Seam-Adaptive Structure-Preserving Image Stitching for Drone Images

Jiaxue Li and Yicong Zhou, *Senior Member, IEEE*

***Abstract*— Drones have been widely used for remote sensing applications. To perform high-quality drone image stitching, this article first proposes a local and global structure-preserving alignment (LGSPA) method that aligns drone images from local dual feature-based and global pixel-based alignment perspectives, while maintaining local linear and global collinear image structures. To enable an optimal image stitching performance, we then propose a seam-adaptive weighting (SAW) scheme to enhance the local alignment accuracy under the guidance of a seam prior. On the ground of LGSPA and SAW, we further develop a seam-adaptive structure-preserving (SASP) image stitching framework to generate the final stitched drone images. Both qualitative and quantitative experimental results demonstrate that LGSPA and SASP are capable of generating higher quality alignment and stitching results than several state-of-the-art methods over multiple challenging aerial scenarios, including low textures, repetitive textures, large parallax, wide baseline, and occlusions.**

***Index Terms*— Drone images, image alignment, image stitching, mesh deformation, multiple challenging scenarios.**

I. Introduction

TAKING advantages of lightweight, flexible mobility, and easy deployment, drones have advanced the development of various remote sensing applications, such as ocean monitoring [1], disaster investigation [2], and traffic management [3]. Nevertheless, captured drone images usually suffer from the limited imaging width. For the sake of covering more regions of interest, image stitching serves as a fundamental task aiming to combine multiple drone images into a single larger image. However, stitching drone images usually encounters kinds of realistic challenges, such as low textures, repetitive textures, large parallax, wide baseline, and occlusions.

For drone image stitching, the existing methods have been proposed based on feature detection, image alignment, and seamline detection [4], [5], [6], [7], [8]. The stitching results are typically assessed from two aspects [9]: 1) alignment accuracy across the entire overlapping area or along the stitching seamline and 2) naturalness of warped images. However, it remains difficult to obtain high-quality stitching results that simultaneously satisfy the aforementioned two aspects especially when confronted with multiple challenging aerial scenarios.

Traditional methods for image alignment are generally built upon a single homography model [10]. However, their alignment accuracy requires strict conditions: pure rotational camera motion or planar scenes. For drone image stitching, both the conditions are prone to be violated, and hence misalignment artifacts and distortions appear easily. To tolerate large parallax and wide baseline, seam-driven methods [11], [12], [13], [14] have been proposed for local alignment. Based on the observation that achieving the best-fitting global alignment does not necessarily ensure optimal image stitching performance, these methods propose to precisely align only a local region that allows an optimal seamline to stitch images seamlessly. Compared with the global alignment methods, they are more flexible to achieve promising results when handling large parallax and wide baseline. However, finding a sensible local alignment (or a local region free from parallax) in challenging scenarios is not an easy task. Furthermore, when the overlapping area between two adjacent drone images is narrow, the estimated local alignment probably introduces uncomfortable projective distortions.

To enhance image naturalness of stitching results, recent methods have been proposed following two mainstreams: spatially varying transformation and mesh deformation. Generally, they partition images into uniform meshes to estimate multiple continuous local warps to embody an accurate global alignment. To mitigate projective distortions, some methods [15], [16], [17] exploit a hybrid transformation strategy to spatially combine the homography model with the similarity transformation. Some other methods [18], [19], [20], [21] introduce additional feature lines to preserve the linear structures of image scenes, suppressing certain distortions. These methods have gained good performance in improving global alignment accuracy and preserving image naturalness. However, they rely on features heavily to estimate their warping models. When feature correspondences are mismatched or exhibit inconsistency between adjacent drone images owing to significant parallax, severe distortions will arise in the overlapping area. Moreover, when feature points and lines are insufficient in drone images, the estimated warping model would be biased, leading to a degradation of alignment accuracy.

Received 31 March 2024; revised 8 August 2024 and 14 October 2024; accepted 26 November 2024. Date of publication 11 December 2024; date of current version 24 December 2024. This work was supported in part by the Science and Technology Development Fund, Macau SAR, under Grant 0049/2022/A1 and Grant 0050/2024/AGJ; and in part by the University of Macau Development Foundation under Grant MYRG-GRG2024-00181-FST-UMDF. *(Corresponding author: Yicong Zhou.)*

Jiaxue Li was with the Department of Computer and Information Science, University of Macau, Macau, China. She is now with the School of Information Engineering, China University of Geosciences, Beijing 100000, China (e-mail: lijiaxue7@gmail.com).

Yicong Zhou is with the Department of Computer and Information Science, University of Macau, Macau, China (e-mail: yicongzhou@um.edu.mo).

Digital Object Identifier 10.1109/TGRS.2024.3515111

Motivated by the above issues, we propose a seam-adaptive structure-preserving (SASP) image stitching method to improve image alignment quality, seamline quality, and drone image stitching strategy, comprehensively. The contributions of our work are summarized as follows.

1) We propose a local and global structure-preserving alignment (LGSPA) model to achieve superior alignment quality with high precision and good naturalness. Leveraging feature points, lines, and color pixels to fit the alignment model, LGSPA also demonstrates robustness against low-texture and repetitive-texture drone image scenes.
2) We propose a seam-adaptive weighting (SAW) scheme to enhance the precision of local alignment along the seamline, enabling the seamline quality to be improved essentially.
3) We further develop an SASP image stitching framework. In this framework, we use the proposed SAW scheme to guide the optimization of the LGSPA model to directly learn an accurate local alignment along a seam prior, thereby achieving the final optimal drone image stitching performance. *To the best of our knowledge, this is the first work to consider the use of suture quality to guide the optimization process of drone image alignment, so as to achieve high-quality drone image stitching results in the remote sensing field. It is highly practical and has high application value.*

Benefited from these improvements, our work is able to handle kinds of challenges in real-world applications, including low textures, repetitive textures, large parallax, wide baseline, and occlusions.

The rest of this article is organized as follows. Section II reviews the related literature. Section III introduces the LGSPA model. Section IV describes the SAW scheme. Section V elaborates the SASP framework. The experimental results on multiple challenging drone image samples and one ablation study are presented in Section VI. Finally, the conclusion is drawn in Section VII.

II. Related Work

This section briefly reviews the related literature to this article from the following three aspects.

A. Local Alignment Methods for Better Stitching

Taking seamline stitching quality into account, seam-driven methods were proposed aiming to find an optimal local alignment for seamless stitching. Gao et al. [11] first estimated a set of candidate local alignments to derive different seamlines and then selected the best one on the basis of the evaluation of seamline stitching qualities. Zhang and Liu [12] randomly estimated a plausible local alignment as their prealignment and applied the mesh-based warping method to locally adjust the alignment accuracy and image naturalness. Lin et al. [13] proposed an iterative seam-guided process to optimize each local alignment hypothesis, respectively, enabling the seamline quality to be improved as well. Recently, pixel-based color image alignment was developed in the quaternion domain [14] and achieved higher alignment accuracy than feature-based alignment methods. To achieve the final optimal color image stitching performance, Li and Zhou [22] further proposed an automatic color image stitching (ACIS) framework to simultaneously learn the local alignment and seamline in an iterative fashion.

B. Global Alignment Methods for Better Naturalness

To encourage the flexibility of traditional parametric transformations, spatially varying transformations have been studied. Zaragoza et al. [23] proposed the as-projective-as-possible (APAP) warp that comprised multiple moving projective transformations. To get rid of projective distortions, Chang et al. [15] proposed the shape-preserving half-projective (SPHP) warp smoothly extrapolating projective transformations into similarity counterparts beyond the overlapping area. The quasi-homography (QH) warp was studied to mitigate the projective distortion of homography by squeezing its scale linearly [9]. Li et al. [24] formulated the local alignment as the triangular facet transformation (TFT) decomposed from the homography model, allowing more flexibility to adaptively fit an accurate stitching field.

Mesh-based warping methods were proposed to manipulate the vertices of the constructed mesh grid cells over entire images. Using matched feature points and lines, Li et al. [19] proposed a dual-feature warping (DFW) model to improve the alignment accuracy of mesh warping. Based on DFW and QH, Liao and Li [20] further proposed the single-perspective warp (SPW) to simultaneously address the issues of alignment, naturalness, and distortions. To preserve prominent structures for images with large parallax, Jia et al. [21] proposed the line-point consistent strategy for global line-guided mesh deformation. Recently, Nie et al. [25] proposed the unsupervised deep image stitching (UDIS) method to remove parallax errors by reconstructing stitching results from feature to pixel level.

C. Drone Image Stitching Methods

The aforementioned image stitching methods have been widely applied to the existing drone image stitching technologies. Wan et al. [5] proposed a mesh-based drone image warping method that minimized parallax errors between matched feature points for alignment and used the similarity transformation as a constraint for naturalness. However, they failed to accurately align drone images that have rich linear structures. Guo et al. [26] explored the SPHP warp [15] to formulate a similarity term for natural mesh deformation. Similar to the APAP warp [23], Xu et al. [27] partitioned UAV images into uniform grid cells, facilitating the adaptive estimation of multiple local homographies. Zhang et al. [28] proposed a robust elastic warp (REW) incorporating a weighted optimization algorithm [29] to combine the homography with the similarity transformation together for alignment.

Although these drone image stitching methods have gained satisfactory results, several limitations persist. First, all these methods belong to single feature-based methods, relying solely on feature points to estimate alignment models. When feature

points are mismatched or distributed unevenly, the estimated alignment models may be biased, resulting in misalignments or distortions. Moreover, obtaining a sufficient number of reliable feature points in natural environments is also a challenging task. Second, these methods use the similarity transformation to mitigate projective distortions for image naturalness, potentially overlooking perspective distortions and failing to preserve salient image structures. Third, they adhere to the traditional image stitching pipeline to optimize image alignment and seamline, sequentially and independently. However, as observed in [11], the best-fitting global alignment cannot necessarily ensure to generate the best seamline in the subsequent steps. To address these issues, we propose an SASP method in this article.

III. Local and Global Structure-Preserving Alignment

In this section, we propose the LGSPA method. To begin with, a prealignment based on quaternion rank-1 alignment (QR1A) is introduced in Section III-A. Then, we formulate the drone image alignment issue as a mesh deformation problem in Section III-B and define each term of our objective function concretely in Section III-C. Finally, the optimal solution is given in Section III-D.

A. Prealignment With QR1A

To provide the best geometric fitting for global alignment, in contrast to harnessing sparse feature correspondences, we work on dense color pixels across the whole overlapping area to estimate the homography model in pixel level.

Let $\{\dot{\mathbf{I}}_i\}_{i=1}^2$ be two drone images described in quaternion representation, $\circ$ indicate image warping operation, and τ denote the homographic transformation. Then, $\{\dot{\mathbf{I}}_i \circ \tau_i\}_{i=1}^2$ implies aligned drone images. We use $P_\Omega(\cdot)$ to extract the image overlaps and assume that well-aligned image overlaps are linearly correlated, since the image contents within the overlapping area correspond to the same aerial scenes. Consequently, when we convert each image overlap into a vector to stack into a matrix, i.e., $\dot{\mathbf{D}} \circ \tau = [\mathrm{vec}(P_\Omega(\dot{\mathbf{I}}_1 \circ \tau_1)), \mathrm{vec}(P_\Omega(\dot{\mathbf{I}}_2 \circ \tau_2))]$, this matrix is supposed to have an underlying rank-1 component $\dot{\mathbf{L}}$. In addition, some unavoidable misalignment artifacts such as occlusions can be modeled as the sparse component $\dot{\mathbf{S}}$. Based on this observation, the QR1A model is defined as

$$\begin{aligned} &\min_{\dot{\mathbf{L}},\dot{\mathbf{S}},\tau} \|\dot{\mathbf{S}}\|_1 \\ &\text{s.t. } \dot{\mathbf{D}} \circ \tau = \dot{\mathbf{L}} + \dot{\mathbf{S}}, \quad \mathrm{rank}(\dot{\mathbf{L}}) = 1. \end{aligned} \tag{1}$$

By minimizing (1), we obtain an estimated homography, i.e., H_{QR1A}, as our prealignment. For the detailed minimization process, read our previous work [14].

B. Mesh Deformation Problem Formulation

We begin with partitioning each input drone image into a uniform grid mesh. Our aim is to estimate the optimal vertex coordinates of deformed mesh cells, such that input drone images can be warped for image alignment.

Mathematically, let $\mathbf{I}$ and $\hat{\mathbf{I}}$ be the input drone image and the corresponding warped drone image, respectively. The set of mesh vertices in $\mathbf{I}$ and $\hat{\mathbf{I}}$ is denoted as $\mathbf{V} = [x_1, y_1, x_2, y_2, \ldots, x_n, y_n]^{\mathrm{T}} \in \mathbb{R}^{2n}$ and $\hat{\mathbf{V}} = [\hat{x}_1, \hat{y}_1, \hat{x}_2, \hat{y}_2, \ldots, \hat{x}_n, \hat{y}_n]^{\mathrm{T}} \in \mathbb{R}^{2n}$, where (x, y) and $(\hat{x}, \hat{y})$ are the vertex coordinates in $\mathbf{I}$ and $\hat{\mathbf{I}}$, respectively; and n is the total number of vertices. Then, given any sampling point $p \in \mathbf{I}$, it can be represented as

$$\tau(p) = \sum_{i=1}^{4} c_i V_i \tag{2}$$

where $\{V_i\}_{i=1}^4$ are four vertices enclosing the grid cell where p is located, and $\{c_i\}_{i=1}^4$ are their corresponding linear combination coefficients that can be calculated using the inverse bilinear interpolation technique [30]. Under the assumption that $\{c_i\}_{i=1}^4$ are known and invariant, we can further represent the corresponding warped point $\hat{p} \in \hat{\mathbf{I}}$ as

$$\tau(\hat{p}) = \sum_{i=1}^{4} c_i \hat{V}_i \tag{3}$$

where $\{\hat{V}_i\}_{i=1}^4$ are unknown coordinates of the corresponding warped grid cell to be estimated. Therefore, the mesh deformation problem can be defined as an optimization function within the variable set $\hat{\mathbf{V}}$. From the optimized $\hat{\mathbf{V}}$, we obtain the warped drone image $\hat{\mathbf{I}}$ using the texture mapping technique.

C. LGSPA Model

To achieve a high-precision alignment while avoiding unnatural distortions in multiple challenging aerial scenarios, we define our LGSPA model as the following energy minimization function:

$$E(\hat{\mathbf{V}}) = \lambda_1 E_{\mathrm{LA}}(\hat{\mathbf{V}}) + \lambda_2 E_{\mathrm{GA}}(\hat{\mathbf{V}}) + \lambda_3 E_{\mathrm{SP}}(\hat{\mathbf{V}}) + E_{\mathrm{DC}}(\hat{\mathbf{V}}) \tag{4}$$

where $E_{\mathrm{LA}}(\hat{\mathbf{V}})$ is the local dual feature-based alignment term, $E_{\mathrm{GA}}(\hat{\mathbf{V}})$ denotes the global pixel-based alignment term, and $E_{\mathrm{SP}}(\hat{\mathbf{V}})$ and $E_{\mathrm{DC}}(\hat{\mathbf{V}})$ represent the structures' preserving term and distortions' control term, respectively. λ_1, λ_2, and λ_3 are weighting parameters to balance the efficacy of each term.

1) Local Dual Feature-Based Alignment Term: Facing large parallax and occlusions, detected and matched feature points are often distributed nonuniformly, leading to a biased estimation of the alignment model. To alleviate this problem, additional feature lines are incorporated to compensate for more robust feature correspondences. Therefore, given M feature point correspondences $\{(p_i, p_i')\}_{i=1,2,\ldots,M}$ and N feature line matchings $\{(l_j, l_j')\}_{j=1,2,\ldots,N}$, we define a local dual feature-based alignment term as follows:

$$E_{\mathrm{LA}}(\hat{\mathbf{V}}) = \sum_{i=1}^{M} \|\tau(\hat{p}_i) - p_i'\|^2 + \sum_{j=1}^{N} \|\mathrm{dis}(\tau(\hat{l}_j^{s,e}), l_j')\|^2 \tag{5}$$

where $\hat{l}_j^{s,e}$ denotes the start point and the endpoint of the jth warped line $\hat{l}$, and $\mathrm{dis}(\cdot)$ is an operator to calculate the distance from the point to the line. Specifically, we represent the reference line as $l' = ax + by + c$, then $\mathrm{dis}(\cdot) = (|a\hat{x} + b\hat{y} + c|/(a^2 + b^2)^{1/2})$, where $(\hat{x}, \hat{y})$ are the coordinates of the warped point in the reference coordinate system.

2) Global Pixel-Based Alignment Term: Aligning drone images solely with the aforementioned feature-based alignment term is insufficient. This is because its constraint is only applied to a subset of vertices within grid cells containing feature correspondences. In the regions where no features are detected or matched, the coordinates of mesh vertices remain unconstrained, resulting in notable distortions.

To address this issue, the existing methods generally perform a single feature-based global alignment before mesh deformation. However, in low-texture natural environments, it is difficult to obtain sufficient reliable feature points or lines to guarantee an accurate estimation for feature-based alignment. Moreover, the matched feature points and lines are less accurate than pixels in general. This also limits the improvements of alignment accuracy in the following mesh deformation. To solve these problems, we define a global pixel-based alignment term E_{GA} as

$$E_{\mathrm{GA}}\left(\hat{\mathbf{V}}\right) = \sum_{i=1}^{n} \left\| \tau\left(\hat{v}_i\right) - H_{\mathrm{QR1A}}(v_i) \right\|^2. \tag{6}$$

The QR1A model fully uses dense color pixels to align drone images, achieving higher alignment precision. Thus, we impose the corresponding QR1A constraint on mesh vertices, aiming to alleviate local instabilities caused by the absence of feature correspondences. Moreover, taking advantages of the pixel-based nature, it exhibits robustness against natural drone images, particularly those characterized by low-texture and repetitive-texture scenes.

3) Structures Preserving Term: Compared with feature points, the detected feature lines are more prominent in aerial scenarios, e.g., urban and roads. They could provide extra geometric structure information to attract more human attention. Thus, preserving feature lines between the warped and unwarped images is beneficial to improving the naturalness quality of mesh deformation. Apart from line segments, global geometric structures have been studied recently by merging multiple collinear line segments together [21]. Consequently, to preserve both local and global linear structures in drone images, the structures' preserving term is defined as

$$\begin{aligned} E_{\mathrm{SP}}\left(\hat{\mathbf{V}}\right) = &\sum_{i=1}^{P}\sum_{s=1}^{U_i-1} \left\| \left\langle \tau\left(\hat{p}^l_{i,s+1}\right) - \tau\left(\hat{p}^l_{i,s}\right), \overrightarrow{\hat{\mathbf{n}}^l_i} \right\rangle \right\|^2 \\ &+ \sum_{j=1}^{Q}\sum_{t=1}^{Z_j-1} \left\| \left\langle \tau\left(\hat{p}^g_{j,t+1}\right) - \tau\left(\hat{p}^g_{j,t}\right), \overrightarrow{\hat{\mathbf{n}}^g_j} \right\rangle \right\|^2 \end{aligned} \tag{7}$$

where P and Q are the numbers of local and global line segments, respectively. We uniformly sample the ith local line segment with U_i points $\{p^l_{i,s}\}_{s=1,2,\ldots,U_i}$, and the jth global line segment with Z_j points $\{p^g_{j,t}\}_{t=1,2,\ldots,Z_j}$, respectively. $\overrightarrow{\hat{\mathbf{n}}}$ represents the normal vector of the corresponding warped line, and $\langle\cdot\rangle$ is an operator to calculate the dot product of two vectors.

4) Distortions Control Term: Similar to the local dual feature-based alignment term, the above structures' preserving term can only impose linear constraints on the mesh vertices within the image area having feature lines. Moreover, the way preserving lines can prevent images from undergoing perspective distortions, but cannot avoid the projective distortions happening. For example, objects are enlarged nonuniformly in the nonoverlapping area. Thus, it remains necessary to control the potential distortions over the entire drone image area, especially the overlapping regions without linear structures and the nonoverlapping area.

According to the analyses of homographic transformation in [20], there exists a unique set of inherent lines that keep parallel, horizontally and vertically, before and after the projective transformation. Mathematically, given a homography prior $H = [h_1\ h_2\ h_3; h_4\ h_5\ h_6; h_7\ h_8\ 1]$, the slope of vertical parallel lines is $k_1 = -(h_7/h_8)$ before warping and $k_2 = ((h_4h_8 - h_5h_7)/(h_1h_8 - h_2h_7))$ after warping, respectively. Afterward, the slope of corresponding horizontal parallel lines can be derived by $-(1/k_1)$ and $-(1/k_2)$. Consequently, by preserving the slope of these inherent parallel lines, we are able to mitigate the perspective distortion of mesh deformation globally.

On the other hand, it was also observed that the scale of these parallel lines is transformed nonuniformly [9]. When facing the large parallax, projective distortion may appear in the nonoverlapping area. To alleviate such projective distortion, we further constrain these horizontal and vertical parallel lines to be scaled in a uniform fashion. In this regard, the distortions' control term consists of a perspective-control subterm E_{pes} and a projective-control subterm E_{pro}

$$E_{\mathrm{dc}}\left(\hat{\mathbf{V}}\right) = \mu_1 E_{\mathrm{pes}}\left(\hat{\mathbf{V}}\right) + \mu_2 E_{\mathrm{pro}}\left(\hat{\mathbf{V}}\right) \tag{8}$$

where μ_1 and μ_2 are their weighting constants. Specifically, we start by establishing a series of crossing lines, including vertical parallel lines $\{l^v_i\}_{i=1,2,\ldots,K}$ and horizontal parallel lines $\{l^u_j\}_{j=1,2,\ldots,R}$ for the target image. Next, similar to (7), these lines are uniformly sampled with F_i points $\{p^v_{i,s}\}_{s=1,2,\ldots,F_i}$ and G_j points $\{p^u_{j,t}\}_{t=1,2,\ldots,G_j}$, respectively. To this end, we are able to effectively control the global perspective by preserving these crossing-line features, and thus the subterm E_{pes} is defined as

$$\begin{aligned} E_{\mathrm{pes}}\left(\hat{\mathbf{V}}\right) = &\sum_{i=1}^{D}\sum_{s=1}^{F_i-1} \left\| \left\langle \tau\left(\hat{p}^v_{i,s+1}\right) - \tau\left(\hat{p}^v_{i,s}\right), \overrightarrow{\hat{\mathbf{n}}^v_i} \right\rangle \right\|^2 \\ &+ \sum_{j=1}^{W}\sum_{t=1}^{G_j-1} \left\| \left\langle \tau\left(\hat{p}^u_{j,t+1}\right) - \tau\left(\hat{p}^u_{j,t}\right), \overrightarrow{\hat{\mathbf{n}}^u_j} \right\rangle \right\|^2 \end{aligned} \tag{9}$$

where D and W are the numbers of constructed vertical and horizontal parallel lines, respectively, and $\overrightarrow{\hat{\mathbf{n}}^v_i}$ and $\overrightarrow{\hat{\mathbf{n}}^u_j}$ are the normal vectors of these lines after projective warping. Meanwhile, we also impose the equidistant-scaling constraint on those sampled points to balance projective distortion. Accordingly, the subterm E_{pro} is defined as follows:

$$\begin{aligned} E_{\mathrm{pro}}\left(\hat{\mathbf{V}}\right) = &\sum_{i=1}^{D}\sum_{s=1}^{F_i-2} \left\| \tau\left(\hat{p}^v_{i,s}\right) + \tau\left(\hat{p}^v_{i,s+2}\right) - 2\tau\left(\hat{p}^v_{i,s+1}\right) \right\|^2 \\ &+ \sum_{j=1}^{W}\sum_{t=1}^{G_j-2} \left\| \tau\left(\hat{p}^u_{j,t}\right) + \tau\left(\hat{p}^u_{j,t+2}\right) - 2\tau\left(\hat{p}^v_{j,t+1}\right) \right\|^2. \end{aligned} \tag{10}$$

Notably, we here use H_{QR1A} as our homography prior.

D. Optimization

All the above-defined terms and subterms are integrated together to constitute (4). Consequently, our objective function can be reformulated in the following matrix multiplication form:

$$\begin{aligned} E(\hat{\mathbf{V}}) = {} & \lambda_1 \left\| \mathbf{C_p}\hat{\mathbf{V}} - \mathbf{P} \right\|^2 + \lambda_1 \left\| \mathbf{C_l}\hat{\mathbf{V}} + \mathbf{E} \right\|^2 \\ & + \lambda_2 \left\| \hat{\mathbf{V}} - \mathbf{V_g} \right\|^2 \\ & + \lambda_3 \left\| \mathbf{C_{sl}}\hat{\mathbf{V}} \right\|^2 + \lambda_3 \left\| \mathbf{C_{sg}}\hat{\mathbf{V}} \right\|^2 + \mu_1 \left\| \mathbf{C_{pes}}\hat{\mathbf{V}} \right\|^2 \\ & + \mu_2 \left\| \mathbf{C_{pro}}\hat{\mathbf{V}} \right\|^2 \end{aligned} \tag{11}$$

where $\mathbf{C_p} \in \mathbb{R}^{2M\times 2n}$, $\mathbf{P} \in \mathbb{R}^{2M}$, $\mathbf{C_l} \in \mathbb{R}^{2N\times 2n}$, $\mathbf{E} \in \mathbb{R}^{2N}$, $\mathbf{V_g} \in \mathbb{R}^{2n}$, $\mathbf{C_{sl}} \in \mathbb{R}^{\sum_{i=1}^{P}(U_i-1)\times 2n}$, $\mathbf{C_{sg}} \in \mathbb{R}^{\sum_{j=1}^{Q}(Z_j-1)\times 2n}$, $\mathbf{C_{pes}} \in \mathbb{R}^{[\sum_{i=1}^{D}(F_i-1)+\sum_{j=1}^{W}(G_j-1)]\times 2n}$, and $\mathbf{C_{pro}} \in \mathbb{R}^{[2\sum_{i=1}^{D}(F_i-2)+2\sum_{j=1}^{W}(G_j-2)]\times 2n}$ are derived from (5) to (10), respectively. Obviously, (11) is a quadratic function, and thus it can be further expressed as the following form:

$$E(\hat{\mathbf{V}}) = \left\| \mathbf{A}\hat{\mathbf{V}} - \mathbf{B} \right\|^2 \tag{12}$$

where

$$\mathbf{A} = \begin{bmatrix} \sqrt{\lambda_1}\mathbf{C_p} \\ \sqrt{\lambda_1}\mathbf{C_l} \\ \sqrt{\lambda_2} \\ \sqrt{\lambda_3}\mathbf{C_{sl}} \\ \sqrt{\lambda_3}\mathbf{C_{sg}} \\ \sqrt{\mu_1}\mathbf{C_{pes}} \\ \sqrt{\mu_2}\mathbf{C_{pro}} \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} \sqrt{\lambda_1}\mathbf{P} \\ -\sqrt{\lambda_1}\mathbf{E} \\ \sqrt{\lambda_2}\mathbf{V_g} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Finally, (12) boils down to a sparse linear matrix equation, and it can be effectively optimized by the sparse linear solver.

E. Complexity Analysis

The LGSPA model comprises four terms. The computational complexity of constructing 1) term E_{LA} is $\mathcal{O}(M+N)$; 2) term E_{GA} is $\mathcal{O}(n)$; 3) term E_{SP} is $\mathcal{O}(\sum_{i=1}^{P}(U_i-1)+\sum_{j=1}^{Q}(Z_j-1))$; and 4) term E_{DC} is $\mathcal{O}(\sum_{i=1}^{D}(F_i-1)+\sum_{j=1}^{W}(G_j-1)+\sum_{i=1}^{D}(F_i-2)+\sum_{j=1}^{W}(G_j-2))$, respectively. Afterward, we use the conjugate gradient method to optimize the LGSPA model, and the cost is $\mathcal{O}(n^2)$.

IV. SAW Scheme

In the process of remote sensing image stitching, seamline detection constitutes a crucial step for achieving satisfactory stitching results. Taking into account that the seamline intersects only a portion of overlapping area, the seamline quality is predominantly determined by the alignment accuracy of its adjacent local regions, i.e., local alignment quality, rather than the global alignment quality. For example, if a seamline cuts through misaligned local regions, stitching artifacts will appear. Therefore, enhancing the stitching capability of seamline necessitates optimizing its local alignment quality as a priority. To this end, we propose the SAW scheme.

We first derive a seamline from the prealignment as the seam prior. According to the specific location of this seam prior, we adaptively calculate location-dependent weights for sparse features in the overlapping area. The rationale behind this weighting scheme is to accentuate the importance of local alignment constraints near the seamline, while allowing for relaxation of constraints farther away from the seamline to prevent unnecessary distortions. To ensure smooth transitions in the values of these weights, we define our SAW scheme as follows:

$$w = \left(1 + \frac{\alpha d^2}{\delta}\right)^{-\frac{\delta+1}{2}} \tag{13}$$

where d denotes the distance from the feature to the seam prior, and α and δ are constant parameters.

In experiments, the seam prior can be visualized as a polyline segment. Thus, the distance from the feature point to the seam prior equals to the shortest geometric distance from the point to the polyline segment. More specifically, we first determine the point p_1 in the polyline segment that has the minimum distance away from the feature point p_2. Then, $d = ((x_1-x_2)^2+(y_1-y_2)^2)^{1/2}$, where (x_1, y_1) and (x_2, y_2) are the coordinates of p_1 and p_2 in the image coordinate system. Similarly, when the feature is a line segment, we compute the shortest geometric distance from the endpoint of the line segment to the seam prior. In addition, α mainly adjusts the values of weight at different distances (i.e., the active range of local alignment), δ controls the decay speed on the whole, and $w \in [0, 1]$. In this article, the seam prior is derived from H_{QR1A}, and we set $\delta = 1$ by default. The value of α will be discussed in the following ablation studies.

V. SASP Image Stitching Framework

Building upon the proposed LGSPA model and SAW scheme, we develop the SASP framework for high-quality drone image stitching.

A. Our Motivation

Due to the mobility of drone devices, adjacent drone images are often captured under wide-baseline conditions. Consequently, large parallax and occlusions frequently occur in aerial scenes with nonplanar surfaces. Current drone image stitching methods typically follow the conventional pipeline of optimizing image alignment and seamline sequentially and independently. However, achieving a perfect global alignment over the entire overlapping area is rarely feasible under these challenging conditions.

In addition, it has been demonstrated that the best-fitting global alignment does not necessarily ensure optimal seamline quality [11]. For instance, an alternative alignment might yield a higher quality seamline despite less accurate geometric fitting. The optimal seamline intersects only a partial region within the overlapping area, so the final stitching performance primarily depends on local alignment accuracy along the seamline rather than global alignment accuracy. Nonetheless, finding a sensible local alignment or a local region free from parallax is also difficult in challenging scenarios.

These insights motivate us to investigate the relationship between image alignment and seamline detection to achieve

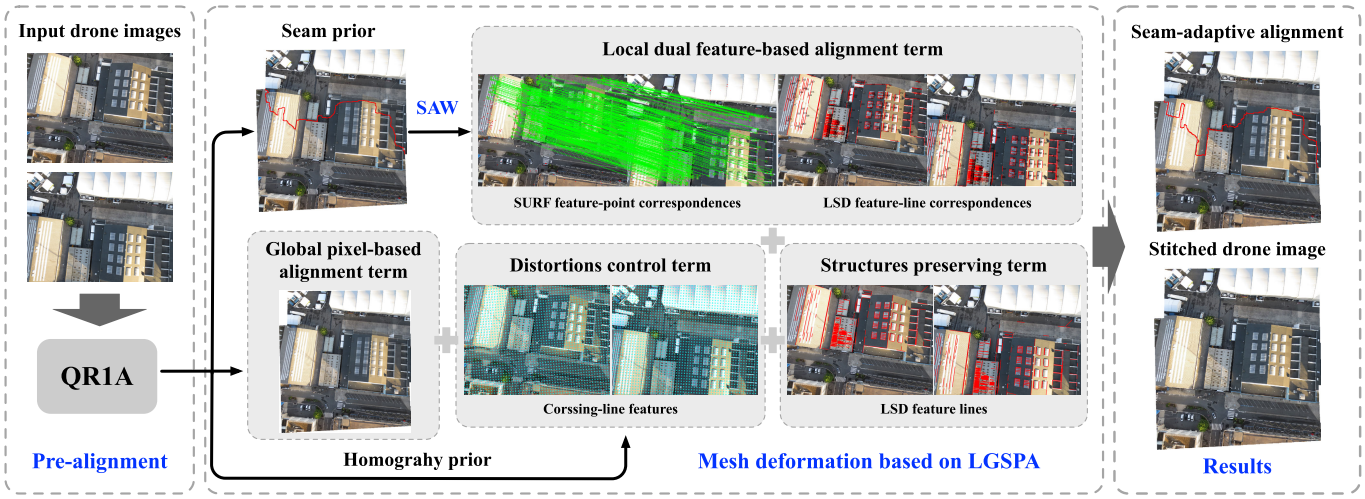

Fig. 1. SASP image stitching framework. Input drone images are initially used to generate a prealignment based on the QR1A model that will serve as the role of: 1) the global pixel-based alignment constraint; 2) a homography prior; and 3) deriving a seam prior. Afterward, we construct each term of LGSPA, respectively; and LGSPA is optimized under the guidance of the SAW scheme to obtain a seam-adaptive alignment to generate the final stitching result.

high-quality drone image stitching in the field of remote sensing.

B. SASP Framework

The developed SASP framework is illustrated in Fig. 1. Given two adjacent drone images with wide baseline, we start by optimizing the QR1A model, which fully uses dense color pixels to achieve a precise estimation of global alignment, i.e., H_{QR1A}. In this framework, H_{QR1A} serves the following three purposes: 1) generating a corresponding seam prior for SAW; 2) providing a robust global alignment constraint on mesh vertices for LGSPA; and 3) acting as a homography prior to explore inherent crossing-line features.

Afterward, we proceed to the mesh deformation stage. Technically, we first use the perception-based seam-cutting (PBSC) approach [31] to generate a seam prior to calculate SAW. Then, we construct four respective terms of the LGSPA model. For the local dual feature-based alignment term, both SURF feature points [32] and LSD line segments [33] are detected. The pairwise dual-feature correspondences are established using the RANSAC algorithm [34] and the coplanar line-points invariant matching strategy [35]. Incorporating the SAW scheme, the local dual feature-based alignment term E_{LA} is then ameliorated as

$$E_{\mathrm{LA}}(\hat{\mathbf{V}}) = \sum_{i=1}^{M} w_i \left\| \tau(\hat{p}_i) - p_i' \right\|^2 + \sum_{j=1}^{N} w_j \left\| \mathrm{dis}\left(\tau\left(\hat{l}_j^{s,e}\right), l_j'\right) \right\|^2 \tag{14}$$

where we use i and j as the number indexes of feature points and line segments. Accordingly, w_i and w_j denote the coefficients of the ith feature point and the jth feature line segment in their corresponding calculations. We derive the coefficients w_i and w_j from our SAW scheme, i.e., (13). It is notable that when $\alpha = 0$ in (13), the active range of local alignment will be spread over the whole overlapping area, and thus (14) turns out to be (5) equally. For the structures' preserving term, we adopt detected LSD line segments as our local linear structures and apply the line-merging algorithm [21] to obtain global collinear structures. We maintain a constant slope for these line segments to preserve both local and global linear structures in drone images.

Following the construction of four terms, we can derive the optimal vertex coordinates of warped mesh grids by optimizing (12). Aligned drone images are thereby obtained using the texture mapping technique [30]. Finally, we use PBSC and the Poisson blending method [36] to compose aligned drone images together.

C. Discussion

In this framework, we align drone images to create continuous well-aligned local regions (i.e., seam-adaptive local alignment) for optimal stitching. We estimate the initial seamline information from prealignment to serve as a seam prior, guiding the optimization of LGSPA for the desired local alignment. Note that this seam prior is not used for stitching images but as a cue to identify a sensible local alignment in the overlapping area. After obtaining the desired local alignment, we conduct seamline detection to generate the final optimal seamline to stitch images seamlessly. Although seamline detection is performed twice in our framework, the total time consumption increases only 1–2 s. Thus, its impact on overall efficiency and practicality is within an acceptable range.

Unlike the existing drone image stitching methods that align images over the entire overlapping area, SASP focuses on optimizing an accurate local alignment. This approach provides us greater flexibility in handling challenging drone images with large parallax, wide baselines, and occlusions. Compared with the existing image stitching methods that find an accurate local alignment from various alignment candidates, SASP directly learns the sensible local alignment for optimal stitching. Hence, SASP is more effective in achieving optimal image stitching outcomes and has significant application value. In addition, this method substantially preserves the naturalness of drone images.

VI. Experiments

To validate the effectiveness of the proposed LGSPA model and SASP framework, this section compares them with the related state-of-the-art image alignment and stitching methods, qualitatively and quantitatively. Furthermore, an ablation study is also designed to analyze the efficacy of the proposed SAW scheme intuitively.

A. Experimental Setup

1) Parameters' Setting: In (11), the five parameters are selected from the range of [1, 100] to determine their respective relative significance. By default, λ_3 and μ_2 are set to 100, because we hope to preserve original image structures and minimize projective distortions to the greatest extent possible. Then, we balance the tradeoff between image alignment and concealing perspective distortion. Empirically, to reduce the computational burden on parameter tuning in our experiments, we first fix μ_1 to 50, i.e., the middle value of the range as a baseline. To cater to various challenging aerial scenarios, we then leverage a nested iterative strategy to fine-tune λ_1 and λ_2 from [1, 100] to achieve the final optimal drone image alignment performance.

2) Datasets: We conduct experiments on a set of challenging drone images that are collected from the *HeJiadong* dataset by the Chinese Academy of Sciences, as well as two public drone image datasets [37], i.e., *Airport* and *Highway*. Concretely, the *HeJiadong* dataset predominantly consists of low-texture and repetitive-texture image scenes. The *Airport* dataset contains abundant salient structures but with large parallax and occlusions. The *Highway* dataset involves the wide baseline, salient structures, and moving vehicles.

3) Comparison Methods: To assess image alignment quality, we consider the SPHP warp [15] as the baseline, the SPW [20], the REW [28], and the TFT [24] as the cutting-edge competitive methods. For image stitching performance evaluation, we regard the PBSC approach [31] as our baseline, the line-point consistence (LPC) stitching method [21], the UDIS method [25] and the ACIS-QR1A method [14] as the state-of-the-art competitive methods. The results of these methods are generated by their source codes.

4) Quantitative Evaluation Criteria: To objectively assess image alignment and stitching performance for drone images, we use the classical peak signal-to-noise ratio (PSNR), structural similarity measure (SSIM) [38], and zero-normalized cross correlation (ZNCC) score [13] as our quantitative evaluation metrics.

B. Alignment Performance of LGSPA Model

To verify the alignment ability of LGSPA, we carry out comparative experiments on the *HeJiadong* dataset. The alignment results are analyzed in terms of two aspects: 1) the global alignment accuracy over the whole overlapping area and 2) the naturalness of warped drone images. After alignment, the linear blending technique is uniformly used to compose overlapped drone images.

TABLE I
Quantitative Image Alignment Qualities of Different Algorithms on the HeJiadong Datasets

Image Set	Metric	SPHP [15]	SPW [20]	TFT [24]	REW [28]	LGSPA
No.1	PSNR	25.88	25.94	22.85	**28.89**	28.46
	SSIM (%)	82.06	81.09	80.90	**90.83**	89.70
	ZNCC score	0.018	0.018	0.040	**0.010**	**0.010**
No.2	PSNR	27.12	27.53	24.26	29.18	**30.23**
	SSIM (%)	87.91	87.63	86.65	91.66	**93.86**
	ZNCC score	0.014	0.012	0.026	0.009	**0.006**
No.3	PSNR	28.87	28.50	**31.45**	31.10	29.89
	SSIM (%)	85.57	83.75	**93.15**	90.79	88.68
	ZNCC score	0.015	0.016	0.011	**0.010**	0.011
Average	PSNR	27.29	27.32	26.18	**29.72**	29.52
	SSIM (%)	85.18	84.15	86.90	**91.09**	90.74
	ZNCC score	0.015	0.015	0.025	0.010	**0.009**

1) Qualitative Evaluation: Fig. 2 shows performance comparisons of different alignment algorithms. To present misalignment artifacts clearly, we sampled the local regions with red and blue rectangles and magnified them aside. Meanwhile, we display the reference drone image of each set as our ground truth. From the reference image, we can observe that the captured aerial scenes are roughly planar; however, it might be difficult to obtain adequate reliable features from the large-area plains and forests.

As highlighted in magnified insets, SPHP cannot accurately align these drone images due to poor feature-matching execution. SPW introduces additional feature lines for compensation and hence could align the local linear structures more accurately. However, the low-texture and repetitive-texture areas, such as the winding river, farmlands, and green woods, are still misaligned apparently. TFT exhibits a higher alignment precision by partitioning the landscape into multiple triangular facets for more adaptive alignment. However, it is overly flexible, resulting in ghosting artifacts and black holes in the overlapping area. Besides, extreme distortions appear in the nonoverlapping area. REW adopts the mTopKRP [39] algorithm to remove mismatched feature correspondences, improving alignment accuracy. However, due to a lack of features, the river framed in the first image set is misaligned, and the forest shown in the second image set suffers from slight ghosting artifacts. By comparison, LGSPA not only aligns the winding river, farmlands, and green woods accurately in the overlapping area but also preserves the naturalness of the nonoverlapping image contents. Therefore, LGSPA demonstrates superior performance in aligning drone images captured in natural environments.

2) Quantitative Evaluation: To quantify the global alignment quality, we measure aligned pixels over the entire overlapping area. Table I records the alignment scores of different algorithms for the above image sets. The best scores are highlighted in bold, and the second best scores are underlined. As reported in Table I, LGSPA outperforms SPHP and CPW with distinct advantages and consistently obtains remarkable scores under different metrics. TFT yields some best scores in the aspect of higher alignment accuracy in the No. 3 case. However, it is clear from Fig. 2(d) that the nonoverlapping image areas are distorted seriously, which cannot be anticipated by those scores computed in the overlapping area. REW achieves competitive scores in terms of global

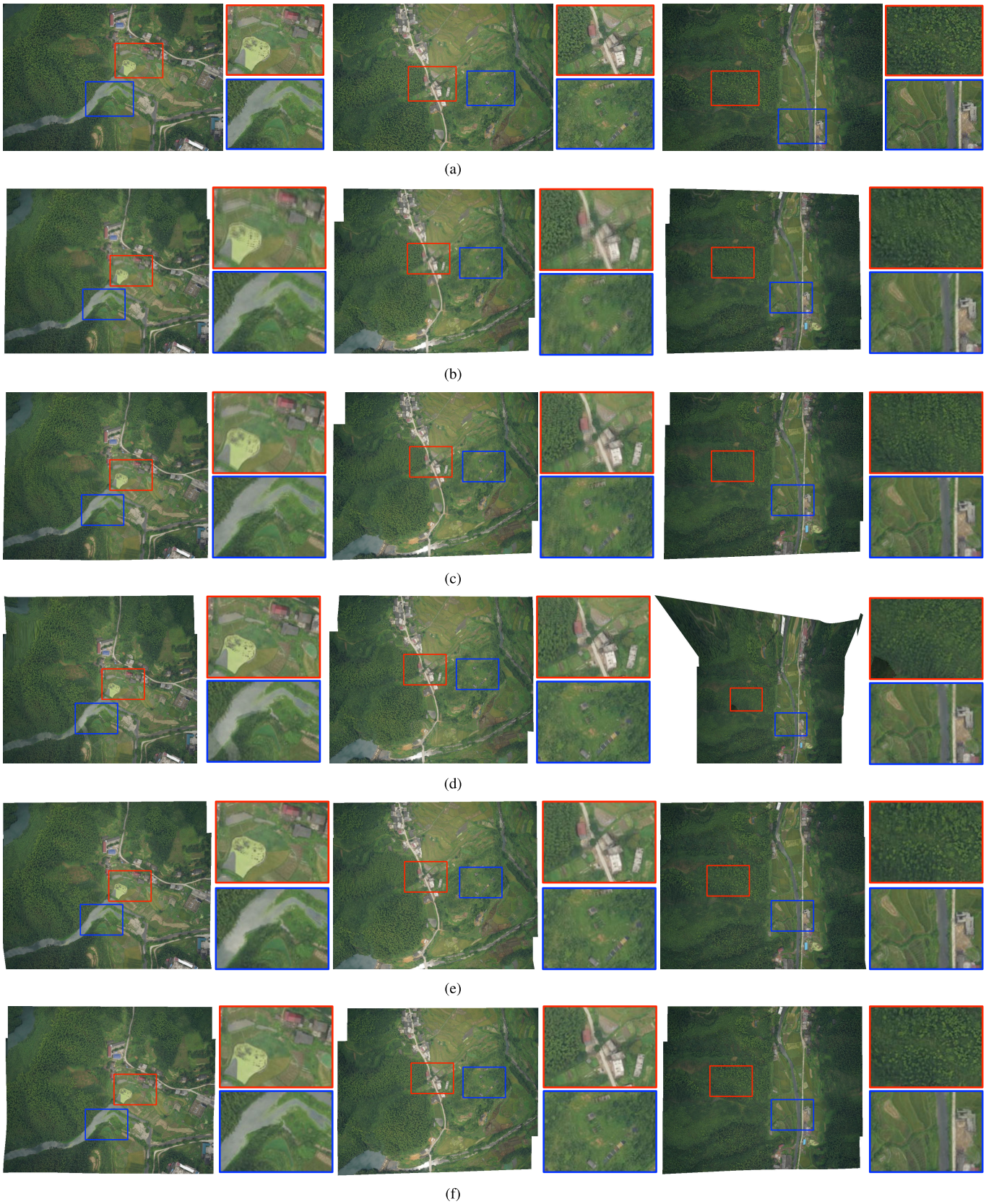

Fig. 2. Qualitative comparisons of different alignment algorithms on the *HeJiadong* dataset. (a) Reference image. Aligned results of (b) SPHP [15], (c) SPW [20], (d) TFT [24], (e) REW [28], and (f) LGSPA. For easy observation, several local regions have been sampled and magnified on the right sides.

alignment accuracy. On average, our alignment algorithm achieves the best and the second best scores under different metrics.

In this regard, LGSPA succeeds in improving the naturalness performance of warped drone images without loss of alignment accuracy in the overlapping area on the whole.

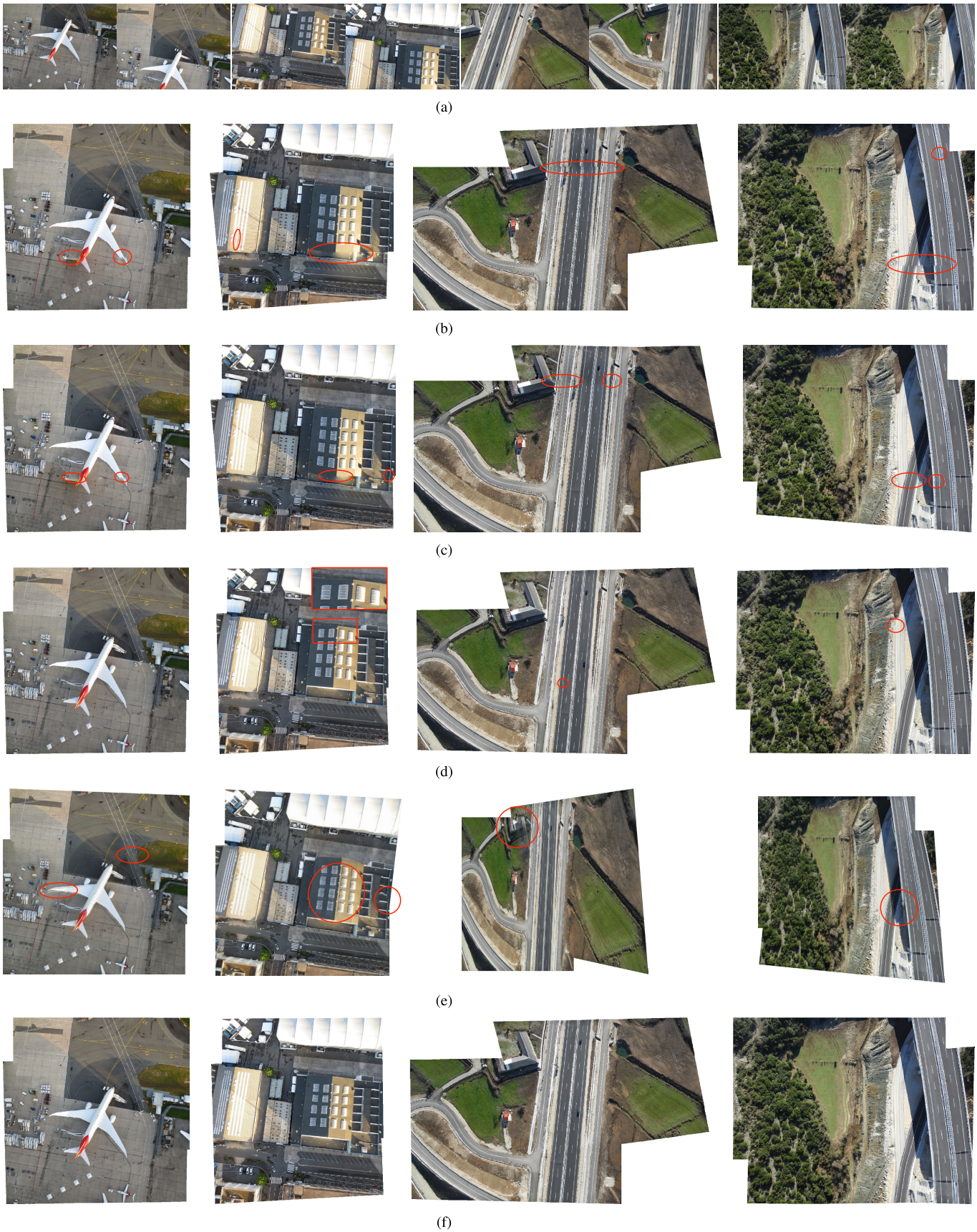

Fig. 3. Qualitative comparisons of different stitching methods on the *Airport* and *Highway* datasets. (a) Input drone images. Stitching results of (b) PBSC [31], (c) LPC [21], (d) ACIS-QR1A [14], (e) UDIS [25], and (f) SASP. For easy observation, several stitching artifacts have been highlighted in red circles and rectangles.

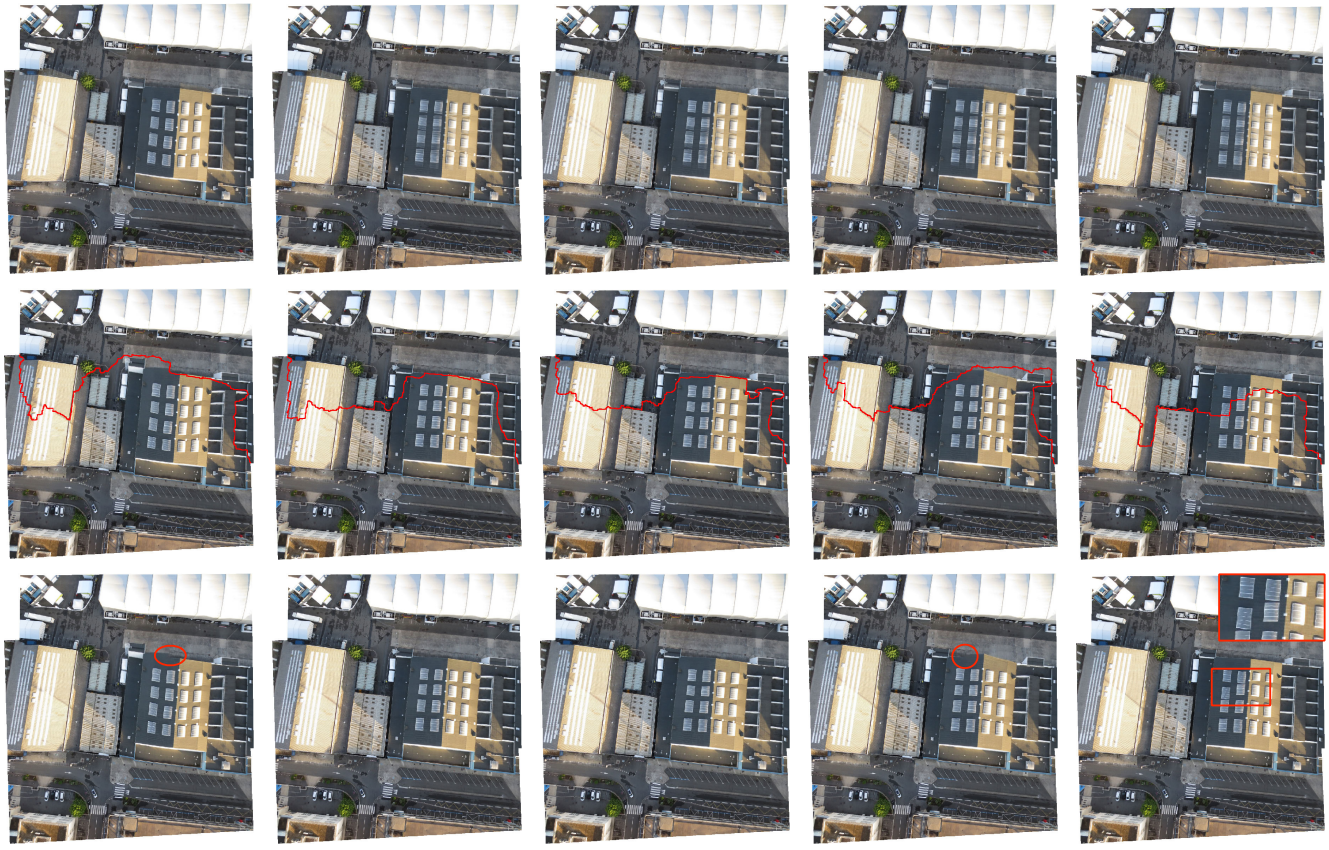

Fig. 4. Ablation study on the proposed SAW scheme. When (a) $\alpha = 0$, (b) $\alpha = 0.03$, (c) $\alpha = 0.05$, (d) $\alpha = 0.07$, and (e) $\alpha = 0.10$, (top row) corresponding alignment results of LGSPA, (middle row) generated seamlines, and (bottom row) stitching results of SASP are illustrated, respectively. For easy observation, stitching artifacts have been highlighted in red circles and rectangles.

C. Stitching Performance of SASP Framework

To demonstrate the stitching ability of SASP in dealing with challenging remote sensing scenarios, we conduct comparative experiments on the *Airport* and *Highway* datasets. The stitching results are analyzed concerning two aspects: 1) the local alignment accuracy along the seamline and 2) the naturalness of stitched drone images.

1) Qualitative Evaluation: Fig. 3 presents the final stitched drone images using different image stitching methods. From input drone images, we can observe that the captured scenes consist of salient structures, which is beneficial to obtaining sufficient feature points and lines. However, the scenes cannot be viewed as planar due to large parallax, occlusions, and wide baseline. These challenges also present obstacles in generation of high-quality stitching results. We use red circles and rectangles to indicate several notable stitching artifacts.

As can be seen, PBSC cannot provide an accurate local alignment for its seamline, producing obvious stitching errors and artifacts. LPC adopts the line-guided mesh deformation to improve the global alignment accuracy and preserve the naturalness of salient structures. However, the local alignment accuracy along the seamline is not being ameliorated effectively, and thus stitching artifacts are retained as well. ACIS-QR1A exploits the seam-driven stitching strategy to iteratively optimize the local alignment, which significantly improves the stitching quality of seamlines. However, it ignores the naturalness issue when warping drone images, as shown in the magnified region. UDIS adopts an unsupervised deep learning method to reconstruct image stitching results in both feature and pixel levels. However, the reconstructed images suffer from unnatural artifacts, e.g., the structures of the highway are squeezed seriously. Compared with them, the stitching results of SASP are free from such errors and artifacts. Therefore, SASP exhibits the robust stitching ability to multiple challenging factors, including large parallax, occlusions, and wide baseline.

2) Quantitative Evaluation: To measure the stitching quality of stitched drone images numerically, we extract the local regions (15×15 pixels) along the seamline to quantify its surrounding local alignment accuracy. The stitching scores of different methods for all the above image sets are detailed in Table II. For the *Airport* dataset, all the methods show comparable scores under different metrics. Nevertheless, the stitching errors of PBSC, LPC, and ACIS-QR1A are all notable, as shown in Fig. 2(b)–(d). For the *Highway* dataset, SASP is consistently superior to other methods under all the metrics. Moreover, SASP keeps the best stitching scores on average performance. Particularly, our average SSIM score is 13% higher than that of LPC, and our average ZNCC score is 16% lower than that of ACIS-QR1A.

TABLE II
QUANTITATIVE IMAGE STITCHING QUALITIES OF DIFFERENT METHODS ON PUBLIC AERIAL DATASETS

Dataset	Metric	PBSC [31]	LPC [21]	ACIS-QR1A [14]	SASP
Airport No.1	PSNR	21.64	21.41	**22.50**	21.35
	SSIM (%)	**75.76**	64.85	72.24	69.24
	ZNCC score	**0.135**	0.215	0.179	0.173
Airport No.2	PSNR	18.59	22.10	**25.98**	25.41
	SSIM (%)	57.71	69.68	**78.42**	77.45
	ZNCC score	0.194	**0.101**	0.128	0.119
Highway No.1	PSNR	18.40	19.11	23.00	**24.29**
	SSIM (%)	62.07	53.25	70.93	**72.08**
	ZNCC score	0.163	0.206	0.174	**0.156**
Highway No.2	PSNR	22.76	21.69	21.14	**23.20**
	SSIM (%)	76.47	72.23	62.07	**76.93**
	ZNCC score	0.140	0.113	0.187	**0.110**
Average	PSNR	20.34	21.07	23.15	**23.56**
	SSIM (%)	68.00	65.00	70.91	**73.92**
	ZNCC score	0.158	0.158	0.167	**0.139**

D. Ablation Study on SAW Scheme

To explore the effects of the proposed SAW scheme on LGSPA and SASP intuitively, we devise an ablation study on the *Airport* dataset to figure out the corresponding alignment and stitching results with the varying values of α in SAW.

As shown in Fig. 4, we set $\alpha = \{0, 0.003, 0.05, 0.07, 0.10\}$ in (13) and illustrate the intermediate results of our image stitching process, respectively. Stitching artifacts of the final results are highlighted in red circles. It can be observed that the best-fitting global alignment is achieved when $\alpha = 0$, since LGSPA tries to accurately align the whole overlapping area. However, due to inconsistent feature correspondences, the salient structures are distorted severely and the stitching performance is also not plausible in the SASP result. When $\alpha = 0.03$ and 0.05, the global alignment accuracy decreases, but the local alignment precision is enhanced along the seamline. Consequently, the final stitched drone images are free from stitching errors and the distortion problem is also well-addressed. When $\alpha = 0.07$, the active range of the local alignment constraint is not large enough to be precisely imposed for the seamline, and thus stitching artifacts appear again.

It can be inferred that the influence of the local alignment constraint should be controlled in a sensible range along the seamline. In this article, we set $\alpha \in [0.03, 0.05]$ in SAW to control the active range of local alignment, enabling SASP to have a much higher tolerance against those challenging stitching factors such as large parallax, occlusions, and wide baseline.

VII. CONCLUSION

This article first proposed LGSPA for drone image alignment with high precision and good naturalness. Then, we proposed SAW to improve seamline quality essentially. Using LGSPA and SAW, we further developed SASP to generate final drone image stitching results against different kinds of realistic challenges. With SASP, we optimized the desired well-aligned local alignment for seamline detection, thereby generating the final optimal seamline to stitch images seamlessly. Qualitative and quantitative experiments have been conducted to verify that LGSPA and SASP are capable of generating higher quality alignment and stitching results than several state-of-the-art methods over multiple challenging aerial scenarios, including low textures, repetitive textures, large parallax, wide baseline, and occlusions. Our future work will develop new seamline detection methods to improve the drone image stitching performance.

REFERENCES

[1] R. Ma, R. Wang, G. Liu, W. Meng, and X. Liu, "UAV-aided cooperative data collection scheme for ocean monitoring networks," *IEEE Internet Things J.*, vol. 8, no. 17, pp. 13222–13236, Sep. 2021.

[2] B. Wang, Y. Sun, D. Liu, H. M. Nguyen, and T. Q. Duong, "Social-aware UAV-assisted mobile crowd sensing in stochastic and dynamic environments for disaster relief networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 1070–1074, Jan. 2020.

[3] F. Ho et al., "Decentralized multi-agent path finding for UAV traffic management," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 2, pp. 997–1008, Feb. 2022.

[4] Y. Mo, X. Kang, P. Duan, and S. Li, "A robust UAV hyperspectral image stitching method based on deep feature matching," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5517514.

[5] Q. Wan, J. Chen, L. Luo, W. Gong, and L. Wei, "Drone image stitching using local mesh-based bundle adjustment and shape-preserving transform," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 7027–7037, Aug. 2021.

[6] Y. Yuan, F. Fang, and G. Zhang, "Superpixel-based seamless image stitching for UAV images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1565–1576, Feb. 2021.

[7] J. W. P. Sun, H. Li, W. Li, X. Meng, C. Ge, and Q. Du, "Low-rank and sparse representation for hyperspectral image processing: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 1, pp. 10–43, Jun. 2021.

[8] Y. Zhang, Y. Wang, X. Chen, X. Jiang, and Y. Zhou, "Spectral–spatial feature extraction with dual graph autoencoder for hyperspectral image clustering," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 12, pp. 8500–8511, Dec. 2022.

[9] N. Li, Y. Xu, and C. Wang, "Quasi-homography warps in image stitching," *IEEE Trans. Multimedia*, vol. 20, no. 6, pp. 1365–1375, Jun. 2018.

[10] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *Int. J. Comput. Vis.*, vol. 74, no. 1, pp. 59–73, Aug. 2007.

[11] J. Gao, Y. Li, T.-J. Chin, and M. S. Brown, "Seam-driven image stitching," *Eurographics 2013—Short Papers*. Girona, Spain, 2013, pp. 45–48.

[12] F. Zhang and F. Liu, "Parallax-tolerant image stitching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3262–3269.

[13] K. Lin, N. Jiang, L. F. Cheong, M. N. Do, and J. Lu, "SEAGULL: Seam-guided local alignment for parallax-tolerant image stitching," in *The 14th Eur. Conf. Comput. Vis.*, Oct. 2016, pp. 370–385.

[14] J. Li and Y. Zhou, "Automatic color image stitching using quaternion rank-1 alignment," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 19720–19729.

[15] C.-H. Chang, Y. Sato, and Y.-Y. Chuang, "Shape-preserving half-projective warps for image stitching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3254–3261.

[16] C.-C. Lin, S. U. Pankanti, K. N. Ramamurthy, and A. Y. Aravkin, "Adaptive as-natural-as-possible image stitching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1155–1163.

[17] Y.-S. Chen and Y.-Y. Chuang, "Natural image stitching with the global similarity prior," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2016, pp. 186–201.

[18] K. Joo, N. Kim, T. Oh, and I. S. Kweon, "Line meets as-projective-as-possible image stitching with moving DLT," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 1175–1179.

[19] S. Li, L. Yuan, J. Sun, and L. Quan, "Dual-feature warping-based motion model estimation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 4283–4291.

[20] T. Liao and N. Li, "Single-perspective warps in natural image stitching," *IEEE Trans. Image Process.*, vol. 29, pp. 724–735, 2020.

[21] Q. Jia et al., "Leveraging line-point consistence to preserve structures for wide parallax image stitching," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2021, pp. 12186–12195.

[22] J. Li and Y. Zhou, "Automatic quaternion-domain color image stitching," *IEEE Trans. Image Process.*, vol. 33, pp. 1299–1312, 2024.
[23] J. Zaragoza, T.-J. Chin, M. S. Brown, and D. Suter, "As-projective-as-possible image stitching with moving DLT," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2339–2346.
[24] J. Li, B. Deng, R. Tang, Z. Wang, and Y. Yan, "Local-adaptive image alignment based on triangular facet approximation," *IEEE Trans. Image Process.*, vol. 29, pp. 2356–2369, 2020.
[25] L. Nie, C. Lin, K. Liao, S. Liu, and Y. Zhao, "Unsupervised deep image stitching: Reconstructing stitched features to images," *IEEE Trans. Image Process.*, vol. 30, pp. 6184–6197, 2021.
[26] D. Guo, J. Chen, L. Luo, W. Gong, and L. Wei, "UAV image stitching using shape-preserving warp combined with global alignment," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
[27] Q. Xu, J. Chen, L. Luo, W. Gong, and Y. Wang, "UAV image mosaicking based on multiregion guided local projection deformation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3844–3855, 2020.
[28] Y. Zhang, Z. Wan, X. Jiang, and X. Mei, "Automatic stitching for hyperspectral images using robust feature matching and elastic warp," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3145–3154, 2020.
[29] K. Sun and Y. Wan, "Image stitching with weighted elastic registration," in *Proc. 10th Int. Conf. Commun., Circuits Syst. (ICCCAS)*, Dec. 2018, pp. 403–407.
[30] P. S. Heckbert, "Fundamentals of texture mapping and image warping," Univ. California, Berkeley, CA, USA, Tech. Rep. UCB/CSD 89/516, 1989.
[31] N. Li, T. Liao, and C. Wang, "Perception-based seam cutting for image stitching," *Signal, Image Video Process.*, vol. 12, no. 5, pp. 967–974, Feb. 2018.
[32] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, Jun. 2008.
[33] R. G. Von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall, "LSD: A line segment detector," *Image Process. Line*, vol. 2, pp. 35–55, Mar. 2012.
[34] M. A. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
[35] Q. Jia, X. Gao, X. Fan, Z. Luo, H. Li, and Z. Chen, "Novel coplanar line-points invariants for robust line matching across views," in *Proc. 14th Eur. Conf. Comput. Vis.*, Jan. 2016, pp. 599–611.
[36] P. Pérez, M. Gangnet, and A. Blake, "Poisson image editing," *Seminal Graph. Papers, Pushing Boundaries*, vol. 2, pp. 577–582, Aug. 2023.
[37] AgEagle. (2023). *Drones, Sensors and Software for Automated Aerial Intelligence*. [Online]. Available: https://ageagle.com/resources/?filter_by=data-set
[38] W. Zhou, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[39] X. Jiang, J. Jiang, A. Fan, Z. Wang, and J. Ma, "Multiscale locality and rank preservation for robust feature matching of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6462–6472, Sep. 2019.

Jiaxue Li received the Ph.D. degree from the University of Macau, Macau, China, in 2024.

She is a Lecturer with China University of Geosciences, Beijing, China. Her research interests include image stitching and remote sensing image processing.

Yicong Zhou (Senior Member, IEEE) received the B.S. degree in electrical engineering from Hunan University, Changsha, China, in 1992, and the M.S. and Ph.D. degrees in electrical engineering from Tufts University, Medford, MA, USA, in 2008 and 2010, respectively.

He is a Professor with the Department of Computer and Information Science, University of Macau, Macau, China. His research interests include image processing, computer vision, machine learning, and multimedia security.

Dr. Zhou is a fellow of the Society of Photo-Optical Instrumentation Engineers (SPIE) and was recognized as one of "Highly Cited Researchers" in 2020, 2021, 2023, and 2024. He serves as a Senior Area Editor for IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY and an Associate Editor for IEEE TRANSACTIONS ON CYBERNETICS, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, and IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.