

Seamless-Through-Breaking: Rethinking Image Stitching for Optimal Alignment

KuanYan Chen^[0009–0005–4360–4079], Atik Garg^[0000–0001–6444–9550], and
Yu-Shuen Wang^[0000–0003–2550–2990] yushuen@cs.nycu.edu.tw

National Yang Ming Chiao Tung University, Hsinchu, Taiwan

Abstract. In this paper, we introduce a novel concept called *seamless-through-breaking* to tackle the challenges that arise in image stitching. Conventional methods attempt to maintain warping continuity while stitching two images together to avoid visible breaks in the final output. However, we propose that content alignment and warping continuity are mutually exclusive, especially when a significant depth gap exists between the foreground and the background. To solve this issue, we use optical flow to warp the source image into the target image’s domain, which allows the creation of holes in the source image. Considering that optical flow estimators are trained on synthetic data, we fine-tune the estimator using real-world data to improve its accuracy in practical applications. Once the images are aligned within the same domain, we fill these holes with content from the target image. Additionally, as no optical flow estimators are perfect, directly copying pixels from the target image to fill the holes may create visual artifacts. To avoid this issue, we apply an image inpainting technique around the edges of the holes to smooth out alignment discrepancies, ensuring that the stitched image looks as natural as if it were captured in one shot.

Keywords: Image Stitching · Image Warping · Optical Flow

1 Introduction

Image stitching is a widely used technique with diverse applications, ranging from everyday use and medical science to satellite imaging. Its main purpose is to combine several photographs, each with a limited field of view (FOV), into one image with an extensive FOV. This is especially important in indoor settings where the photographer is constrained by proximity to objects, preventing the capture of the entire scene in a single shot. Therefore, stitching methods are essential for obtaining a complete view.

Traditional image stitching techniques heavily rely on identifying spatial relationships between images through geometric features such as scale-invariant feature transform (SIFT) [23] and Kanade-Lucas-Tomasi (KLT) [30]. This process then derives a series of spatial transformations for seamless image integration. Additionally, line segments have also been extensively used as geometric features to maintain the consistency of global structures [6, 10, 20, 35]. However,

geometric features are not always present, especially in images with homogeneous backgrounds or where feature matching is challenging. As a result, recent trends [12, 13, 25, 26, 29, 31, 32] have shifted towards employing deep learning models that utilize high-level semantic features to find spatial transformations, significantly reducing the visual artifacts produced during stitching.

During the stitching process, one image is typically designated as the fixed target, and another image (the source) is warped into the target image’s space to achieve optimal alignment in the overlapping region. Earlier methods assume that images are 2D planes in a 3D space and apply homographic transformations for mapping. However, given the parallax effect resulting from camera position changes, simple homography cannot accurately align objects with large depth variations. To address this, recent advancements [14, 22, 37, 38] have introduced local warping techniques to mitigate this issue, though these methods still operate under the postulation of image smoothness and continuity.

In this study, we propose an innovative approach that introduces holes in the source image, which we refer to as *controlled image breaking* due to occlusion during warping, as a solution to alignment issues. To achieve this, we calculated pixel-level correspondences by focusing on the overlapping areas between the source and target with the help of FlowFormer [8]. It allows our method to break the source image in a controlled manner during warping, ensuring that corresponding pixels occupy the same position after alignment. A large portion of missing pixels resulting from image breakage within these areas, can be directly copied from the target image, given that the source and target images are already aligned within the same coordinate system. A small fraction of pixels near the overlapping boundary are repaired using the image inpainting technique since optical flows for aligning source and target images are not guaranteed perfect.

FlowFormer’s calculation is confined to the overlapping sections. For areas outside this overlap, we utilize the thin-plate spline (TPS) [1] algorithm to ensure a smooth transition near the overlap boundary. Specifically, we strategically place control points inside the boundary of the overlap region for warping non-overlapping areas. These points are distributed evenly and in zones of higher color gradients. This placement is crucial for aligning object structures between overlapping and non-overlapping regions after warping, providing a comprehensive solution that bridges the discontinuity between these distinct areas.

In addition to the seamless-through-breaking approach, we fine-tune the optical flow estimator using real-world datasets to enhance the accuracy of content alignment in practice. Since annotating optical flows of real photos for training is expensive, we fine-tune the estimator based on the color difference of aligned pixels. Observing that colors of homogeneous regions are similar, to prevent image flipping, we apply regularization terms to prevent drastic offsets of neighboring pixels if these pixels are visible both in source and target images. Furthermore, since optical flows can only warp the overlapping area of the source image, we encourage the pixels along the boundary to move smoothly to achieve a seamless transition to the non-overlapping area when using a TPS warp. The fine-tuning

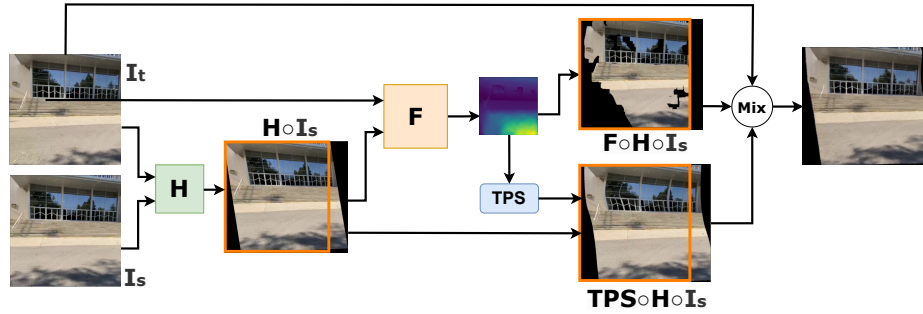


Fig. 1. We warp the source image I_s to the domain of the target image I_t to seamlessly stitch the two images. To explain the process, we use an orange rectangle to represent the target image’s domain in each transformed source image. The warping is initially achieved using a homography matrix H . Next, the area of I_s that overlaps with I_t is warped using optical flows F , whereas the non-overlapping region is warped using the thin-plate-spline (TPS) method. After the warping is complete, we merge the target image I_t with the warped images, $F \circ H \circ I_s$ and $TPS \circ H \circ I_s$, to generate the final stitched result.

achieves both precise content alignment and seamless transition between source and target images.

We have presented a method to merge images into a unified and visually continuous image. To validate the effectiveness of our approach, we applied it to several public datasets and conducted both quantitative and qualitative comparisons with current state-of-the-art methods [10, 20, 28]. The results demonstrate our method’s superiority across multiple metrics, confirming its efficacy. Notably, our approach does not require fine-tuning on new datasets, saving time and memory in practical applications and indicating that the model training details need not be disclosed to all subsequent users, greatly enhancing the model’s practicality post-deployment. This research not only provides a new scheme for image stitching but also offers valuable insights and references for future studies in related fields.

2 Related Works

Image stitching relies on the overlapping regions between images to identify corresponding SIFT features [23], facilitating a transformation that warps one image to the domain of another, thereby achieving a stitched outcome [2, 6, 16, 17, 22, 37]. Given that such approaches often neglect larger scale features such as straight lines [6, 10, 20, 35], which requires additional methods to preserve their shape, Du et al. [3] explored the representation of geometric structures within images through multiple edges, ensuring their shape is maintained during warping. Another approach involves leveraging depth maps [21] and semantic maps [15]

during image alignment to preserve high-level structures. Recent developments have used deep neural networks to compute homographies directly from input images [12, 13, 25, 26, 29, 31, 32], which minimizes mismatched alignment even when images contain large homogeneous backgrounds. This innovative approach pushes the boundaries of image stitching technologies further.

Traditional methods face limitations with global transformations, such as homography, which operates under the assumption that scenes are planar surfaces within a 3D space. This assumption often fails in complex scenes where distances between objects and the camera vary, leading to inaccuracies in aligning corresponding features. To address these challenges, Gao et al. [6] introduced a method that segments the scene into distant and ground planes, computing a homography for each and interpolating between the two to achieve nonlinear warping. This approach mitigates the limitations of a singular homography. Similarly, alternative techniques divide the scene into numerous cell-like subregions, each with its own homography. The different homographies of neighboring regions are then smoothed [14, 37] to enhance the coherence of the stitch. Diverging from the strategy of calculating homography before smoothing, Lin et al. [22] proposed optimizing a smoothly varying affine stitching field based directly on corresponding features, yielding impressive results. In a different approach, Zheng et al. [38] argued against the uniform spatial division of images for warping, pointing out the flaw in applying the same transformation to all pixels within a local region regardless of depth variance. They suggested a method that divides the overlapping regions into projective-consistent planes based on the orientation of normal vectors and reprojection errors, allowing for more accurate alignment in areas with significant parallax.

These stitching methods strive to maintain continuity in warping various regions within an image. However, this strategy encounters difficulties in areas where the depth of pixels is discontinuous. In such regions, changes in the camera’s position can result in background pixels being occluded by the foreground. Insisting on warping continuity in these instances may prevent the correct alignment of the same objects across images. An exceptional solution to this challenge involves pixel-wise warping [12]. This method leverages 3D virtual environments to generate images from different views and their stitched outcomes. Nevertheless, the viewpoints necessary for stitched results are not always available, even in virtual environments, and the outcomes of stitching are not unique. This limitation leads to a model’s potential overfitting to its training dataset, adversely affecting its ability to generalize to new data.

Our method distinguishes itself by grounding in optical flow estimation [8, 9, 33, 34], which focuses on the precise alignment of corresponding pixels within the overlapping regions. This approach allows for flexibility in warping non-overlapping areas, requiring only that the boundary between overlapping and non-overlapping regions maintains a similar warping to avoid discontinuity artifacts. This strategy offers a novel solution to the challenge of image stitching, particularly in complex scenes where large parallax is prevalent.

3 Seamless Warping Algorithm

3.1 Overview

Our method aims to warp the source image into the domain of the target image before seamlessly merging them through seam-cutting [11]. It begins with the application of homography to warp the source image, effectively reducing the perspective differences between the two images. This is followed by a second warping process using optical flow [8] to precisely align corresponding pixels within the overlapping regions. The two-step warping process is essential due to the typically limited search range of optical flow estimation, which may overlook distant corresponding pixels.

Unlike homography warping, which utilizes a transformation of the entire image, optical flow specifically targets the warping of areas within the source image that overlap with the target image. We introduce control points along the borders of the overlapping regions and employ the thin-plate spline (TPS) warping algorithm [1] to ensure a smooth transition with non-overlapping areas, thus avoiding discontinuity artifacts. However, due to variations in pixel depths, the image warped by optical flows may exhibit holes. To remedy this, we can directly retrieve some of the missing pixels from the target image, since the source image has been transformed into the domain of the target image. Considering that optical flow estimators are imperfect, we also incorporate an image inpainting technique to minimize the artifacts that can arise from directly filling missing pixels with data from the target image.

3.2 Optical Flow Estimation

After the initial homography warping, we employ FlowFormer [8] to compute the optical flow, which further refines the warping of the source image to better align corresponding objects across varying depths. Since FlowFormer is initially trained using synthetic data, its performance may diminish when applied directly to real-world images. To mitigate this issue, we fine-tune FlowFormer’s parameters using the UDIS-D dataset [27], which offers a more representative set of real-world scenarios.

Color consistency. To fine-tune FlowFormer in the absence of ground-truth optical flows from real data, we minimize the color differences between the warped source image and the target image. This method effectively utilizes the inherent visual information present in the actual images for training purposes. The loss function is defined as follows:

$$\mathcal{L}_c = \|\mathbf{I}_t \cdot \varphi(\mathbf{1}, \mathbf{F} \circ \mathbf{H}) - \varphi(\mathbf{I}_s, \mathbf{F} \circ \mathbf{H})\|_1, \quad (1)$$

Here, \mathbf{I}_s and \mathbf{I}_t represent the source and target images, respectively. \mathbf{H} and \mathbf{F} are the homography and the optical flow fields. $\varphi(\cdot, \cdot)$ is the warping function, and $\mathbf{1}$ is a mask containing only 1s, which is used to indicate the overlapping region after warping.



Fig. 2. We fill in the holes after warping the source image using optical flows. The left side shows the source and target images. In the middle left, the source image contains holes. In the middle right, we start by filling the middle part of the hole by directly copying pixels from the target image, since the two images are in the same coordinate system. Finally, on the right, we use an inpainting technique to fill the gaps around the hole’s boundary to minimize potential visual artifacts, as no optical flow estimators are perfect.

Regularization. Considering that color differences may not provide sufficient guidance for optical flow computation in homogeneous regions – where adjacent pixels have similar colors – we introduce a regularization term. This term encourages the optical flows of neighboring pixels to be similar, promoting smooth warping and preventing issues such as image flipping, which can disrupt the visual continuity of the stitched image. Specifically, we minimize

$$\mathcal{L}_r = \sum_{i=1}^h \sum_{j=1}^w \mathcal{M}_{i,j} (\|\mathbf{F}_{i,j} - \mathbf{F}_{i+1,j}\| + \|\mathbf{F}_{i,j} - \mathbf{F}_{i,j+1}\|), \quad (2)$$

where the subscripts i and j denote the pixel coordinate, and $\mathcal{M}_{i,j}$ is an occlusion mask indicating whether pixel at (i, j) will be occluded after being transformed to the domain of the target image. Let $\mathbf{F}_{s \rightarrow t}$ be the warping field from the source image to the target image. In our implementation, we set $\mathcal{M}_{i,j} = 1$ if the pixel at (i, j) can retain its position after a forward and a backward warping $\mathbf{F}_{t \rightarrow s} \circ \mathbf{F}_{s \rightarrow t}$, and $\mathcal{M}_{i,j} = 0$ otherwise. This ensures that we do not regularize optical flows of occluded regions.

Boundary Smoothness. Since the warping of overlapping and non-overlapping areas is conducted using different techniques, the boundary between these regions should not distort excessively. Otherwise, it can lead to visible artifacts in the non-overlapping areas. Let \mathcal{B} be an operation that extracts the boundary of the overlapping region. To mitigate this, we incorporate a boundary term that maintains the rigidity of these transitions, which is written as:

$$\mathcal{L}_b = \|\mathcal{B}(\mathbf{F})\|_1. \quad (3)$$

This additional constraint ensures that the integrity of image geometry is preserved across the stitched panorama, providing a cohesive and visually appealing final result.

	PSNR↑				SSIM↑			
	Easy	Moderate	Hard	Average	Easy	Moderate	Hard	Average
$I_{3 \times 3}$	15.87	12.76	10.68	12.86	0.530	0.286	0.146	0.303
SIFT+RANSAC [5]	28.75	24.08	18.55	23.27	0.916	0.833	0.636	0.779
APAP [36]	27.96	24.39	20.21	23.79	0.901	0.837	0.682	0.794
ELA [18]	29.36	25.10	19.19	24.01	0.917	0.855	0.691	0.808
SPW [20]	26.98	22.67	16.77	21.60	0.880	0.758	0.490	0.687
LPC [10]	26.94	22.63	19.31	22.59	0.878	0.764	0.610	0.736
UDIS [27]	25.16	20.96	18.36	21.17	0.834	0.669	0.495	0.648
UDIS++ [28]	30.19	25.84	21.57	25.43	0.933	0.875	0.739	0.838
FlowFormer++ [8]	30.57	26.89	23.50	26.62	0.945	0.909	0.838	0.891
Ours	31.40	27.65	24.09	27.34	0.961	0.932	0.873	0.917

Table 1. We have evaluated the preciseness of the alignment at the overlapping area of our method and the baseline algorithms on the UDIS-D dataset. Additionally, we have compared our method with FlowFormer++, which can be considered a variant of our method. The only difference is that the optical flow estimator in FlowFormer++ was not fine-tuned using real-world datasets.

3.3 Hole Filling

Utilizing optical flows for image warping can inevitably result in holes, especially in areas where depth discontinuities occur. Since the source image is already warped to align with the target image’s domain, most of these holes can be effectively filled using pixels from the target image. However, given that optical flow estimation is not flawless, we employ inpainting techniques to mend these gaps. Specifically, for the centers of the holes, we directly copy pixels from the target image. For the peripheries of the holes, where blending with the overlapping and non-overlapping regions is crucial, we apply the reference-guided inpainting technique [19] to integrate the areas seamlessly. The combination of the two strategies is particularly beneficial here, as it helps to avoid the creation of visual artifacts that often occur with larger repairs using inpainting. In our implementation, we perform an erosion operation on the hole mask to delineate the specific areas where pixels should be copied directly from the target image. The remaining regions are then addressed with inpainting.

3.4 Warp on Non-Overlapping Region

Optical flow fields are only capable of aligning the overlapping regions of the source and target images. This means that objects not present in the target image cannot be warped, which may result in the source image being broken into segments. To prevent discontinuity artifacts, we utilize the TPS [1] algorithm to achieve a smooth transition. The practical implementation involves strategically placing control points along the boundary between the overlapping and non-overlapping areas. These control points are influenced by the optical flows and can guide the warping of the non-overlapping area.



Fig. 3. In this example, source and target images are represented using a bitmap form and a rectangle. (Left) Control points are uniformly sampled at the boundary of the overlapping area. (Right) Points are adjusted to high gradient positions for better alignment around structural regions. Note that control points are used to warp the outer region of a source image, and the points in the bottom left corner are unnecessary.

When using the TPS method for warping, pixels that do not overlap with the control points are moved based on interpolation. To ensure better alignment around structural regions, we relocate the control points, which are uniformly distributed along the boundary of the overlapping region, towards areas with higher color gradients. This adjustment is done by allowing each control point to move within a range of $D/2$ when locating it on a high gradient position. Here, D represents the average distance between neighboring control points. As shown in Figure 3, this critical adjustment helps align structural objects across the boundary more precisely. In our implementation, we first sample points along the target image’s boundary, then remove the points if they do not overlay on the source image.

3.5 Training Details

We fine-tuned the FlowFormer++ model [8] on the UDIS-D dataset [27] using an NVIDIA RTX A6000 GPU card. To prevent overfitting, we mixed the synthetic data, which was initially used to train FlowFormer++, with the UDIS-D dataset during fine-tuning. The synthetic data training was supervised using ground-truth optical flows while the UDIS-D dataset training was based on the proposed objective function. We used the AdamW optimizer to update network parameters, with a batch size of 2 and a learning rate of 3×10^{-5} . Additionally, we set $\lambda_a=1.0$, $\lambda_b=1.0$, $\lambda_c=1.0$, $\lambda_d=0.5$, and $\lambda_e=0.3$ for the AdamW optimizer. For inpainting, we fine-tuned the reference-guided TransRef method [19] on 512×512 image sizes using the UDIS-D dataset [27]. We followed the default hyperparameter settings as outlined in the paper, except for the learning rate, which we adjusted to 2×10^{-5} .



Fig. 4. We compared our image stitching results with several baselines, including APAP [36], ELA [18], SPW [20], LPC [10] [10], UDIS [27] and UDIS++ [28]. It is evident that ghosting effects are present in the results generated by the baseline methods due to misalignment of the source and target images.

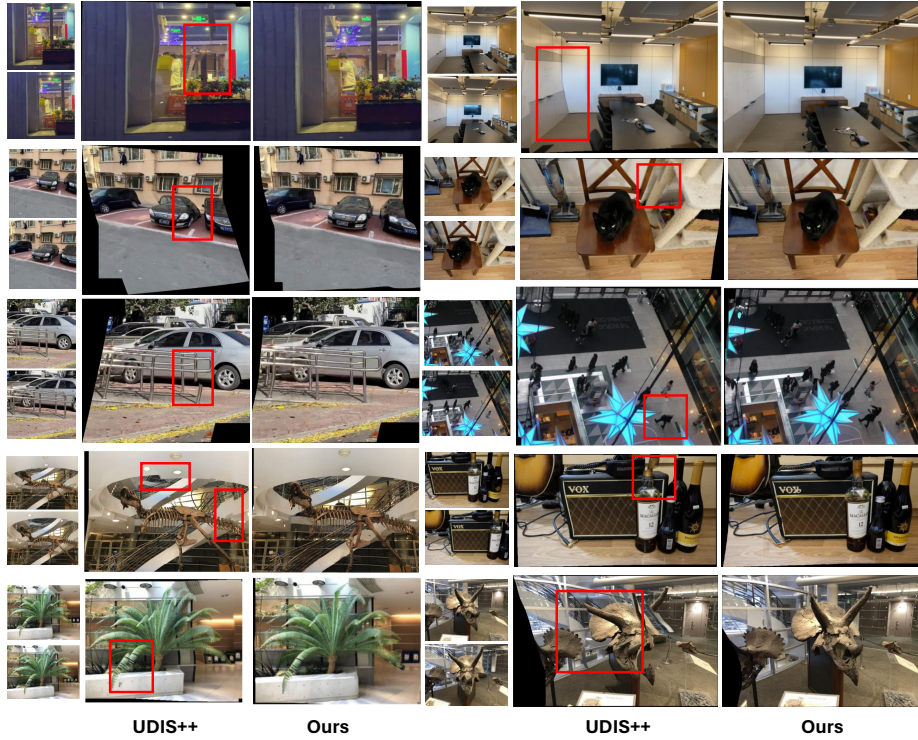


Fig. 5. UDIS++ [28] uses seam-cutting [11] to avoid ghosting effects, but this results in discontinuity artifacts due to image alignment issues and imperfect stitching. These artifacts are highlighted in rectangles. In the figure, original images are shown on the left, while the stitched results are compared on the right.

4 Experiments

We conducted evaluations accordingly to evaluate our seamless image stitching method. Following the approach of [28], we distinguished between the training and testing sets, with all subsequent experimental results being derived from the testing set. Additionally, we tested our method on other public datasets [4, 7, 24] without further network training to evaluate its generalization ability.

4.1 Warp Comparison

Quantitative evaluation. We initially conducted a quantitative comparison on the UDIS-D dataset [27] to evaluate whether the content in the overlapping sections of the source and target images was correctly aligned. Specifically, we measured the difference in pixel colors using PSNR and SSIM. The dataset categorizes each test sample by difficulty level – easy, moderate, and hard. During the

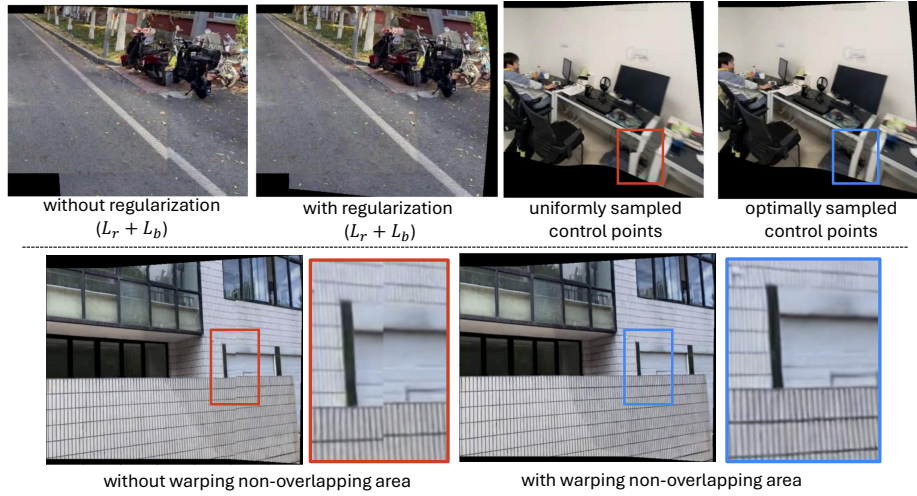


Fig. 6. We evaluated the effectiveness of fine-tuning FlowFormer++ with regularization terms (top left), the sampling strategy of control points used in TPS warp (top right), and warping non-overlapping regions. All of these tactics reduce discontinuity artifacts when compositing images.

comparison, since traditional methods may fail to generate homography transformations on challenging samples due to a lack of geometric features, in such cases, we employed an $I_{3 \times 3}$ identity matrix as a substitute metric to transform the source image. As demonstrated in Table 1, our method showed significant improvements across tasks of varying difficulty levels. Specifically, when pixel colors of the two images were measured by PSNR, our method achieved a 4% improvement in easy scenarios and an 11% improvement in hard scenarios. These results highlight the critical role of image breaking in handling parallax effects effectively, showcasing its robustness across different levels of complexity within image stitching challenges.

Qualitative evaluation. The image stitching results generated by the baseline and our algorithms are displayed in Figure 4. The overlapping pixels are blended linearly to facilitate a visual comparison of the warping quality. It is apparent that previous methods have struggled to align foreground objects with large parallax in their attempt to preserve warping continuity. This often results in ghosting artifacts in the blended regions of the source and target images. Our seamless-through-breaking approach effectively addresses this issue.

4.2 Comparison of the Composition Results

An alternative method for combining images is seam-cutting. This technique involves identifying a seam that has the least color inconsistencies [11], which passes through the overlapping area of the aligned source and target images.

While this method can be used in combination with all image stitching techniques, we have mainly compared our algorithm with UDIS++ [28], which is a SOTA that provides precise alignment of images. Traditional methods of finding the seam involve dynamic programming or graph cut algorithms. However, we have used the composition network of UDIS++ to compose the source and target images. The reason for doing so is that traditional methods produce discrete seams, while the seams extracted by the composition network are continuous. As shown in Figure 5, when there is a large area of misalignment between the images, it may not be possible to find perfect seams that seamlessly blend the images together. While seam-cutting can reduce ghosting effects caused by inaccurate alignment, it can produce discontinuity artifacts. These artifacts are unavoidable because maintaining the continuity of the image and aligning pixels with large differences in depth conflict with each other. Our algorithm, on the other hand, eliminates these problems and produces better results.

The results shown in Figures 4 and 5 demonstrate a seamless transition between overlapping and non-overlapping areas, as their boundaries are hardly noticeable despite the two regions being altered in two separate steps using different methods. Moreover, the inpainting does not introduce any unexpected objects that were not present in the original images since it only synthesizes pixels near the periphery of the holes, leaving the rest of the image intact.

4.3 Ablation Study

We have evaluated the presented method from two different perspectives. Firstly, we assessed the accuracy of content alignment by fine-tuning FlowFormer++ [8] using a real-world dataset. The comparative results are presented in Table 1, where our fine-tuned FlowFormer++ improved in terms of PSNR and SSIM in all difficulty conditions. Secondly, we evaluated the continuity of image composition, as the final output is a composition of two images, and overlapping and non-overlapping regions are warped using different methods. In Figure 6, we have presented comparisons with and without the regularization ($\mathcal{L}_r + \mathcal{L}_b$) terms (including boundary smoothness) when fine-tuning FlowFormer++, the sampling strategy of control points used in TPS, and with and without warping the non-overlapping regions. The results indicate that all our presented strategies are necessary to achieve high-quality stitched images.

4.4 Performance

In our experiments, our proposed warping process takes an average of approximately 0.9 seconds per a pair of images on the UDIS++ test dataset [28]. The subsequent inpainting step requires roughly 0.04 seconds to complete. This workflow showcases the efficiency and compatibility of our warping and inpainting methodologies in addressing complex tasks within computational constraints.



Fig. 7. (Left) Our approach does not change lighting conditions while stitching images. Discontinuities may appear if the two images (top) significantly differ in brightness. (Right) Even state-of-the-art optical flow estimation fails to warp images with large homogeneous backgrounds.

4.5 Limitations

We have demonstrated the strengths of our algorithm in the comparisons, but there is still room for improvement. Firstly, our approach merges images without altering their lighting conditions. Hence, visual artifacts may still be present even if we accurately align the images, as shown in Figure 7 (left). This issue can be addressed using Poisson blending, and we consider it to be our future work. Secondly, when using the reference-guided inpainting technique to precisely fill holes after warping the source image, the filled pixels may not accurately represent the objects in the original scene. Lastly, while our method benefits from advanced optical flow estimation techniques, the estimation is not always correct (Figure 7 right). We also had to limit the image resolution to 512×512 for image stitching due to large memory consumption when computing optical flows.

5 Conclusions

We have developed a seamless-through-breaking strategy to stitch images together and create a seamless output with a wider view. Unlike traditional methods, our approach prioritizes precise content alignment over image continuity when these two objectives conflict. Specifically, we allow images to break and then fill in the gaps. Our comparative study shows that this method is effective. While we used only two images in our demonstration, the technique is extendable to stitching more than two images together to create panoramas. In this case, the previous stitched result becomes the target image, and the new image is the source image. We will share the codes for public use soon.

Acknowledgments

We thank the reviewers for their constructive comments. This study is based upon work partially supported by the National Science and Technology Council (NSTC), Taiwan, under Contract No. 111-2221-E-A49 -129 -MY3, Contract No. 113-2425-H-A49 -001 - and Contract No. 113-2221-E-A49 -149 -MY3.

References

1. Bookstein, F.L.: Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on pattern analysis and machine intelligence* **11**(6), 567–585 (1989)
2. Brown, M., Lowe, D.G.: Automatic panoramic image stitching using invariant features. *International journal of computer vision* **74**, 59–73 (2007)
3. Du, P., Ning, J., Cui, J., Huang, S., Wang, X., Wang, J.: Geometric structure preserving warp for natural image stitching. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 3688–3696 (2022)
4. Du, P., Ning, J., Cui, J., Huang, S., Wang, X., Wang, J.: Geometric structure preserving warp for natural image stitching. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 3688–3696 (June 2022)
5. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* **24**(6), 381–395 (1981)
6. Gao, J., Kim, S.J., Brown, M.S.: Constructing image panoramas using dual-homography warping. In: *CVPR 2011*. pp. 49–56. IEEE (2011)
7. Herrmann, C., Wang, C., Bowen, R.S., Keyder, E., Zabih, R.: Object-centered image stitching. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. pp. 821–835 (2018)
8. Huang, Z., Shi, X., Zhang, C., Wang, Q., Cheung, K.C., Qin, H., Dai, J., Li, H.: Flowformer: A transformer architecture for optical flow. In: *European conference on computer vision*. pp. 668–685. Springer (2022)
9. Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., Brox, T.: FlowNet 2.0: Evolution of optical flow estimation with deep networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 2462–2470 (2017)

10. Jia, Q., Li, Z., Fan, X., Zhao, H., Teng, S., Ye, X., Latecki, L.J.: Leveraging line-point consistence to preserve structures for wide parallax image stitching. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12186–12195 (2021)
11. Kwatra, V., Schödl, A., Essa, I., Turk, G., Bobick, A.: Graphcut textures: Image and video synthesis using graph cuts. *Acm transactions on graphics (tog)* **22**(3), 277–286 (2003)
12. Kweon, H., Kim, H., Kang, Y., Yoon, Y., Jeong, W., Yoon, K.J.: Pixel-wise deep image stitching. *arXiv preprint arXiv:2112.06171* (2021)
13. Lai, W.S., Gallo, O., Gu, J., Sun, D., Yang, M.H., Kautz, J.: Video stitching for linear camera arrays. *arXiv preprint arXiv:1907.13622* (2019)
14. Lee, K.Y., Sim, J.Y.: Warping residual based image stitching for large parallax. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8198–8206 (2020)
15. Li, A., Guo, J., Guo, Y.: Image stitching based on semantic planar region consensus. *IEEE Transactions on Image Processing* **30**, 5545–5558 (2021)
16. Li, J., Deng, B., Tang, R., Wang, Z., Yan, Y.: Local-adaptive image alignment based on triangular facet approximation. *IEEE Transactions on Image Processing* **29**, 2356–2369 (2019)
17. Li, J., Wang, Z., Lai, S., Zhai, Y., Zhang, M.: Parallax-tolerant image stitching based on robust elastic warping. *IEEE Transactions on multimedia* **20**(7), 1672–1687 (2017)
18. Li, J., Wang, Z., Lai, S., Zhai, Y., Zhang, M.: Parallax-tolerant image stitching based on robust elastic warping. *IEEE Transactions on Multimedia* **20**(7), 1672–1687 (2018). <https://doi.org/10.1109/TMM.2017.2777461>
19. Liao, L., Liu, T., Chen, D., Xiao, J., Wang, Z., Lin, C.W.: Transref: Multi-scale reference embedding transformer for reference-guided image inpainting. *arXiv preprint arXiv:2306.11528* (2023)
20. Liao, T., Li, N.: Single-perspective warps in natural image stitching. *IEEE transactions on image processing* **29**, 724–735 (2019)
21. Liao, T., Li, N.: Natural image stitching using depth maps. *arXiv preprint arXiv:2202.06276* (2022)
22. Lin, W.Y., Liu, S., Matsushita, Y., Ng, T.T., Cheong, L.F.: Smoothly varying affine stitching. In: CVPR 2011. pp. 345–352. IEEE (2011)
23. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International journal of computer vision* **60**, 91–110 (2004)
24. Mildenhall, B., Srinivasan, P.P., Ortiz-Cayon, R., Kalantari, N.K., Ramamoorthi, R., Ng, R., Kar, A.: Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. *ACM Transactions on Graphics (TOG)* **38**(4), 1–14 (2019)
25. Nie, L., Lin, C., Liao, K., Liu, M., Zhao, Y.: A view-free image stitching network based on global homography. *Journal of Visual Communication and Image Representation* **73**, 102950 (2020)
26. Nie, L., Lin, C., Liao, K., Liu, S., Zhao, Y.: Unsupervised deep image stitching: Reconstructing stitched features to images. *IEEE Transactions on Image Processing* **30**, 6184–6197 (2021)
27. Nie, L., Lin, C., Liao, K., Liu, S., Zhao, Y.: Unsupervised deep image stitching: Reconstructing stitched features to images. *IEEE Transactions on Image Processing* **30**, 6184–6197 (2021). <https://doi.org/10.1109/TIP.2021.3092828>

28. Nie, L., Lin, C., Liao, K., Liu, S., Zhao, Y.: Parallax-tolerant unsupervised deep image stitching. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 7399–7408 (2023)
29. Nie, L., Lin, C., Liao, K., Zhao, Y.: Learning edge-preserved image stitching from multi-scale deep homography. *Neurocomputing* **491**, 533–543 (2022)
30. Shi, J., et al.: Good features to track. In: 1994 Proceedings of IEEE conference on computer vision and pattern recognition. pp. 593–600. IEEE (1994)
31. Song, D.Y., Lee, G., Lee, H., Um, G.M., Cho, D.: Weakly-supervised stitching network for real-world panoramic image generation. In: European Conference on Computer Vision. pp. 54–71. Springer (2022)
32. Song, D.Y., Um, G.M., Lee, H.K., Cho, D.: End-to-end image stitching network via multi-homography estimation. *IEEE Signal Processing Letters* **28**, 763–767 (2021)
33. Sun, D., Yang, X., Liu, M.Y., Kautz, J.: Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 8934–8943 (2018)
34. Teed, Z., Deng, J.: Raft: Recurrent all-pairs field transforms for optical flow. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16. pp. 402–419. Springer (2020)
35. Xiang, T.Z., Xia, G.S., Bai, X., Zhang, L.: Image stitching by line-guided local warping with global similarity constraint. *Pattern recognition* **83**, 481–497 (2018)
36. Zaragoza, J., Chin, T.J., Brown, M., Suter, D.: As-projective-as-possible image stitching with Moving DLT. In: In Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (2013)
37. Zaragoza, J., Chin, T.J., Brown, M.S., Suter, D.: As-projective-as-possible image stitching with moving dlt. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2339–2346 (2013)
38. Zheng, J., Wang, Y., Wang, H., Li, B., Hu, H.M.: A novel projective-consistent plane based image stitching method. *IEEE Transactions on Multimedia* **21**(10), 2561–2575 (2019)