A Survey on Feature-Based and Deep Image Stitching

Sima Soltanpour[a] and Chris Joslin[b]
School of Information Technology, Carleton University, 1125 Colonel By Dr, Ottawa, Canada
{simasoltanpour, chrisjolain}@cunet.carleton.ca

Keywords: Image Stitching, Parallax, Homography, Deep Learning, Survey.

Abstract: Image stitching is a process of merging multiple images with overlapped parts to generate a wide-view image. There are many applications in a variety of fields for image stitching such as 360-degree cameras, virtual reality, photography, sports broadcasting, video surveillance, street view, and entertainment. Image stitching methods are divided into feature-based and deep learning algorithms. Feature-based stitching methods rely heavily on accurate localization and distribution of hand-crafted features. One of the main challenges related to these methods is handling parallax problems. In this survey, we categorize feature-based methods in terms of parallax tolerance which has not been discovered in the existing survey papers. Moreover, considerable research efforts have been dedicated to applying deep learning methods for image stitching. In this way, we also comprehensively review and compare the different types of deep learning methods for image stitching and categorize them into three different groups including deep homography, deep features, and deep end-to-end framework.

1 INTRODUCTION

Image stitching is a popular research area that has been well studied in the past decades and it has numerous applications. Multimedia (Gaddam et al., 2016), medical imaging (Li et al., 2017a), motion detection (Sreyas et al., 2012), video surveillance (Wang et al., 2017), and virtual reality (Kim et al., 2019) are some of the important areas that image stitching is creating remarkable impacts. Image stitching is defined as a process to combine multiple images captured from different viewing positions with overlapping fields of view (FOV) to produce a panoramic image with a wider field of view.

Several surveys have been published in image stitching during the last decades. A survey by Shashank et al. (Shashank et al., 2014) focuses on the introduction and general summarization of the stitching algorithms. Adel et al. (Adel et al., 2014) divides methods to stitch two or multiple images into two general approaches: direct and feature-based techniques. Image pixel intensities are compared using direct methods. These methods are computationally expensive and are not robust to lighting changes and large motions. While feature-based methods aim to find a relationship between the images by extracting distinguished features. The last approach is more robust against scene movement compared with the direct method. Image mosaicing techniques are discussed by Ghosh et al. (Ghosh and Kaabouch, 2016). Algorithms are classified based on registration and blending along with their advantages and disadvantages. There are three major steps for traditional image stitching methods. Feature detection and matching, image registration and warping, and image blending. In the first step, the corresponding relationships between the original images are calculated. Then image registration is applied to estimate a transformation model from the target image plane to the reference image plane. Usually, a homography transformation defined by a 3 by 3 matrix is used for warping. Since an image contains objects with different depth levels, applying only a global homography produces some artifacts and ghosting effects. To reduce unpleasant seams or projective distortions, a blending algorithm is applied. A survey by Wei et al. (Wei et al., 2019) reviews image/ video stitching algorithms and classifies them into two categories including pixel-based methods and feature-based methods. In pixel-based methods, image information such as intensity, color, gradient, and geometry is used to register multiple images. In contrast to pixel-based methods, feature-based methods are defined by estimating a 2D motion model with sparse feature points. A survey on feature-

[a] https://orcid.org/0000-0002-2131-8902
[b] https://orcid.org/0000-0002-6728-2722

Soltanpour, S. and Joslin, C.
A Survey on Feature-Based and Deep Image Stitching.
DOI: 10.5220/0013368500003912

In *Proceedings of the 20th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2025) - Volume 3: VISAPP*, pages 777-788
ISBN: 978-989-758-728-3; ISSN: 2184-4321

based methods is presented in (Wang and Yang, 2020) by describing and evaluating image registration and seam removal techniques. A review on panoramic image stitching techniques is presented by (Abbadi et al., 2021) for feature-based methods. A comparative study on feature-based techniques is presented by Megha and Rajkumar (Megha and Rajkumar, 2022). They analyze stitching methods in two different categories including spatial-domain and frequency domain methods. Recently, A comparative analysis of feature detectors and descriptors for image stitching is presented by (Sharma et al., 2023). Also, recent reviews by Fu et al. (Fu et al., 2023) and (Yan et al., 2023) focus on image stitching techniques based on camera types. However, in this survey paper, we focus on a main challenge related to feature-based methods which is parallax and also we propose three different categories for deep learning based methods which have not been presented in the existing review papers.

Image stitching algorithms encounter some challenges including wide baseline, real time applications, low texture, and large parallax. The most challenging task for image stitching is handling the parallax problem. To this end, this survey categorizes traditional methods into two different categories: parallax intolerant methods and parallax tolerant methods. To the best of our knowledge, this is the first time that a survey focuses on the traditional stitching methods in terms of parallax problems.

Algorithms of deep learning to solve geometric computer vision problems have been applied in various tasks such as deep neural network based homography computation (DeTone et al., 2016), and deep homography mixture (Yan et al., 2023). These algorithms have outperformed traditional methods. Inspired by that improvement, recent papers in image stitching focus on developing deep learning based methods. In this way, algorithms based on the deep neural network for image stitching are comprehensively discussed in this survey paper. We divide these algorithms into three different categories including deep features, deep homography, and deep framework which have not been discovered in the existing reviews. Figure 1 illustrates our classification of image stitching algorithms.

The remainder of this paper is organized as follows. Feature-based image stitching methods are reviewed and categorized along with their strength and weakness in Section 2. Section 3 provides a comprehensive survey of deep learning based methods for image stitching, including methods categorization and description. Challenges and potential future research directions are discussed in Section 4. Finally, Section 5 concludes this paper.

2 FEATURE-BASED METHODS

Feature-based methods rely on keypoints extraction in each image using local invariant hand-crafted features. Then, feature matching is applied to establish feature correspondences between the two sets of keypoints. There are many approaches to detect keypoints and describe feature vectors such as SIFT (Scale Invariant Feature Transform) (Lowe, 2004), SURF (Speeded Up Robust Features) (Bay et al., 2006), ORB (Oriented FAST and Rotated BRIEF) (Rublee et al., 2011), and etc.

The performance of the stitching algorithms can be influenced by parallax. This survey categorizes feature-based methods into two different categories in terms of parallax handling. Since feature-based methods were discussed in some survey papers, we provide this categorization on the most famous and latest methods in this area.

2.1 Parallax Intolerant Methods

Parametric transforms such as homography, affine, and perspective are very popular for traditional image stitching among researchers. They can produce correct stitched images while the scenes are planar or camera motion between source frames is parallax-free. These methods are useful to source frames taken from the same physical location. There is no difference between overlapping and non-overlapping regions while applying these algorithms. We divide these algorithms into three different groups including global homography, mesh-based local homography, and hybrid methods.

2.1.1 Global Homography

These methods apply a single transformation matrix to align the entire image. They work effectively where depth variations are minimal. So they cannot handle large parallax scenarios. In some methods, an optimal global transformation is estimated for the whole input image. AutoStitch (Brown and Lowe, 2007) is a representative example proposed by Brown et al. in this field. A global homography transformation is estimated to align images from the same plane. To handle complicated applications containing multiple planes, a Dual Homography Warping (DHW) is proposed by Gao et al. (Gao et al., 2011). Two predominate planes including a distant back plane and a ground plane define a panoramic scene. SIFT keypoints are clustered in two groups. For each group, a global homography is estimated as distant plane homography and grand plane homography. A weight map is calculated to combine two homographies. Recently, Li et al. (Li

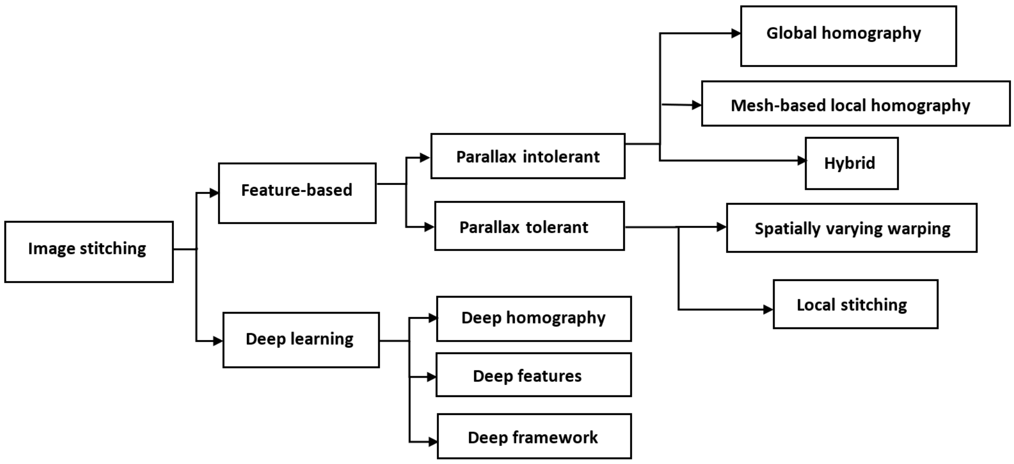


Figure 1: Classification of image stitching algorithms.

et al., 2024) proposed a Local-Peak Scale-Invariant Feature Transform to compute the homography matrix.

2.1.2 Mesh-Based Local Homography

To solve alignment errors related to the global transformation, a local adaptive field is constructed using points correspondences between input images. Mesh-based local homography methods divide the image into smaller regions and apply local homography related to each part. One of the famous methods in this area is as-projective-as-possible (APAP) (Zaragoza et al., 2013). A local homography for each image patch is computed to reach an accurate local alignment. Inspired by the Moving Least Squares (MLS) method (Alexa et al., 2003), Moving Direct Linear Transformation (DLT) is introduced for warping. This algorithm works by considering the global homography while the camera translation is zero. However, local homography for images captured under camera translation is helpful in reaching a more accurate alignment. Figure 2 demonstrates image stitching using the APAP method. A mesh-based framework is introduced by Zhang et al. (Zhang et al., 2016) to optimize alignment. They propose a scale-preserving term for image alignment optimization with local perspective correction. A seam-cut model is applied to reduce visual artifacts that are caused by misalignment. A seamless stitching method based on multiple homography matrix is proposed in (Tengfeng, 2018). A-KAZE feature point detection algorithm (Alcantarilla and Solutions, 2011) is applied in this work. A projection model is determined using Direct Linear Transform (DLT) on 25 by 25 blocks. To obtain the seamless stitching of multiple images, the Min-Cut/Max-Flow of the edge detection operator and Laplacian multiresolution fusion algorithm is added to this work.

2.1.3 Hybrid Methods

To improve stitching performance some methods combine global and local homography approaches. A spatial combination of a projective transformation and a similarity transformation is proposed by Chang et al. (Chang et al., 2014). The method is called shape-preserving half-projective (SPHP). This method preserves images' original perspective and aligns them globally. A combination of local homography and global similarity transformations is applied in Adaptive As-Natural-As-Possible (ANAP) image stitching (Lin et al., 2015). The preservative distortion in non-overlapping areas is handled by homography linearization and slightly changing to the global similarity.

The previous methods such as SPHP and ANAP, apply global similarity to handle projective distortion. The main problem related to those methods is perspective distortion in the non-overlapping areas. A quasi-homography warp is proposed by Li et al. (Li et al., 2017c) to balance the perspective distortion against the projective distortion. The above-discussed methods apply SIFT keypoints for image stitching. Error in finding matched keypoints results in distortion errors for larger input image set. To improve the quality of panorama, an algorithm based on the A-KAZE feature (Alcantarilla and Solutions, 2011) is proposed by Qu et al. (Qu et al., 2019) which uses a binary tree for image stitching. The input image set is considered the leaf node set of the binary tree and the bottom-up approach is applied to construct a complete binary tree. The final stitched image is obtained from the root node image. This method enhances the accuracy of feature point detection compared to SIFT keypoints and improves the quality of the stitching

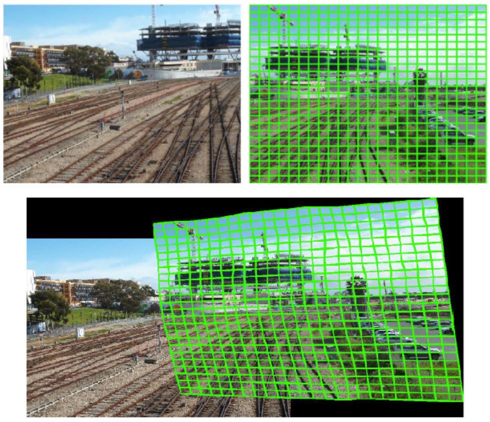


Figure 2: Image stitching using APAP warping (Zaragoza et al., 2013). The input images are related to views under different rotation and translation.

process. Furthermore, the panoramic distortion is improved by their automatic image straightening model. They applied a bi-directional KNN (K Nearest Neighbor) matching strategy. The binary tree is also applied by Qu et al. (Qu et al., 2020) along with an estimated overlapping area to solve time-consuming problems during unordered image stitching. Another image stitching method using the A-KAZE features is proposed by Sharma et al. (Sharma and Jain, 2020). Their algorithm is divided into the following steps: feature points detection and descriptors by A-KAZE, finding matching pairs by KNN algorithm, removing false matched points using MSAC (M-estimator SAmple Consensus) algorithm, and finding homography matrix from correct matches.

2.2 Parallax Tolerant Methods

Image stitching under parallax is still a challenging task. It is very critical to generate high-resolution stitched images and videos in various applications such as surveillance (Gaddam et al., 2016) and virtual reality (Anderson et al., 2016). Methods discussed in the previous section cannot handle significant depth variations and camera translation. In this section, we review traditional methods that apply advanced transformations and warping techniques to address parallax. Figure 3 illustrates a comparison of parallax intolerant and parallax tolerant methods for stitching two input images captured under different viewpoints. We divide different stitching algorithms that can handle parallax into two groups including spatially-varying warping and local stitching methods.

Figure 3: Comparison of parallax intolerant and tolerant methods (Li et al., 2017b). Left: two images of railtracks database (Zaragoza et al., 2013), Centre: Undesirable artifacts by applying global alignments, Right: Elimination of misalignment.

2.2.1 Spatially-Varying Warping Methods

These methods utilize adaptive transformations that vary across the image to align features effectively, addressing depth and perspective variations. In (Lin et al., 2011) a global affine transformation is replaced with a smoothly varying affine stitching over the entire coordinate frame. Every point has its affine parameter. In this way, the affine stitching field is very smooth and can be extended over the non-overlapping areas. This method is suitable to address small parallax. A dual-feature warping based on the sparse feature matches and line correspondences is proposed by Li et al. (Li et al., 2015). Geometrical and structural information can be obtained from line segments specifically in low texture conditions. A structure-preserving warping is proposed by Lin et al. (Lin et al., 2016) to deal with challenging image stitching with large parallax. They propose a seam-guided local alignment (SEAGULL) scheme which performs seam-guided feature re-weighting to look for a good local alignment iteratively. A local similarity refinement (LSR) approach is proposed in (Li et al., 2018) to handle parallax in combination with SPHP (Chang et al., 2014). Deconvolution is applied to enhance geometry matching. SPHP is used to address distortion in non-overlapping areas. Adaptive pixel warping is another method to handle large parallax proposed by Lee and Sim (Lee and Sim, 2018). The epipolar geometry is applied to warp multiple foreground objects, distant backgrounds, and ground planes adaptively. Warping is performed by an off-plane pixel in a target image to a reference image using its ground plane pixel (GPP). Optimal GPPs are estimated for the foreground objects by applying the spatio-temporal feature matches. Energy minimizing is applied to refine the initially obtained GPPs. Figure 4 shows two inputs captured under different viewpoints and a comparison of stitching using homography, APAP (Zaragoza et al., 2013), and adaptive pixel warping. As the figure illustrates parallax adaptive stitching could address the parallax challenge and generate a stitched image without artifacts compared to homography-based and APAP methods. A recent

Table 1: Parallax intolerant image stitching methods.

Algorithm	Descriptor	Blending algorithm	Strength	Weakness
Global Homography methods				
AutoStitch (Brown and Lowe, 2007)	SIFT	Multi-band blending	Fully automated approach	Single axis of rotation
DHW (Gao et al., 2011)	SIFT	Alpha blending (Shum and Szeliski, 2001)	Two planes	Multiple planes
LP-SIFT (Li et al., 2024)	LP-SIFT	-	Fast	Multiple planes
Mesh-based local homography methods				
APAP (Zaragoza et al., 2013)	SIFT	Pyramid blending (Szeliski, 2006)	Global and local transformations	Distortion in non-overlapped areas
Multi-viewpoint (Zhang et al., 2016)	SIFT	Average blending	Wide-baseline images	Local distortion
Multiple homography matrix (Tengfeng, 2018)	A-KAZE	Laplacian fusion	Less fracture	Multiple planes
Hybrid methods				
SPHP (Chang et al., 2014)	SIFT	Linear blending	Projective transformation of the overlapping regions into the non-overlapping regions	Sensitive to parameter selection, Multiple distinct planes
ANAP (Lin et al., 2015)	SIFT	-	Robust to parameter selection	Large motion
Quasi-homography (Li et al., 2017c)	SIFT	Seam-cutting (Boykov et al., 2001)	Parameter free, To handle perspective distortion	Different planes, Time consuming
Binary tree(Qu et al., 2019)	A-KAZE	-	Less distortion	Multiple planes
Binary tree and (Qu et al., 2020) an estimated overlapping area	A-KAZE	-	Time efficiency, Unordered images	Brightness difference
A-KAZE-based (Sharma and Jain, 2020)	A-KAZE	Weighted average	Less distortion	Computational cost

seam-based image stitching is proposed by (Zhang et al., 2025) that applies dense flow estimation generated by Local Feature Matching with Transformers (LoFTR) (Sun et al., 2021). A spatial smooth warping model is estimated by weighting point pairs.

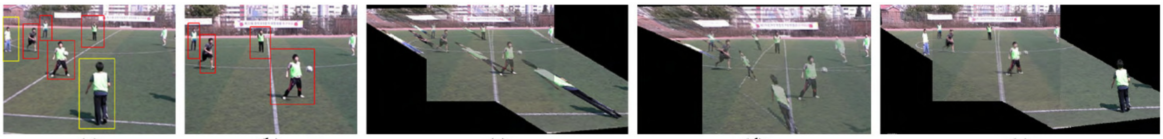
2.2.2 Local Stitching Methods

These methods divide the image into smaller regions and apply localized transformations to achieve accurate alignment in the presence of parallax. As we can see from the previous section, spatially varying warping can handle parallax better than homography for image stitching. However, it still is not robust under large parallax. A method based on the local alignment is proposed by Zhang et al. (Zhang and Liu, 2014) for optimal stitching. A hybrid alignment model is adopted to combine homography and content-preserving warping (CPW) to handle parallax and prevent objectionable local distortion. Figure 5 shows the stitching pipeline of the method presented in (Zhang and Liu, 2014).

A line-guided local warping method with a global similarity constraint is proposed by Xiang et al. (Xiang et al., 2018) for image stitching. A stitching algorithm inspired by Zaragoza's local projection (Zaragoza et al., 2013) is proposed by Li et al. (Li et al., 2017b) called robust elastic warping. The quality of image alignment can be effectively improved by local homography estimation and projection correction. However, handling local misalignment from local warping is a challenging task. To handle this problem, Li et al. (Li et al., 2019a) proposed a deviation-corrected warping with global similarity constraints called As-Aligned-As-Possible (AAAP) image stitching. First outliers are removed from matched points to correct pixel offsets. Then local homography and global similarity are used for warping. For more improvement, a three-dimensional mesh interpolation model is adopted and a local projection deviation of the local warping model is described. Two single-perspective warps have been proposed by Liao and Li (Liao and Li, 2019) including a parametric warp and mesh-based warp. Wen et al. (Wen et al., 2022) proposed a hybrid warping model based on local and global homography to handle large parallax. They estimated the homographies of different depth regions by dividing matching features into multiple layers. A seam-based parallax tolerant image stitching method is proposed by Zhang et al. (Zhang et al., 2024). They introduce an iterative algorithm to select inliers and solve the mesh warping model. The quaternion representation of the color image is applied in a recent stitching method by Li et al. (Li and Zhou, 2024). This method presents the joint optimization strategy of local alignment and seamline iteratively. Recently, a method (Zhang and Xiu, 2024) based on the human visual system and SIFT algorithm is proposed for image stitching. Dynamic programming is applied to find the optimal seamline. A semantic-based method (Zhang and Jiang, 2025) based on mesh optimization has been proposed recently to preserve global and local structures of the stitched images.

2.3 Summary

We divided feature-based methods into two different categories based on their robustness under parallax. The main problem related to these algorithms is they are not suitable for real applications and are evaluated on standard datasets. The presence of valid feature point pairs is critical for any feature-based stitching method. However, in practical applications, it is very common to have low-texture areas in the captured images due to different capturing scenarios which causes incorrect feature point matching and stitching errors. Tables 1 and 2 show a detailed comparison of feature-based methods discussed in this section. As the tables show most of the methods apply SIFT keypoints



(a) (b) (c) (d) (e)

Figure 4: Image stitching (Lee and Sim, 2018). (a)-(b) Input images with large parallax, (c) Using homography based warping, (d) APAP (Zaragoza et al., 2013), (e) Parallax adaptive stitching.

Table 2: Parallax tolerant image stitching methods.

Algorithm	Descriptor	Blending algorithm	Strength	Weakness
Spatially-varying warping methods				
SVA (Lin et al., 2011)	SIFT	Poisson blending with optimal seam (Chan and Efros, 2007)	To handle most kinds of motions	Affine Inconsistency
Dual-feature warping (Li et al., 2015)	SIFT	Linear blending	To handle low-texture images	Large parallax
SEAGULL (Lin et al., 2016)	SIFT	-	Large parallax	Computational cost
LSR (Li et al., 2018)	SIFT	Linear blending	Robust under noise	Large parallax
Adaptive warping (Lee and Sim, 2018)	SIFT	Average blending	Large parallax	Moving cameras
Seam-based warping (Zhang et al., 2025)	LOFTR	-	large parallax	Real time applications
Local stitching methods				
CPW (Zhang and Liu, 2014)	SIFT	Multi-band blending (Burt and Adelson, 1983)	Time efficient	Depends on salient structures
REW (Li et al., 2017b)	SIFT	Pyramid blending (Burt and Adelson, 1983)	Flexibility and computational efficiency	Occlusion handling
Line-guided local warping(Xiang et al., 2018)	Line	Intensity average	To handle low-textured images	Unstable under broken lines
AAAP (Li et al., 2019a)	SIFT	Linear-based pixel smoothing model	To handle local misalignment	Computational cost
Single-perspective warps(Liao and Li, 2019)	SIFT	Linear blending	Naturalness	Large parallax
Hybrid warping(Wen et al., 2022)	SIFT	-	Large parallax	Computational cost
Seam-based (Zhang et al., 2024)	SIFT, LOFTR	Linear blending	Large parallax	Relies on features, Illumination
AQCIS (Li and Zhou, 2024)	Quaternion representation of the color image	Poisson blending	Large parallax, low textures	Projective distortions
HVS(Zhang and Xiu, 2024)	SIFT	-	To handle brightness difference and contrast	Large parallax
Semantics-preserving(Zhang and Jiang, 2025)	SIFT	-	Large parallax	Significant disparities, and limited texture

for the feature extraction step. A few applied the A-KAZE descriptor. Some of these methods (Brown and Lowe, 2007), (Gao et al., 2011), and (Zaragoza et al., 2013) that rely on a transformation method like homography cannot distinguish between overlapping and non-overlapping regions to handle distortion. Local and hybrid methods (Li et al., 2019a), (Wen et al., 2022), (Zhang et al., 2024), and (Li and Zhou, 2024) are possible solutions to handle misalignment and artifacts in feature-based image stitching methods. These algorithms rely on extracted feature points, so insufficient points may result in misalignment. One possible solution can be applying deep learning based feature methods like LoFTR (Sun et al., 2021) used by (Zhang et al., 2024) and (Zhang et al., 2025).

3 DEEP LEARNING BASED METHODS

Image stitching using deep learning is still in development compared to traditional methods. A survey paper by Fu et al. (Fu et al., 2023) provides a review of some deep learning based methods for image stitching. They did not classify these methods and only discussed them in their survey. Another survey by Yan et al. (Yan et al., 2023) focuses on deep learning methods based on camera types. However, we divide deep stitching methods into three different categories. The first category is related to the algorithms that estimate homography using deep learning. The second group relies on detecting features using deep learning. As discussed in the previous section, feature extraction is one of the important steps in image stitching algorithms. We review recent works that apply deep learning to improve feature extraction performance for stitching task. The learned features are more flexible than hand-crafted features like SIFT, A-KAZE, etc. to take advantage of the image information. Finally, the main focus of the third category is on performing all steps of the image stitching within a single deep model.

3.1 Deep Homography Based Methods

Traditional homography estimation methods heavily depend on sparse feature correspondences resulting in poor robustness in low-textured images. To enhance the performance of homography estimation a learning model called homographyNet is proposed by Detone et al. (DeTone et al., 2016) for the first time. The homographyNet is a deep CNN that estimates

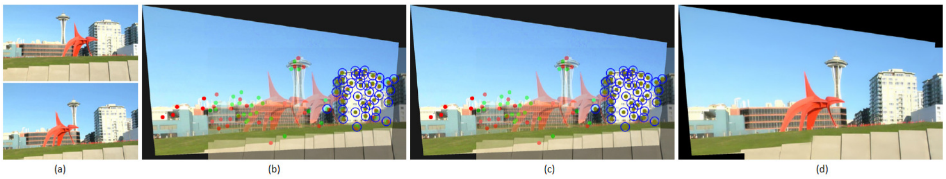


Figure 5: Local stitching pipeline (Zhang and Liu, 2014). (a) Input images with large parallax, (b) Optimal local homography alignment, (c) Locally alignment refinement using content preserving, (d) Final stitching result after seam cutting and multi-band blending.

the homography between two images. This method does not need feature detection and correspondence. Homography calculation steps and all parameters are trained in an end-to-end scheme using a large labeled dataset. Deep neural network homography estimation is also proposed in (Nguyen et al., 2018) and (Wang et al., 2018). Since the above methods apply relatively simple network architecture, the stitching result using those homography estimation methods results in distortion and artifacts. They only work well for images with small displacement and large overlap. Inspired by homographyNet, a deep homography based on semantic alignment network (Rocco et al., 2017) is proposed by Zhao et al. (Zhao et al., 2021) for stitching images with small parallax. The architecture of this network is shown in figure 6. As the figure illustrates first a rough homography estimation based on the low-resolution feature map is computed. Then a refinement is performed according to the feature maps with progressively increased resolution. Moreover, a new loss function is also proposed in their work to take image content into consideration. To handle images with low overlap rates a context correlation layer (CCL) is designed by et al. (Nie et al., 2021a). The long range correlation within feature maps can be effectively captured and applied in a learning framework. Multi-grid homography from global to local is proposed to handle depth varying images with parallax. They introduced a depth-aware shape-preserved loss, to add depth perception capability to their network. A content-aware unsupervised deep homography estimation is proposed by Liu et al. (Liu et al., 2022). The algorithm learns an outlier mask to only select reliable regions for homography estimation. A novel triplet loss is customized for their network to achieve unsupervised training. A Recurrent Elastic Warp (REwarp) is proposed by Kim et al. (Kim et al., 2024) to estimate homography and thin-plane spline using two recurrent neural networks. This approach provides an elastic image alignment for parallax tolerant image stitching.

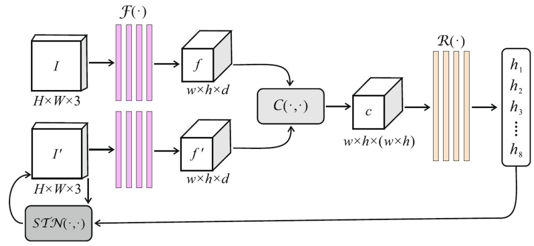


Figure 6: The architecture of the network proposed in (Zhao et al., 2021).

3.2 Methods Based on Deep Feature Extraction

This section describes stitching methods that design CNNs to extract feature points. A method is proposed by Hoang et al. (Hoang et al., 2020) that directly estimates feature locations between two images by maximizing an image patch similarity metric. They collect a large dataset containing high resolution images and videos from natural tourism scenes to train the network and evaluation step. Inspired by convolution neural attack, a method based on the semantic feature extraction is proposed by et al. (Shi et al., 2020) for image mosaic. A neural network is used to compute and quantify the semantic features of each pixel in an image. The flow of image mosaic based on feature semantic extraction is demonstrated in figure 7. A multi-scenario stitching algorithm for autonomous driving application is proposed by Wang et al. (Wang et al., 2020) that applies convolutional neural networks to extract features. This feature extraction network contains two paths including a dimensionality-reduced feature extraction and a precisely located symmetrical decoder. Image stitching using matched dominant semantic planar regions extracted with deep Convolutional Neural Network (CNN) is proposed by Li et al. (Li et al., 2021). A mesh-based optimization method is used to stitch images. A method proposed by Du et al. (Du et al., 2022) employs deep learning-based edge detection to represent geometric structures. They introduce a GEometric Structure preserving (GES) energy which is added into the Global Similarity Prior (GSP) stitching

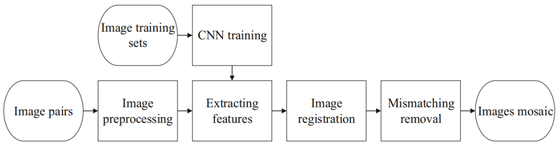


Figure 7: Flow of image mosaic proposed in (Shi et al., 2020).

model called GES-GSP for natural image stitching.

3.3 Methods Based on Complete Deep Learning Framework

There are stitching methods that perform all steps of image stitching using a deep CNN model (Lai et al., 2019) and (Shen et al., 2019). Lai et al. (Lai et al., 2019) proposed a video stitching network that warps input images gradually to obtain the output. This method is designed for a specific stitching situation such as a fixed camera position. Shen at al. (Shen et al., 2019) proposed a panorama generative adversarial network (PanoGAN) for real-time image stitching. However, they cannot handle distortions related to the depth differences since they do not use depth information. End-to-end networks are proposed to stitch images from fixed view in (Li et al., 2019c). This algorithm is designed for surveillance videos application. However, it is not suitable for arbitrary view point image stitching. End-to-end image stitching network is proposed by Song et al. (Song et al., 2021) using multi-homography estimation. Since this method estimates multiple homographies to cover the depth differences in the overlapped images, it is robust under parallax. Multiple homographies generate global warping maps which can be adjusted by local displacement maps. Warping maps are applied to warp an input image multiple times and weight maps create the final result. A deep image stitching framework to handle large parallax is proposed by Kweon et al. (Kweon et al., 2021). The framework contains two modules including the Pixel-wise Warping Module (PWM) and Stitched Image Generating Module (SIGMo). An optical flow estimation model is employed by PWM to relocate the pixels of the target based on the obtained warp field. Warped images and reference image are fed into SIGMo for blending and distortion removal. An edge Guided Composition Network (EGCNet) is designed by Dai et al. (Dai et al., 2021) for the composition stage in image stitching. A whole composition stage is considered as an image blending problem. Two pre-registered images are inputs to the network to predict blending weights during training. A perceptual edge branch is built to improve the performance by providing edge guidance. An unsupervised deep image stitching consisting of two stages is proposed by Nie et al. (Nie et al., 2021b). First, an ablation-based loss is designed for an unsupervised homography network that can handle large baseline scenes. Second, an unsupervised image reconstruction network is designed to eliminate the artifacts from features to pixels. The pipeline of this method is shown in figure 8. As the figure illustrates input images are warped in the course image alignment stage using single homography. The warped images are applied to reconstruct the stitched image from feature to pixel. Song et al. (Song et al., 2022) proposed a training end-to-end model to take some fisheye images and make a panorama image. Nie et al. (Nie et al., 2023) proposed a parallax-tolerant unsupervised image stitching by presenting a seam-inspired composition and a simple iterative warping method. A deep image stitching framework is proposed by Kweon et al. (Kweon et al., 2023) to exploit pixel-wise warp filed to handle large parallax issues. The deep learning-based framework consists of a Pixel-wise Warping Module (PWM) and a Stitched Image Generating Module (SIGMo). A deep network is proposed by Tchinda et al. (Nghonda Tchinda et al., 2023) for semi-supervised image stitching. A fast unsupervised image stitching model is proposed by Ni et al. (Ni et al., 2024). An adaptive feature extraction module (FEM) for deformation with an unsupervised alignment network is proposed. Also, they proposed a stitching restoration network to remove the redundant sampling operations. Jiang et al. (Jiang et al., 2024) proposed a framework consisting of pyramid-based residual homography estimation and global-aware reconstruction modules for infrared and visible image-based multispectral image stitching. Infrared images are less affected by environmental factors and can be applied to improve the accuracy of the stitching framework.

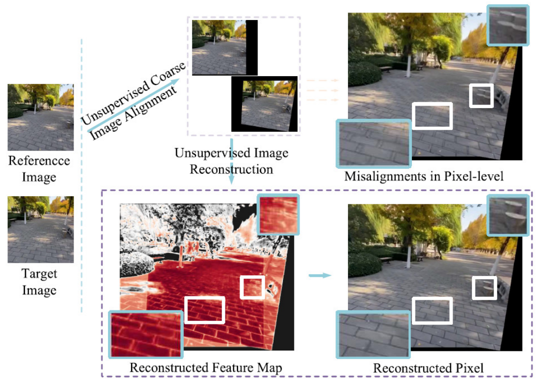


Figure 8: Unsupervised deep image stitching pipeline proposed in (Nie et al., 2021b).

Table 3: Deep image stitching methods.

Algorithm	Training dataset	Strength	Weakness
Deep homography			
(Zhao et al., 2021)	Places365 (Zhou et al., 2017a)	Computationally efficient	Parallax
(Nie et al., 2021a)	MS-COCO (DeTone et al., 2016) UDIS-D (Nie et al., 2021b)	Parallax	Grids number is limited by the network architecture and data size
(Liu et al., 2022)	Their own DB	Unsupervised learning, handle depth disparity issue	Large Parallax
(Kim et al., 2024)	UDIS-D	Real-time applications	Large parallax
Deep features			
(Hoang et al., 2020)	Their own DB	Efficient features	Parallax
(Shi et al., 2020)	ImageNet2012	No need to shallow features	Parallax
(Wang et al., 2020)	COCO	Fixed view	Real-time performance
(Li et al., 2021)	ADE20k(Zhou et al., 2017b)	Exploit regional information	Textureless areas
GES-GSP(Du et al., 2022)	50 datasets	Structure preserving	Spatial constraints
End-to-End framework			
(Lai et al., 2019)	CARLA simulator	Strong parallax	Requires camera calibration
(Li et al., 2019c)	CROSS dataset (Li et al., 2019b)	Fixed view	Arbitrary View
PanoGAN (Shen et al., 2019)	Their own DB	Real-time	Parallax
(Song et al., 2021)	CARLA simulator (Dosovitskiy et al., 2017)	Parallax	Real application
(Kweon et al., 2021)	Their own DB	Parallax	Low resolution
EGCNet(Dai et al., 2021)	RISD	Parallax,Object movement	Needs Brightness adjustment
(Nie et al., 2021b)	Their own DB	Unsupervised learning	Large parallax
(Song et al., 2022)	Their own DB, CROSS dataset	Fixed view	Real-world applications
(Nie et al., 2023)	UDIS-D	Unsupervised learning, Large parallax	Needs GPUs to be efficient
(Kweon et al., 2023)	PDIS (thier own DB), UDIS	Large parallax	Real-time applications
(Nghonda Tchinda et al., 2023)	Their own DB	Unstructured camera arrays	Large parallax
(Ni et al., 2024)	MS-COCO, UDIS-D	Unsupervised learning, Fast	Complex scenes, Real-time applications
(Jiang et al., 2024)	Their own DB, RoadScene	A board spectrum of parallax and illumination	Repeated structures, symmetrical scenes

3.4 Summary

Deep image stitching is described in three different categories in this section. Table 3 summarizes the discussed methods with their advantages and weaknesses. Training databases are listed in the table to provide some ideas regarding large set databases. Methods based on deep homography calculation still suffer from misalignment and distortion problem. Deep feature-based methods outperform conventional methods like SIFT. However, warping and blending tasks are still important in these methods. Finally, the end-to-end deep framework is a possible solution to all challenges related to image stitching and is still under development.

4 CHALLENGES AND FUTURE WORKS

Image stitching is an interesting research area that has many applications and has been studied in the last decades. Most proposed methods work well under natural baseline and small parallax. Some methods provide more efficient algorithms to handle wide baselines and large parallax. However, developing more sophisticated methods that can handle most practical application scenarios still needs research. The main challenge related to the discussed methods is handling large parallax for real applications. The number of object classes in a scene, low-textured images, depth estimation for dynamic scenes, and computational cost are other challenges that are important to be considered for practical and real-time applications. Recently, end-to-end deep image stitching networks attracted researchers' attention while large datasets covering practical applications' challenges are one crucial requirement to train those networks effectively. Therefore, developing image stitching methods that can perform well under diverse scenes handling low texture images, wide baseline, and large parallax with low computational cost is a future research direction in this field.

5 CONCLUSION

This survey provides a comprehensive discussion on image stitching methods from hand-crafted features to deep stitching methods. We divide feature-based methods into two categories including parallax intolerant, and parallax tolerant. Deep learning-based methods are categorized in three groups including deep homography, deep features, and end-to-end framework. Besides the brief summary and explanation of the main steps of each method, we summarize their weakness and strengths and provide future research directions.

ACKNOWLEDGEMENTS

This work was supported by the NSERC Discovery Grant #RGPIN-2021-04248.

REFERENCES

Abbadi, N. K. E., Al Hassani, S. A., and Abdulkhaleq, A. H. (2021). A review over panoramic image stitching techniques. In *Journal of Physics: Conference Series*, volume 1999, page 012115. IOP Publishing.

Adel, E., Elmogy, M., and Elbakry, H. (2014). Image stitching based on feature extraction techniques: a survey. *International Journal of Computer Applications*, 99(6):1–8.

Alcantarilla, P. F. and Solutions, T. (2011). Fast explicit diffusion for accelerated features in nonlinear scale spaces. *IEEE Trans. Patt. Anal. Mach. Intell*, 34(7):1281–1298.

Alexa, M., Behr, J., Cohen-Or, D., Fleishman, S., Levin, D., and Silva, C. T. (2003). Computing and rendering point set surfaces. *IEEE Transactions on visualization and computer graphics*, 9(1):3–15.

Anderson, R., Gallup, D., Barron, J. T., Kontkanen, J., Snavely, N., Hernández, C., Agarwal, S., and Seitz, S. M. (2016). Jump: virtual reality video. *ACM Transactions on Graphics (TOG)*, 35(6):1–13.

Bay, H., Tuytelaars, T., and Gool, L. V. (2006). Surf: Speeded up robust features. In *European conference on computer vision*, pages 404–417. Springer.

Boykov, Y., Veksler, O., and Zabih, R. (2001). Fast approximate energy minimization via graph cuts. *IEEE Transactions on pattern analysis and machine intelligence*, 23(11):1222–1239.

Brown, M. and Lowe, D. G. (2007). Automatic panoramic image stitching using invariant features. *International journal of computer vision*, 74(1):59–73.

Burt, P. J. and Adelson, E. H. (1983). A multiresolution spline with application to image mosaics. *ACM Transactions on Graphics (TOG)*, 2(4):217–236.

Chan, L. H. and Efros, A. A. (2007). Automatic generation of an infinite panorama. *Technical Report*.

Chang, C.-H., Sato, Y., and Chuang, Y.-Y. (2014). Shape-preserving half-projective warps for image stitching. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3254–3261.

Dai, Q., Fang, F., Li, J., Zhang, G., and Zhou, A. (2021). Edge-guided composition network for image stitching. *Pattern Recognition*, 118:108019.

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

Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., and Koltun, V. (2017). Carla: An open urban driving simulator. In *Conference on robot learning*, pages 1–16. PMLR.

Du, P., Ning, J., Cui, J., Huang, S., Wang, X., and Wang, J. (2022). Geometric structure preserving warp for natural image stitching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3688–3696.

Fu, M., Liang, H., Zhu, C., Dong, Z., Sun, R., Yue, Y., and Yang, Y. (2023). Image stitching techniques applied to plane or 3-d models: a review. *IEEE Sensors Journal*, 23(8):8060–8079.

Gaddam, V. R., Riegler, M., Eg, R., Griwodz, C., and Halvorsen, P. (2016). Tiling in interactive panoramic video: Approaches and evaluation. *IEEE Transactions on Multimedia*, 18(9):1819–1831.

Gao, J., Kim, S. J., and Brown, M. S. (2011). Constructing image panoramas using dual-homography warping. In *CVPR 2011*, pages 49–56. IEEE.

Ghosh, D. and Kaabouch, N. (2016). A survey on image mosaicing techniques. *Journal of Visual Communication and Image Representation*, 34:1–11.

Hoang, V.-D., Tran, D.-P., Nhu, N. G., Pham, V.-H., et al. (2020). Deep feature extraction for panoramic image stitching. In *Asian Conference on Intelligent Information and Database Systems*, pages 141–151. Springer.

Jiang, Z., Zhang, Z., Liu, J., Fan, X., and Liu, R. (2024). Multispectral image stitching via global-aware quadrature pyramid regression. *IEEE Transactions on Image Processing*, 33:4288–4302.

Kim, H. G., Lim, H.-T., and Ro, Y. M. (2019). Deep virtual reality image quality assessment with human perception guider for omnidirectional image. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(4):917–928.

Kim, M., Lee, Y., Han, W. K., and Jin, K. H. (2024). Learning residual elastic warps for image stitching under dirichlet boundary condition. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 4016–4024.

Kweon, H., Kim, H., Kang, Y., Yoon, Y., Jeong, W., and Yoon, K.-J. (2021). Pixel-wise deep image stitching. *arXiv preprint arXiv:2112.06171*.

Kweon, H., Kim, H., Kang, Y., Yoon, Y., Jeong, W., and Yoon, K.-J. (2023). Pixel-wise warping for deep image stitching. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 1196–1204.

Lai, W.-S., Gallo, O., Gu, J., Sun, D., Yang, M.-H., and Kautz, J. (2019). Video stitching for linear camera arrays. *arXiv preprint arXiv:1907.13622*.

Lee, K.-Y. and Sim, J.-Y. (2018). Stitching for multi-view videos with large parallax based on adaptive pixel warping. *IEEE Access*, 6:26904–26917.

Li, A., Guo, J., and Guo, Y. (2021). Image stitching based on semantic planar region consensus. *IEEE Transactions on Image Processing*, 30:5545–5558.

Li, D., He, Q., Liu, C., and Yu, H. (2017a). Medical image stitching using parallel sift detection and transformation fitting by particle swarm optimization. *Journal of Medical Imaging and Health Informatics*, 7(6):1139–1148.

Li, H., Wang, L., Zhao, T., and Zhao, W. (2024). Local-peak scale-invariant feature transform for fast and random image stitching. *Sensors*, 24(17):5759.

Li, J., Jiang, P., Song, S., Xia, H., and Jiang, M. (2019a). As-aligned-as-possible image stitching based on deviation-corrected warping with global similarity constraints. *IEEE Access*, 7:156603–156611.

Li, J., Wang, Z., Lai, S., Zhai, Y., and Zhang, M. (2017b). Parallax-tolerant image stitching based on robust elastic warping. *IEEE Transactions on multimedia*, 20(7):1672–1687.

Li, J., Yu, K., Zhao, Y., Zhang, Y., and Xu, L. (2019b). Cross-reference stitching quality assessment for 360 omnidirectional images. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 2360–2368.

Li, J., Zhao, Y., Ye, W., Yu, K., and Ge, S. (2019c). Attentive deep stitching and quality assessment for 360 omnidirectional images. *IEEE Journal of Selected Topics in Signal Processing*, 14(1):209–221.

Li, J. and Zhou, Y. (2024). Automatic quaternion-domain color image stitching. *IEEE Transactions on Image Processing*, 33:1299–1312.

Li, N., Xu, Y., and Wang, C. (2017c). Quasi-homography warps in image stitching. *IEEE transactions on multimedia*, 20(6):1365–1375.

Li, S., Yuan, L., Sun, J., and Quan, L. (2015). Dual-feature warping-based motion model estimation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4283–4291.

Li, W., Jin, C.-B., Liu, M., Kim, H., and Cui, X. (2018). Local similarity refinement of shape-preserved warping for parallax-tolerant image stitching. *IET Image Processing*, 12(5):661–668.

Liao, T. and Li, N. (2019). Single-perspective warps in natural image stitching. *IEEE transactions on image processing*, 29:724–735.

Lin, C.-C., Pankanti, S. U., Natesan Ramamurthy, K., and Aravkin, A. Y. (2015). Adaptive as-natural-as-possible image stitching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1155–1163.

Lin, K., Jiang, N., Cheong, L.-F., Do, M., and Lu, J. (2016). Seagull: Seam-guided local alignment for parallax-tolerant image stitching. In *European conference on computer vision*, pages 370–385. Springer.

Lin, W.-Y., Liu, S., Matsushita, Y., Ng, T.-T., and Cheong, L.-F. (2011). Smoothly varying affine stitching. In *CVPR 2011*, pages 345–352. IEEE.

Liu, S., Ye, N., Wang, C., Zhang, J., Jia, L., Luo, K., Wang, J., and Sun, J. (2022). Content-aware unsupervised deep homography estimation and its extensions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):2849–2863.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110.

Megha, V. and Rajkumar, K. (2022). A comparative study on different image stitching techniques. *Int. J. Eng. Trends Technol*, 70(4):44–58.

Nghonda Tchinda, E., Panoff, M. K., Tchuinkou Kwadjo, D., and Bobda, C. (2023). Semi-supervised image stitching from unstructured camera arrays. *Sensors*, 23(23):9481.

Nguyen, T., Chen, S. W., Shivakumar, S. S., Taylor, C. J., and Kumar, V. (2018). Unsupervised deep homography: A fast and robust homography estimation model. *IEEE Robotics and Automation Letters*, 3(3):2346–2353.

Ni, J., Li, Y., Ke, C., Zhang, Z., Cao, W., and Yang, S. X. (2024). A fast unsupervised image stitching model based on homography estimation. *IEEE Sensors Journal*, 24:29452—29467.

Nie, L., Lin, C., Liao, K., Liu, S., and Zhao, Y. (2021a). Depth-aware multi-grid deep homography estimation with contextual correlation. *IEEE transactions on circuits and systems for video technology*, 32(7):4460–4472.

Nie, L., Lin, C., Liao, K., Liu, S., and Zhao, Y. (2021b). Unsupervised deep image stitching: Reconstructing stitched features to images. *IEEE Transactions on Image Processing*, 30:6184–6197.

Nie, L., Lin, C., Liao, K., Liu, S., and Zhao, Y. (2023). Parallax-tolerant unsupervised deep image stitching. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 7399–7408.

Qu, Z., Li, J., Bao, K.-H., and Si, Z.-C. (2020). An unordered image stitching method based on binary tree and estimated overlapping area. *IEEE Transactions on Image Processing*, 29:6734–6744.

Qu, Z., Wei, X.-M., and Chen, S.-Q. (2019). An algorithm of image mosaic based on binary tree and eliminating distortion error. *Plos one*, 14(1):e0210354.

Rocco, I., Arandjelovic, R., and Sivic, J. (2017). Convolutional neural network architecture for geometric matching. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6148–6157.

Rublee, E., Rabaud, V., Konolige, K., and Bradski, G. (2011). Orb: An efficient alternative to sift or surf. In *2011 International conference on computer vision*, pages 2564–2571. IEEE.

Sharma, S. K. and Jain, K. (2020). Image stitching using akaze features. *Journal of the Indian Society of Remote Sensing*, 48(10):1389–1401.

Sharma, S. K., Jain, K., and Shukla, A. K. (2023). A comparative analysis of feature detectors and descriptors for image stitching. *Applied Sciences*, 13(10):6015.

Shashank, K., SivaChaitanya, N., Manikanta, G., Balaji, C. N., and Murthy, V. (2014). A survey and review over image alignment and stitching methods. the. In *International Journal of Electronics & Communication Technology*, volume 5, pages 50–52.

Shen, C., Ji, X., and Miao, C. (2019). Real-time image stitching with convolutional neural networks. In *2019 IEEE International Conference on Real-time Computing and Robotics (RCAR)*, pages 192–197. IEEE.

Shi, Z., Li, H., Cao, Q., Ren, H., and Fan, B. (2020). An image mosaic method based on convolutional neural network semantic features extraction. *Journal of Signal Processing Systems*, 92(4):435–444.

Shum, H.-Y. and Szeliski, R. (2001). Construction of panoramic image mosaics with global and local alignment. In *Panoramic vision*, pages 227–268. Springer.

Song, D.-Y., Lee, G., Lee, H., Um, G.-M., and Cho, D. (2022). Weakly-supervised stitching network for real-world panoramic image generation. In *European Conference on Computer Vision*, pages 54–71. Springer.

Song, D.-Y., Um, G.-M., Lee, H. K., and Cho, D. (2021). End-to-end image stitching network via multi-homography estimation. *IEEE Signal Processing Letters*, 28:763–767.

Sreyas, S. T., Kumar, J., and Pandey, S. (2012). Real time mosaicing and change detection system. In *Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing*, pages 1–8.

Sun, J., Shen, Z., Wang, Y., Bao, H., and Zhou, X. (2021). Loftr: Detector-free local feature matching with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8922–8931.

Szeliski, R. (2006). Image alignment and stitching. In *Handbook of mathematical models in computer vision*, pages 273–292. Springer.

Tengfeng, W. (2018). Seamless stitching of panoramic image based on multiple homography matrix. In *2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, pages 2403–2407. IEEE.

Wang, L., Yu, W., and Li, B. (2020). Multi-scenes image stitching based on autonomous driving. In *2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, volume 1, pages 694–698. IEEE.

Wang, M., Cheng, B., and Yuen, C. (2017). Joint coding-transmission optimization for a video surveillance system with multiple cameras. *IEEE Transactions on Multimedia*, 20(3):620–633.

Wang, X., Wang, C., Bai, X., Liu, Y., and Zhou, J. (2018). Deep homography estimation with pairwise invertibility constraint. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, pages 204–214. Springer.

Wang, Z. and Yang, Z. (2020). Review on image-stitching techniques. *Multimedia Systems*, 26(4):413–430.

Wei, L., Zhong, Z., Lang, C., and Yi, Z. (2019). A survey on image and video stitching. *Virtual Reality & Intelligent Hardware*, 1(1):55–83.

Wen, S., Wang, X., Zhang, W., Wang, G., Huang, M., and Yu, B. (2022). Structure preservation and seam optimization for parallax-tolerant image stitching. *IEEE Access*, 10:78713–78725.

Xiang, T.-Z., Xia, G.-S., Bai, X., and Zhang, L. (2018). Image stitching by line-guided local warping with global similarity constraint. *Pattern recognition*, 83:481–497.

Yan, W., Tan, R. T., Zeng, B., and Liu, S. (2023). Deep homography mixture for single image rolling shutter correction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9868–9877.

Zaragoza, J., Chin, T.-J., Brown, M. S., and Suter, D. (2013). As-projective-as-possible image stitching with moving dlt. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2339–2346.

Zhang, F. and Liu, F. (2014). Parallax-tolerant image stitching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3262–3269.

Zhang, G., He, Y., Chen, W., Jia, J., and Bao, H. (2016). Multi-viewpoint panorama construction with wide-baseline images. *IEEE Transactions on Image Processing*, 25(7):3099–3111.

Zhang, J. and Jiang, N. (2025). Image stitching algorithm based on semantics-preserving warps. *Signal, Image and Video Processing*, 19(1):1–11.

Zhang, J. and Xiu, Y. (2024). Image stitching based on human visual system and sift algorithm. *The Visual Computer*, 40(1):427–439.

Zhang, Z., He, J., Shen, M., Shi, J., and Yang, X. (2024). Multimodel fore-/background alignment for seam-based parallax-tolerant image stitching. *Computer Vision and Image Understanding*, 240:103912.

Zhang, Z., He, J., Shen, M., and Yang, X. (2025). Seam estimation based on dense matching for parallax-tolerant image stitching. *Computer Vision and Image Understanding*, 250:104219.

Zhao, Q., Ma, Y., Zhu, C., Yao, C., Feng, B., and Dai, F. (2021). Image stitching via deep homography estimation. *Neurocomputing*, 450:219–229.

Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., and Torralba, A. (2017a). Places: A 10 million image database for scene recognition. *IEEE transactions on pattern analysis and machine intelligence*, 40(6):1452–1464.

Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., and Torralba, A. (2017b). Scene parsing through ade20k dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 633–641.