*remote sensing*

MDPI

*Article*

# A Second-Order Spatial Guided Sampling Method for Outlier Removal in UAV Remote Sensing Image Registration

**Yifan Deng [1,*], Xiang Chen [2], Zhigao Chen [3] and Jiaxing Zhou [3]**

[1] Faculty of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China
[2] Shanghai Institute of Satellite Engineering, Shanghai 201109, China; 2152.050350@163.com
[3] School of Electrical Engineering and Automation, Xiamen University of Technology, Xiamen 361024, China; 2422141004@stu.xmut.edu.cn (Z.C.); 2021000082@xmut.edu.cn (J.Z.)
* Correspondence: ivandunne@xjtu.edu.cn

## Highlights

### What are the main findings?

- We propose SPGSC++, a second-order spatial guided method for outlier removal in UAV remote sensing image registration, which uses common-neighbor structural quality to suppress false correspondences under complex geometric distortions.
- The method achieves state-of-the-art F-scores on four challenging datasets (CIAP, SUIRD, UAV, and ZHANGZHOU), demonstrating robust performance in scenes with repetitive textures, viewpoint changes, and local deformations.

### What are the implications of the main findings?

- The PCA-based dimensionality reduction and second-order compatibility scoring enable efficient and reliable robust model estimation without requiring additional training data or ground-truth annotations.
- The framework can be integrated as a plug-in module after tentative feature matching, improving the registration accuracy and robustness in practical UAV remote sensing applications.

## Abstract

Image registration is fundamental to UAV remote sensing data processing, but repetitive textures, viewpoint and scale variations, terrain relief, and local deformation can introduce numerous outliers into tentative correspondences. The existing random-sampling and spatial-consistency methods commonly rely on pointwise confidence or isolated pairwise relations and therefore cannot reliably determine whether a locally compatible relation receives stable group-level structural support. To address this limitation, we propose SPGSC++, a second-order spatial guided sampling method for outlier removal. SPGSC++ first uses PCA to reduce the dimensionality of tentative correspondences, thereby reducing the computational burden of subsequent spatial relation construction. It then derives second-order structural quality from the common-neighbor support of directly compatible edges and uses this quality for initial data subset selection and residual-informed data subset refinement to generate and optimize a homography. Comparative experiments show that, on the five datasets CIAP, SUIRD, UAV, ZHANGZHOU, and PAN, the proposed method achieves mean Precision, Recall, F-Score, Runtime, and mean reprojection error values of 0.9627, 0.9797, 0.9702, 0.3724 s, and 1.6454, respectively. Relative to the original SPGSC, SPGSC++ increases the mean F-Score by 0.0391 and reduces Runtime and mean reprojection error by 32.39% and 58.12%, respectively. Overall, the innovation of SPGSC++ lies in using second-order common-neighbor structural quality to guide initial subset

Check for updates

Academic Editors: Pablo Rodríguez-Gonzálvez and Ming-Der Yang

Received: 17 August 2026
Revised: 12 October 2026
Accepted: 26 October 2026
Published: 2 November 2026

**Copyright:** © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

selection and residual-informed refinement for robust estimation in UAV remote sensing image registration.

**Keywords:** UAV remote sensing image; image registration; outlier removal; second-order spatial relation; robust model estimation

## 1. Introduction

With the rapid development of unmanned aerial vehicle (UAV) remote sensing, high-resolution remote sensing imagery has been widely used in disaster assessment [1], agricultural monitoring [2], and target localization [3] because of its flexible acquisition, high timeliness, and fine spatial resolution. The joint analysis of multitemporal, multiview, and multisensor remote sensing data generally relies on accurate image registration. However, changes in the flight attitude, imaging angle, terrain relief, and illumination may introduce scale changes, rotation, parallax, and even local nonlinear deformation between images, making stable correspondences difficult to establish. Improving the accuracy and robustness of UAV remote sensing image registration in complex scenes is therefore an important task in UAV remote sensing data processing.

Local-feature-based methods remain the predominant techniques for remote sensing image registration. In general, registration comprises two stages. First, tentative feature correspondences are generated between two images according to descriptor similarity. Second, outliers are removed, while inliers are retained to improve registration accuracy. UAV remote sensing images often contain large areas of repetitive land-cover patterns and texture-poor regions. Buildings, roads, and farmland, for example, can create many visually similar areas in which features at different locations have similar descriptors. Complex imaging conditions further reduce local-feature stability; so, the tentative correspondence set inevitably contains a high proportion of outliers [4]. Outlier removal is therefore critical to the accuracy and reliability of the final correspondences.

Random sample consensus (RANSAC) is widely used for robust model estimation in image registration. It repeatedly draws a minimal sample set, fits a candidate model, and classifies inliers according to residuals [5]. Traditional RANSAC, however, samples uniformly and does not exploit the spatial distribution of correspondences. When the outlier rate is high or the inliers are unevenly distributed, the probability that a minimal sample set contains only inliers decreases, making reliable model hypotheses difficult to obtain. Recent studies have consequently sought to improve consensus assessment by incorporating spatial structures among feature points and constructing neighborhood relations to increase the reliability of model estimation.

Spatial-consistency constraints improve correspondence discrimination by analyzing relations among matches rather than treating each feature independently. In complex remote sensing scenes, however, terrain relief and viewpoint variation may cause local distortion and allow some outliers to satisfy local spatial constraints by chance. Repetitive textures and unstable local features may further produce structured false correspondences. Moreover, many approaches concatenate paired feature coordinates into four-dimensional vectors, increasing the cost of subsequent relation construction as the tentative-correspondence set grows. Reducing this computational burden while determining whether a locally compatible relation receives stable group-level support therefore remains a key challenge.

To address these issues, we propose SPGSC++ based on the second-order-proximity-guided sampling framework of SPGSC for UAV remote sensing image registration. First, PCA is used to compress the four-dimensional joint coordinate representation of tentative correspondences into a two-dimensional representation to reduce the burden of subsequent relation construction. An empirical first-order affinity is then used to generate candidate edges, and common-neighbor support among directly connected edges yields second-order structural quality. The resulting quality guides the initial data subset selection and is combined with residual-informed data subset refinement and a structural-quality stopping criterion to perform homography estimation. The overall workflow is shown in Figure 1.

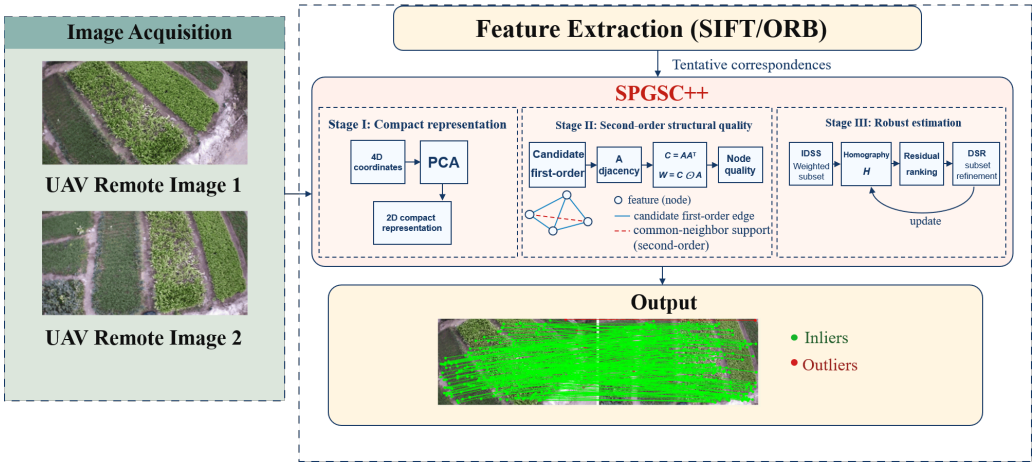



**Figure 1.** Overall workflow of SPGSC++.

The main contributions of this paper are as follows:

(1) We construct a second-order spatial-quality representation for UAV remote sensing image registration using candidate first-order edges and their common-neighbor support. This representation extends isolated pairwise compatibility to group-level structural verification and reduces the influence of accidental local relations on correspondence-quality assessment.

(2) We introduce a compact PCA-based correspondence representation that projects the original four-dimensional coordinates into two dimensions, reducing the computational burden of subsequent relation construction.

(3) We integrate second-order structural quality with the initial data subset selection and residual-informed data subset refinement so that structural evidence contributes to model initialization, updating, and termination. Five-dataset comparisons, UAV-domain ablation, parameter-sensitivity analysis, and evaluation across varying outlier ratios jointly assess the effectiveness and applicability boundaries of the extended framework.

The remainder of this paper is organized as follows. Section 2 reviews UAV remote sensing image registration, robust model estimation, and spatial-consistency-based correspondence removal. Section 3 describes SPGSC++ in detail. Section 4 introduces the datasets and evaluation metrics. Section 5 presents the comparative experiments, ablation studies, and qualitative analysis. Section 6 discusses the method and its limitations, and Section 7 concludes the paper and outlines future work.

## 2. Related Work

### *2.1. UAV Remote Sensing Image Registration*

Image registration, which establishes correspondences between two overlapping images, underpins a range of applications including UAV image stitching, three-dimensional reconstruction, and multisource data fusion. Compared with ordinary natural images, which often exhibit relatively limited viewpoint and scale changes, UAV remote sensing images cover larger areas and can vary substantially in imaging altitude, flight attitude, and viewing direction. Terrain relief induces local parallax, which, together with variations in illumination, shadows, seasons, and land-cover states, collectively gives rise to complex local non-rigid deformations. Consequently, a single global transformation model is insufficient to accurately characterize the full local imaging relationships between images. Moreover, repetitive textures and similar structures commonly occur in built-up areas, roads, and agricultural fields, while water bodies and bare land tend to be textureless—both of which undermine the discriminative power of local feature descriptors.

Traditional local-feature methods, including SIFT [6], SURF [7], and ORB [8], generally detect keypoints with scale, rotation, or affine adaptability, construct descriptors from local intensity, gradient, or binary patterns, and establish tentative correspondences according to descriptor distance. These methods require no task-specific training and are broadly applicable, but their performance is sensitive to cross-view deformation, repetitive textures, and local appearance changes. Learning-based approaches, such as SuperPoint [9], utilize neural networks for the joint learning of keypoints, descriptors, and matching confidence. While they exhibit strong representational power on data distributions akin to the training set, their cross-sensor, cross-region, and cross-temporal generalization remains critically dependent on the sufficiency and diversity of training samples. Regardless of whether handcrafted or learned features are used, tentative correspondences may contain false matches. This paper therefore does not redesign the detector or descriptor; instead, it addresses outlier removal and robust model estimation after tentative feature matching.

### *2.2. Random-Sampling-Based Robust Model Estimation*

RANSAC follows a sample–fit–verify framework. It draws a minimal sample set from the tentative correspondences, estimates a candidate geometric model, and selects the best model over repeated iterations. This framework estimates parameters from a minimal sample set even in the presence of outliers and can be combined with homography, affine, or fundamental-matrix models. Its effectiveness, however, depends strongly on the composition of the minimal sample set. The required number of samples increases exponentially with the outlier rate and minimal-sample size, and the model quality deteriorates further when the inliers are unevenly distributed or a degenerate sample is drawn [10].

To reduce the probability of drawing a contaminated minimal subset, guided-sampling variants of RANSAC rank or weight tentative correspondences using matching confidence [11], prior ordering [12], or local geometric consistency [13]. By contrast, local-optimization variants [14] refine an initially generated hypothesis by expanding, reranking, or re-estimating its consensus set. These strategies target different stages of the sample–fit–verify process, but their effectiveness still depends on the reliability of the sampling or the initial hypothesis [15]. In contrast to sampling based on the confidence of an individual correspondence, SPGSC++ evaluates the sampling quality from the common-neighbor support of directly connected candidate edges and subsequently incorporates the resulting structural quality into data subset refinement.

### *2.3. Spatial Consistency and Graph-Structured Correspondence Removal*

Spatial-consistency methods exploit the tendency of correct matches to exhibit coherent geometric changes between two images. Pairwise compatibility is commonly established from distance [16], displacement [17], scale [18], or local neighborhood structure [19]. Recent correspondence pipelines can also be divided into frontend correspondence generation and backend verification and model estimation. Such pairwise relations are essentially first-order associations. Unlike individual correspondence decisions based solely on descriptor similarity, these first-order relations introduce mutual constraints among correspondences and typically require no training data [20]. This benefit, however, comes with a limitation: a pairwise relation indicates only whether two correspondences are compatible under a given threshold—it does not reveal whether this relation receives collective support from other correspondences. In remote sensing scenes with repetitive textures or local distortions, some false matches may accidentally satisfy geometric constraints and thus form spurious compatibility edges. Learned methods such as SuperPoint and LightGlue mainly detect, describe, or generate tentative correspondences, whereas SPGSC++ starts from an existing correspondence set, evaluates the structural quality, removes false correspondences, and estimates a geometric model.

Recent spatial-consistency methods differ in their structural units and in how structural evidence is used. LPM [21] preserves local neighborhood relations, LGSC [22] evaluates local graph-structure agreement, TRESAC [19] uses triplet relations to guide robust sampling, TRGDC [23] combines triplet relations with density-based correspondence discrimination, and IMSCAN [24] uses spatial-clustering information for deterministic geometric-model fitting. SPGSC++ obtains node quality from the common-neighbor support of directly connected candidate edges and passes the same quality signal to both IDSS and residual-informed DSR.

SPGSC is the direct methodological predecessor of SPGSC++. SPGSC introduced second-order proximity to guide seed and initial minimal-subset generation and used its original subset-improvement mechanism. SPGSC++ inherits the principle that a reliable correspondence should receive group-level structural support and extends the framework for UAV remote sensing registration through a compact PCA correspondence representation and the integration of the candidate graph and common-neighbor quality with IDSS, residual-informed DSR, and a structural-quality stopping criterion.

### *2.4. Summary*

Table 1 summarizes the main distinctions among existing outlier-rejection methods in terms of the information they exploit and how it is incorporated into model estimation.

**Table 1.** Integration strategies of existing correspondence-removal methods. ✓ indicates the method possesses this property; × indicates it does not.

| Method Category | Pointwise Information | Pairwise Relation | Higher-Order/ Common-Neighbor Relation | Guided Sampling | Training-Free |
|---|---|---|---|---|---|
| Descriptor or pointwise confidence | ✓ | × | × | Partial | Method-dependent |
| Random-sampling robust estimation | Partial | × | × | Partial | ✓ |
| First-order spatial consistency | Partial | ✓ | × | Partial | ✓ |
| Graph or higher-order structure | Partial | ✓ | ✓ | Partial | Method-dependent |
| SPGSC++ | × | ✓ | ✓ | ✓ | ✓ |

In summary, pointwise quality estimates are vulnerable to outliers, random sampling has a low probability of obtaining an effective minimal sample set at high outlier rates, and first-order spatial consistency cannot determine whether a compatibility edge has stable common-neighbor support. SPGSC++ is a structure-guided robust-estimation module applied after tentative feature matching. Its key distinction is that it constructs second-order spatial quality from the common neighbors of directly compatible edges and uses this quality in both initial sample selection and subsequent model refinement.

## 3. Proposed Method

SPGSC++ comprises two stages: spatial-structure construction and robust model estimation. Given two UAV remote sensing images $I$ and $I_2$ to be registered, feature detection, description, and tentative matching produce $N$ tentative correspondences $M = \{m_i\}_{i=1}^{N}$, $m_i = (x_i, x'_i)$, where $\mathbf{x}_i = (x_i, y_i)^\top$ and $\mathbf{x}'_i = (x'_i, y'_i)^\top$ are the coordinates of correspondence $i$ in the two images. The coordinates are combined into a four-dimensional (4D) joint representation and compressed into a two-dimensional (2D) principal-component space using principal component analysis (PCA). A first-order compatibility graph is then constructed from pairwise distance consistency. Common-neighbor support between directly compatible node pairs defines a second-order structural quality score for each correspondence. Finally, these scores guide initial data subset selection (IDSS), and residual-rank-band-driven data subset refinement (DSR) iteratively updates the model and classifies the correspondences as inliers or outliers.

### *3.1. Construction of the Second-Order Spatial Relation*

For each correspondence, we construct a 4D joint coordinate. Eigenvalue decomposition is performed on the centered matrix, and the two directions of largest variance are retained to obtain a 2D representation. PCA is used here as a representation-compression technique rather than as a robust outlier estimator. Its purpose is only to reduce the dimensional burden of subsequent relation construction. Because classical PCA can be influenced by the outlier distribution, SPGSC++ does not classify a correspondence from its PCA coordinates; the representation is used only to generate candidate first-order compatibility edges.

For correspondence $m_i$, we construct the 4D joint coordinate

$$z_i = [x_i, y_i, x'_i, y'_i]^{\mathrm{T}} \tag{1}$$

all correspondences into matrix $Z \in \mathbb{R}^{N\times 4}$. Let $\boldsymbol{\mu}$ denote its mean vector and $Z_c = Z - 1\boldsymbol{\mu}^{\mathrm{T}}$ the centered matrix. Eigenvalue decomposition is performed on $Z_c$; the loading matrix $V_2 \in \mathbb{R}^{4\times 2}$ associated with the two directions of largest variance is retained to obtain

$$\tilde{Z} = Z_c V_2, \tilde{z}_i = [\tilde{z}_{i1}, \tilde{z}_{i2}]^{\mathrm{T}} \tag{2}$$

This step compresses the 4D joint correspondences into a compact 2D representation and provides a common input for subsequent relation construction.

The two coordinates of the 2D principal-component representation are treated as two scalar directions. For any correspondences $m_i$ and $m_j$, we compute:

$$d_{ij}^{(1)} = |\tilde{z}_{i1} - \tilde{z}_{j1}|, d_{ij}^{(2)} = |\tilde{z}_{i2} - \tilde{z}_{j2}| \tag{3}$$

If the pairwise distances along the two directions are similar, the correspondence pair is considered to have strong first-order compatibility [25]. The corresponding compatibility score is defined as

$$s_{ij} = \exp\left[-\frac{\left(d_{ij}^{(1)} - d_{ij}^{(2)}\right)^2}{2\sigma^2}\right] \tag{4}$$

where $\sigma$ is an empirical parameter controlling the sensitivity of the Gaussian function to the distance difference. Equation (4) measures the relative balance between the coordinate variations of two tentative correspondences along the two retained principal directions. It defines an empirical first-order affinity in the PCA-reduced representation rather than a strict geometric invariant between the original images under arbitrary transformations. In particular, similar distances along the two retained directions do not imply identical physical scaling along the original image axes. The adjacency matrix

$$A_{ij} = \begin{cases} 1, & s_{ij} \geq \tau, \\ 0, & s_{ij} < \tau. \end{cases} \tag{5}$$

is then constructed. Because $s_{ij} = s_{ji}$, matrix $A$ is a symmetric binary matrix.

To distinguish stable spatial relations from candidate first-order edges formed by chance, SPGSC++ further computes common-neighbor support.

$$C = A \cdot A^{\mathrm{T}} \tag{6}$$

where $C_{ij} = \sum_{h=1}^{N} A_{ih} A_{hj}$ is the number of length-2 paths between nodes $i$ and $j$, as illustrated in Figure 2. When $A$ is symmetric, this value is also the number of common neighbors of the two nodes. To exclude node pairs with no direct compatibility from the quality calculation, a Hadamard product is used for element-wise masking:

$$W = C \odot A \tag{7}$$

Consequently, only directly connected candidate edges supported by common neighbors contribute to the second-order structural quality. An accidental edge that satisfies Equation (4) but remains isolated from the surrounding structure generally receives limited support, whereas an edge repeatedly supported by multiple common neighbors obtains a higher quality score. Supported edges are aggregated into node quality and used by IDSS and DSR for model generation and refinement.

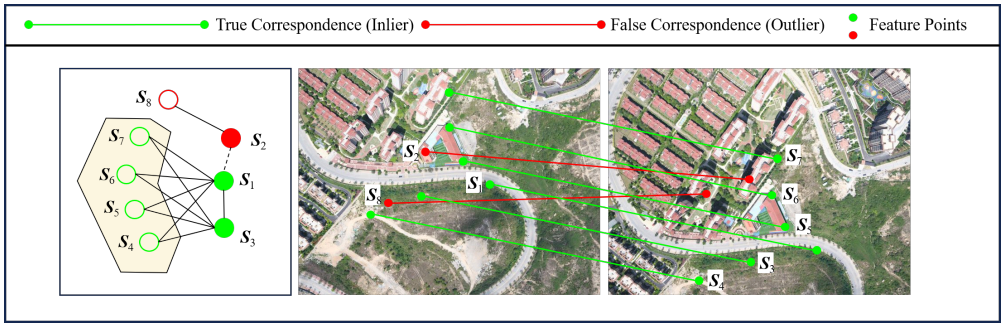



**Figure 2.** Principle of second-order proximity.

### *3.2. Second-Order-Quality-Guided Initial Data Subset Selection*

After obtaining the node quality score $q_i$ for each feature correspondence, we use second-order quality-guided initial data subset selection (IDSS) to select a minimal sample set for generating a model hypothesis. The strategy converts each second-order quality score into a sampling probability and applies sequential weighted sampling without replacement. Geometrically reliable correspondences are consequently prioritized during hypothesis construction, improving sampling efficiency and providing a high-quality initial solution for subsequent model refinement.

Let $W = \{q_i\}_{i=1}^{N}$ be the set of all node quality scores. The softmax function first converts the compatibility-score set $W$ into the probability distribution $P = \{\pi_i\}_{i=1}^{N}$:

$$\pi_i = \frac{q_i}{\sum_{n=1}^{N} q_n}, \sum_{i=1}^{N} \pi_i = 1 \tag{8}$$

The probability distribution prioritizes nodes with high quality scores while retaining exploration, providing a mathematically defined input for efficient and robust weighted minimal-set generation. To prevent repeated selection of the same sample, a subset is constructed by sequential weighted sampling without replacement. The candidate index set $P^{(0)} = \{1, \ldots, N\}$ is initialized first. At the $t$-th draw, the probability distribution $\pi_i^{t-1} = \frac{q_i}{\sum_{j \in C^{(t-1)}} q_n}$ is renormalized over the current candidate set $P^{(t-1)}$, a sample $i_t$ is drawn accordingly, and $i_t$ is removed from the candidate set. Each subsequent probability distribution is therefore recalculated over a smaller set. This prevents a high-scoring correspondence from being selected repeatedly while preserving its priority and introducing sample diversity. The resulting subset is more reliable under the geometric constraints and provides a robust basis for constructing model hypotheses.

### *3.3. Data Subset Refinement*

Initial data subset selection prevents most outliers from entering the sample. When the inlier ratio among the tentative correspondences is low, however, some outliers may still enter the subset and reduce the accuracy of model estimation. SPGSC++ therefore uses a residual-rank-band-driven data subset refinement strategy. Let the subset at iteration $t$ be $S^{(t)}$. A model $\theta^{(t)}$ is fitted from this subset, and residuals are computed for all correspondences and sorted in ascending order:

$$r_{i_1^{(t)}}(\theta^{(t)}) \leq r_{i_2^{(t)}}(\theta^{(t)}) \leq \cdots \leq r_{i_N^{(t)}}(\theta^{(t)}) \tag{9}$$

The subset for the next iteration is constructed from a residual-rank band of width $p$ ending at the $k$-th ordered residual:

$$S^{(t+1)} = \{m_{i_{k-p+1}^{(t)}}, \ldots, m_{i_k^{(t)}}\} \tag{10}$$

The cost of a candidate model is defined as the sum of residuals within this band:

$$J^{(t)} = \sum_{h=k-p+1}^{k} r_{i_h^{(t)}}(\theta^{(t)}) \tag{11}$$

The model with the smallest $J^{(t)}$ observed during the iterations is retained as the current best hypothesis. Unlike repeatedly selecting the $p$ samples with the smallest residuals, the fixed-width band focuses each update on samples near a prescribed consensus-set boundary. Second-order structural quality also controls termination of the refinement process. Let $q_{i_k^{(t)}}$ denote the quality score of the correspondence associated with the $k$-th ordered residual, and let $\bar{q}(S^{(t)})$ and $\bar{q}(S^{(t-1)})$ be the mean qualities of the subsets from the two preceding iterations. Refinement terminates when $q_{i_k^{(t)}} > \bar{q}(S^{(t)}), q_{i_k^{(t)}} > \bar{q}(S^{(t-1)})$ is satisfied; otherwise, the subset is updated using the new residual-rank band. This procedure yields an effective subset with a high inlier ratio and coherent geometry. The resulting model separates inliers from outliers and provides a reliable basis for robust model estimation. The complete SPGSC++ procedure is summarized in Algorithm 1.

All experiments use a homography model. The parameter $p$ denotes the model-dependent minimal sampling/fitting subset size rather than a freely tuned residual-band width; therefore, $p = 4$ for the homography used in this study. The parameter $k$ denotes

the residual-rank position used during DSR updating, and $T$ is the maximum number of refinement iterations. The common configuration selected from the sensitivity analysis is $\sigma = 0.4$, $\tau = 0.4$, $k = 20$, $T = 20$ and $p = 4$.

**Algorithm 1** SPGSC++ for second-order spatial guided correspondence removal.

**Input:** tentative correspondence set $\mathcal{M}$, minimal-sample size $p$, residual rank $k$, maximum iterations $T$
**Output:** estimated model $\boldsymbol{\theta}^*$ and inlier set $\mathcal{I}$

1. Construct a 4D joint matrix $Z$ from the correspondence coordinates and obtain its 2D PCA representation $\widetilde{Z}$;
2. Compute pairwise distance differences $s_{ij}$ along the two principal-component directions and threshold them to obtain $A$;
3. Compute $C = A \cdot A^{\mathrm{T}}$, $W = C \odot A$, and $q_i = \sum_j W_{ij}$;
4. Normalize $q$ into $p$ and draw $s$ samples without replacement;
5. For $t = 1, \ldots, T$:
(1) Fit $\theta^{(t)}$ from $S^{(t)}$, then compute and rank all residuals;
(2) Compute the residual-rank-band cost $J^{(t)}$ and retain the current model if the cost is lower;
(3) Evaluate the second-order structural-quality stopping criterion and terminate if it is satisfied;
(4) Use samples with residual ranks from $k - p + 1$ to $k$ to construct $S^{(t+1)}$;
6. Set $\hat{\theta}$ to the best-so-far model and return the inlier set $\hat{I}$.

## 4. Datasets and Evaluation Metrics

### *4.1. Datasets*

The evaluation comprises image pairs from five datasets. All methods use the same tentative correspondences, homography setting, and evaluation thresholds. These datasets cover different degrees of viewpoint variation, scale variation, texture change, and local geometric distortion, allowing the inlier-identification performance of the algorithm to be assessed under challenging tentative correspondences. All methods take the same precomputed tentative correspondences as input and classify them as inliers or outliers according to their estimated models. The datasets are described below.

(1) CIAP: The dataset contains 54 pairs of 700 × 700 color-infrared aerial images acquired over eastern Illinois, USA. All image pairs are orthorectified and are suitable for a rigid-transformation model, although their overlapping regions are relatively small. CIAP is commonly used in tasks such as feature-matching-based image stitching [26].
(2) UAV: The UAV dataset contains 41 image pairs of 600 × 337 pixels acquired by a rotary-wing UAV over farmland. Low-altitude flight and variations in imaging angle introduce marked projective distortion. The dataset is suitable for evaluating feature-matching methods in practical applications such as automated UAV-based crop monitoring [27].
(3) SUIRD: This dataset contains 60 pairs of low-altitude remote sensing images captured by small UAVs in natural landscapes and urban streets. Viewpoint changes during acquisition produce low overlap ratios, nonrigid image deformation, and extreme viewing-angle differences. According to the type of viewpoint variation, the dataset is divided into five groups: Horizontal, Vertical, Scaling, Mixture, and Extreme [28].
(4) ZHANGZHOU: This self-constructed UAV aerial-image dataset contains 50 high-resolution images of 11,648 × 8736 pixels. The images were collected during a continuous UAV flight over the northern area of the Zhangzhou Development Zone, Fujian Province, China, and were used for a preliminary evaluation of matching performance on consecutive frames [29].

(5) PAN: This dataset consists of 33 pairs of panchromatic aerial images with resolutions of 561 × 518 or 600 × 700. These image pairs are captured at different times using a frame camera and often include terrain relief variations or large viewpoint changes. This dataset is commonly used in change detection tasks and presents significant challenges to the stability of feature matching algorithms under complex geometric conditions [30].

### *4.2. Evaluation Metrics*

Precision, Recall, F-Score, Runtime, and mean reprojection error (MRE) are used for evaluation. The first three metrics characterize correspondence-classification performance, Runtime measures the computational efficiency of the false-correspondence-removal, and MRE measures the geometric accuracy of the homography estimated from the retained correspondences. All methods use identical tentative correspondences, homography settings, and evaluation thresholds. Runtime covers only the false-correspondence removal and excludes feature detection, description, and tentative matching. Each experiment is repeated 50 times, and the arithmetic mean Runtime is reported. Experiments are conducted in MATLAB R2018b on a 64-bit Windows 10 computer equipped with an Intel(R) Core(TM) i5-1035G1 CPU @ 1.00 GHz and 8 GB RAM.

Let $TP$ denote the number of true inliers correctly classified as inliers, $FP$ the number of true outliers incorrectly classified as inliers, and $FN$ the number of true inliers incorrectly removed. The metrics are defined as

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{12}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{13}$$

$$\text{F-Score} = \frac{2\,\text{Precision}\,\text{Recall}}{\text{Precision} + \text{Recall}} \tag{14}$$

Precision is the proportion of predicted inliers that are correct, Recall is the proportion of true inliers retained, and F-Score is their harmonic mean. Runtime is the mean processing time for false-correspondence removal on one test case. MRE is the one-way mean Euclidean reprojection error from the first image to the second: evaluation correspondences are projected using the estimated homography and homogeneous normalization, and the Euclidean distance to the corresponding coordinates in the second image is averaged over valid projections. MRE is reported in pixels. Higher Precision, Recall, and F-Score and lower Runtime and MRE indicate better performance.

## 5. Experimental Design and Results

### *5.1. Qualitative Comparison*

Figure 3 separately presents tentative correspondences before optimization and the correspondences retained after SPGSC++ processing. Figure 4 shows the checkerboard images of each example after registration, which are used to observe the matching quality in regions with repetitive texture, viewpoint variation, or local deformation. Green and red lines denote ground-truth inliers and outliers, respectively.

The image pairs include small overlaps, scale changes, illumination differences, and local geometric distortions, all of which can generate numerous outliers during tentative feature matching. The input correspondences are generated using SIFT and generally contain many false correspondences and relatively few true matches. Across the ten example image pairs, the numbers of feature pairs and proportions of inliers are shown as Table 2. SPGSC++ retains most true correspondences in these challenging cases while removing

all but a small number of outliers. As is shown in Table 3, these results indicate robust performance across different geometric transformations and high outlier ratios.

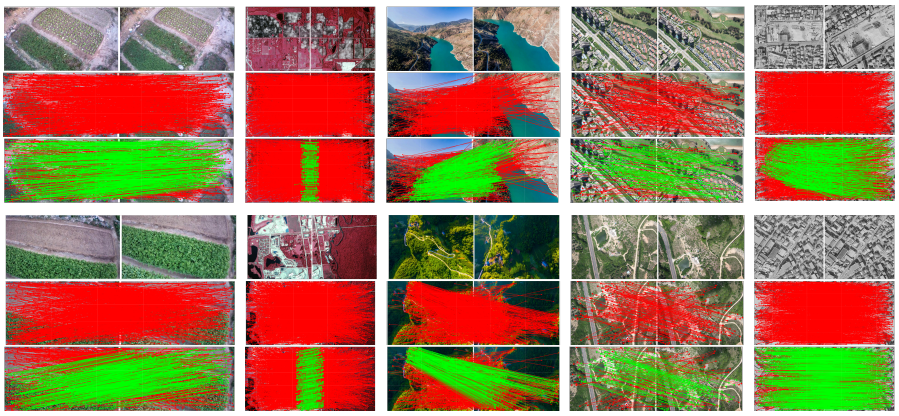


**Figure 3.** Qualitative matching results of SPGSC++. Red lines represent the initial matching pairs, and green lines represent the inliers retained after outlier removal.

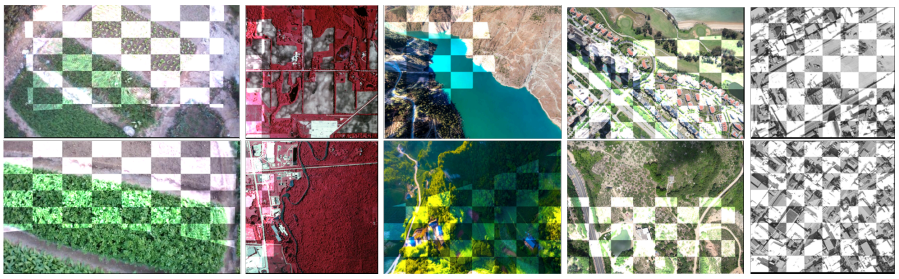


**Figure 4.** Checkerboard results of SPGSC++. The fixed image and the registered moving image are interleaved in a checkerboard pattern. Each pair of adjacent tiles originates from the two different images. Continuous structures and textures across tile boundaries indicate accurate alignment.

**Table 2.** The numbers of feature pairs and proportions of inliers in representative image pairs.

| Dataset | Image Pairs | Feature Pairs Number | Inliers Proportions |
|---|---|---|---|
| CIAP | 39 | 908 | 42.18% |
| | 45 | 885 | 40.11% |
| SUIRD | 5 | 2297 | 13.32% |
| | 39 | 2280 | 19.30% |
| UAV | 30 | 812 | 55.05% |
| | 35 | 873 | 27.38% |
| ZHANGZHOU | 10 | 249 | 59.04% |
| | 15 | 194 | 57.73% |
| PAN | 24 | 506 | 26.99% |
| | 25 | 870 | 50.49% |

**Table 3.** The qualitative matching results of SPGSC++ in representative image pairs.

| Dataset | Image Pairs | Precision | Recall | F-Score |
|---|---|---|---|---|
| CIAP | 39 | 98.18% | 100% | 99.08% |
| | 45 | 92.74% | 98.85% | 95.70% |
| SUIRD | 5 | 98.31% | 100% | 99.15% |
| | 39 | 99.47% | 100% | 99.74% |
| UAV | 30 | 98.88% | 99.37% | 99.12% |
| | 35 | 98.48% | 100% | 99.23% |

**Table 3.** *Cont*.

| Dataset | Image Pairs | Precision | Recall | F-Score |
|---|---|---|---|---|
| ZHANGZHOU | 10 | 94.19% | 99.32% | 96.69% |
| | 15 | 94.07% | 99.11% | 96.52% |
| PAN | 24 | 96.69% | 99.88% | 98.26% |
| | 25 | 98.02% | 96.92% | 97.47% |

### *5.2. Quantitative Comparison*

The comparison includes RANSAC, TRESAC, LPM, LGSC, TRGDC, IMSCAN, the original SPGSC, and SPGSC++. RANSAC provides a uniform-sampling baseline; TRESAC represents relation-guided sampling; LPM, LGSC, TRGDC, and IMSCAN represent local-neighborhood, graph-structured, spatial-clustering and spatial-based correspondence-removal strategies, respectively; and SPGSC serves as the direct predecessor baseline for assessing the incremental contribution. All methods use identical tentative correspondences, homography settings, and evaluation thresholds.

As shown in Table 4, across the five datasets, SPGSC++ achieves mean Precision, Recall, F-Score, Runtime, and MRE values of 0.9627, 0.9797, 0.9702, 0.3724 s, and 1.6454 pixels, respectively. Its MRE values on CIAP, SUIRD, UAV, ZHANGZHOU, and PAN are 0.1724, 1.8878, 2.8210, 1.8676, and 1.4781 pixels. SPGSC++ obtains the lowest MRE on four datasets, whereas IMSCAN obtains the lowest value on SUIRD.

Relative to the original SPGSC, SPGSC++ improves the five-dataset mean Precision, Recall, and F-Score by 0.0306, 0.0416, and 0.0391, respectively. The mean Runtime decreases from 0.5508 to 0.3724 s, and the mean MRE decreases from 3.9292 to 1.6454 pixels, corresponding to reductions of 32.39% and 58.12%. Relative to TRESAC, SPGSC++ increases the mean F-Score from 0.9393 to 0.9702, reduces the mean Runtime from 0.6352 to 0.3724 s, and reduces the mean MRE from 2.8323 to 1.6454 pixels. These results demonstrate an overall gain over both the direct predecessor and the added guided-sampling baseline.

On PAN, SPGSC++ achieves a Precision of 0.9728, a Recall of 0.9944, an F-Score of 0.9827, a Runtime of 0.0123 s, and an MRE of 1.4781 pixels. It obtains the highest Recall and F-Score and the lowest MRE, whereas TRGDC achieves a higher Precision. The method therefore achieves a favorable balance between performance and efficiency, with its runtime generally falling in the middle range among the evaluated methods.

### *5.3. Parameter Sensitivity Analysis*

A one-factor-at-a-time protocol is used to evaluate the sensitivity to $\sigma$, $\tau$, $k$ and $T$. One parameter is varied, while all remaining parameters are fixed at their reference values. CIAP, UAV, SUIRD, and ZHANGZHOU use the same candidate ranges and fixed settings without dataset-specific tuning. Because $p = 4$ is determined by the homography model, it is not treated as a free hyperparameter. Figure 5 reports the F-Score for each setting. The highest four-dataset mean F-Score for k is 0.9818 at $k = 20$, with values of 0.9679–0.9818 over $k = 19$–$22$. The highest mean F-Score for $T$ is 0.9805 at $T = 50$, with values of 0.9760–0.9805 over $T = 48$–$52$. The highest mean F-Scores for $\tau$ and $\sigma$ are 0.9860 and 0.9817, both at 0.4; the values remain between 0.9703 and 0.9817 over $\sigma = 0.3$–$0.5$. These results indicate local stability around the selected configuration. Therefore, the common configuration selected from the sensitivity analysis is $\sigma = 0.4$, $\tau = 0.4$, $k = 20$, $T = 50$ and $p = 4$.

**Table 4.** Quantitative comparison on five datasets.

| Dataset | Method | Runtime (s) | Precision | Recall | F-Score | MRE (Pixel) |
|---|---|---|---|---|---|---|
| CIAP | RANSAC | 8.0606 | 0.7343 | 0.7786 | 0.7491 | 6.4728 |
| | LPM | 0.0112 | 0.9819 | 0.9988 | 0.9898 | 0.1817 |
| | LGSC | 0.0583 | 0.9346 | 0.9985 | 0.9612 | 0.8085 |
| | TRGDC | 0.4081 | 0.9417 | 0.9977 | 0.9660 | 0.7401 |
| | IMSCAN | 0.7537 | 0.9987 | 0.9284 | 0.9543 | 0.9814 |
| | TRESAC | 0.9403 | 0.9738 | 0.9797 | 0.9761 | 0.4405 |
| | SPGSC | 0.9198 | 0.9743 | 0.9864 | 0.9798 | 0.3937 |
| | SPGSC++ | 0.8160 | 0.9885 | 0.9970 | 0.9926 | 0.1724 |
| SUIRD | RANSAC | 1.4870 | 0.8448 | 0.9367 | 0.8807 | 4.7507 |
| | LPM | 0.0110 | 0.8892 | 0.9748 | 0.9172 | 3.1676 |
| | LGSC | 0.1109 | 0.9193 | 0.9370 | 0.9088 | 2.8576 |
| | TRGDC | 0.4210 | 0.9308 | 0.6981 | 0.7546 | 5.0406 |
| | IMSCAN | 1.0010 | 0.9096 | 0.9254 | 0.9149 | 0.9336 |
| | TRESAC | 1.0994 | 0.9071 | 0.9487 | 0.9222 | 2.4884 |
| | SPGSC | 0.9364 | 0.8958 | 0.9096 | 0.8927 | 4.3799 |
| | SPGSC++ | 0.7438 | 0.9238 | 0.9434 | 0.9317 | 1.8878 |
| UAV | RANSAC | 2.9877 | 0.8825 | 0.9151 | 0.8971 | 5.9372 |
| | LPM | 0.0049 | 0.9323 | 0.9959 | 0.9594 | 3.4303 |
| | LGSC | 0.0907 | 0.9370 | 0.9980 | 0.9630 | 3.0768 |
| | TRGDC | 0.2702 | 0.9492 | 0.9829 | 0.9635 | 3.1036 |
| | IMSCAN | 0.7394 | 0.9547 | 0.9341 | 0.9371 | 3.6929 |
| | TRESAC | 0.5313 | 0.9770 | 0.7991 | 0.8662 | 4.8860 |
| | SPGSC | 0.3848 | 0.9150 | 0.8735 | 0.8866 | 5.5653 |
| | SPGSC++ | 0.2560 | 0.9778 | 0.9700 | 0.9728 | 2.8210 |
| ZHANGZHOU | RANSAC | 1.2969 | 0.9020 | 0.9713 | 0.9345 | 4.9139 |
| | LPM | 0.0037 | 0.9087 | 0.9961 | 0.9487 | 2.7384 |
| | LGSC | 0.0257 | 0.9349 | 0.9971 | 0.9640 | 2.2068 |
| | TRGDC | 0.0858 | 0.9433 | 0.9823 | 0.9615 | 2.2900 |
| | IMSCAN | 0.1500 | 0.9844 | 0.9350 | 0.9527 | 2.9511 |
| | TRESAC | 0.0714 | 0.9931 | 0.9394 | 0.9611 | 3.8573 |
| | SPGSC | 0.0689 | 0.9469 | 0.9809 | 0.9630 | 4.8781 |
| | SPGSC++ | 0.0342 | 0.9507 | 0.9937 | 0.9713 | 1.8676 |
| PAN | RANSAC | 0.9872 | 0.9909 | 0.7656 | 0.8315 | 5.1105 |
| | LPM | 2.4955 | 0.9439 | 0.9634 | 0.9534 | 3.0218 |
| | LGSC | 0.1444 | 0.9555 | 0.9394 | 0.9332 | 3.6066 |
| | TRGDC | 0.4947 | 0.9968 | 0.9330 | 0.9527 | 3.3165 |
| | IMSCAN | 1.2616 | 0.9748 | 0.9755 | 0.9715 | 2.1902 |
| | TRESAC | 0.5338 | 0.9942 | 0.9561 | 0.9708 | 2.4895 |
| | SPGSC | 0.4443 | 0.9287 | 0.9401 | 0.9335 | 4.4290 |
| | SPGSC++ | 0.0123 | 0.9728 | 0.9944 | 0.9827 | 1.4781 |

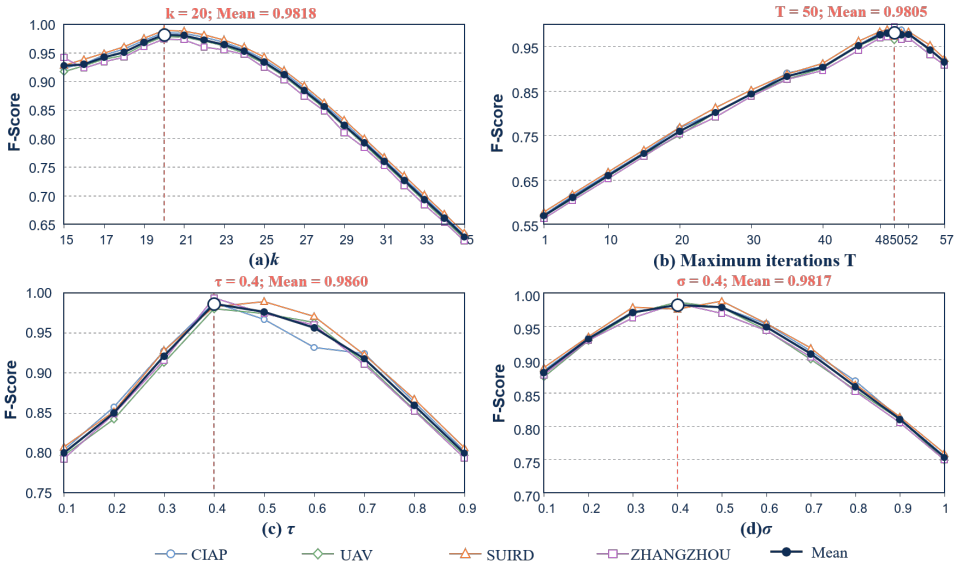



**Figure 5.** Parameter sensitivity of SPGSC++.

### *5.4. Ablation Study*

The ablation study uses UAV and treats the full SPGSC++ method as the reference. All module variants are re-evaluated on the UAV image pairs using identical tentative correspondences, homography settings, parameter configurations, and evaluation protocols. The ablated variants are as follows: SPGSC++ without PCA; Pairwise+IDSS+DSR, in which are SPGSC++ without PCA; Pairwise+IDSS+DSR, which replaces the common-neighbor second-order relation with the candidate first-order relation; SPG+DSR without IDSS; SPG+IDSS without DSR; and the complete SPGSC++. In this subsection, SPG denotes the second-order spatial relation.

As shown in Table 5, the complete SPGSC++ achieves a Precision of 0.9778, a Recall of 0.9700, an F-Score of 0.9728, and a Runtime of 0.2560 s. Removing PCA increases the Runtime to 0.3941 s and reduces the F-Score to 0.9520. The complete SPGSC++ (with PCA) therefore reduces the Runtime by 35.04% relative to the no-PCA variant while maintaining or even improving correspondence-removal performance. This finding supports PCA as a compact representation mechanism.

**Table 5.** Results of the ablation study.

| Method | Runtime (s) | Precision | Recall | F-Score |
|---|---|---|---|---|
| SPGSC++ without PCA | 0.3941 | 0.9563 | 0.9548 | 0.9520 |
| Pairwise+IDSS+DSR | 0.2803 | 0.8509 | 0.8621 | 0.8335 |
| SPG+DSR (without IDSS) | 0.2486 | 0.8711 | 0.6509 | 0.7202 |
| SPG+IDSS (without DSR) | 0.2389 | 0.8290 | 0.8272 | 0.8182 |
| SPGSC++ | 0.2706 | 0.9355 | 0.9030 | 0.9155 |

Replacing the second-order common-neighbor relation with the first-order relation reduces the F-Score to 0.8633; the complete SPGSC++ improves the F-Score by 0.1095 while reducing the Runtime from 0.2805 to 0.2560 s. Removing IDSS produces the largest decrease among the structural and sampling variants, reducing the F-Score to 0.7791, mainly because the Recall decreases from 0.9700 to 0.7159. Removing DSR reduces the F-Score to 0.8677. The variants without IDSS and DSR save only approximately 1.96% and 0.25% of the runtime, respectively, but incur substantial performance losses. These results show that the improvement arises from the complementary effects of second-order structural verification, structure-quality-guided initialization, and residual-informed refinement. Figure 6 separately reports the F-Score and Runtime of the five variants.

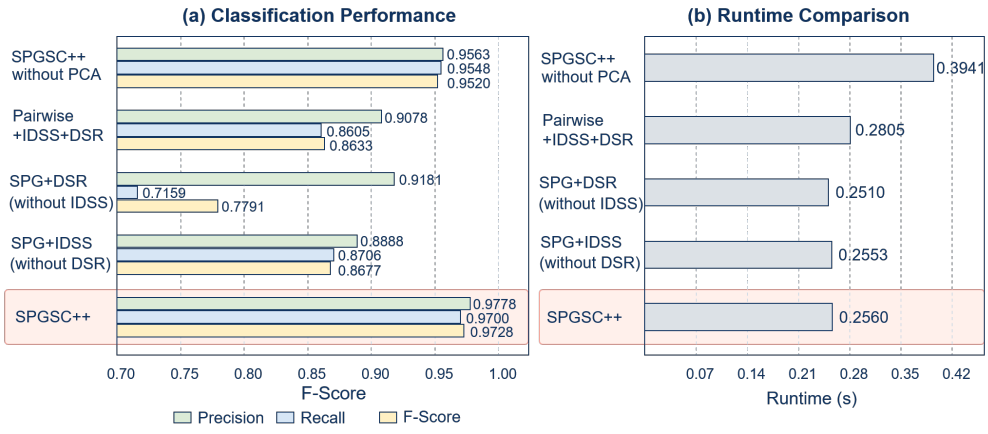



**Figure 6.** Ablation Study of SPGSC++.

### *5.5. Robustness Under Varying Outlier Ratios and Statistical Significance Analysis*

To characterize the empirical operating range of the complete SPGSC++ pipeline, 20 image pairs are selected to cover outlier ratios from 2.03% to 91.22% and are evaluated using the same configuration without pair-specific tuning. As is shown in Figure 7, across the 20 pairs, the mean Precision, Recall, and F-Score are 0.9639, 0.9726, and 0.9673, respectively, and the individual F-Scores range from 0.8379 to 1.0000. For the four pairs with outlier ratios of at least 79%, the mean Precision, Recall, and F-Score are 0.9731, 0.9847, and 0.9787. At outlier ratios of 88.18% and 91.22%, the F-Scores remain 0.9728 and 0.9614.

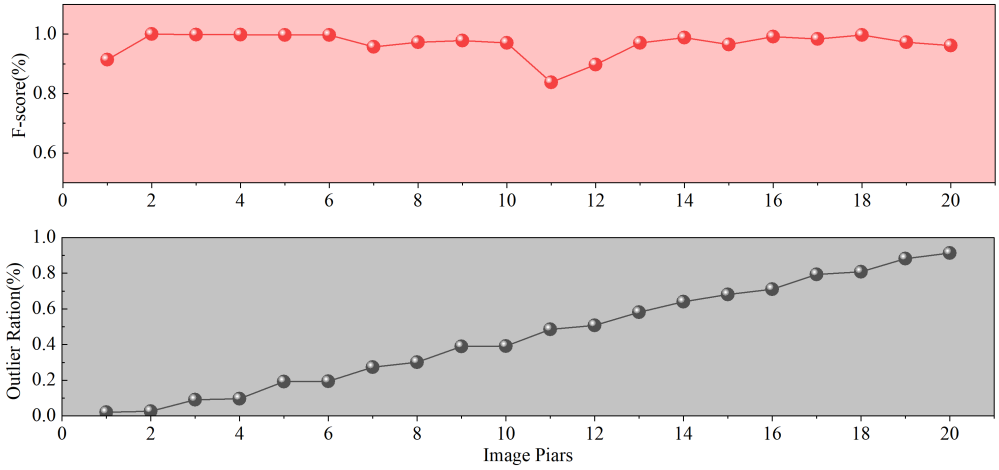



**Figure 7.** Performance robustness of SPGSC++ under different outlier ratios.

To assess the stability and statistical significance of the algorithm performance, the Precision, Recall, F-Score, and Runtime are statistically evaluated on five datasets, and their means, standard deviations, and 95% confidence intervals are calculated. In addition, a two-sided Wilcoxon signed-rank test is used to evaluate the paired performance differences between SPGSC++ and each comparison method at a significance level of 0.05. Detailed statistics are presented in Tables 6–10, with Figures 8–12 showing boxplots comparing SPGSC++ with the baseline methods.

The statistical analysis shows that, on the CIAP dataset, SPGSC++ achieves mean Precision and F-Score values of 0.9885 and 0.9926, with standard deviations of 0.0229 and 0.0134 and narrow 95% confidence-interval half-widths of $\pm$0.0062 and $\pm$0.0037, respectively, indicating limited variation across image pairs. The Wilcoxon tests show that SPGSC++ is significantly better than RANSAC and LGSC in both Precision and F-Score ($p < 0.05$). For the Recall, SPGSC++ reaches 0.9970, 0.9937, and 0.9944 on CIAP, ZHANGZHOU, and PAN, respectively; on PAN, its advantages over RANSAC, LGSC, and TRGDC are statistically significant. On SUIRD and UAV, SPGSC++ achieves F-Scores of 0.9317 and 0.9728, respectively; the former differs significantly from LPM, LGSC, and TRGDC, whereas the latter differs significantly from IMSCAN and TRESAC. These results indicate that SPGSC++ does not obtain high scores simply by enlarging the retained set. Instead, second-order common-neighbor support validates candidate edges, IDSS prioritizes correspondences with high structural quality, and DSR further refines the data subset, thereby suppressing false correspondences while preserving high recall. In terms of the Runtime, LPM has the lowest mean runtime on the first four datasets, whereas SPGSC++ is the fastest method on PAN, requiring 0.0123 s on average. SPGSC++ is significantly faster than RANSAC on all five datasets ($p < 0.05$). Compared with its predecessor SPGSC, its runtime decreases from 0.3848 to 0.2560 s on UAV and from 0.4443 to 0.0123 s on PAN, with both differences being statistically significant. This finding indicates that PCA-based compact representation

and the subsequent structure-guided sampling reduce the computational burden in some scenarios without sacrificing matching quality.

Comprehensive analysis across all datasets indicates that SPGSC++ achieves the highest mean F-Score on each of the five datasets, together with relatively small standard deviations and confidence intervals. Overall, the joint action of second-order spatial relations, IDSS, and DSR provides a stable balance between Precision and Recall, whereas PCA primarily supplies a compact representation and improves computational efficiency. These results collectively support the effectiveness of SPGSC++ in matching quality, robustness, and efficiency.

**Table 6.** Statistical significance analysis of benchmark methods versus the proposed method on the CIAP dataset.

| Metric | Statistic | RANSAC | LPM | LGSC | TRGDC | IMSCAN | TRESAC | SPGSC | SPGSC++ |
|---|---|---|---|---|---|---|---|---|---|
| Precision | Mean | 0.7343 | 0.9819 | 0.9346 | 0.9417 | 0.9987 | 0.9738 | 0.9743 | 0.9885 |
| | Std | 0.4020 | 0.0408 | 0.1186 | 0.0999 | 0.0076 | 0.1013 | 0.0914 | 0.0229 |
| | 95%CI | ±0.1097 | ±0.0111 | ±0.0324 | ±0.0273 | ±0.0021 | ±0.0276 | ±0.0250 | ±0.0062 |
| | *p*-value | $3.73 \times 10^{-5}$ | 0.582 | 0.032 | 0.213 | $4.48 \times 10^{-7}$ | 0.004 | 0.279 | |
| Recall | Mean | 0.7786 | 0.9988 | 0.9985 | 0.9977 | 0.9284 | 0.9797 | 0.9864 | 0.9970 |
| | Std | 0.3486 | 0.0029 | 0.0035 | 0.0041 | 0.1483 | 0.0775 | 0.0565 | 0.0066 |
| | 95%CI | ±0.0951 | ±0.0008 | ±0.0010 | ±0.0011 | ±0.0405 | ±0.0211 | ±0.0154 | ±0.0018 |
| | *p*-value | $4.77 \times 10^{-4}$ | 0.048 | 0.212 | 0.833 | $2.50 \times 10^{-6}$ | 0.044 | 0.697 | |
| F-Score | Mean | 0.7491 | 0.9898 | 0.9612 | 0.9660 | 0.9543 | 0.9761 | 0.9798 | 0.9926 |
| | Std | 0.3862 | 0.0225 | 0.0713 | 0.0576 | 0.1140 | 0.0889 | 0.0764 | 0.0134 |
| | 95%CI | ±0.1054 | ±0.0062 | ±0.0195 | ±0.0157 | ±0.0311 | ±0.0243 | ±0.0208 | ±0.0037 |
| | *p*-value | $9.60 \times 10^{-5}$ | 0.510 | 0.028 | 0.250 | 0.187 | 0.095 | 0.105 | |
| Runtime | Mean | 8.0606 | 0.0112 | 0.0583 | 0.4081 | 0.7537 | 0.9403 | 0.9198 | 0.8160 |
| | Std | 10.2402 | 0.0098 | 0.0542 | 0.2977 | 0.8289 | 1.6176 | 1.0325 | 1.1852 |
| | 95%CI | ±2.7950 | ±0.0027 | ±0.0148 | ±0.0813 | ±0.2262 | ±0.4415 | ±0.2818 | ±0.3235 |
| | *p*-value | $1.63 \times 10^{-10}$ | $1.63 \times 10^{-10}$ | $1.63 \times 10^{-10}$ | 0.990 | 0.695 | 0.005 | 0.820 | |

**Table 7.** Statistical significance analysis of benchmark methods versus the proposed method on the SUIRD dataset.

| Metric | Statistic | RANSAC | LPM | LGSC | TRGDC | IMSCAN | TRESAC | SPGSC | SPGSC++ |
|---|---|---|---|---|---|---|---|---|---|
| Precision | Mean | 0.8448 | 0.8892 | 0.9193 | 0.9308 | 0.9096 | 0.9071 | 0.8958 | 0.9238 |
| | Std | 0.2329 | 0.1942 | 0.1379 | 0.1498 | 0.1936 | 0.1854 | 0.1884 | 0.0448 |
| | 95%CI | ±0.0662 | ±0.0552 | ±0.0392 | ±0.0426 | ±0.0550 | ±0.0527 | ±0.0536 | ±0.0127 |
| | *p*-value | 0.237 | 0.124 | 0.028 | 0.002 | 0.033 | 0.063 | 0.051 | |
| Recall | Mean | 0.9367 | 0.9748 | 0.9369 | 0.6981 | 0.9254 | 0.9487 | 0.9096 | 0.9434 |
| | Std | 0.1358 | 0.0416 | 0.1903 | 0.3355 | 0.1631 | 0.1164 | 0.2151 | 0.0768 |
| | 95%CI | ±0.0386 | ±0.0118 | ±0.0541 | ±0.0954 | ±0.0464 | ±0.0331 | ±0.0611 | ±0.0218 |
| | *p*-value | 0.085 | 0.066 | 0.181 | $1.35 \times 10^{-4}$ | 0.511 | 0.288 | 0.293 | |
| F-Score | Mean | 0.8807 | 0.9172 | 0.9088 | 0.7546 | 0.9149 | 0.9222 | 0.8927 | 0.9317 |
| | Std | 0.2038 | 0.1582 | 0.1993 | 0.3019 | 0.1778 | 0.1569 | 0.2018 | 0.0499 |
| | 95%CI | ±0.0579 | ±0.0449 | ±0.0566 | ±0.0858 | ±0.0505 | ±0.0446 | ±0.0574 | ±0.0142 |
| | *p*-value | 0.149 | 0.036 | 0.024 | 0.015 | 0.090 | 0.088 | 0.218 | |
| Runtime | Mean | 1.4870 | 0.0110 | 0.1109 | 0.4210 | 1.0010 | 1.0994 | 0.9364 | 0.7438 |
| | Std | 0.3147 | 0.0055 | 0.0534 | 0.4521 | 0.6204 | 0.1715 | 0.7159 | 0.6895 |
| | 95%CI | ±0.0894 | ±0.0016 | ±0.0152 | ±0.1285 | ±0.1763 | ±0.0487 | ±0.2035 | ±0.1960 |
| | *p*-value | $1.72 \times 10^{-7}$ | $7.56 \times 10^{-10}$ | $7.56 \times 10^{-10}$ | $2.48 \times 10^{-8}$ | $2.78 \times 10^{-6}$ | $2.31 \times 10^{-5}$ | 0.229 | |

**Table 8.** Statistical significance analysis of benchmark methods versus the proposed method on the UAV dataset.

| Metric | Statistic | RANSAC | LPM | LGSC | TRGDC | IMSCAN | TRESAC | SPGSC | SPGSC++ |
|---|---|---|---|---|---|---|---|---|---|
| Precision | Mean | 0.8825 | 0.9323 | 0.9370 | 0.9492 | 0.9547 | 0.9770 | 0.9150 | 0.9778 |
| | Std | 0.2438 | 0.1108 | 0.1085 | 0.0744 | 0.0638 | 0.1093 | 0.1893 | 0.0372 |
| | 95%CI | ±0.0770 | ±0.0350 | ±0.0343 | ±0.0235 | ±0.0201 | ±0.0345 | ±0.0598 | ±0.0117 |
| | *p*-value | 0.245 | 0.362 | 0.250 | 0.275 | 0.160 | 0.020 | 0.110 | |
| Recall | Mean | 0.9151 | 0.9959 | 0.9980 | 0.9829 | 0.9341 | 0.7991 | 0.8735 | 0.9700 |
| | Std | 0.2134 | 0.0070 | 0.0043 | 0.0418 | 0.1313 | 0.2171 | 0.2432 | 0.0562 |
| | 95%CI | ±0.0674 | ±0.0022 | ±0.0014 | ±0.0132 | ±0.0415 | ±0.0685 | ±0.0768 | ±0.0178 |
| | *p*-value | 0.347 | $7.42 \times 10^{-4}$ | $8.13 \times 10^{-5}$ | 0.270 | 0.469 | $3.18 \times 10^{-5}$ | 0.178 | |
| F-Score | Mean | 0.8971 | 0.9594 | 0.9630 | 0.9635 | 0.9371 | 0.8662 | 0.8866 | 0.9728 |
| | Std | 0.2305 | 0.0666 | 0.0652 | 0.0431 | 0.0827 | 0.1707 | 0.2205 | 0.0383 |
| | 95%CI | ±0.0728 | ±0.0210 | ±0.0206 | ±0.0136 | ±0.0261 | ±0.0539 | ±0.0696 | ±0.0121 |
| | *p*-value | 0.470 | 0.781 | 0.650 | 0.220 | 0.046 | $4.15 \times 10^{-4}$ | 0.085 | |
| Runtime | Mean | 2.9877 | 0.0049 | 0.0907 | 0.2702 | 0.7394 | 0.5313 | 0.3848 | 0.2560 |
| | Std | 2.3775 | 0.0022 | 0.0494 | 0.1378 | 0.4434 | 0.2095 | 0.1517 | 0.2345 |
| | 95%CI | ±0.7504 | ±0.0007 | ±0.0156 | ±0.0435 | ±0.1400 | ±0.0661 | ±0.0479 | ±0.0740 |
| | *p*-value | $2.42 \times 10^{-8}$ | $2.42 \times 10^{-8}$ | $5.86 \times 10^{-8}$ | 0.505 | $2.42 \times 10^{-8}$ | 0.001 | 0.002 | |

**Table 9.** Statistical significance analysis of benchmark methods versus the proposed method on the ZHANGZHOU dataset.

| Metric | Statistic | RANSAC | LPM | LGSC | TRGDC | IMSCAN | TRESAC | SPGSC | SPGSC++ |
|---|---|---|---|---|---|---|---|---|---|
| Precision | Mean | 0.9020 | 0.9087 | 0.9349 | 0.9433 | 0.9844 | 0.9931 | 0.9469 | 0.9507 |
| | Std | 0.1191 | 0.0761 | 0.0604 | 0.0903 | 0.0260 | 0.0135 | 0.0986 | 0.0368 |
| | 95%CI | ±0.0339 | ±0.0216 | ±0.0172 | ±0.0257 | ±0.0074 | ±0.0038 | ±0.0280 | ±0.0105 |
| | *p*-value | 0.002 | $3.17 \times 10^{-4}$ | 0.186 | 0.175 | $1.73 \times 10^{-6}$ | $2.54 \times 10^{-9}$ | 0.026 | |
| Recall | Mean | 0.9713 | 0.9961 | 0.9971 | 0.9823 | 0.9350 | 0.9394 | 0.9809 | 0.9937 |
| | Std | 0.0892 | 0.0064 | 0.0047 | 0.0479 | 0.1340 | 0.1148 | 0.0781 | 0.0101 |
| | 95%CI | ±0.0254 | ±0.0018 | ±0.0013 | ±0.0136 | ±0.0381 | ±0.0326 | ±0.0222 | ±0.0029 |
| | *p*-value | 0.864 | 0.189 | 0.031 | 0.318 | 0.006 | $1.40 \times 10^{-6}$ | 0.356 | |
| F-Score | Mean | 0.9345 | 0.9487 | 0.9640 | 0.9615 | 0.9527 | 0.9611 | 0.9630 | 0.9713 |
| | Std | 0.1052 | 0.0445 | 0.0339 | 0.0698 | 0.0847 | 0.0811 | 0.0882 | 0.0200 |
| | 95%CI | ±0.0299 | ±0.0126 | ±0.0096 | ±0.0198 | ±0.0241 | ±0.0231 | ±0.0251 | ±0.0057 |
| | *p*-value | 0.008 | $6.33 \times 10^{-4}$ | 0.279 | 0.285 | 0.702 | 0.325 | 0.067 | |
| Runtime | Mean | 1.2969 | 0.0037 | 0.0257 | 0.0858 | 0.1500 | 0.0714 | 0.0689 | 0.0342 |
| | Std | 0.0367 | 0.0030 | 0.0047 | 0.0266 | 0.0490 | 0.0448 | 0.1159 | 0.0141 |
| | 95%CI | ±0.0104 | ±0.0009 | ±0.0013 | ±0.0076 | ±0.0139 | ±0.0127 | ±0.0329 | ±0.0040 |
| | *p*-value | $7.55 \times 10^{-10}$ | $7.55 \times 10^{-10}$ | $9.76 \times 10^{-9}$ | $8.03 \times 10^{-10}$ | $7.56 \times 10^{-10}$ | $2.51 \times 10^{-9}$ | 0.076 | |

**Table 10.** Statistical significance analysis of benchmark methods versus the proposed method on the PAN dataset.

| Metric | Statistic | RANSAC | LPM | LGSC | TRGDC | IMSCAN | TRESAC | SPGSC | SPGSC++ |
|---|---|---|---|---|---|---|---|---|---|
| Precision | Mean | 0.9909 | 0.9439 | 0.9850 | 0.9968 | 0.9748 | 0.9942 | 0.9287 | 0.9728 |
| | Std | 0.0150 | 0.1483 | 0.0298 | 0.0037 | 0.0362 | 0.0056 | 0.2162 | 0.0548 |
| | 95%CI | ±0.0053 | ±0.0526 | ±0.0106 | ±0.0013 | ±0.0128 | ±0.0020 | ±0.0767 | ±0.0194 |
| | *p*-value | $3.90 \times 10^{-4}$ | 0.001 | 0.003 | $1.82 \times 10^{-5}$ | 0.244 | $3.90 \times 10^{-4}$ | 0.024 | |
| Recall | Mean | 0.7656 | 0.9634 | 0.9697 | 0.9330 | 0.9755 | 0.9561 | 0.9401 | 0.9944 |
| | Std | 0.2787 | 0.1410 | 0.1691 | 0.1699 | 0.0982 | 0.1117 | 0.2318 | 0.0140 |
| | 95%CI | ±0.0988 | ±0.0500 | ±0.0600 | ±0.0602 | ±0.0348 | ±0.0396 | ±0.0822 | ±0.0050 |
| | *p*-value | $4.97 \times 10^{-6}$ | 0.054 | 0.001 | $2.72 \times 10^{-4}$ | 0.158 | 0.063 | 0.158 | |
| F-Score | Mean | 0.8315 | 0.9534 | 0.9631 | 0.9527 | 0.9715 | 0.9708 | 0.9335 | 0.9827 |
| | Std | 0.2190 | 0.1445 | 0.1640 | 0.1307 | 0.0642 | 0.0710 | 0.2266 | 0.0326 |
| | 95%CI | ±0.0777 | ±0.0512 | ±0.0582 | ±0.0463 | ±0.0228 | ±0.0252 | ±0.0804 | ±0.0115 |
| | *p*-value | $3.55 \times 10^{-5}$ | 0.210 | $7.04 \times 10^{-4}$ | 0.893 | 0.721 | 0.666 | 0.548 | |
| Runtime | Mean | 0.9872 | 2.4955 | 0.1524 | 0.4947 | 1.2616 | 0.5338 | 0.4443 | 0.0123 |
| | Std | 0.9812 | 4.3045 | 0.0967 | 0.3634 | 0.9872 | 0.2109 | 0.4339 | 0.0124 |
| | 95%CI | ±0.3479 | ±1.5263 | ±0.0343 | ±0.1289 | ±0.3500 | ±0.0748 | ±0.1539 | ±0.0044 |
| | *p*-value | $5.39 \times 10^{-7}$ | $5.39 \times 10^{-7}$ | $5.91 \times 10^{-7}$ | $5.39 \times 10^{-7}$ | $5.39 \times 10^{-7}$ | $5.39 \times 10^{-7}$ | $5.39 \times 10^{-7}$ | |

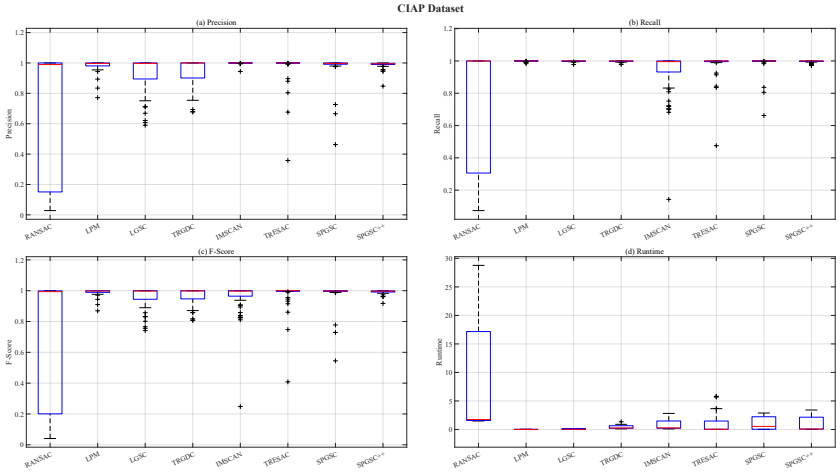



**Figure 8.** Boxplots comparing benchmark methods with the proposed method on the CIAP dataset. The blue box denotes the interquartile range (IQR, 25th–75th percentiles), the red horizontal line inside the box marks the median, and the dashed whiskers extend to the most extreme non-outlier values within 1.5 × IQR. Red '+' markers indicate outliers. Lower runtime and higher precision/recall/F-score indicate better performance. Same explanation applies to the following figures of the same type.

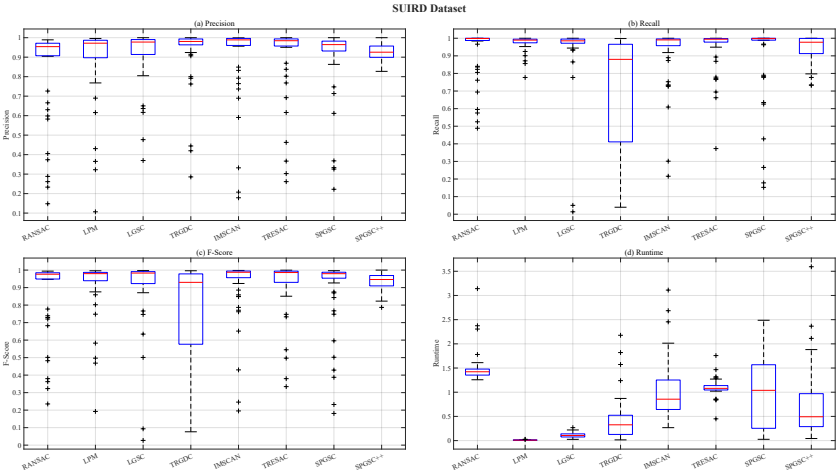



**Figure 9.** Boxplots comparing benchmark methods with the proposed method on the SUIRD dataset.

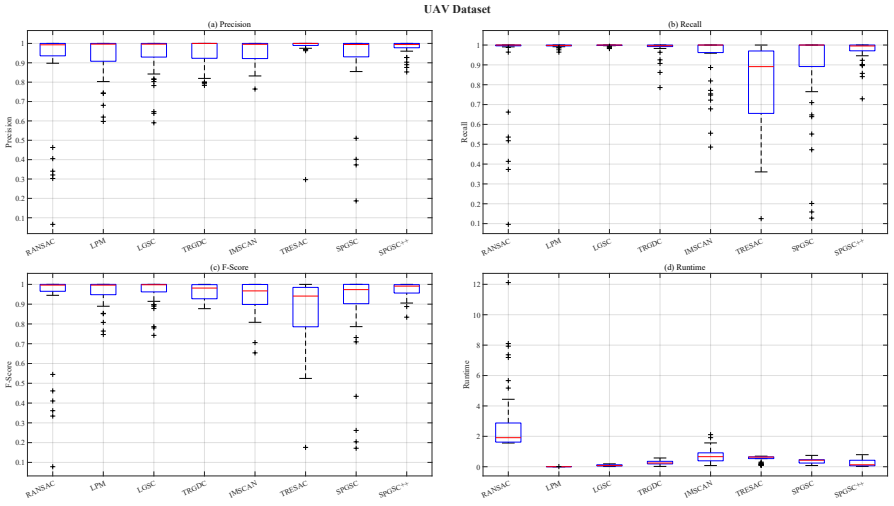



**Figure 10.** Boxplots comparing benchmark methods with the proposed method on the UAV dataset.

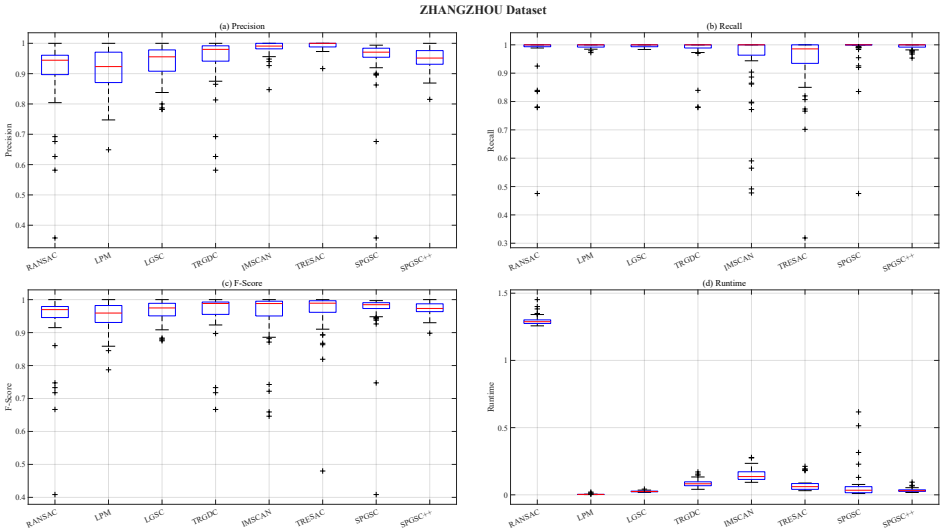



**Figure 11.** Boxplots comparing benchmark methods with the proposed method on the ZHANGZHOU dataset.

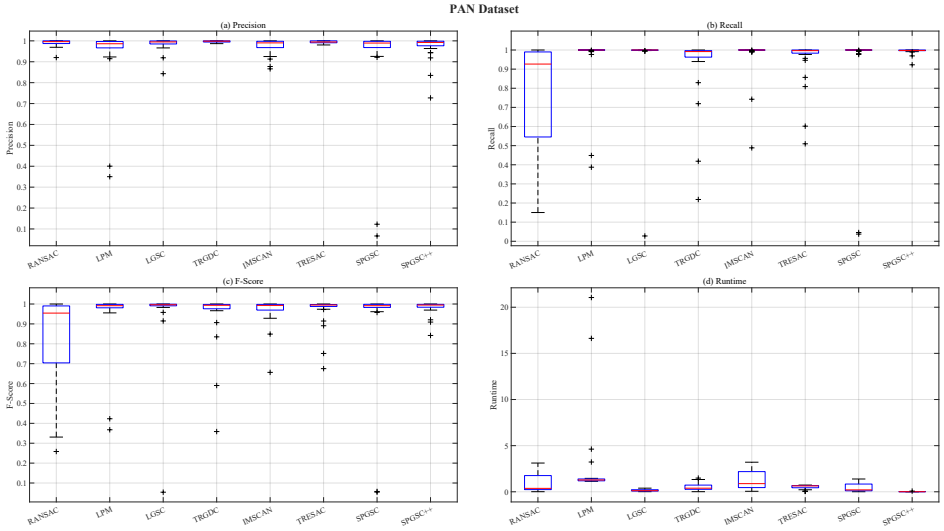



**Figure 12.** Boxplots comparing benchmark methods with the proposed method on the PAN dataset.

## 6. Discussion

SPGSC++ is based on the premise that a correct correspondence should not only be pairwise compatible with individual correspondences but should also lie within a local structure supported by multiple correspondences. The components of SPGSC++ form a sequentially coupled estimation chain that addresses different failure modes. PCA provides a compact correspondence representation and reduces the burden of relation construction, but it neither classifies correspondences nor supplies robustness by itself. The empirical first-order affinity retains candidate edges with plausible pairwise behavior, whereas common-neighbor support examines whether these edges receive repeated support from the surrounding group. The resulting second-order evidence is aggregated into node quality and used by IDSS to reduce the probability of initializing the model from a contaminated minimal subset. Because an initially plausible topology may still contain systematic false relations, DSR introduces model-residual information after hypothesis generation. Structural information therefore improves initialization when residuals are not yet reliable, while residual-informed refinement reduces dependence on imperfections in the initial topology.

In addition, correspondence-classification accuracy and geometric registration accuracy are related but not equivalent. Even if a high proportion of true matches is retained, the estimated transformation may still be poorly conditioned when these matches are concentrated within a limited region or provide insufficient spatial coverage. SPGSC++ achieved the lowest overall mean MRE across the five datasets but was not the best method on SUIRD. This is because a sufficiently large, nondegenerate, and spatially distributed inlier set must remain in the tentative correspondences, enough true relations must survive candidate-edge construction to form common-neighbor support, and the image pair must also be reasonably approximated by one dominant global homography. The performance may degrade under insufficient overlap or texture, strongly clustered inliers, severe spatially varying parallax caused by terrain relief, multiple planes or motions, or nonrigid deformation. Highly repetitive structures may also form a dense false graph, while structured high-variance outliers may distort the PCA representation. The common-neighbor stage can suppress isolated accidental candidate edges, but it cannot recover a true edge already removed during the first-order candidate-graph construction stage.

Evaluation with one frozen parameter configuration across five datasets provides empirical evidence of parameter transfer across the evaluated sources, but it does not establish universal cross-domain generalization. The absence of offline training does not guarantee unconditional cross-domain robustness. The current dense implementation stores adjacency and common-neighbor matrices and therefore has memory complexity; common-neighbor computation may also dominate the Runtime for large correspondence sets. Accurate registration may also serve as geometric preprocessing and quality control for constructing bitemporal UAV remote sensing change-captioning datasets by reducing pseudo-changes caused by misregistration. This potential application requires dedicated validation in terms of downstream annotation noise and change-captioning performance [31,32].

## 7. Conclusions

This study presented SPGSC++, a second-order spatial guided sampling method for UAV remote sensing image registration. SPGSC++ links empirical candidate first-order relations, common-neighbor-supported second-order structural quality, structure-guided initial data subset selection, and residual-informed data subset refinement within a unified robust-estimation pipeline. PCA compresses the four-dimensional correspondence coordinates into a compact two-dimensional representation to reduce the burden of relation construction. Across five datasets, SPGSC++ achieved a mean F-Score of 0.9702, a mean Runtime of 0.3724 s, and a mean MRE of 1.6454 pixels. Relative to the original SPGSC, the mean F-Score increased by 0.0391, while the Runtime and MRE decreased by 32.39% and 58.12%, respectively. Ablation experiments showed substantial degradation when the second-order structure was replaced with first-order relations or when IDSS or DSR was removed, indicating that the improvement arises from complementary structural and residual evidence.

These results demonstrate empirical effectiveness on the evaluated datasets and under the evaluated protocol, but they do not imply that all complex UAV registration problems have been solved. Classical PCA may be influenced by structured high-variance outliers, the candidate-edge affinity is not a strict invariant to arbitrary perspective distortion or terrain-induced parallax, and common-neighbor support may assign high quality to a dense false structure. Severe parallax and nonrigid deformation may also violate the assumption of one global homography. Future work will investigate robust or incremental dimensionality reduction, adaptive candidate-edge and neighborhood parameters, fusion of appearance, orientation, and scale information, sparse common-neighbor computation, and local, piecewise, or multi-model transformations. We will also evaluate combinations of learned

frontends with the SPGSC++ backend under larger-scale parameter evaluation protocols and examine its role in constructing UAV remote sensing change-captioning datasets.

**Author Contributions:** Y.D.: Conceptualization, methodology, software, formal analysis, investigation, writing—original draft preparation, visualization; X.C.: validation, resources, data curation; Z.C.: validation, investigation; J.Z.: supervision, project administration, funding acquisition. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China (grant number 12302061) and the Fujian Provincial Department of Science and Technology (grant number 2026J0011535).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data are contained within the article.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Zhou, Q.; Su, Q.; Wei, X.; Mao, W.; Ren, Y.; Chen, Y.; Bi, J.; Zhao, M.; Tsukada, M. CLIFF: A Multi-Modal Remote Sensing Model for Geological Hazard Monitoring Based on Bitemporal UAV Images. *Remote Sens.* **2026**, *18*, 2432. [CrossRef]
2. Liu, Y.; Liu, Y.; Wang, J. RFM-GAN: Robust feature matching with GAN-based neighborhood representation for agricultural remote sensing image registration. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 6006105. [CrossRef]
3. Zhou, J.; Chen, Z.; Gao, D.; Yao, Y.; Deng, Z.; Zhu, T. Fast and accurate target localization in unmanned aerial vehicle reconnaissance based on improved feature matching. *J. Electron. Imaging* **2025**, *34*, 053035. [CrossRef]
4. Guo, H.; Chen, W.; Deng, Z.; Chen, S.; Hong, X.; Chen, S.; Wang, D.H. Sampling consensus by neighborhood interaction information for remote sensing image matching. *IEEE Geosci. Remote Sens. Lett.* **2024**, *22*, 6000205. [CrossRef]
5. Fischler, M.A.; Bolles, R.C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **1981**, *24*, 381–395.
6. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
7. Bay, H.; Ess, A.; Tuytelaars, T.; Van Gool, L. Speeded-up robust features (SURF). *Comput. Vis. Image Underst.* **2008**, *110*, 346–359. [CrossRef]
8. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An efficient alternative to SIFT or SURF. In *Proceedings of the 2011 International Conference on Computer Vision*; IEEE: New York, NY, USA, 2011; pp. 2564–2571.
9. DeTone, D.; Malisiewicz, T.; Rabinovich, A. Superpoint: Self-supervised interest point detection and description. In *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*; IEEE: New York, NY, USA, 2018; pp. 337–33712.
10. Guo, H.; Lu, Y.; Chen, H.; Luo, H.; Xiao, G.; Zhang, H.; Wang, H. Motion consistency guided robust geometric model fitting with severe outliers. *IEEE Trans. Ind. Electron.* **2021**, *69*, 4065–4075. [CrossRef]
11. Chum, O.; Matas, J. Matching with PROSAC-progressive sample consensus. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*; IEEE: New York, NY, USA, 2005; Volume 1, pp. 220–226.
12. Wong, H.S.; Chin, T.J.; Yu, J.; Suter, D. Dynamic and hierarchical multi-structure geometric model fitting. In *Proceedings of the 2011 International Conference on Computer Vision*; IEEE: New York, NY, USA, 2011; pp. 1044–1051.
13. Kanazawa, Y.; Kawakami, H. Detection of planar regions with uncalibrated stereo using distributions of feature points. In *Proceedings of the BMVC*; BMVA Press: Durham, UK, 2004; pp. 1–10.
14. Chum, O.; Matas, J.; Kittler, J. Locally optimized RANSAC. In *Proceedings of the Joint Pattern Recognition Symposium*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 236–243.
15. Lai, T.; Wang, W.; Liu, Y.; Li, Z.; Lin, S. Robust model estimation by using preference analysis and information theory principles. *Appl. Intell.* **2023**, *53*, 22363–22373. [CrossRef]
16. Cheng, C.M.; Lai, S.H. A consensus sampling technique for fast and robust model fitting. *Pattern Recognit.* **2009**, *42*, 1318–1329. [CrossRef]
17. Xia, Y.; Jiang, J.; Lu, Y.; Liu, W.; Ma, J. Robust feature matching via progressive smoothness consensus. *ISPRS J. Photogramm. Remote Sens.* **2023**, *196*, 502–513. [CrossRef]
18. Lai, T.; Wang, H.; Yan, Y.; Chin, T.J.; Zheng, J.; Li, B. Accelerated guided sampling for multistructure model fitting. *IEEE Trans. Cybern.* **2019**, *50*, 4530–4543. [CrossRef] [PubMed]

19. Guo, H.; Lu, Y.; Xiao, G.; Lin, S.; Wang, H. Triplet relationship guided sampling consensus for robust model estimation. *IEEE Signal Process. Lett.* **2022**, *29*, 817–821. [CrossRef]
20. Chen, Z.; Zhou, J.; Gao, D.; Chen, R.; Huang, G. SCUGSC: Robust Model Fitting via Spatial Consensus Unit Guided Sampling Consensus for Feature Matching. *Digit. Signal Process.* **2026**, *184*, 106446. [CrossRef]
21. Ma, J.; Zhao, J.; Jiang, J.; Zhou, H.; Guo, X. Locality preserving matching. *Int. J. Comput. Vis.* **2019**, *127*, 512–531. [CrossRef]
22. Jiang, X.; Xia, Y.; Zhang, X.P.; Ma, J. Robust image matching via local graph structure consensus. *Pattern Recognit.* **2022**, *126*, 108588. [CrossRef]
23. Zhou, J.; Yao, Y.; Chen, X.; Guo, H.; Li, Q.; Deng, Z. Triplet relationship guided density clustering for feature matching with a large number of outliers. *Clust. Comput.* **2025**, *28*, 314. [CrossRef]
24. Xiao, G. Spatial clustering guided two-view multi-structural deterministic geometric model fitting. *IEEE Trans. Image Process.* **2025**, *34*, 6016–6028. [CrossRef] [PubMed]
25. Guo, H.; Xiao, G.; Su, L.; Li, T.; Wang, D.H.; Wang, H. Second-order proximity guided sampling consensus for robust model fitting. *IEEE Trans. Circuits Syst. Video Technol.* **2024**, *34*, 10704–10717. [CrossRef]
26. Ma, J.; Jiang, J.; Zhou, H.; Zhao, J.; Guo, X. Guided locality preserving feature matching for remote sensing image registration. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 4435–4447. [CrossRef]
27. Guo, H.; Xiao, G.; Su, L.; Zhou, J.; Wang, D.H. Local neighbor propagation on graphs for mismatch removal. *Inf. Sci.* **2024**, *653*, 119749. [CrossRef]
28. Liu, J.; Liang, A.; Zhao, E.; Pang, M.; Zhang, D. Homography matrix-based local motion consistent matching for remote sensing images. *Remote Sens.* **2023**, *15*, 3379. [CrossRef]
29. Chen, Z.; Zhou, J.; Gao, D.; Lin, J.; Huang, G.; Deng, Z. A UAV scene matching method based on triplet relationship and sliding mode control for GPS-denied environments. *Measurement* **2026**, *267*, 120581. [CrossRef]
30. Jiang, X.; Jiang, J.; Fan, A.; Wang, Z.; Ma, J. Multiscale locality and rank preservation for robust feature matching of remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 6462–6472. [CrossRef]
31. Hoxha, G.; Chouaf, S.; Melgani, F.; Smara, Y. Change captioning: A new paradigm for multitemporal remote sensing image analysis. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5627414. [CrossRef]
32. Liu, C.; Zhao, R.; Chen, H.; Zou, Z.; Shi, Z. Remote sensing image change captioning with dual-branch transformers: A new method and a large scale dataset. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5633520. [CrossRef]


**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.