Image Stitching based on Feature Detection and Extraction: An Analysis

Nan Zhao

Computer Science, University of Liverpool, Liverpool, L693BX, United Kingdom
zhaonan1118@yzpc.edu.cn

Abstract. Image stitching is a popular research area in the fields of computer vision and computer graphics. The feature points of images provide crucial information for this process. The accurate extraction of these features is essential to minimize misalignment and defects in the final stitched image. This paper extensively discusses the application of deep neural network-based feature detection algorithms in image stitching. Initially, it introduces several commonly used feature detection algorithms such as scale invariant feature transform (SIFT), speed up robust feature (SURF), before delving into deep learning-based methods, specifically focusing on convolutional neural network-based feature detectors. The paper elaborates on the operational mechanisms of these algorithms in image stitching, emphasizing the efficient extraction of key feature points from images and the subsequent matching of these points for seamless stitching. Moreover, a comparative analysis of the advantages and limitations of these modern methods relative to conventional approaches is provided. The paper concludes with a concise overview of the current challenges encountered in the realm of image stitching, including issues related to feature extraction and matching in complex scenes, as well as performance and efficiency constraints when dealing with large-scale image datasets. In summary, the paper offers insights into the advancements in image stitching techniques and highlights potential areas for future research and development.

Keyword: Image Stitching, Features Detection, Features Extraction.

1 Introduction

Image stitching is a crucial and demanding task in the field of computer vision. It has seen rapid advancements over the past few decades, enabling the creation of panoramic images with a wider field of view by combining images captured from different viewing positions. This technique finds applications in various fields such as biology, medicine, surveillance video, autonomous driving, and virtual reality [1]. Image stitching is a process that combines multiple images to create a seamless, high-resolution panorama or photograph. When there is overlap between two images, they are merged to form a single frame [1]. The alignment of images in a set is crucial for successful stitching, as lens distortion and parallax errors can impact the final result. Using advanced alignment methods, such as those found in large camera systems, helps to overcome these

Y. Wang (ed.), *Proceedings of the 2024 2nd International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI 2024)*, Advances in Computer Science Research 115,
https://doi.org/10.2991/978-94-6463-540-9_72

obstacles and produce accurate panoramic images. Panoramic image mosaic involves stitching together numerous images to create a composite image with a wider field of view than a standard camera can capture [1]. These image sets may consist of multiple digital images taken at different times, from various sensors, and at different angles of the same scene. Attention to detail and precise alignment are essential to ensure the seamless merging of images and the creation of a flawless panoramic image [1].

The process of image stitching involves the use of image alignment algorithms and image blending techniques. These algorithms are essential for creating mosaics, summaries, and stabilized videos by establishing correspondence between images with varying degrees of overlap [1]. Panorama creation algorithms ensure seamless alignment by employing registration algorithms to blend images smoothly and address potential issues like parallax, scene movement, and exposure changes that can result in blurriness or ghosting. Image alignment algorithms are typically used to generate seamless panoramic images from handheld cameras. The alignment process begins by determining a mathematical model that correlates the pixel coordinates of two images [2]. Translation angles are easily estimated, with the most straightforward method involving exhaustive testing of all possible alignments [2]. Image blending involves adjusting calculations during calibration, such as image remapping, color correction, and merging images to create large, seamless images that minimize visible seams. This process aims to reduce intensity differences between overlapping pixels in two images [3]. Two commonly used image blending methods are Alpha Feathering and Gaussian Pyramid. Alpha blending is effective when images have well-aligned pixels with only intensity variations, while the Gaussian Pyramid involves merging images from different frequency bands and filtering them together [3].

This paper focuses on the description of feature detection algorithms and feature extraction based on deep neural methods in image stitching. Finally, the paper also has a simple description of existing challenges in image stitching. The framework of image stitching methods is shown in figure 1, which includes feature detection methods and deep network methods.

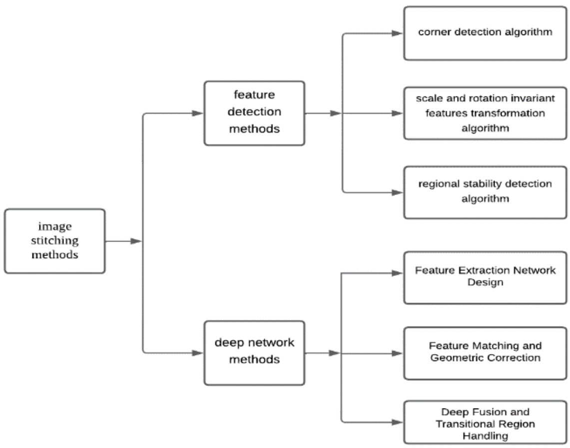

Fig. 1 The framework of image stitching methods

2 Feature Detection Methods

2.1 Corner Detection Algorithm

Corner detection algorithms mainly focus on finding points of significant changes in the image, which are called corner points [2]. The corner detection algorithm uses the local gradient information of the image and the contrast of pixels to detect corner points [2]. The corner points are usually important feature points in the image, which are suitable for target tracking and image matching [3].

The corner detection algorithm is an important class of methods in the field of image processing, which aims to identify corner points in images [3]. Corner points in images have significant changes in local areas, which are usually represented as key feature points in the image. The corner detection algorithm is widely used in image matching and target tracking [3]. Classic corner detection algorithms have Good Features to Track (GFTT), Features from Accelerated Segment Test (FAST), and Solenoidal Tracker at

Rhic Detector [3]. Firstly, the GFTT algorithm is a corner detection method based on local gradient changes, which uses a feature response function to evaluate the gradient change in the area of the pixel to determine the corner point. The advantage of the GFTT algorithm is simplicity and robustness, and it stably detects corner points in different scenarios [4]. Secondly, the FAST algorithm is a fast corner detection method, which quickly detects the circular area around the pixel to determine whether a corner point exists. The FAST algorithm detects corners by comparing the brightness values of pixels, thereby achieving high-speed corner detection. Although the FAST algorithm performs well on speed, it has challenges in dealing with noise and different scales [4]. In addition, the Solenoidal Tracker at Relativistic Heavy Ion Collider is a large-scale, high-energy nuclear physics international collaboration established based on the Solenoidal Tracker At RHIC (STAR), which relies on the Relativistic Heavy Ion Collider (RHIC) in the United States. STAR detector is known as the Center Surround Extrema feature detector (CenSurE), which is a corner detection method that combines scale invariance and stability. STAR detector uses a special feature detection strategy that combines brightness differences between pixels and gradient information around pixels to determine corner points. STAR detector provides stability and adapts to image scale changes, which performs well in a variety of applications [4]. GFTT, FAST, and STAR detector have respective characteristics. GFTT focuses on simplicity and stability, FAST emphasizes speed and efficiency, and STAR detector takes into account both scale invariance and stability. In practical applications, an appropriate corner detection algorithm can be selected based on task requirements and scene characteristics [4]. In this paper, the principles and characteristics of other types of image feature extraction algorithms are analyzed in detail.

2.2 Scale and Rotation Invariant Features Transformation Algorithm

Scale-invariant feature algorithms are a class of algorithms dedicated to extracting features from images under varying scales and rotation conditions. Their key goal is to extract image feature points that remain stable under these conditions, to cope with changes in images under different conditions. The algorithms utilize local features and scale space information to describe features in images [5].

Scale Invariant Feature Transform (SIFT) is one of the classical algorithms. It detects key points in images and calculates their local feature descriptors to achieve scale and rotation invariance. It finds wide application in tasks like object recognition and image registration. Speed up robust feature (SURF), based on SIFT, is an accelerated version of feature extraction algorithms. It utilizes fast box filters for acceleration while maintaining good scale and rotation invariance, making it advantageous in real-time systems. Additionally, Binary Robust Invariant Scalable (BRISK) combines binary descriptors with rotation invariance, showing high computational efficiency and robustness in real-time image processing [5]. Oriented FAST and Rotated BRIEF (ORB) combines FAST key point detection with Binary Robust Independent Elementary Features (BRIEF) descriptors, demonstrating efficiency and good scale invariance, suitable for image processing tasks on real-time systems and mobile devices [5].

In recent years, improvements in scale-invariant feature algorithms have shown diversity and innovation, aiming to enhance algorithm performance and applicability. One significant improvement targets the issue of mismatching during feature point matching. In practical applications, factors like image noise, occlusion, or viewpoint changes may lead to mismatching of feature points, affecting the accuracy of the algorithm. Therefore, some researchers propose improved matching strategies, such as geometric constraint-based matching algorithms or machine learning-based matching methods [5]. These methods identify correct matching points and mismatching points by introducing geometric constraints or utilizing machine learning algorithms, thus improving matching accuracy and robustness [5].

On the other hand, research efforts have also been devoted to improving the robustness and generalization ability of algorithms in specific scenarios. For instance, in the problem of image feature extraction under low-light conditions, some scholars propose adaptive feature extraction methods with strong adaptability. These methods can extract stable feature points even under significant changes in lighting conditions by appropriately enhancing images or adjusting parameters of feature extraction algorithms, thereby enhancing the robustness of the algorithm. Moreover, studies are focusing on improving feature descriptors, such as RootSIFT and Fast Retina Key-point (FREAK). These feature descriptors, while maintaining scale invariance, further improve the accuracy and robustness of feature matching, providing stronger support for the performance of the algorithm in practical applications [5].

These specific improvement methods and technologies have brought significant performance improvements and broader application possibilities for scale-invariant feature algorithms.

2.3 Regional Stability Detection Algorithm

The main objective of region stability detection algorithms is to detect stable regions or region boundaries in images to extract representative region features. These algorithms typically utilize the intrinsic properties of image regions or differences between pixels to detect these stable regions [5]. For example, a classic region stability detection algorithm is Maximally Stable Extremal Regions (MSER) which extracts regions with maximal stability from images. The MSER algorithm extracts features by finding stable extremal regions in images that remain stable across different scales. Similarly, there is the Maximally Stable Diverticulum (MSD) algorithm, another stability detection algorithm based on extremal regions, which effectively extracts stable image regions [5].

The characteristics of these region stability detection algorithms lie in their ability to extract stable image regions, which are crucial for tasks such as image segmentation and object detection. However, when faced with complex backgrounds or noise interference, these algorithms may be affected to some extent, resulting in less accurate detection results. Therefore, improving the robustness and stability of these algorithms is one of the important research directions [5].

To address this issue, recent research has proposed new methods and techniques. For example, some scholars have proposed improved region stability detection algorithms. The DeepMSER algorithm utilizes deep learning techniques for feature extraction and classification of stable regions, thereby enhancing the accuracy and robustness of stable region detection. Additionally, AdaptiveMSD combines adaptive feature extraction and machine learning techniques to effectively address challenges in complex scenes. At the same time, there are also studies focusing on optimizing the performance of algorithms by accelerating the algorithm's execution speed through parallelization and optimization techniques, thereby enhancing the ability to process large-scale data. These improvements and optimizations provide important support for the further development of region stability detection algorithms, enabling them to better address challenges and demands in complex scenes [5].

3 Deep Network Methods

This section delves into various aspects of image stitching algorithms based on deep learning, including feature extraction network design, feature matching and geometric correction, deep fusion, and transitional region handling.

3.1 Feature Extraction Network Design

Based on Convolutional Neural Network (CNN) Architecture. In the task of image stitching, selecting a suitable CNN architecture is crucial for feature extraction. For instance, some researchers opt for ResNet-50 as the feature extraction network. By fine-tuning ResNet-50 and leveraging its pre-training on large-scale image datasets, researchers can adapt it to the stitching task by lowering the learning rate. The deep-level feature maps of ResNet-50 provide rich semantic information, thereby enhancing the accuracy and stability of stitching. Additionally, EfficientNet-B3 is another common choice [6,7]. The EfficientNet series models excel in lightweight and efficiency, thus achieving good performance with lower computational costs in image stitching tasks. Moreover, some researchers explore novel architectures like Adaptive CNNs. These adaptive CNNs adjust the size and structure of convolutional kernels adaptively based on input images, thus achieving better feature extraction performance across different stitching scenarios [8].

Based on Feature Pyramid Network (FPN) Design. FPN plays a crucial role in addressing the feature extraction needs at different scales. For example, some researchers apply FPN to construct feature pyramids. By extracting features at different scales from various layers, FPN enhances the algorithm's adaptability to scale changes in image stitching. This network structure enables simultaneous attention to both local image details and global semantic information, thereby improving the quality and stability of stitching. Additionally, Path Aggregation Network (PANet) is another common choice. PANet with its top-down and bottom-up pathways can extract rich multi-scale features across multiple levels, achieving good performance in image stitching. Furthermore, some researchers explore feature pyramid network designs combined with self-attention mechanisms. This network can adaptively adjust the weights of different scale features to enhance the representational power and stitching performance of the feature pyramid [8].

3.2 Feature Matching and Geometric Correction

Matching Network Design. Accurate feature matching requires the design of dedicated matching networks. For instance, some researchers adopt Siamese network-based feature matching methods. The Siamese network structure learns the similarity between image features, thus achieving accurate feature matching. Its shared parameter structure enables the network to fully utilize training data for learning, thereby improving matching accuracy and robustness [8]. Additionally, some researchers explore methods based on optical flow estimation and dense feature matching. Optical flow estimation captures pixel displacement information in images, aiding feature matching in dynamic scenes. Dense feature matching strategies achieve more precise and stable matching results by uniformly sampling feature points on images [8].

Geometric Correction Module Design. Geometric correction modules are responsible for geometric transformations of images to achieve alignment and stitching. For example, some researchers adopt geometric correction methods based on the Random sample consensus (RANSAC) algorithm [8]. The RANSAC algorithm effectively estimates geometric transformation relationships between images, thus achieving precise image alignment and stitching. Additionally, some researchers propose learning-based geometric correction methods [8]. This method learns geometric transformation relationships between images through an end-to-end learning framework and optimizes correction parameters by minimizing reprojection errors, thus achieving accurate image stitching [8]. Moreover, geometric correction methods based on image segmentation are also common choices. By segmenting images into different regions and performing individual geometric correction for each region, this method improves corrections accuracy and stability [8].

3.3 Deep Fusion and Transitional Region Handling

Deep Fusion Network Construction. Deep fusion networks smooth transitional regions between stitched images and ensure depth consistency [9]. For example, some researchers employ generative adversarial network (GAN) based deep fusion networks. The networks combine the mechanisms of generative adversarial networks to learn transitional relationships between images and generate realistic stitching results [9]. Additionally, some researchers utilize structures like CycleGAN for deep fusion. CycleGAN learns mapping relationships between images and applies them to stitching tasks, achieving natural and continuous deep fusion effects. Furthermore, deep fusion network designs based on self-attention mechanisms are also receiving attention. This network adaptively adjusts the weights of different regions to enhance the quality and effectiveness of deep fusion. Overall, deep fusion networks play a crucial role in image stitching, effectively improving the visual quality and continuity of stitched images[9].

Transitional Region Generation Methods. In the process of image stitching, transitional region handling is a critical step that requires appropriate methods to seamlessly integrate transitional regions with stitched images [9]. For example, some researchers adopt transitional region generation methods based on image editing. They use image to edit techniques to post-process stitched images, generating smooth and natural transitional regions, further enhancing the visual quality of stitched images [9]. Additionally, some researchers explore methods utilizing GANs to generate transitional regions [9]. By training GANs to learn the distribution of real images, realistic transitional regions can be generated, thus improving the continuity and realism of stitched images [8]. Transitional region generation methods play a crucial role in achieving natural and continuous image stitching effects and are indispensable in image stitching algorithms [8,9].

4 Existing Problems of Image Stitching Algorithms

- **Feature matching problem.**

The keys of image stitching find similar feature points between different images. However, factors such as lighting, viewing angle, and deformation in the images may cause feature points to be inaccurate or unmatched. For example, lens distortion causes the shape and scale of the image to be inconsistent in the taking images, which can affect the accuracy of stitching [10].

- **Processing of overlapping areas.**

Image stitching usually involves overlapping areas between multiple images. Overlapping areas that achieve smooth transitions is a challenge, overlapping areas cause ambiguity in data interpretation. For example, the overlapping of two objects might challenge how accurately to ensure the boundaries of each object [10].

- **Movement and deformation.**

Movement of objects and deformation of the scene exist during the image stitching process, discontinuous and distorted stitching results may occur. The situation is caused by handheld shooting, vibration, and unstable camera settings. Wide-angle lenses often cause distortion in the stitching images, such as barrel distortion and pincushion distortion, which cause misalignment. Objects of a scene and the camera exist at different distances, and the related position of images is changed, which causes ghosting or discontinuities in the image stitching [10].

- **Error propagation.**

Each image has errors that may accumulate and expand during the stitching process, which decreases the quality of the final stitched result [10]. Errors in feature detection and extraction make images incorrectly align, which produces misalignment artifacts in the output of image stitching. Feature detection and extraction are inaccurate, subsequent stages of the image-stitching process are affected by the error. Estimation errors of the homography matrix that represents the geometric transformation of images can lead to deformations of the final image stitching. Inaccurate estimation of homography may propagate errors to subsequent stages, which affects the overall quality of the image stitching [10].

- **Occlusion problem.**

Objects occluding in the image, such as trees, buildings, etc., cause missing and discontinuous in the procession stitching [10]. Objects movement of frames and capturing scenes with complex foreground elements, which are difficult to accurately align and stitch images.

The relative position of image objects changes because of the movement of the viewpoint, which causes parallax. The parallax makes misalignments and inconsistencies in image stitching, especially along the edges of the object. Depth differences in a scene make obvious seams in image stitching, especially objects of foreground hide elements of the background. The depth discontinuities are important to create a visually pleasing image composition.

5 Conclusion

This paper extensively delves into various feature detection algorithms used in image stitching and discusses the current challenges in this field. Aside from examining common algorithms like SIFT, SURF, and ORB, it also focuses on deep learning-based methods, particularly those utilizing convolutional neural networks for feature detection. By elucidating how these algorithms play a crucial role in extracting and matching key feature points for image stitching, the potential of deep learning in image processing becomes evident. While acknowledging the significant contribution of these algorithms, it is crucial to underscore the challenges that still persist. Feature matching issues, handling overlapping areas, and addressing object movement and scene deformation are key obstacles that need to be overcome. Improving matching accuracy to identify more feature points is essential to enhance stitching quality. Moreover, developing sophisticated algorithms for seamless transitions and natural stitching effects in overlapping areas is imperative. Efforts are also required to accurately capture and rectify object movement and scene distortion to prevent discontinuities and distortions in the final stitched images.

To tackle these challenges, future research can focus on several key areas. This includes optimizing feature matching algorithms for increased accuracy and efficiency, exploring novel techniques for handling overlapping regions for better stitching results, and studying recognition and processing methods for object movement and scene deformation to enhance stability and accuracy in stitching. Furthermore, integrating the strengths of deep learning with traditional approaches could further elevate the quality and efficiency of image stitching processes.

In conclusion, while advancements in image stitching algorithms driven by deep learning have been remarkable, significant challenges remain. Through continuous research and innovation, it is anticipated that image stitching technology will continue to progress, offering more precise, efficient, and natural solutions for practical applications.

References

1. Nie L, Lin C, Liao K, et al. Unsupervised deep image stitching: Reconstructing stitched features to images. IEEE Transactions on Image Processing, 30: 6184-6197 (2021).
2. Mehrotra, R., Nichani, S., & Ranganathan, N. Corner detection. Pattern recognition, 23 (11), 1223-1233 (1990).
3. Chen, J., Zou, L. H., Zhang, J., & Dou, L. H.The Comparison and Application of Corner Detection Algorithms. Journal of multimedia, 4(6) (2009).
4. Danielsson, J., Jägemar, M., Behnam, M., Sjödin, M., & Seceleanu, T. Measurement-based evaluation of data-parallelism for opencv feature-detection algorithms. In 2018 IEEE 42nd Annual Computer Software and Applications Conference 1, 701-710, (2018).
5. Tao, Y., Xia, Y., Xu, T., & Chi, X. Research Progress of the Scale Invariant Feature Transform (SIFT) Descriptors. Journal of Convergence Information Technology, 5 (1), 116-121 (2010).
6. Lou, J., Zhu, W., Wang, H., & Ren, M. Small target detection combining regional stability and saliency in a color image. Multimedia Tools and Applications, 76, 14781-14798, (2017).
7. Yang, A., Yang, X., Wu, W., Liu, H., & Zhuansun, Y. Research on feature extraction of tumor image based on convolutional neural network. IEEE access, 7, 24204-24213, (2019).
8. Yao, R., Liu, C., Zhang, L., & Peng, P. Unsupervised anomaly detection using variational auto-encoder based feature extraction. In 2019 IEEE International Conference on Prognostics and Health Management, 1-7, (2019).
9. Yu, S., Chen, H., Garcia Reyes, E. B., & Poh, N. Gaitgan: Invariant gait feature extraction using generative adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 30-37), (2017).
10. Szeliski, R. Image alignment and stitching: A tutorial. Foundations and Trends® in Computer Graphics and Vision, 2(1), 1-104, (2007).