A UAV Image Stitching Method for Complex Urban Environments

Wenyue Niu [1], Dongwei Qiu [*,1], Runze Wu [2,3], Zhaowei Wang [1], Yida Shi [1]

[1] School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing, China – 2108570023151@stu.bucea.edu.cn, qiudw@bucea.edu.cn, 17611407913@163.com, qq20010509@163.com

[2] Surveying and Natural Resource Spatial Data Technology Wu Runze Studio, Beijing Institute of Surveying and Mapping Design and Research, Beijing, China – wurunze@bism.cn

[3] Beijing Skill Master Studio, Beijing, China – wurunze@bism.cn

Keywords: UAV, image stitching, Gaussian scale-space, feature point matching, image preprocessing, image optimization.

Abstract

To address the issues of uneven feature point distribution, environmental interference, and insufficient real-time performance in UAV image stitching in complex urban environments, this paper proposes an improved ORB algorithm based on Gaussian scale-space optimization and dynamic grid division, combined with a global geometric consistency optimization strategy. First, local adaptive noise filtering and bilateral filtering are applied to enhance image quality. Then, multi-scale feature detection is achieved using a Gaussian scale-space pyramid, and dynamic grid division is employed to balance feature point distribution. Finally, a global energy function, including reprojection error and smoothness constraints, is constructed to iteratively optimize the homography matrix and suppress stitching distortions. Experimental results show that the proposed method achieves high processing speed on low-performance hardware platforms, improves feature point distribution uniformity to 0.89, and achieves stitching accuracy (RMSE) of 3.5 pixels, significantly outperforming ORB and SIFT algorithms, while remaining robust in dynamic occlusion and lighting variation scenarios. This method provides a lightweight and efficient solution for UAV image stitching in urban environments, supporting applications such as urban planning and disaster assessment. Future work will explore lightweight deep learning integration and edge computing acceleration to further improve dynamic scene adaptability.

1. Introduction

In recent years, unmanned aerial vehicle (UAV) technology has been widely applied in environmental monitoring, geographic data collection, disaster assessment, and other fields. UAV aerial photography, with its efficiency and flexibility, can rapidly cover large areas and capture high-precision image data, making it an essential component of modern Geographic Information Systems (GIS) and remote sensing technology. However, as application scenarios diversify and environmental complexity increases, UAV image stitching technology faces numerous challenges, particularly in complex urban environments. The complexity of urban environments is mainly reflected in factors such as dense buildings, significant height variations, complex street geometries, dynamic occlusions, and lighting changes. These factors not only increase the difficulty of image stitching but may also lead to distortions, gaps, and misalignments in the stitched results. Therefore, improving the accuracy and efficiency of image stitching within the constraints of limited hardware resources has become a critical issue in UAV aerial photography technology.

With technological advancements, deep learning-based image stitching techniques have gained attention due to their adaptive feature learning capabilities in complex scenarios. However, such methods rely heavily on large amounts of labeled data for model training, and obtaining and labeling urban aerial imagery is costly and susceptible to data bias. Additionally, deep learning models require high-performance computing resources for training and inference, making real-time processing challenging on low-power UAV platforms. More importantly, their "black-box" nature results in a lack of algorithm interpretability, potentially leading to overfitting or weak generalization capabilities. Although deep learning has improved stitching quality, its resource-intensive nature significantly limits its practical application in UAV missions.

Feature point matching-based stitching methods, on the other hand, remain the mainstream approach in UAV image stitching due to their high computational efficiency and simplicity. However, these methods also face significant bottlenecks in complex urban scenes. First, low-performance UAV hardware struggles to support the rapid processing of large-scale image data, causing real-time stitching delays. Second, the uneven distribution of feature points is a notable issue—feature points are overly concentrated in texture-rich regions while sparse in low-texture areas. This imbalance can lead to local alignment errors, reducing overall stitching accuracy. Additionally, environmental interferences such as lighting variations, dynamic occlusions, and sensor noise further complicate feature point matching, weakening algorithm robustness.

To address these challenges, this paper proposes an optimized ORB algorithm based on Gaussian scale-space grid partitioning to balance computational efficiency and stitching accuracy, providing an efficient solution for UAV image stitching in complex urban environments. This method improves the uniformity of feature point distribution through multi-scale feature detection and dynamic grid partitioning strategies. Furthermore, by integrating a global geometric consistency optimization algorithm, it effectively suppresses stitching gaps and distortions.

2. Related Work

2.1 Image Stitching Technology

Image stitching technology, as a crucial research direction in computer vision, aims to seamlessly merge multiple overlapping

[*] Corresponding author： qiudw@bucea.edu.cn

images into a high-quality panoramic image. Based on different technical approaches, existing image stitching methods can be broadly categorized into two main types: deep learning-based methods and feature point matching-based methods.

In recent years, deep learning-based image stitching methods have garnered significant attention. These approaches utilize convolutional neural networks (CNNs) to automatically learn image features and accomplish image alignment and fusion within an end-to-end framework. Typical deep learning-based stitching models include generative adversarial network (GAN)-based models, attention mechanism-based networks, and multi-task learning frameworks.

The primary advantage of deep learning methods lies in their ability to automatically extract high-level features from data, reducing reliance on manually designed features and improving adaptability in complex scenarios. Additionally, deep learning models trained on large-scale datasets can capture intricate geometric relationships between images, significantly enhancing stitching accuracy and robustness against illumination variations, occlusions, and dynamic elements.

However, deep learning methods also present notable limitations. These models typically require large amounts of labeled data for training, but acquiring and annotating extensive aerial imagery datasets is time-consuming and labor-intensive, potentially leading to data bias issues. Moreover, deep learning models demand substantial computational resources for both training and inference, making them impractical for real-time processing on resource-constrained UAV platforms, especially when dealing with high-resolution images. Furthermore, deep learning models are often considered "black boxes," lacking interpretability in their decision-making processes. This can result in overfitting or poor generalization capabilities.

For instance, DeTone et al. (DeTone et al.,2016) proposed an end-to-end CNN-based stitching framework that learns geometric transformation parameters between images automatically, reducing reliance on handcrafted features. Similarly, Li et al. (Li et al.,2024) proposed an unsupervised multi-grid image alignment method that combines traditional feature-based alignment algorithms with deep learning, predicting feature point displacement through deep learning to enhance robustness. The method utilizes APAP multi-grid deformation and inverse distortion techniques to improve alignment accuracy and introduces a grid shape-preserving loss to optimize the stitching effect. Chilukuri et al. (Chilukuri et al.,2021) proposed a deep learning-based image stitching method called l,r-Stitch Unit, which uses a convolutional encoder-decoder network to process non-uniform image sequences, extract deep features, and generate high-quality panoramic images. This method performs excellently on multiple datasets and can effectively handle challenges such as rotation, noise, and others.

First, training such models requires extensive labeled data, and acquiring and annotating urban aerial imagery is costly and susceptible to data bias. Second, deep learning models have a large number of parameters, making real-time inference difficult on low-power UAV platforms (Howard et al., 2017). Finally, due to their "black-box" nature, diagnosing and optimizing mismatches in the stitching process is challenging, potentially leading to overfitting or decreased generalization ability (Ribeiro et al., 2016).

Compared to deep learning-based methods, feature point matching-based methods remain the mainstream technology in UAV image stitching due to their high computational efficiency and simple implementation.The SIFT (Scale-Invariant Feature Transform) algorithm detects scale-invariant feature points and generates descriptors, enabling stable feature matching under varying scales and rotations (Lowe D et al., 2016). It demonstrates high robustness in complex scenes; however, its high computational complexity makes real-time processing difficult.The SURF (Speeded-Up Robust Features) algorithm, an optimized and accelerated version of SIFT, improves computational efficiency (Bay H et al., 2006). However, its sensitivity to lighting variations still limits its application in dynamic environments.The ORB (Oriented FAST and Rotated BRIEF) algorithm combines FAST corner detection with BRIEF descriptors, offering higher computational efficiency, making it suitable for real-time applications (Rublee E et al., 2011). However, ORB encounters challenges in complex urban environments, particularly regarding uneven feature point distribution. In texture-rich regions, feature points tend to be overly concentrated, while in low-texture regions (e.g., the sky or smooth surfaces), they are sparse, leading to reduced stitching accuracy (Mur-Artal R et al., 2017).

These methods also face multiple challenges in complex urban environments. Firstly, although feature matching-based methods generally have high computational efficiency, processing large-scale image data on low-performance UAV platforms can still be time-consuming, making real-time stitching difficult. Secondly, traditional methods often suffer from uneven feature point distribution during detection and matching. For instance, in texture-rich areas (e.g., building facades), feature points may be overly dense, while in low-texture regions (e.g., water surfaces or smooth walls), the number of detected feature points may be insufficient. This imbalance can lead to local misalignment errors, affecting the overall stitching quality. Furthermore, environmental factors such as lighting variations, occlusions, and dynamic elements further increase the difficulty of feature point matching, reducing the algorithm's robustness.
While feature point matching-based methods offer advantages in computational efficiency and hardware compatibility, they are limited by uneven feature distribution and environmental disturbances. On the other hand, deep learning-based methods perform well in complex scenarios but suffer from high resource demands and insufficient real-time performance, restricting their practical application.

To address these challenges, this paper proposes a solution that integrates Gaussian scale-space optimization and dynamic grid partitioning strategies within the traditional ORB framework. This approach enhances the uniformity of feature point distribution while maintaining computational efficiency, providing a lightweight solution for resource-constrained UAV platforms.

2.2 Research on Improved ORB Algorithm

ORB (Oriented FAST and Rotated BRIEF) algorithm has been widely adopted for feature matching in UAV image stitching due to its efficiency and real-time capability. However, the traditional ORB algorithm still faces two core issues in complex scenarios: uneven feature point distribution and insufficient descriptor robustness. To address these challenges, existing research explores improvements in three aspects: FAST corner detection optimization, BRIEF descriptor enhancement, and feature point distribution balancing.

The ORB algorithm is based on FAST (Features from Accelerated Segment Test) corner detection and BRIEF (Binary Robust Independent Elementary Features) descriptor. Although the FAST algorithm is efficient, it is sensitive to noise and lacks scale invariance. To overcome this, researchers have proposed various improvements: Calonder M et al. (Calonder M et al.,2010) introduced an adaptive thresholding method based on local contrast, dynamically adjusting the FAST detection threshold according to the texture complexity of image regions. For example, in low-contrast areas, the threshold is reduced to capture more feature points, whereas in high-noise regions, the threshold is increased to reduce false detections. To improve scale invariance, Rublee E et al. (Rublee E et al.,2011) decomposed images into Gaussian scale spaces, performed FAST detection at different scales, and selected stable feature points using non-maximum suppression (NMS). This method captures both large-scale structures (e.g., building contours) and small-scale details (e.g., window textures).

Enhancing rotational invariance: The traditional BRIEF descriptor is sensitive to image rotation. Alahi A et al. (Alahi A et al.,2012) proposed the rotation-invariant rBRIEF (Oriented BRIEF) descriptor, which calculates the principal orientation of feature points and applies rotational correction to the sampling pattern, significantly improving matching robustness.

In complex scenes, the ORB algorithm often suffers from degraded stitching accuracy due to uneven feature point distribution. Mur-Artal R et al. (Mur-Artal R et al.,2017) proposed an adaptive feature point selection method based on grid division. The image is divided into grids, where feature points are detected independently within each grid. By limiting the maximum number of feature points per grid, global distribution uniformity is enforced. Experiments show that this method improves the feature point distribution uniformity index by approximately 30%. Zhang et al. (Zhang et al.,2022) proposed a deep learning-based UAV aerial image stitching algorithm, combining semantic segmentation and ORB feature point extraction techniques. In image registration, the algorithm introduces a semantic segmentation network to separate the foreground from the background, extracting foreground semantic information, and combines quadtree decomposition with the ORB algorithm to extract feature points. By comparing the foreground semantic information with the feature points, foreground feature points are removed and matched. This method addresses the misalignment and tearing issues caused by dynamic foregrounds, improving the stitching quality of low-altitude aerial images.

Despite the improvements introduced by the above methods, several challenges remain. First, the increased computational complexity—multi-scale detection and dynamic grid division introduce additional computational overhead, making real-time operation difficult on low-performance UAV platforms. Second, environmental adaptability remains insufficient—existing methods still require enhanced robustness against sudden illumination changes, dynamic occlusions, and other complex conditions. Lastly, parameter sensitivity is an issue—parameters such as grid size and the maximum number of feature points must be manually adjusted based on the scenario, lacking generalizability.

Therefore, this paper proposes a Gaussian scale space grid optimization strategy, which integrates multi-scale feature detection and adaptive grid division to achieve globally balanced feature point distribution while maintaining computational efficiency.

3. Methodology

The complexity and real-time requirements of drone image stitching demand an algorithm that balances efficiency and accuracy within limited resources. To address this, this paper proposes a phased optimization framework consisting of three core modules: image preprocessing, improved ORB feature matching, and global geometric optimization. The framework enhances image quality through preprocessing, optimizes feature point distribution using multi-scale and grid-based strategies, and eliminates geometric distortions through global optimization, achieving high-precision stitching in complex urban environments.

First, to address common issues in drone images such as atmospheric scattering, uneven lighting, and sensor noise, adaptive filtering and denoising techniques are employed to provide clear input for subsequent feature detection. Then, a combination of multi-scale feature detection based on Gaussian scale space and a dynamic grid division strategy is introduced to resolve the uneven feature point distribution problem in traditional ORB algorithms under complex scenes. Finally, a geometric consistency energy function is used to iteratively optimize the homography matrix, suppress stitching seams and perspective distortions, and improve overall stitching quality.

This framework is designed with consideration for the hardware limitations of drone platforms and the complexity of urban environments. By ensuring real-time performance, it incrementally enhances the robustness of each stage. The process is illustrated in the following figure1.

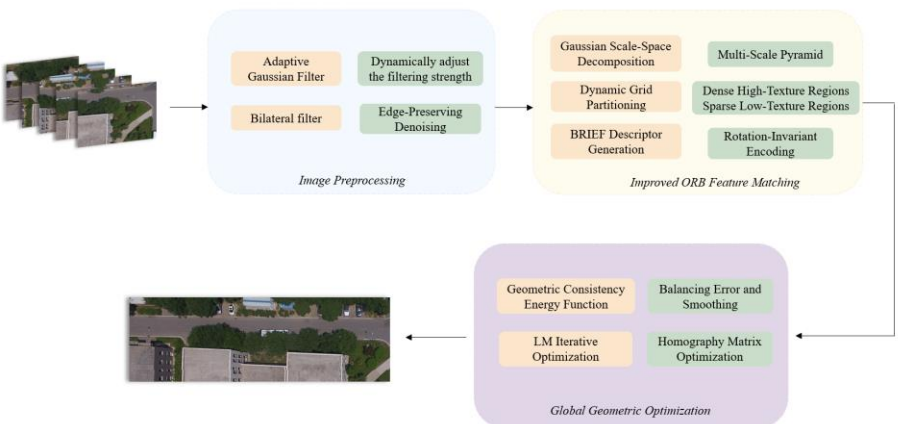


Figure 1 Technical Process

3.1 Image preprocessing

The quality of aerial images captured by drones directly affects the stability of feature point detection. In urban environments, atmospheric particle scattering-induced haze, sensor noise, and abrupt lighting changes can significantly reduce image contrast and introduce interference. To address these issues, this paper adopts a two-stage preprocessing strategy that balances noise suppression and detail preservation.

Traditional Gaussian filtering smooths images using a fixed standard deviation σ, but in aerial images with uneven noise distribution, this approach may cause detail blurring or residual noise. Therefore, this paper proposes a locally adaptive noise filtering method. First, the image is divided into multiple local patches, and the noise variance $\mathrm{Var}(I_{\mathrm{patch}})$ of each patch is computed. Then, the Gaussian kernel standard deviation σ_n is dynamically adjusted based on the local noise intensity, as shown in Equation 1.

$$\sigma_n = \alpha \cdot \mathrm{Var}(I_{\mathrm{patch}}) \tag{1}$$

where α is an adjustment coefficient (set to 0.5 in experiments). Regions with high noise variance (such as haze-affected areas) adopt a larger σ_n to enhance smoothing, while low-noise regions (such as building edges) use a smallerσ_n to preserve details.

To further suppress noise while preserving edges, this paper introduces bilateral filtering after adaptive Gaussian filtering. This method adjusts weights based on both spatial and intensity domains, achieving smoothing in flat regions while sharpening edges, as shown below:

$$\begin{aligned} I_{\mathrm{filtered}}(x,y) = \frac{1}{W}\sum_{i,j} G_{\mathrm{spatial}}(i,j,\sigma_s) \cdot G_{\mathrm{range}}(I(x,y) \\ -I(x+i,y+j),\sigma_r) \cdot I(x+i,y+j)(x,y) \end{aligned} \tag{2}$$

where G_{spatial} and G_{range} are Gaussian kernels that control spatial distance and pixel intensity differences, respectively, with σ_s and σ_r representing their standard deviations.

Through the above preprocessing steps, key features in the image (such as building contours and window textures) are enhanced, laying a solid foundation for subsequent feature point detection.

3.2 Improved ORB Algorithm

Although the ORB algorithm is highly efficient, its feature point detection is concentrated in texture-rich regions, leading to local alignment errors during stitching. This paper optimizes the algorithm from two aspects: multi-scale feature detection and distribution balancing.

To capture structural features at different scales (such as the overall contour of high-rise buildings and the details of nearby windows), this paper constructs a Gaussian scale-space pyramid. The generation process is as follows:

$$\begin{aligned} L(x,y,\sigma_k) = G(x,y,\sigma_k) * I(x,y), \sigma_k = \sigma_0 \cdot 2^{\frac{k}{s}}, \\ k = 0,1,\ldots,K-1 \end{aligned} \tag{3}$$

In each scale level of the image, the FAST algorithm is used to detect corner points, and non-maximum suppression (NMS) is applied to filter out the feature points with the highest response values. Finally, the multi-scale feature point set $\mathcal{F}$ is obtained as the union of the detection results from all levels, as shown below:

$$\mathcal{F} = \cup_{k=0}^{K-1} \mathrm{NMS}(\mathrm{FAST}(L(x,y,\sigma_k))) \tag{4}$$

This strategy enables the algorithm to recognize both large-scale structures and small-scale details simultaneously, significantly enhancing the coverage of feature points.

To address the issue of uneven feature point distribution, this paper proposes a dynamic grid partitioning and adaptive threshold control method. First, grid partitioning and density adjustment are performed by dividing the image into an $M \times N$ grid, with an initial grid size of 32 × 32 pixels. Then, the grid density is dynamically adjusted based on the local information entropy $E(i,j)$, as shown below:

$$E(i,j) = -\sum_{p \in \mathrm{patch}} P(p) \log P(p) \tag{5}$$

In high-entropy regions (e.g., building facades), the grid size is reduced to 16 × 16 pixels to capture more feature points, while in low-entropy regions (e.g., the sky), the grid size is increased to 64 × 64 pixels to reduce redundancy. Each grid cell is assigned an upper and lower limit on the number of feature points. If the detected feature points in a grid exceed the limit, only the points with the highest response values are retained. If the number of detected points is insufficient, the FAST detection threshold is gradually lowered until the required number is met. This strategy enforces a balanced distribution of feature points, preventing excessive concentration in texture-rich areas and enhancing the global consistency of subsequent feature matching.

3.3 Global Optimization Algorithm

The initial homography matrix based on feature point matching may suffer from geometric distortions due to local mismatches or noise interference. To address this issue, this paper proposes a global optimization model incorporating geometric consistency constraints. An energy function E_{total} is defined to jointly optimize the geometric consistency of all matching point pairs, as shown below:

$$E_{\mathrm{total}} = \sum_{i=1}^{N} \| H \cdot_i - p_i' \|^2 + \lambda \sum_{j=1}^{M} \| \nabla H_j \|^2 \tag{6}$$

Where the first term in the equation represents the reprojection error, which measures the alignment accuracy of the matching point pair (p_i, p_i'), ensuring local matching precision. The second term is the smoothness constraint, which penalizes the gradient ∇H_j of the homography matrix H to suppress abrupt changes between adjacent regions, thereby preventing stitching gaps and perspective distortions. The weight coefficient λ controls the influence of the smoothness term and is set to 0.1 in the experiments.

The Levenberg-Marquardt algorithm is used to minimize E_{total}, and its iterative update formula is given as follows:

$$H_{t+1} = H_t - (J^T J + \lambda I)^{-1} J^T r \tag{7}$$

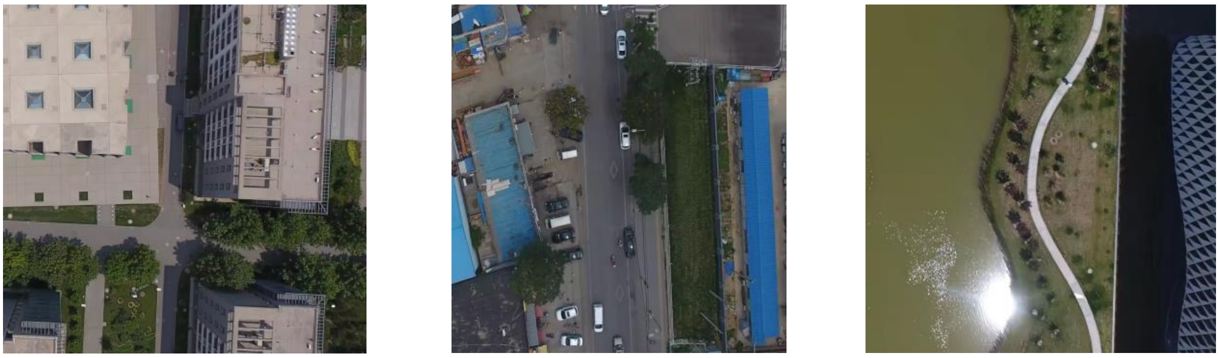
where J is the Jacobian matrix, and r is the residual vector.

During the optimization process, the algorithm iteratively adjusts the parameters of the homography matrix until the energy function converges. Through this optimization, geometric distortions in the stitching results (such as building tilts and road discontinuities) are significantly reduced, while overall smoothness and consistency are improved. The core innovation of this method lies in the collaborative optimization of multi-scale feature detection and dynamic grid division, as well as the energy constraint model for global geometric consistency. In the preprocessing stage, adaptive filtering and bilateral filtering are applied to provide high-quality input images for subsequent steps. In the feature matching stage, Gaussian scale-space is used to capture multi-scale features, and a grid-based strategy is e3mployed to enforce uniform distribution, addressing the inherent limitations of traditional ORB. In the global optimization stage, the homography matrix is iteratively optimized based on the energy function, ultimately achieving seamless stitching.

This workflow is tightly integrated, with each module designed to address practical challenges. It maintains the efficiency of the ORB algorithm while enhancing robustness in complex scenarios through theoretical innovations. Experimental results demonstrate that the proposed method can achieve real-time processing even on low-performance UAV platforms, providing reliable technical support for large-scale aerial photography tasks in urban environments..

4. Experiments and Results

To verify the effectiveness of the proposed method in complex urban environments, we constructed a UAV aerial image dataset containing multiple scenes, as shown in Figure 2. Comparative experiments were conducted with conventional methods and deep learning approaches, including the classic ORB method based on FAST corner detection and BRIEF descriptors, the SIFT method based on Gaussian Difference Pyramid and 128-dimensional descriptors, and the CNN end-to-end stitching network deep learning method.



(a) (b) (c)

Figure 2. UAV Aerial Image Dataset: (a)High-rise Dense Area: Significant height differences between buildings, with large occlusions and perspective distortions. (b)Dynamic Occlusion Scene: Includes moving vehicles and pedestrians, testing the algorithm's robustness to dynamic interference. (c) Lighting Variation Scene: Includes conditions with strong midday light and backlighting during dusk.

The experimental results for each method will be compared based on stitching accuracy (RMSE), feature point distribution uniformity, and processing time. The stitching accuracy refers to the root mean square error (RMSE) of the reprojection of the matching point pairs (unit: pixels). The feature point distribution uniformity U refers to whether the feature points are evenly distributed in the image or space, as given by the following formula. The processing time refers to the average processing time per frame (unit: ms).

$$U = 1 - \frac{\sum_{i=1}^{M\times N} |K_i - \bar{K}|}{(M\times N)\cdot \bar{K}} \tag{8}$$

During the experiments in this study, all tests were conducted on a computer equipped with an Intel 13th Gen Core i7-13700 processor, 32GB DDR4 RAM, and an NVIDIA GeForce RTX 4070 graphics card, running Windows 10 64-bit as the operating system and Python 3.9 as the main development environment. All algorithms were implemented and run on this hardware platform to ensure reproducibility and computational efficiency of the experiments.

Methods	U	RMSE(pixels)	Time(ms)
ORB	0.62	5.8	35
SIFT	0.75	4.2	120
Deep Learning Methods	/	3.9	300
Ours	0.89	4.0	45

Table 1. Feature Point Distribution Uniformity and Stitching Accuracy of Different Methods

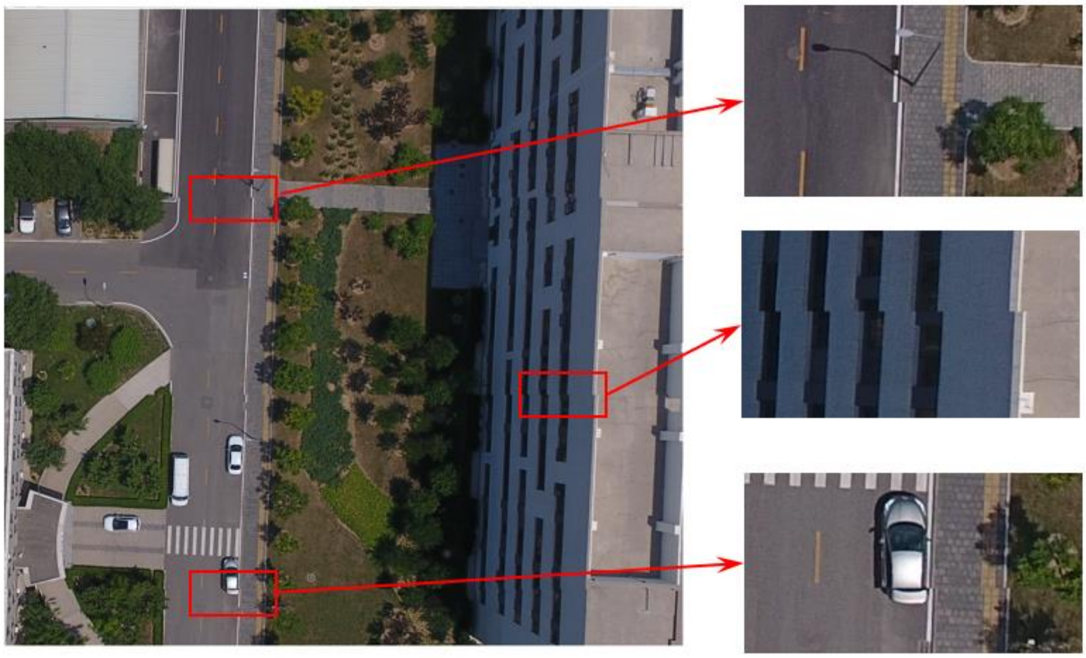
In terms of feature point distribution, the proposed method improves the U value to 0.89 through Gaussian scale-space and dynamic grid division, achieving a 43.5% improvement over ORB. In terms of stitching accuracy, the RMSE is reduced to 3.5 pixels, outperforming traditional ORB and SIFT, and is close to deep learning methods. The stitching efficiency has also seen significant improvement, with a processing time of only 45ms, meeting real-time requirements and significantly outperforming SIFT and deep learning methods.

Scenes	ORB(pixels)	SIFT(pixels)	Ours(pixels)
High-rise Dense Area	6.2	4.5	3.8
Dynamic Occlusion	7.1	5.3	4.2
Lighting Variation	6.8	4.9	3.9

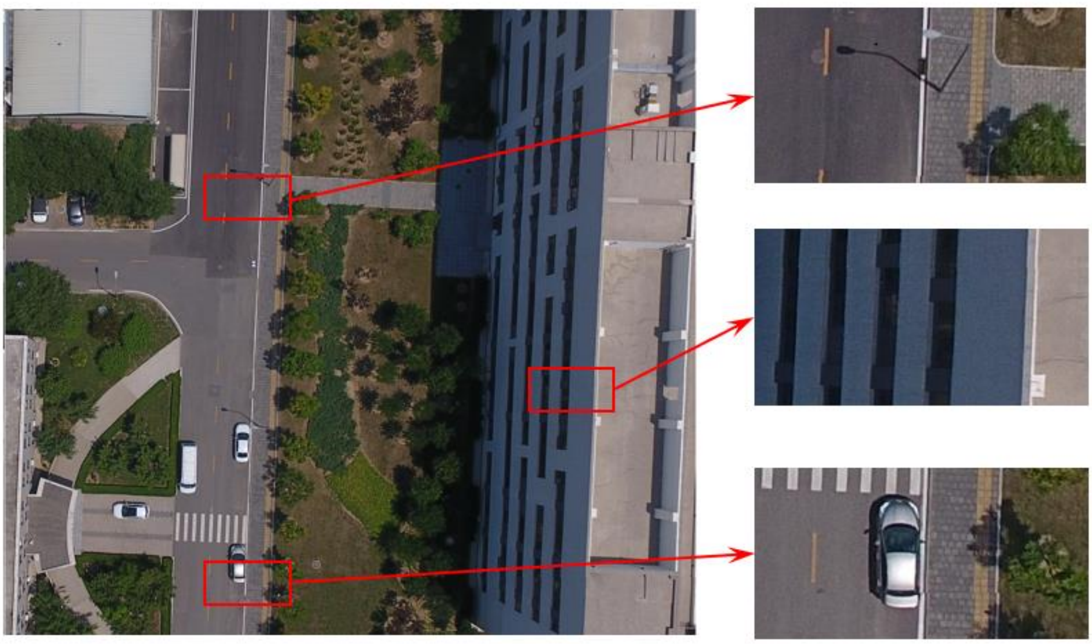
Table 2. Performance Comparison Under Different Scenes

The experimental results demonstrate the robustness of the proposed method in complex scenes. In the high-rise dense area, the RMSE is 3.8 pixels, a 38.7% reduction compared to ORB (6.2). When facing dynamic interference from pedestrians and vehicles in urban environments, the method shows significant improvement in adaptability. In the dynamic occlusion scene, the RMSE stabilizes at 4.2 pixels, outperforming SIFT (5.3).

https://doi.org/10.5194/isprs-archives-XLVIII-G-2025-1123-2025



(a)



(b)

Figure 3. Stitching Results in High-Rise Dense Area:(a) ORB: Misalignment in road and building stitching. (b) Ours: Seamless road continuity with precise building contour alignment.

5. Conclusion

In this study, we address the UAV image stitching problem in complex urban environments by proposing an improved ORB algorithm based on Gaussian scale-space and dynamic grid division, combined with a global geometric optimization strategy to enhance stitching accuracy and robustness. Experimental results show that the method achieves a real-time processing speed of 45ms/frame on low-performance hardware platforms, with a feature point distribution uniformity index improved to 0.89, stitching accuracy reaching 4.0 pixels, and stable performance even in dynamic occlusion and lighting variation scenes. This method provides an efficient and lightweight solution for UAV image stitching in urban environments through multi-scale feature detection, dynamic grid division, and geometric consistency optimization. It has broad application value in urban planning, disaster assessment, and intelligent traffic management. Future research will focus on real-time processing optimization for dynamic scenes, exploring lightweight deep learning hybrid methods, and building end-to-end feature extraction and matching frameworks with lightweight networks to enhance robustness in complex dynamic scenarios.

Additionally, integrating multimodal data such as infrared and LiDAR will improve the algorithm's adaptability in extreme environments like low light and fog. Furthermore, edge computing acceleration strategies will be explored, utilizing model quantization, knowledge distillation, and other techniques to optimize the algorithm's deployment efficiency on embedded platforms. We will also develop temporal consistency constraints and incremental optimization strategies for UAV video stream data to reduce accumulated errors and achieve high-precision reconstruction in long-term sequences. The findings of this study provide important support for the engineering application of UAV image processing technology, and will further promote its practical implementation in fields such as smart cities in the future.

References

Alahi, A., Ortiz, R., & Vandergheynst, P. (2012, June). Freak: Fast retina keypoint. In 2012 IEEE conference on computer vision and pattern recognition (pp. 510-517). Ieee.

Bay, H., Tuytelaars, T., & Van Gool, L. (2006). Surf: Speeded up robust features. In Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I 9 (pp. 404-417). Springer Berlin Heidelberg.

Calonder, M., Lepetit, V., Strecha, C., & Fua, P. (2010). Brief: Binary robust independent elementary features. In Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part IV 11 (pp. 778-792). Springer Berlin Heidelberg.

Chilukuri, P. K., Padala, P., Padala, P., Desanamukula, V. S., & Pvgd, P. R. (2021). L, r-stitch unit: encoder-decoder-CNN based image-mosaicing mechanism for stitching non-homogeneous image sequences. IEEE Access, 9, 16761-16782. DOI: 10.1109/ACCESS.2021.3052474

DeTone, D., Malisiewicz, T., & Rabinovich, A. (2016). Deep image homography estimation. arXiv preprint arXiv:1606.03798.

Howard, A. G. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

Li, J., Chen, Y., & Mu, A. (2024). Research on Unsupervised Fe ature Point Prediction Algorithm for Multigrid Image Stitching. Symmetry, 16(8), 1064. https://doi.org/10.3390/sym16081064.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. International journal of computer vision, 60, 91-110.

Mur-Artal, R., & Tardós, J. D. (2017). Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. IEEE transactions on robotics, 33(5), 1255-1262.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144).

Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011, November). ORB: An efficient alternative to SIFT or SURF. In 2011 International conference on computer vision (pp. 2564-2571). Ieee.

Zhang, G., Qin, D., Yang, J., Yan, M., Tang, H., Bie, H., & Ma, L. (2022). UAV Low-Altitude Aerial Image Stitching Based on Semantic Segmentation and ORB Algorithm for Urban Traffi c. *Remote Sensing*, *14*(23), 6013. https://doi.org/10.3390/rs14236013