Supplementary Material:
Geometric Structure Preserving Warp for Natural Image Stitching

Peng Du[1], Jifeng Ning[1†], Jiguang Cui[1], Shaoli Huang[2], Xinchao Wang[3] and Jiaxin Wang[1]
[1]Northwest A&F University, China, [2]Tencent AI-Lab, [3]National University of Singapore, Singapore
{peng-du, jiaxinwang123}@nwafu.edu.cn, njf@nwsuaf.edu.cn,
cuijiguang1997@163.com, shaolihuang@tencent.com, xinchao@nus.edu.sg

This supplementary document consists two sections. Section 1 provides supplemental experiment and discussion, which consists of more experiments, quantitative comparison and the limitation of the proposed method. Section 2 presents the derivation for the proposed method.

1. Supplemental Experiment and Discussion

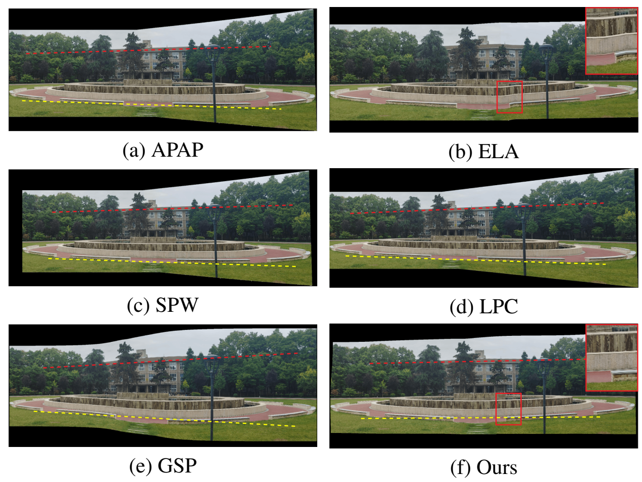
In this section, we present the results of the proposed GES-GSP in three subsections, which are more experimental results, quantitative comparison and its limitation, respectively.

1.1. More experimental results

We illustrate the effectiveness of the proposed method by more examples. The proposed method compares with AutoStitch [1], APAP [9], ELA [5], SPW [6], LPC [4], GSP [3].

In Figure 1, the results of APAP [9], SPW [6] and LPC [4], the right of the stitched result is severely stretched and non-uniformly enlarged. The scale of the right of the GSP [3] result does not match that of the left image. The results of the above methods show obvious left-right asymmetry and result in obvious damage to the overall structure of the fountain. The result of ELA [5] does not show the problem of asymmetry of left and right deformation, but there is obvious bending (in the red closeup). Our method clearly preserves the global structure of the scene where not only the overall structure of the fountain has not been destroyed, but also buildings in the distance and the fountain are on the same horizon (as indicated by the red and yellow lines).

In Figure 2, the results of APAP [9], SPW [6] and LPC [4] show that the unreasonable projections appear in the bottom half of the image (in the blue box). The result of ELA [5] exhibits many misalignments (in the red circles) and distortion (in the red closeup). Although the GSP [3]'s result is more natural, the building is bent (as indicated by the red lines). Our result shows that the building structures are preserved without projection distortion.



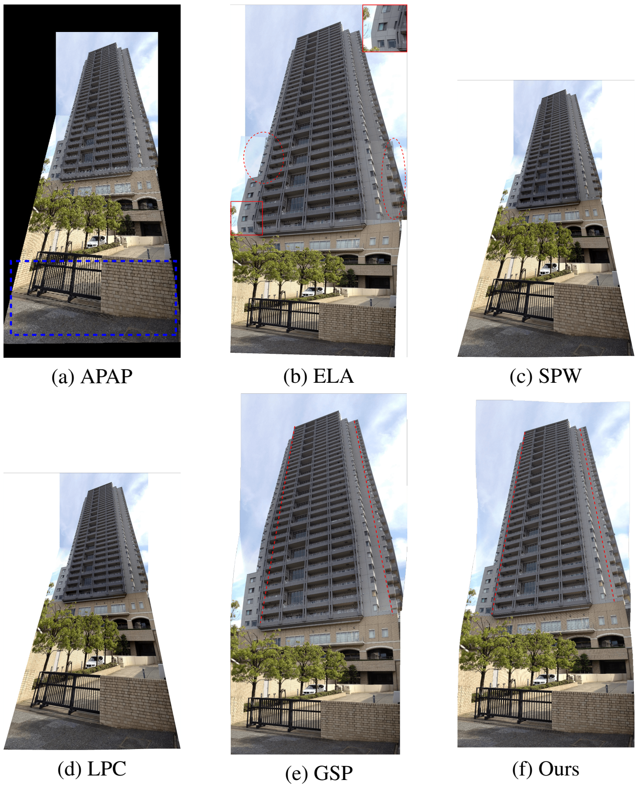
(a) APAP (b) ELA (c) SPW (d) LPC (e) GSP (f) Ours

Figure 1. An example of stitching two images.

In Figure 3, the results of APAP [9], SPW [6] and LPC [4] show that the roof and the ground are not on the same horizon (as indicated by the red and yellow lines), and the unreasonable projections appear in the right half of the image (in the blue box). The result of ELA [5] exhibits misalignments (in the red circle) and distortion (in the red closeup). The GSP [3]'s result shows that it solves the problem of different horizons, but ignores the structure-preserving, *e.g.*, the building is bent (as indicated by the red lines). Our result shows that the building structures are well preserved.

In Figure 4, ELA [5] plane projection failed and the results of AutoStitch [1], ELA [5] (Spherical projection) and GSP [3] show that the building is bent (as indicated by the red line). Unreasonable projections appear in the results of APAP [9] and SPW [6]. The proposed method constrains the geometric structure of the stairs and buildings and obtain natural stitching result.

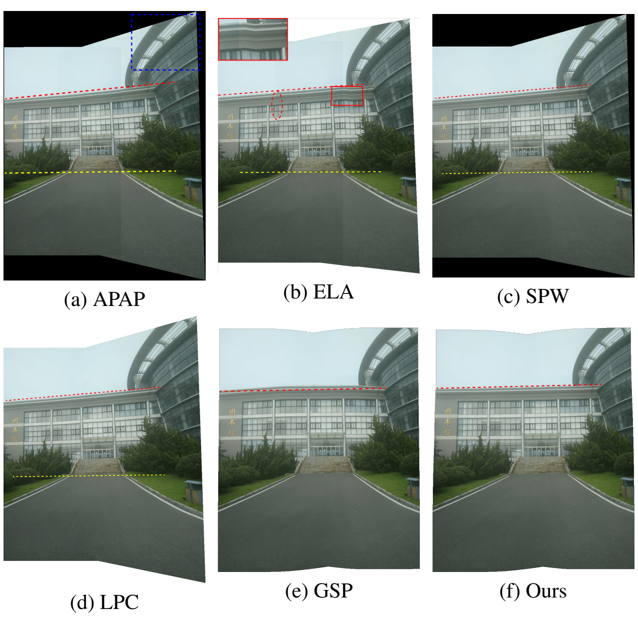
†Corresponding author: Jifeng Ning, njf@nwsuaf.edu.cn



(a) APAP (b) ELA (c) SPW

(d) LPC (e) GSP (f) Ours

Figure 2. An example of stitching two images (from [2]).

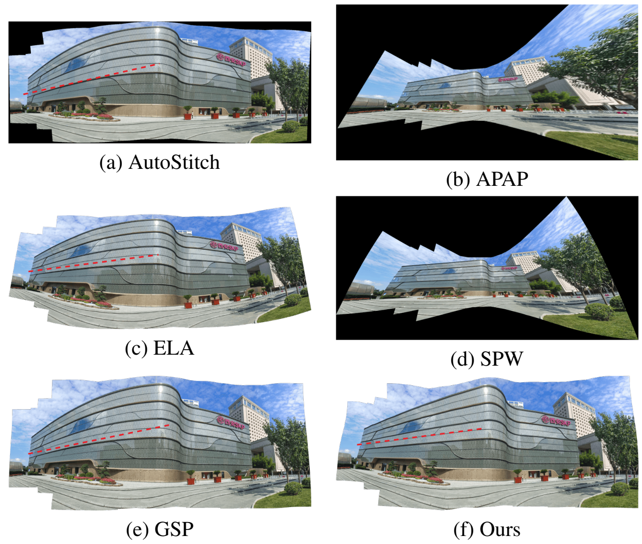


(a) APAP (b) ELA (c) SPW

(d) LPC (e) GSP (f) Ours

Figure 3. An example of stitching two images (from [4]).

In Figure 5, the result of AutoStitch [1], the platform and railway are obviously bent due to spherical projection. There is no obvious bending of line structure in APAP [9] and ELA [5], but the content (in the red closeup) in the left severely stretched and non-uniformly enlarged. The railroad tracks (as indicated by the green line) are bent in the result of SPW [6]. The line structures (as indicated by the red, yellow and green lines), such as railroad tracks, platforms and wire pole, are bent in the result of GSP [3]. Our method protects the global and local geometric structure well which results in better stitching.



(a) AutoStitch (b) APAP

(c) ELA (d) SPW

(e) GSP (f) Ours

Figure 4. An example of stitching 14 images (from [4]).

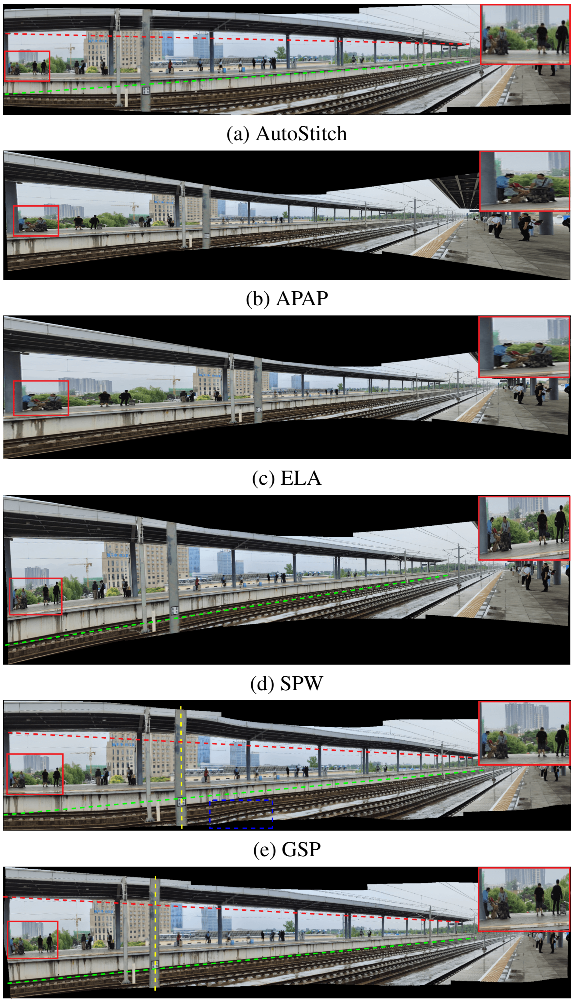
In Figure 6, SPW [6] fails to handle 2D spatial relationship. The result of ELA [5] (Spherical projection) is distorted due to spherical projection. The APAP [9]'s result exhibits severe projection distortion. In GSP [3]'s result, many structures are bent (as indicated by the red lines). Since the proposed method extracts more integrated structures to preserve, our result is more natural.

1.2. Quantitative Evaluation

It is difficult to accurately determine which is the better through subjective evaluation when comparing some similar results. Therefore, it needs quantitatively evaluate the distortion degree of the stitching result.

In general, if an image only undergoes similarity transformation, the distortion of the geometric structure is minimal in the figure [7] [9]. Both GSP and GES-GSP are essentially mesh based optimizations where the quads corresponding to each row or column will be distorted after stitching. The great distortion indicates poor capability of geometric structure preservation.

Let the mesh in Figure 7(a) be deformed to Figure 7(b). Our method fits the grid points of each row and column into multiple straight lines (red lines). The shortest distance



(a) AutoStitch

(b) APAP

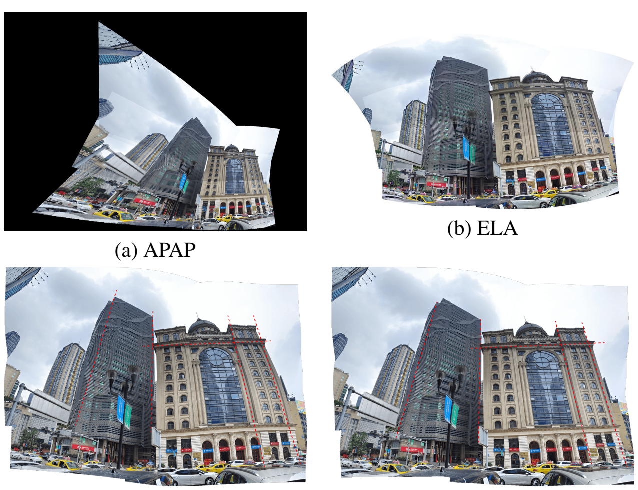
(c) ELA

(d) SPW

(e) GSP

(f) Ours

Figure 5. An example of stitching six images in the scene of station platform.



(a) APAP

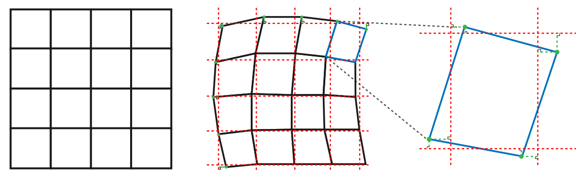
(b) ELA

(c) GSP

(d) Ours

Figure 6. An example of stitching six images.

(green lines in Figure 7(b)) from each grid point to the fitting line is taken as the residual value of this grid point. The root mean square error (RMSE) and standard deviation (SD) of all residual values of grid points are used to judge the degree of grid distortion. Then, an evaluation method called Mean Distorted Residuals (MDR) is proposed to quantitatively evaluate the degree of distortion of geometric structure. It helps us to indirectly evaluate image stitching quality. Our evaluation index consists of two parts: Root Mean Squared Error (RMSE) and Standard Deviation (SD) where smaller values mean better content-preservation.



(a)

(b)

Figure 7. (a). The mesh before deformation. (b) The mesh of (a) after deformation. The red lines represent the lines fitted by the grid points in each row and column. The green lines represent the shortest distance between the grid points and their corresponding fitted line.

The RMSE index evaluating distortion of the meshes in the i-th image is

$$RMSE_i = \frac{1}{2} \times (\sum_{x=0}^{X} \sum_{y=0}^{Y} \frac{1}{X} \sqrt{\frac{dst^2\left(v_{xy}, l_x\right)}{Y}} + \sum_{y=0}^{Y} \sum_{x=0}^{X} \frac{1}{Y} \sqrt{\frac{dst^2\left(v_{xy}, l_y\right)}{X}}) \tag{1}$$

where, l_x, l_y are the fitted straight lines of the grid points in the x-th column and the y-th row, respectively. v_{xy} is the grid point of position (x, y). X, Y are the number of columns and rows of the mesh respectively. $dst(v, l)$ is the shortest distance from point v to line l.

The SD of the i-th image is

$$SD_i = \frac{1}{2} \times (\sum_{x=0}^{X} \sum_{y=0}^{Y} \frac{1}{X} \sqrt{\frac{\left(dst\left(v_{xy}, l_x\right) - avg_x\right)^2}{Y}} + \sum_{y=0}^{Y} \sum_{x=0}^{X} \frac{1}{Y} \sqrt{\frac{\left(dst\left(v_{xy}, l_y\right) - avg_y\right)^2}{X}}) \tag{2}$$

where, avg_x and avg_y are the mean of the residual values of the grid points in the x-th column and y-th row respectively.

The evaluation index of the result is

$$\begin{gathered} RMSE = \frac{1}{n} \sum_{i=1}^{N} RMSE_i \\ SD = \frac{1}{\mathrm{n}} \sum_{i=1}^{N} \mathrm{SD}_i \end{gathered} \tag{3}$$

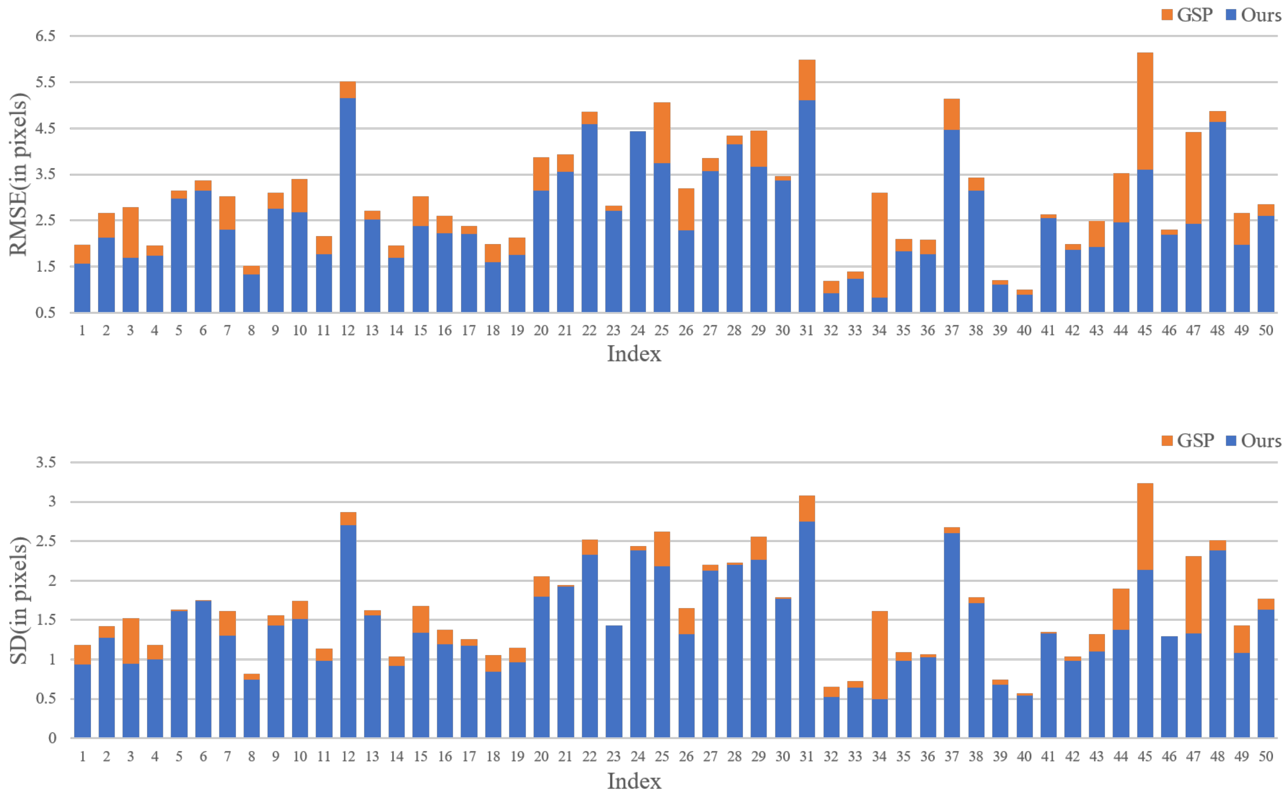


Figure 8. The RMSE and SD comparisons between GSP and proposed GES-SGP on 50-dataset.

where N is the number of images in a dataset.

Figure 8 shows the evaluation results of GSP [3] and our GES-GSP using MDR on 50-dataset. It indicates that our method is superior to GSP in the content-preserving.

1.3. Limitation of the Proposed Method

Generally, the deep learning-based method such as HED [8] provides better overall performance than classical edge detection (e.g., Canny). The performance is further improved when applying post-processing techniques such as thinning, burr removing, and branch cutting. The structure feature extraction is indeed a critical step in our framework, which affects the quality of our stitching results. However, if there are no salient geometric structures in scene, our method performs similarly with the GSP. For instance, Figure 9 shows the result comparison on a farmland image, where few salient features exist in the scene.

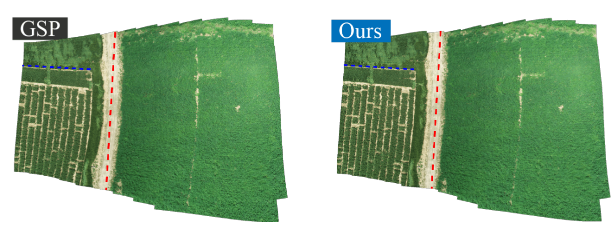


Figure 9. Stitching comparisons between GSP and our method for a group of farmland images with few salient features.

2. Derivation of the Proposed Method

The proposed method is a linear optimization problem, and the optimal mesh vertex set $\hat{V}$ can be obtained as a closed-form solution by solving a sparse linear matrix equation.

The derivation is as follows. Rewriting $\widehat{V} = \underset{\widehat{V}}{\arg\min}\left(\psi_a\left(\widehat{V}\right)+\lambda_l\psi_l\left(\widehat{V}\right)+\psi_g\left(\widehat{V}\right)+\lambda_{ges}\psi_{ges}\left(\widehat{V}\right)\right)$ in matrix form as

$$\widehat{V} = \arg\min_{\widehat{V}} \left\| \begin{bmatrix} A_a \\ A_l \\ A_g \\ A_{ges} \end{bmatrix} \widehat{V} - \begin{bmatrix} 0 \\ 0 \\ b_g \\ 0 \end{bmatrix} \right\|, \tag{4}$$

where A_a, A_l, A_g and A_{ges} corresponds to the matrix form of ψ_a, ψ_l, ψ_g and ψ_{ges} respectively.

Next, we present how to get the matrix A_{ges}. Let one sampling point $\widehat{V}_i\left(v'_{ix}, v'_{iy}\right)$ and two endpoint $\widehat{V}_a\left(v'_{ax}, v'_{ay}\right)$ and $\widehat{V}_b\left(v'_{bx}, v'_{by}\right)$ in the sampling point set S from a contour line, and expand $V_i = V_a + u_i\left(V_b - V_a\right) + h_i \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \left(V_b - V_a\right)$ to

$$\begin{bmatrix} v'_{ix} \\ v'_{iy} \end{bmatrix} = \begin{bmatrix} v'_{ax} \\ v'_{ay} \end{bmatrix} + u_i \begin{bmatrix} v'_{bx} - v'_{ax} \\ v'_{by} - v'_{ay} \end{bmatrix} + h_i \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} v'_{bx} - v'_{ax} \\ v'_{by} - v'_{ay} \end{bmatrix}. \tag{5}$$

Further, Equation (5) is decomposed into

$$\begin{bmatrix} 1 & 0 & -\left(1-u_i\right) & h_i & -u_i & -h_i \\ 0 & 1 & -h_i & -\left(1-u_i\right) & h_i & -u_i \end{bmatrix} \begin{bmatrix} v'_{ix} & v'_{iy} & v'_{ax} & v'_{ay} & v'_{bx} & v'_{by} \end{bmatrix}^T = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{6}$$

Each feature point is represented by the 2D bilinear interpolation of the four vertices. Let the four vertices of the grid where the sampling point $\widehat{V}_i$ is located be $\widehat{V}_{ik}\left(v'_{ikx}, v'_{iky}\right), k = 1, 2, 3, 4$. $\widehat{V}_i\left(v'_{ix}, v'_{iy}\right)$ can be represented by four vertices of the grid

$$\left[\begin{array}{c} v'_{ix} \\ v'_{iy} \end{array}\right] = \sum_{k=1}^{4} c_{ik} \left[\begin{array}{c} v'_{ikx} \\ v'_{iky} \end{array}\right], \tag{7}$$

where $c_{ik} = \frac{\left|d_x^{ik} d_y^{ik}\right|}{\left|d_x^{i1}\times d_y^{i1}\right|+\left|d_x^{i2}\times d_y^{i2}\right|+\left|d_x^{i3}\times d_y^{i3}\right|+\left|d_x^{i4}\times d_y^{i4}\right|}$, c_{ik} represents the weight of vertex $\widehat{V}_{ik}$ corresponding to sampling point $\widehat{V}_i$, $d_x^{ik} = v'_{ix} - v'_{ikx}$ and $d_y^{ik} = v'_{iy} - v'_{iky}$ respectively represents the difference of x, y between sampling point $\widehat{V}_i$ and vertex $\widehat{V}_{ik}$. Similarly, the results of $\widehat{V}_a\left(v'_{ax}, v'_{ay}\right)$,$\widehat{V}_b\left(v'_{bx}, v'_{by}\right)$ can be obtained.

Then, Equation (6) can be rewritten as

$$\left[\begin{array}{ccc} C_{ix} & h_i C_{ay} + \left(u_i - 1\right) C_{ax} & -u_i C_{bx} - h_i C_{by} \\ C_{iy} & \left(u_i - 1\right) C_{ay} - h_i C_{ax} & h_i C_{bx} - u_i C_{by} \end{array}\right] \left[\begin{array}{c} \widehat{V}_i \\ \widehat{V}_a \\ \widehat{V}_b \end{array}\right] = \left[\begin{array}{c} 0 \\ 0 \end{array}\right] \tag{8}$$

where $\widehat{V}_s = \left[\begin{array}{cccccccc} v'_{s1x} & v'_{s1y} & v'_{s2x} & v'_{s2y} & v'_{s3x} & v'_{s3y} & v'_{s4x} & v'_{s4y} \end{array}\right]^T$,
$C_{sx} = \left[\begin{array}{cccccccc} c_{s1} & 0 & c_{s2} & 0 & c_{s3} & 0 & c_{s4} & 0 \end{array}\right]$, $C_{sy} = \left[\begin{array}{cccccccc} 0 & c_{s1} & 0 & c_{s2} & 0 & c_{s3} & 0 & c_{s4} \end{array}\right]$, $s = i, a, b$.

After adding the adapted weight w_i of $\widehat{V}_i$, A_{gp} looks like

$$\left[\begin{array}{ccccccccccc} & & & & & \cdots & & & & & \\ 0 & \cdots & w_i C_{ix} & \cdots & w_i\left(h_i C_{ay} + \left(u_i - 1\right) C_{ax}\right) & \cdots & -w_i\left(u_i C_{bx} + h_i C_{by}\right) & \cdots & 0 \\ 0 & \cdots & w_i C_{iy} & \cdots & w_i\left(\left(u_i - 1\right) C_{ay} - h_i C_{ax}\right) & \cdots & w_i\left(h_i C_{bx} - u_i C_{by}\right) & \cdots & 0 \\ & & & & & \cdots & & & & & \end{array}\right] \tag{9}$$

Finally, the optimal mesh vertex set can be calculated $\widehat{V}$ by general numerical optimization algorithm.

References

[1] Matthew Brown and David G Lowe. Automatic panoramic image stitching using invariant features. *International journal of computer vision*, 74(1):59–73, 2007. 1, 2

[2] Chehan Chang, Yoichi Sato, and Yung-Yu Chuang. Shape-preserving half-projective warps for image stitching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3254–3261, 2014. 2

[3] Yusheng Chen and Yung-Yu Chuang. Natural image stitching with the global similarity prior. In *Proceedings of the European conference on computer vision*, pages 186–201. Springer, 2016. 1, 2, 4

[4] Qi Jia, Zhengjun Li, Xin Fan, Haotian Zhao, Shiyu Teng, Xinchen Ye, and Longin Jan Latecki. Leveraging line-point consistence to preserve structures for wide parallax image stitching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 12186–12195, 2021. 1, 2

[5] Jing Li, Zhengming Wang, Shiming Lai, Yongping Zhai, and Maojun Zhang. Parallax-tolerant image stitching based on robust elastic warping. *IEEE Transactions on multimedia*, 20(7):1672–1687, 2017. 1, 2

[6] Tianli Liao and Nan Li. Single-perspective warps in natural image stitching. *IEEE Transactions on Image Processing*, 29:724–735, 2019. 1, 2

[7] Kaimo Lin, Nianjuan Jiang, Loong-Fah Cheong, Minh Do, and Jiangbo Lu. Seagull: Seam-guided local alignment for parallax-tolerant image stitching. In *European conference on computer vision*, pages 370–385. Springer, 2016. 2

[8] Saining Xie and Zhuowen Tu. Holistically-nested edge detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1395–1403, 2015. 4

[9] Julio Zaragoza, Tat-Jun Chin, Michael S Brown, and David Suter. As-projective-as-possible image stitching with moving dlt. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2339–2346, 2013. 1, 2