Object-level Geometric Structure Preserving for Natural Image Stitching

Wenxiao Cai, Wankou Yang*

Southeast University, 2 SiPaiLou, Nanjing JS 210096, China
{wxcai,wkyang}@seu.edu.cn

Abstract

The topic of stitching images with globally natural structures holds paramount significance, with two main goals: pixel-level alignment and distortion prevention. The existing approaches exhibit the ability to align well, yet fall short in maintaining object structures. In this paper, we endeavour to safeguard the overall OBJect-level structures within images based on Global Similarity Prior (OBJ-GSP), on the basis of good alignment performance. Our approach leverages semantic segmentation models like the family of Segment Anything Model to extract the contours of any objects in a scene. Triangular meshes are employed in image transformation to protect the overall shapes of objects within images. The balance between alignment and distortion prevention is achieved by allowing the object meshes to strike a balance between similarity and projective transformation. We also demonstrate that object-level semantic information is necessary in low-altitude aerial image stitching. Additionally, we propose StitchBench, the largest image stitching benchmark with most diverse scenarios. Extensive experimental results demonstrate that OBJ-GSP outperforms existing methods in both pixel alignment and shape preservation.

Code — https://github.com/RussRobin/OBJ-GSP
Extended version — https://arxiv.org/abs/2402.12677

Introduction

Image stitching aims to align multiple images and create a composite image with a larger field of view. This method is widely utilized across diverse domains, including smartphone panoramic photography (Xiong and Pulli 2009), robotic navigation (Dewangan, Raja, and Singh 2014), and virtual reality (Anderson et al. 2016; Kim, taek Lim, and Ro 2020). In recent years, the problem of alignment has largely been addressed. Methods such as APAP (Zaragoza et al. 2013) and GSP (Chen and Chuang 2016) divide the images into multiple grids, compute local transformation matrices within each grid, and combine them with global transformation information to achieve precise alignment in overlapping regions. Thus, the main concern of image stitching nowadays is to prevent distortion on the basis of good alignment performance.

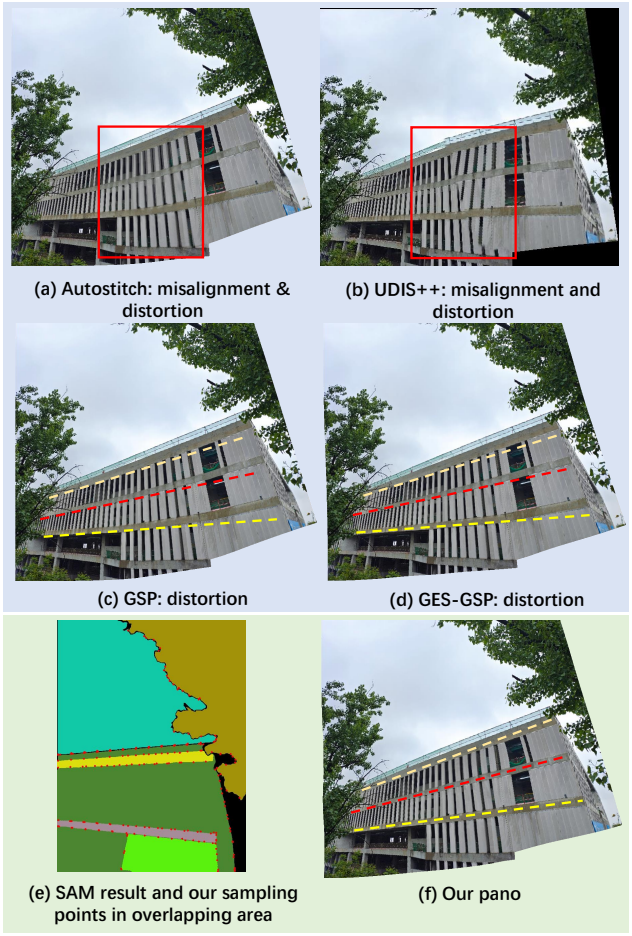

Figure 1: Red boxes indicate blurriness. (a) and (b) are not aligned well. (c) GSP (Chen and Chuang 2016) aligns well but distorts the building. Based on this, (d) GES-GSP (Du et al. 2022) tries to prevent distortion but still fails in this case. (f) our method protects the structure of the building by sampling on object contours extracted by segmentation (e).

Existing works extract lines in images are preserve them in image transformation. LPC (Jia et al. 2021) extracts and

*Correspond to: Wankou Yang

matches lines in alignment. Based on good alignment performance of GSP, GES-GSP (Du et al. 2022) adds the similarity transformation of line structures into considerations. However, (a) they only preserves line structures, ignoring overall and object-level structures, (b) focusing only on individual lines can be quite chaotic and mislead the model (Fig. 2), (c) straight or curved structures do not exist in some scenes.

Since an important criterion for humans to judge whether an image looks natural is the naturalness of the object structures within the image, our key insight is to extract these structures and preserve them during stitching. Nowadays, state-of-the-art segmentation models can identify almost any object with superior performance. We use them to get object shapes, which represents the image structure, and then use triangle meshes to preserve these segmented object shapes during the stitching. We generate triangle meshes within each object. During image transformation, these triangle meshes tend to reach a balance between projection and similarity transformations, effectively preserving the structure of the objects. As demonstrated in Fig. 1 (f), our method excels in maintaining the overall structure of images by preventing distortion of prominent object shapes. OBJ-GSP capitalizes on object-level preserving, and we adopt leverage segmentation models to extract geometric information. As shown in Fig. 1 (e), segmentation models treats objects as cohesive entities, transcending the segmentation of individual lines and curves adopted in previous works (Du et al. 2022; Jia et al. 2021). This allows for a more nuanced understanding of the relationships between individual geometric structures, and superior to previous work, it works even when there are no prominent linear structures in the images.

Previous works often used their own collected images without testing on datasets from other papers. We unified the datasets from previous works and incorporated our own collected hand-held camera and aerial images to create StitchBench, the most complete benchmark to date. We also demonstrate that in low-altitude aerial image stitching, semantic segmentation in OBJ-GSP pipeline is necessary. When the drone flies at a low altitude, the camera moves significantly, and there is a considerable distance difference between the roofs and the ground relative to the camera. These conditions do not satisfy the assumptions of image stitching (Brown and Lowe 2007), which assumes a fixed camera optical center or distant scenes, making stitching unfeasible. In this case, it is necessary to use a semantic segmentation model to identify the houses, then perform orthorectification to project them onto the ground before stitching.

To summarize, the main contributions of the proposed OBJ-GSP include:

- We propose to preserve object contours before and after image transformation to maintain the overall structure of the image. Object shapes are not limited to images with obvious linear structures and are not misled by excessively noisy line structures.
- We introduce the segmentation models into image stitching, facilitating the extraction of any object in the scene. Furthermore, we demonstrate that segmentation and OBJ-GSP are crucial for low-altitude aerial image stitching.
- We collect StitchBench, which is by far the largest and most diverse image stitching benchmark.

Related Work

Grid Based Image Stitching

Autostitch (Brown and Lowe 2007), a pioneering work in image stitching, matches feature points and aligns them by homography transformation. Building upon this foundation, numerous stitching algorithms partition images into grids, compute geometric transformation relationships for each grid, and combine them into a global transformation to align overlapping regions and seamlessly transit the transformation to non-overlapping areas. APAP (Zaragoza et al. 2013), AANAP (Lin et al. 2015a), and GSP (Chen and Chuang 2016) have evolved over time, essentially addressing most alignment problems in images. However, their grid deformation methods have no knowledge of object shapes. They pay too much attention on alignment and thus causes geometric distortion. To address this, LPC (Jia et al. 2021) and GES-GSP (Du et al. 2022) propose to preserve line structures. However, (a) their method only preserves line structures, ignoring the overall structure of objects, (b) an excessive number of lines without object structure information can mislead the model, (c) some scenes do not contain straight or curved structures. We find that the large segmentation models like SAM (Kirillov et al. 2023) can segment all types of objects and provide their contours. This helps image stitching maintain shape consistency, so we have incorporated the family of SAM into our method. We use triangular grids to protect the overall object-level geometric structure, and establish connections between dispersed geometric transformations, achieving superior results.

Geometric Structure Extraction

Previous works employ Line Segment Detector (von Gioi et al. 2012) to detect straight lines in images, and edge detection methods like Canny (Canny 1986) and HED (Xie and Tu 2015) to identify edges. However, these methods require line structures to be present in the image. In cases where textures are unclear or lighting is poor, conventional methods cannot extract lines effectively, whereas large models can still operate successfully in these scenarios. We employ the family of SAM and EfficientSAM (Xiong et al. 2023) to extract object-level structures and preserve them during stitching. It is notable that segmentation models are not limited by line structures and can segment almost any object. In the future, the accuracy and speed of SAM-type methods will both improve (Xiong et al. 2023; Zhang et al. 2023), further enhancing the quality and speed of our image stitching techniques.

Deep Learning Based Stitching

In recent years, several methods (Jia et al. 2023) like UDIS (Nie et al. 2021) have attempted to model certain image stitching steps as unsupervised deep learning problems, leading to notable advances in this field. UDIS++ (Nie et al. 2023) also addresses the distortion problem on the basis of

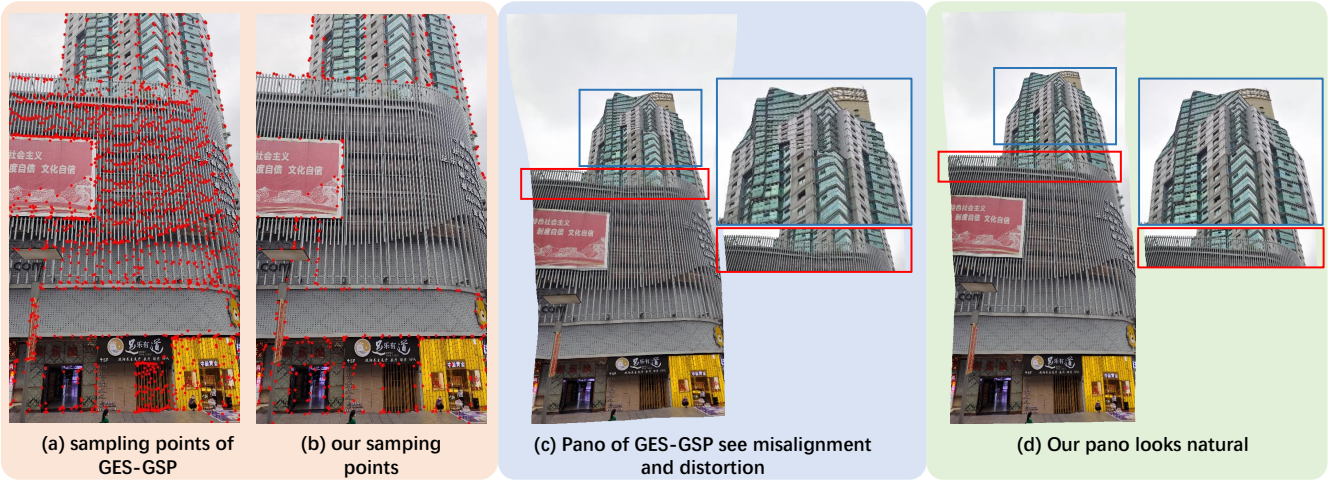

Figure 2: Sampling points in our OBJ-GSP focus more on main structures so we can stitch precisely, as shown in (b)(d).

good alignment performance, which aligns perfectly with our goals. We adhere to the traditional approach in the stitching domain by preserving results in grid transformation, while UDIS++ provides a completely new deep learning-based pipeline, although currently its performance is not as good as ours.

The Proposed Method

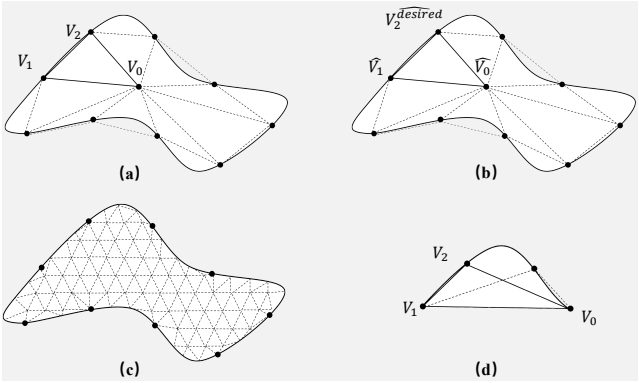

Figure 3: (a) Our triangle mesh. V_0 is the center of object. (b) With every edge undergoing similarity and projection transformation, object in (a) is transformed into (b). (c) Triangle mesh with near-equilateral triangles of similar sizes across the region. (d) Triangle sampling strategy.

OBJ-GSP introduces SAM to segment objects to obtain their structural contours and preserve object-level structures as well as aligning feature points in stitching. Locally, our approach retains the original perspective of each image. On a global scale, it seeks to preserve overall structure (Chen and Chuang 2016). Moreover, at the object-level, we ensure the integrity of objects within the images, preventing distortion. To this end, we take four aspects into consideration: alignment, global similarity, local similarity and object-level shape preservation. A grid mesh is adopted to guide the image deformation, where V and E represent the sets of vertices and edges within the grid mesh, as shown in Fig. 3. Image stitching methods aim to find a set of deformed vertex positions, denoted as $\widetilde{V}$, that minimizes the energy function $\psi(V)$.

Alignment Term extracts feature points p by with an extractor (e.g. SIFT (**?**)) and matches feature point pairs with matcher Φ. For each feature point pair $(p, \Phi(p))$, $\widetilde{v}(p)$ represents the position of p as a linear combination of four vertex positions, and M represents the set of all feature point pairs. The algorithm linearly combines the coordinates of the four vertices of each grid to represent the position of p through bi-linear interpolation. By optimizing the positions of grid vertices after geometric transformation, it aims to bring p as close as possible to $\Phi(p)$. Therefore, the energy equation is defined as:

$$\psi_a(V) = \sum_{p_k \in M} \|\widetilde{v}(p_k) - \widetilde{v}(\Phi(p_k))\|^2. \tag{1}$$

Local Similarity Term aims to ensure that the transition from overlapping to non-overlapping regions is natural. Each grid undergoes a similarity transformation to minimize shape distortion. For an edge (j, k), S_{jk} represents its similarity transformation. Suppose v_j transforms to $\widetilde{v_j}$ after deformation, and the energy function is defined as:

$$\psi_l(V) = \sum_{(j,k) \in E_i} \|(\widetilde{v}_k - \widetilde{v}_j) - S_{jk}(v_k - v_j)\|^2. \tag{2}$$

Global Similarity Term operates on a global scale to ensure the entire image undergoes a similarity transformation. GSP algorithm evaluates the scale s and rotation θ within the global image transformation and computes parameters $c(e)$ and $s(e)$ for similarity. Thus, the energy function is defined as:

$$\psi_g(V) = \sum_{e_j \in E} w(e_j)^2[(c(e_j) - s\cos\theta)^2 + (s(e_j) - s\sin\theta)^2]. \tag{3}$$

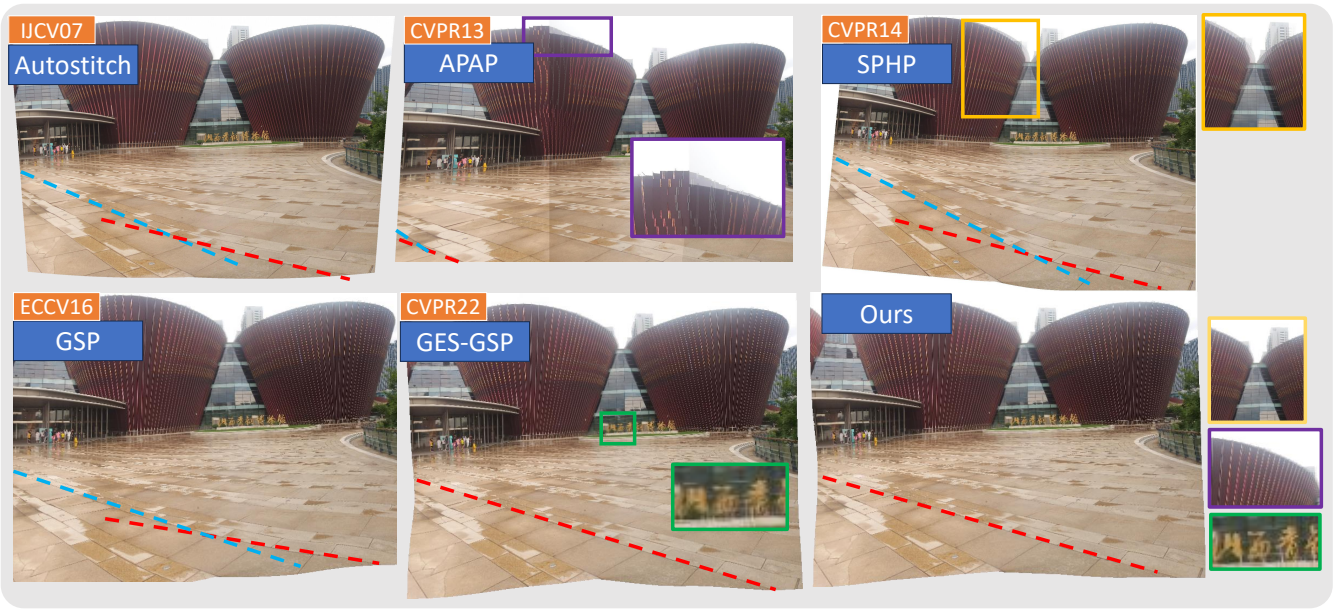

Figure 4: APAP and SPHP see misalignment, and we delineated the indistinct portions using color-coded boxes. Autostitch, APAP, SPHP and GSP sees distortion. The convergence of the blue and red lines is essential, and we signify distortion by the intersection of these two lines. GES-GSP successfully prevents distortion, but it undergoes misalignment. Our method addresses misalignment and distortion well.

After obtaining the contours, we generate a triangular mesh for each semantic object, preserving the shape of the object through similarity transformations within the triangle mesh. Unlike the As-Rigid-As-Possible (ARAP) (Igarashi, Moscovich, and Hughes 2005) method, we simplify computational complexity by directly locating the center of the object and connecting it to sampling points on the object's semantic boundary to form a triangular mesh. In Fig 3, V_0 represents the object's center, while V_1 and V_2 are sampling points on the semantic boundary of the object, forming a triangular mesh with these three points. (x_{01}, y_{01}) refer to the known coordinates of a feature point in the local coordinate plane. One vertex, V_2, of the triangle can be represented using the edges of the triangle $\overrightarrow{V_0V_1}$ and an orthogonal coordinate system obtained by rotating this edge counterclockwise by 90 degrees:

$$V_2 = V_0 + x_{01}\overrightarrow{V_0V_1} + y_{01}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\overrightarrow{V_0V_1}. \quad (4)$$

After the mesh deformation, V_0 and V_1 are transformed into $\widehat{V_0}$ and $\widehat{V_1}$. To preserve the shape of the segmentation result, we aim for the triangle to undergo a similarity transformation, keeping x_{01} and y_{01} unchanged. Therefore, we desire V_2 to transform into:

$$\widehat{V_2^{desired}} = \widehat{V_0} + x_{01}\overrightarrow{\widehat{V_0}\widehat{V_1}} + y_{01}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\overrightarrow{\widehat{V_0}\widehat{V_1}}. \quad (5)$$

The corresponding energy term for the transformed $\widehat{V_2}$ is calculated as:

$$E_{V_2} = \|\widehat{V_2^{desired}} - \widehat{V_2}\|^2. \quad (6)$$

Similar definitions for energy terms are applied to $\widehat{V_0}$ and $\widehat{V_1}$, resulting in the error sum for a triangle:

$$E_{\{V_0,V_1,V_2\}} = \sum_{i=0,1,2} \|\widehat{V_i^{desired}} - \widehat{V_i}\|^2. \quad (7)$$

Initially, our approach constructs the triangular mesh by selecting sampling points and the object's center. Unlike ARAP (Igarashi, Moscovich, and Hughes 2005), we do not employ equilateral triangular meshes, as objects segmented from the image often lead to very small equilateral triangles. Experimental results demonstrate that this approximation not only has no adverse impact on the final outcome but also reduces computational complexity:

$$E_{V_0,V_1,V_2} = \sum_{i=1,2} \|\widehat{V_i^{desired}} - \widehat{V_i}\|^2. \quad (8)$$

We extract N_c semantic object structures from a single image using semantic segmentation, and N_s represents the total number of all sampling points within geometric structure i. Similar to GES-GSP (Du et al. 2022), ω is a coefficient calculated based on the positions of the sampling points. Consequently, the total error equation is as follows:

$$\psi_{obj}(V) = \sum_{\beta=1}^{N_c}\sum_{\alpha=1}^{N_s} \omega_\alpha^\beta E_\alpha^\beta. \quad (9)$$

To conclude, our objective function is given by:

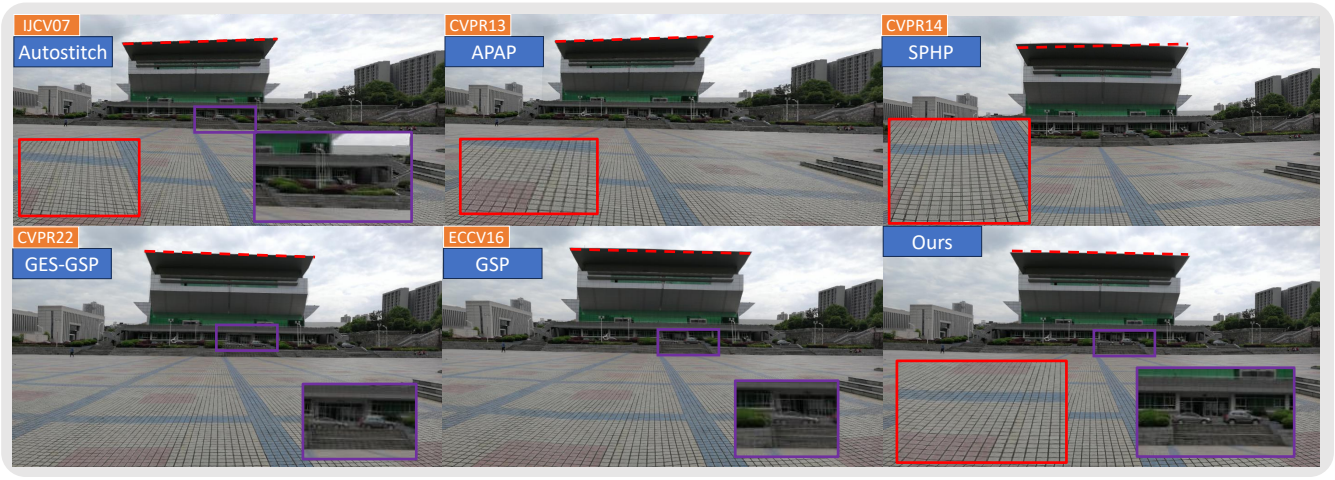

Figure 5: We magnify the ground in the red box and the car in the purple box. OBJ-GSP aligns well but the remaining five methods shows misalignment. Distortion is not observed in this case.

$$\widetilde{V} = \underset{\widetilde{V}}{arg\min} \Big(\psi_a(\widetilde{V}) + \lambda_l \psi_l(\widetilde{V}) + \psi_g(\widetilde{V}) + \lambda_{obj} \psi_{obj}(\widetilde{V}) \Big) . \tag{10}$$

Eq. 10 can be solved with linear optimization. For fair comparison, our parameters are identical to those of GES-GSP: $\lambda_l = 0.75$, $\lambda_{obj} = 1.5$. Our λ_{obj} corresponds λ_{ges} in to GES-GSP.

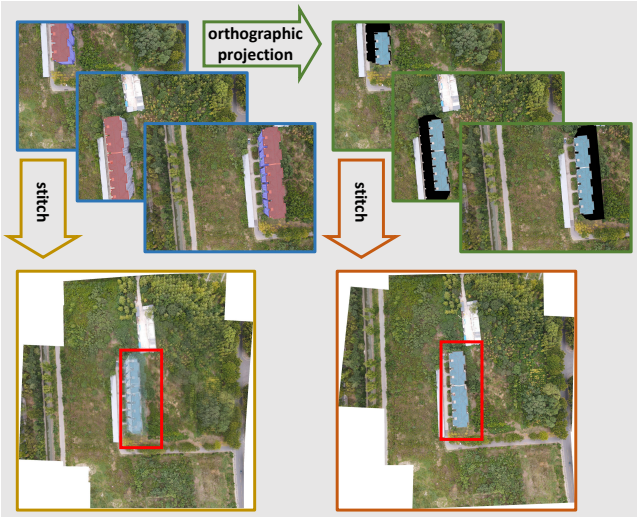

Figure 6: OBJ-GSP for low-altitude drone image stitching. We segment our roofs and walls from aerial images first. Walls are masked out, and roofs are projected to the ground. Then OBJ-GSP stitches the projected images.

Experiments

StitchBench

Previous work often collected a small number of images themselves and performed qualitative tests only. Meanwhile, they have different focuses, such as parallax between the foreground and background, sparse features in natural scenery, precise alignment and no distinct structures to preserve, and distinct line structures, without comprehensively evaluating models' performance in a wide range of scenarios. To address the issue, we present the most extensive image stitching benchmark to date: StitchBench, which include 122 pairs of images from 12 works. We collect 18 pairs of images captured by cameras, in which the preservation of object structures is crucial. StitchBench also includes 7 sets of urban scenes captured by low-altitude drones, featuring tall buildings and requiring the assistance of segmentation models. To overcome our subjective preferences and the limited locations where we collected the images, we also collect test images used in previous state-of-the-art works, namely AANAP (Lin et al. 2015a), APAP (Zaragoza et al. 2013), CAVE (Nomura, Zhang, and Nayar 2007), DFW (Li et al. 2015), DHW (Gao, Kim, and Brown 2011), GES-GSP (Du et al. 2022), LPC (Jia et al. 2021), SEAGULL (Lin et al. 2016), REW (Li et al. 2018), SVA (Lin et al. 2011) and SPHP (Lin et al. 2015b). StitchBench is currently the most comprehensive stitching test dataset. An algorithm should demonstrate general applicability to perform well on all subsets of StitchBench: aligning well and preventing distortion naturally.

Evaluation Metrics. We quantitatively assess the quality of our stitching results from two perspectives: distortion prevention and alignment. First, we employ the Mean Distorted Residuals (MDR) metric to measure the degree of image distortion. In intuitive terms, if the points on the same side of the mesh were originally collinear and remain collinear after stitching, it implies that the stitching result has minimal distortion. Furthermore, we employ the Naturalness Im-

Models / Datasets	OBJ-GSP	AANAP	APAP	CAVE	DFW	DHW
Mean Distorted Residuals (MDR ↓)						
GSP ECCV16	1.15296	1.06183	1.25495	0.90884	0.98457	1.08755
GES-GSP CVPR22	1.14366	1.06213	1.24249	0.90821	0.98034	1.05619
OBJ-GSP (ours)	**1.12229**	**1.05930**	**1.20123**	**0.89731**	**0.97259**	**1.00496**
Improvement(%)	1.9	0.3	3.3	1.2	0.8	4.9
Naturalness Image Quality Evaluator (NIQE ↓)						
UDIS TIP21	3.69421	3.01517	3.69421	-	5.74137	3.28645
UDIS++ ICCV23	3.34003	2.95493	3.56812	4.07702	**5.09680**	3.23392
GSP ECCV16	2.66597	2.84241	3.4356	4.04708	5.61905	2.75485
GES-GSP CVPR22	2.64986	2.77220	3.48713	4.03835	5.71544	2.70838
OBJ-GSP (ours)	**2.54906**	**2.74965**	**3.39280**	**4.01565**	5.69104	**2.60825**
Improvement(%)	3.8	0.8	2.7	0.6	0.4	3.7

Models / Datasets	GES-GSP	LPC	REW	SEAGULL	SVA	SPHP
Mean Distorted Residuals (MDR ↓)						
GSP ECCV16	1.06986	1.30562	1.16192	1.14467	1.51158	1.17784
GES-GSP CVPR22	1.15462	1.11993	**1.47197**	1.14340	1.04473	1.22256
OBJ-GSP (ours)	**0.98288**	**1.10622**	**1.08635**	**1.08296**	1.47813	**1.07699**
Improvement(%)	5.9	9.5	5.9	3.3	-0.4	5.8
Naturalness Image Quality Evaluator (NIQE ↓)						
UDIS TIP21	5.02442	3.76994	3.61888	4.67437	8.02090	4.35149
UDIS++ ICCV23	4.93279	3.66565	3.67661	4.38520	7.5419	4.1248
GSP ECCV16	3.84897	4.28546	3.18549	5.10784	7.00495	3.04781
GES-GSP CVPR22	3.79240	3.35315	**2.75234**	4.69390	6.96670	2.97173
OBJ-GSP (ours)	**3.70041**	**3.23057**	2.81480	**4.08903**	**6.96149**	**2.49712**
Improvement(%)	2.4	3.7	-2.2	12.9	0.5	16.0

Table 1: We report the mean MDR for distortion prevention evaluation, and NIQE to measure alignment performance. UDIS and UDIS++ are not feature point based so we only report NIQE and leave qualitative results in supplementary material. Best results are labeled with **bold text**. Lower MDR and NIQE indicates better stitched panorama. Improvement row compares our proposed method and GES-GSP. Our mean improvement to GES-GSP is **3.5%** in MDR and **3.8%** in NIQE.

age Quality Evaluator (NIQE) (Mittal, Soundararajan, and Bovik 2013) metric to evaluate alignment performance. We argue that NIQE is a more intitutive and better indicator of alignment than RMSE, SSIM and PSNR, as it measures image clarity, and stitching results with misalignment will produce blurry areas, leading to worse NIQE scores.

Baselines

We compare with GSP (Chen and Chuang 2016) and GES-GSP (Du et al. 2022). UDIS (Nie et al. 2021) and UDIS++ (Nie et al. 2023) are famous works in applyig deep learning into image stitching. Since they do not explicitly use feature points, we are unable to measure its quality with MDR. We provide a detailed comparasion between OBJ-GSP and UDIS++ in the supplementary materials.

Results

Quantitative Results. Table 1 shows MDR and NIQE results on datasets used in other stitching algorithms and our own dataset. We outperform GSP and GES-GSP in both alignment and shape preservation. UDIS++ (Nie et al. 2023) is a good try in deep learning based image stitching, but the performance is still no better than ours.

Qualitative Results. In image stitching, human perception is primarily influenced by two key factors: alignment and distortion. Misalignment can result in undesirable artifacts such as ghosting and blurriness, while distortion can make the stitched image appear unnatural. Fig. 1 and 2 elucidates the reasons behind the superior performance of our OBJ-GSP method. With the assistance of semantic segmentation techniques, we place greater emphasis on preserving critical structures and ensuring holistic protection at the object-level for objects within the images. Fig. 4 and 5 illustrates the stitching outcomes of six different methods, where we use straight lines and boxes to demonstrate the effects of alignment and distortion.

Low Altitude Aerial Image Stitching

Image stitching requires meeting one of two conditions (Brown and Lowe 2007): either the camera's optical center remains stationary while the camera rotates, or the scene only consists of objects that are far from the camera. Existing stitching algorithms mainly address the issue of stitching when these conditions are slightly violated. For low-altitude aerial images, where the flight height of the aircraft is around 100 meters but the height of buildings is no less than 20-40 meters, the camera's optical center moves significantly during drone shooting, thus completely failing

Figure 7: GES-GSP and the proposed OBJ-GSP with three Segment Anything Model backbones.

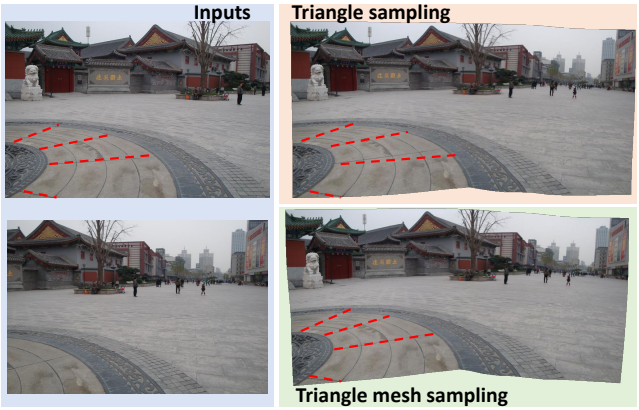

Figure 8: Triangle sampling preserves the shapes of lines on the ground, while triangle mesh sampling fails in preserving.

to meet the two assumptions for stitching. Moreover, if the left and right walls of a building are captured in two separate shots, it would be a logical error to include both walls in a panorama (for a cube, at most three faces can be seen at a time, and it's impossible to see two opposing faces simultaneously). For stitching low-altitude aerial images, we first use a semantic segmentation model to segment the roofs and walls. We then calculate the height of the roofs and orthographically project the buildings onto the ground plane before stitching. In this scenario, the semantic segmentation model is essential for the stitching process (Cai, Du, and Yang 2023). The stitching pipeline and result is in Fig. 6.

Ablation Studies and Discussions

Lightweight SAM Backbones

To assess the influence of semantic segmentation results on the stitching performance, we conducted a comparative analysis across three different backbones of the SAM (Kirillov et al. 2023) model, namely ViT-B, ViT-L, and ViT-H (Dosovitskiy et al. 2020), with a progression from smaller to larger models. Larger models are inherently capable of capturing more fine-grained semantic details. The corresponding stitching results are depicted in Fig. 7. It is worth noting that the models based on ViT-B and ViT-L exhibit some blurriness and minor distortions. From the perspective of MDR, our model, under the three aforementioned backbone configurations, achieved improvements of 1.9%, 2.7%, and 3.6% over the baseline model GES-GSP (Du et al. 2022), respectively. The application of EfficientSAM (Xiong et al. 2023), time consumption of OBJ-GSP, and its real-world applications are included in supplementary materials.

Sampling Strategies

Incorporating SAM (Kirillov et al. 2023) with ViT-H (Dosovitskiy et al. 2020) backbone, triangular mesh sampling yields superior results compared to the triangular sampling proposed in GES-GSP (Du et al. 2022). The triangular mesh preserves the shape of lines but fails to maintain the overall geometric structure of the image. As illustrated in the Fig. 8, we outline the structural elements in the original image using red dashed lines and then superimposed them onto the combined results of triangular sampling and triangular grid sampling. Triangular grid sampling retains the positional relationships between the lines present in the original image.

Conclusion

In this paper, we propose OBJect-level Geometric Structure Preserving, which stands as a novel approach to achieving natural and visually pleasing composite images. OBJ-GSP protects object shapes by first segmenting them out, and then preserve the structures with triangle meshes. We also demonstrate that semantic segmentation is necessary when it comes to low-altitude aerial image stitching. We collect new test image pairs in common scenes and aerial imaging, and choose images from previous works, to establish the most comprehensive image stitching benchmark by far: StitchBench. Detailed experiments with comprehensive baselines in StitchBench demonstrate the effectiveness of OBJ-GSP.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Nos. 62276061 and 62436002. This work was also supported by Shenzhen Science and Technology Program under Nos. JCYJ20230807114659029.

References

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 Trans. Graph.*, 35: 198:1–198:13.

Brown, M. A.; and Lowe, D. G. 2007. Automatic Panoramic Image Stitching using Invariant Features. *International Journal of Computer Vision*, 74: 59–73.

Cai, W.; Du, S.; and Yang, W. 2023. UAV image stitching by estimating orthograph with RGB cameras. *J. Vis. Commun. Image Represent.*, 94: 103835.

Canny, J. F. 1986. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8: 679–698.

Chen, Y.-S.; and Chuang, Y.-Y. 2016. Natural Image Stitching with the Global Similarity Prior. In *European Conference on Computer Vision*.

Dewangan, A. K.; Raja, R.; and Singh, R. 2014. An Implementation of Multi Sensor Based Mobile Robot with Image Stitching Application.

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; Uszkoreit, J.; and Houlsby, N. 2020. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *ArXiv*, abs/2010.11929.

Du, P.; Ning, J.; Cui, J.; Huang, S.; Wang, X.; and Wang, J. 2022. Geometric Structure Preserving Warp for Natural Image Stitching. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3678–3686.

Gao, J.; Kim, S. J.; and Brown, M. S. 2011. Constructing image panoramas using dual-homography warping. *CVPR 2011*, 49–56.

Igarashi, T.; Moscovich, T.; and Hughes, J. F. 2005. As-rigid-as-possible shape manipulation. *ACM SIGGRAPH 2005 Papers*.

Jia, Q.; Feng, X.; Liu, Y.; Fan, X.; and Latecki, L. J. 2023. Learning Pixel-wise Alignment for Unsupervised Image Stitching. *Network*, 1(1): 1.

Jia, Q.; Li, Z.; Fan, X.; Zhao, H.; Teng, S.; Ye, X.; and Latecki, L. J. 2021. Leveraging Line-point Consistence to Preserve Structures for Wide Parallax Image Stitching. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 12181–12190.

Kim, H. G.; taek Lim, H.; and Ro, Y. M. 2020. Deep Virtual Reality Image Quality Assessment With Human Perception Guider for Omnidirectional Image. *IEEE Transactions on Circuits and Systems for Video Technology*, 30: 917–928.

Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A. C.; Lo, W.-Y.; Dollár, P.; and Girshick, R. B. 2023. Segment Anything. *ArXiv*, abs/2304.02643.

Li, J.; Wang, Z.; Lai, S.; Zhai, Y.; and Zhang, M. 2018. Parallax-Tolerant Image Stitching Based on Robust Elastic Warping. *IEEE Transactions on Multimedia*, 20: 1672–1687.

Li, S.; Yuan, L.; Sun, J.; and Quan, L. 2015. Dual-Feature Warping-Based Motion Model Estimation. *2015 IEEE International Conference on Computer Vision (ICCV)*, 4283–4291.

Lin, C.-C.; Pankanti, S.; Ramamurthy, K. N.; and Aravkin, A. Y. 2015a. Adaptive as-natural-as-possible image stitching. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1155–1163.

Lin, C.-C.; Pankanti, S.; Ramamurthy, K. N.; and Aravkin, A. Y. 2015b. Adaptive as-natural-as-possible image stitching. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1155–1163.

Lin, K.; Jiang, N.; Cheong, L. F.; Do, M. N.; and Lu, J. 2016. SEAGULL: Seam-Guided Local Alignment for Parallax-Tolerant Image Stitching. In *European Conference on Computer Vision*.

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

Mittal, A.; Soundararajan, R.; and Bovik, A. C. 2013. Making a "Completely Blind" Image Quality Analyzer. *IEEE Signal Processing Letters*, 20: 209–212.

Nie, L.; Lin, C.; Liao, K.; Liu, S.; and Zhao, Y. 2021. 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. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 7365–7374.

Nomura, Y.; Zhang, L.; and Nayar, S. K. 2007. Scene collages and flexible camera arrays. In *Proceedings of the 18th Eurographics conference on Rendering Techniques*, 127–138.

von Gioi, R. G.; Jakubowicz, J.; Morel, J.-M.; and Randall, G. 2012. LSD: a Line Segment Detector. *Image Process. Line*, 2: 35–55.

Xie, S.; and Tu, Z. 2015. Holistically-Nested Edge Detection. *International Journal of Computer Vision*, 125: 3 – 18.

Xiong, Y.; and Pulli, K. 2009. Sequential image stitching for mobile panoramas. *2009 7th International Conference on Information, Communications and Signal Processing (ICICS)*, 1–5.

Xiong, Y.; Varadarajan, B.; Wu, L.; Xiang, X.; Xiao, F.; Zhu, C.; Dai, X.; Wang, D.; Sun, F.; Iandola, F. N.; Krishnamoorthi, R.; and Chandra, V. 2023. EfficientSAM: Leveraged Masked Image Pretraining for Efficient Segment Anything. *ArXiv*, abs/2312.00863.

Zaragoza, J.; Chin, T.-J.; Tran, Q.-H.; Brown, M. S.; and Suter, D. 2013. As-Projective-As-Possible Image Stitching with Moving DLT. *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2339–2346.

Zhang, C.; Han, D.; Qiao, Y.; Kim, J. U.; Bae, S.-H.; Lee, S.; and Hong, C.-S. 2023. Faster Segment Anything: Towards Lightweight SAM for Mobile Applications. *ArXiv*, abs/2306.14289.