure 5. Orthomosaic Output.

The orthomosaic generation was conducted utilizing Agisoft Metashape software. The two selected adjacent keyframes, the intermediate frames connecting them, and the initial EOP estimated through the preceding pipeline were used as input data. Specifically, to form a stable block geometry during the bundle adjustment process, differential weights were applied to the camera parameters. High accuracies of 30 m for position and 10 degrees for attitude were assigned to the keyframes whose EOP were directly estimated via virtual GCPs, allowing them to serve as anchor points that form the backbone of the entire block. On the other hand, low accuracies of 300 m for position and 30 degrees for attitude were set for the intermediate frames assigned through interpolation to secure flexibility during the optimization process.

Furthermore, as can be seen in Table 1, an average of approximately 24 frames was used for merging the orthomosaics of each segment. Since rapid information provision is essential due to the nature of a disaster situation like a wildfire, frames were sampled at 2-second intervals from the original video to maximize computational efficiency and prevent the adverse effects that numerous frames would have on processing time.

To quantitatively verify the geospatial accuracy of the final generated orthomosaics for each segment, an independent check point evaluation method was adopted. Five points on man-made structures with distinct geometric features such as road intersections and building corners that were not involved in the matching and EOP estimation processes were randomly selected. Subsequently, the RMSE was calculated by measuring the distances between the planar coordinates of the corresponding

Metric	SuperGlue (S2)	SuperGlue (S3)	LightGlue (L1)	LightGlue (L2)
Frame Used	21	28	17	29
Area (m^2)	54,238	464,983	559,963	3,182,658
RMSE (m)	28.30	178.52	198.41	600.37
Max Error (m)	31.71	182.37	250.07	772.36

Table 1. Quantitative geospatial accuracy and spatial coverage evaluation results of orthomosaics for each SuperGlue and LightGlue matching segment.

points on the high-precision reference map and the generated orthomosaics.

According to the RMSE analysis based on independent check points in Table 1, the positional errors were 28.30 m for the S2 segment, 178.52 m for the S3 segment, 198.41 m for the L1 segment, and 600.37 m for the L2 segment. In particular, this suggests that the S2 segment established the highest geospatial accuracy among the four segments with a maximum error within 31.71 m. This demonstrates that such deviation in errors across the segments has a close correlation with the spatial coverage of the target area covered by each segment rather than a direct performance difference of the matching algorithms. Examining the relationship between the generated area of the orthomosaics and the RMSE, a tendency is observed where the positional error increases proportionally as the covered area becomes larger. This is because as the distance between the keyframes at both ends increases and the area widens, minute errors accumulate during the bundle adjustment process estimating the EOP of the intermediate frames, thereby amplifying the geometric distortion of the final orthomosaic.

Therefore, when comparing only the simple absolute RMSE values, it may appear that the SuperGlue model derived generally superior geospatial accuracy compared to the LightGlue model. However, the interpretation changes when considering the error occurrence rate relative to the area covered by the orthomosaic. In the case of the L2 segment showing the largest error of 600.37 m, it is the result of processing an overwhelmingly large surface area as a single block compared to other segments. In other words, rather than being an error caused by the degraded matching ability of the LightGlue algorithm, it is reasonable to view it as a result reflecting the inherent perspective distortion and cumulative error of oblique aerial imagery connecting a wide trajectory.

In conclusion, both models demonstrated that they can provide excellent geospatial information for initial response within 30 m for relatively narrow areas, even under the harsh conditions of utilizing only limited virtual ground control points. This level of accuracy is sufficient for initial wildfire response, where rapid situational awareness is more critical than centimeter-level precision. Simultaneously, it suggests that the error accumulation phenomenon occurring when the area is significantly expanded is a challenge that must be addressed in the future through measures such as the addition of intermediate reference points.

4. Conclusion

This study proposed a computational pipeline that performs automated georeferencing using oblique imagery from wildfire suppression helicopters lacking precise sensor information. The proposed method automatically extracted virtual GCPs by combining coarse position estimation through ResNet50-based image retrieval and state-of-the-art deep learning-based feature matching models, and based on this, it estimated the initial EOPs and performed bundle adjustment with differentially applied weights to successfully generate the final orthomosaics. As a result of the evaluation, it was confirmed that the frame coverage capable of georeferencing could be expanded by combining the complementary strengths of the SuperGlue and LightGlue models, which exhibit different strengths in forested areas lacking distinct landmarks and regions containing man-made structures respectively. Although it has limitations for precision surveying purposes at the cm level because the cumulative error showed a tendency to increase proportionally as the coverage area of the orthomosaic increased, it demonstrated high practical utility in that it can rapidly provide spatial information essential for initial suppression command and fire line identification without manual plotting in urgent disaster situations where sensor metadata is inaccurate. Overall, this study demonstrates that reliable and rapid georeferencing is achievable even under extreme sensor-limited conditions, making it suitable for real-world wildfire response scenarios.

Acknowledgements

This study was carried out with the support of ‘R&D Program for Forest Science Technology (RS-2025-25441578)’ provided by Korea Forest Service(Korea Forestry Promotion Institute).

References

Agisoft LLC, 2026. Agisoft metashape professional software, version 2.2.2. agisoft.com (31 March 2026).

Arribas-Bel, D., Simini, F. et al., 2020. contextily: context geo-tiles in python. GitHub repository. github.com/geopandas/contextily (31 March 2026).

Bao, S., Xiao, N., Lai, Z., Zhang, H., Kim, C., 2015. Optimizing watchtower locations for forest fire monitoring using location models. *Fire Safety Journal*, 71, 100–109. doi.org/10.1016/j.firesaf.2014.11.016.

Calkin, D. E., Cohen, J. D., Finney, M. A., Thompson, M. P., 2014. How risk management can prevent future wildfire disasters in the wildland-urban interface. *Proc. Natl. Acad. Sci. U.S.A.*, 111(2), 746–751. doi.org/10.1073/pnas.1315088111.

DeTone, D., Malisiewicz, T., Rabinovich, A., 2018. Superpoint: Self-supervised interest point detection and description. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 337–348.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.

Ifimov, G., Naprstek, T., Johnston, J. M., Arroyo-Mora, J. P., Leblanc, G., Lee, M. D., 2021. Geocorrection of airborne mid-wave infrared imagery for mapping wildfires without GPS or IMU. *Sensors*, 21(9), 3047. doi.org/10.3390/s21093047.

Jiang, H., Karpur, A., Cao, B., Huang, Q., Araujo, A., 2024. Omniglue: Generalizable feature matching with foundation model guidance. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 19865–19875.

Jiang, S., Jiang, W., 2017. On-board GNSS/IMU assisted feature extraction and matching for oblique UAV images. *Remote Sensing*, 9(8), 813. doi.org/10.3390/rs9080813.

Korea Forest Service, 2026. Pan-government takes preemptive measures against wildfires. forest.go.kr/kfsweb/cop/bbs/selectBoardArticle.do?bbsId=BBSMSTR_1036&mn=NKFS_04_02_01&nttId=3215879 (30 March 2026).

Lindenberger, P., Sarlin, P.-E., Pollefeys, M., 2023. Lightglue: Local feature matching at light speed. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 17581–17592.

Luft, J., Schiewe, J., 2021. Automatic content-based georeferencing of historical topographic maps. *Transactions in GIS*, 25(6), 2888–2906. doi.org/10.1111/tgis.12794.

NASA Jet Propulsion Laboratory, 2013. NASA Shuttle Radar Topography Mission Global 1 arc second. Accessed via USGS EarthExplorer.

Park, G., Seo, Y., Ahn, H., Ko, S. J., Lee, B., Lee, Y., 2025. Semantic segmentation of fire fronts in helicopter imagery using vision transformer models. *Korean Journal of Remote Sensing*, 41(2-1), 375–385. doi.org/10.7780/kjrs.2025.41.2.1.11.

Pautrat, R., Suárez, I., Yu, Y., Pollefeys, M., Larsson, V., 2023. GlueStick: Robust image matching by sticking points and lines together. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 9672–9682.

Rothermel, R. C., 1972. A mathematical model for predicting fire spread in wildland fuels. Research Paper INT-115, USDA Forest Service.

Sarlin, P.-E., DeTone, D., Malisiewicz, T., Rabinovich, A., 2020. Superglue: Learning feature matching with graph neural networks. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4937–4946.

Turner, D., Lucieer, A., Watson, C., 2012. An automated technique for generating georectified mosaics from ultra-high resolution unmanned aerial vehicle (UAV) imagery, based on structure from motion (SfM) point clouds. *Remote Sensing*, 4(5), 1392–1410. doi.org/10.3390/rs4051392.

Verykokou, S., Ioannidis, C., 2016. Automatic rough georeferencing of multiview oblique and vertical aerial image datasets of urban scenes. *The Photogrammetric Record*, 31(155), 281–303. doi.org/10.1111/phor.12156.

Zhuo, X., Koch, T., Kurz, F., Fraundorfer, F., Reinartz, P., 2017. Automatic UAV image geo-registration by matching UAV images to georeferenced image data. *Remote Sensing*, 9(4), 376. doi.org/10.3390/rs9040376.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLIX-B3-2026-1395-2026