Image Stitching by Points Grouping and Mesh Optimization

Mowen Xue Student[1]
xmw300568@buaa.edu.cn
Xudong Li Pro[1, 2]
xdli@buaa.edu.cn
Hongzhi Jiang Pro[1, 2]
jhz1862@buaa.edu.cn
Huijie Zhao Pro[1, 2]
hjzhao@buaa.deu.cn

[1]School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing, China

[2]Beihang University Qingdao Research Institute, Qingdao, China

Abstract

Image stitching is a cost-effective way to expand the field-of-view of imaging system. The traditional homography-based image stitching uses a global homography transformation matrix for image transformation, which is stable, but only works well for flat scenes, relative far scenes or the scenes which are captured by the camera with rotation only. The As-Projective-As-Possible and Content-Preserving-Warping methods, which are realized by mesh optimization, improve the stitching result to a certain degree, but there is obvious ghost in the near scenes or images which have relatively large parallax. In this paper, an image stitching method which utilizes depth information and mesh optimization is proposed. The feature points are detected and then clustered, and the depth information are used to assign weights to each mesh to compute homography for each mesh respectively. Experiments show proposed method has better results than other methods..

1 Introduction

Image stitching is to process two adjacent images with overlapping regions to obtain a wider field-of-view image. It is widely used in many fields, such as medical analysis, industrial parts inspection, virtual reality, etc. In recent years, image stitching has been widely studied, and many excellent image stitching algorithms have emerged, which can be roughly divided as Homography-based method, the Spatially-varying method, the Content-preserving method, and the Post-processing method.

Homography-based. The early methods mainly used a single global homography matrix [1]. This kind of method is simple, fast, and the results are relatively stable. For simple scenes, better stitching results can be achieved. However, for complex non-planar scenes, or the camera has translational motion during capturing, the method will produce large errors due to the limitations of the model itself. Based on the Homography method, Gao et al. proposed a Dual-Homography method [2]. It is assumed that the scene consists of two dominant planes, the background plane and the foreground plane. The scene is distinguished according to the feature points, and then two homography matrices are calculated and used respectively to warp the source image. This method improves the results

of the single matrix method to a certain extent, but there are still much ghost for scenes with large parallax.

Spatially-varying. Since the Homography-based method has poor processing ability for large parallax scenes, a series of spatially-varying method is proposed. This type of method firstly divides the image into uniform meshes, and then calculates the transformation matrix for each part separately. Specifically, Lin et al. proposed the Smoothly-varying-affine (SVAS) method [3], in which global affine transformation matrix is calculated, and then the difference between the matrix of each point and the global matrix is estimated, respectively, to obtain the final transformation matrices of each point. This method improved the stitching result, but would still have much ghost when the scene has large parallax. In order to further improve the accuracy of stitching, Zaragoza et al. proposed the As-projective-as-possible (APAP) method [4], which divides the image into uniform meshes, and then calculates the perspective transform for each mesh according to the distribution of feature points. For most scenes, the APAP method can achieve good result, but it still has some defects: 1) The method depends heavily on the quality and distribution of feature points. In the regions with less feature points, the registration error is large; 2) it may cause distortion in nonoverlapping region; 3) The weights assigned to each mesh depend only on the distance between feature points and meshes, so that when the scene has a large depth changes, it will lead to a significant error. For the problem 1), some methods combined the line features with point features [5] [16]. These methods could reduce the stitching errors in the regions with less feature points. For the problem 2), some methods are proposed to deal with the distortion problem in image transformation [8]. Chang et al. proposed the Shape-preserving half-projective (SPHP) method [7], which reduces the transformation process by dividing the image and using perspective transformation and similar transformation respectively. Lin et al. proposed the Adaptive as-natural-as-possible method [6], which reduces the deformation by linearizing the homography matrix and gradually transforming it into a similar transformation and simultaneously transforming the two images. However, the existing methods do not deal with the problems 3).

Content-Preserving. The Content-preserving warps (CPW) method was originally proposed by Liu et al. and applied to image stitching and video stabilization techniques [9]. This kind of method improves the registration accuracy by dividing the images into meshes and using constraints to obtain the optimal vertices point position while maintaining the original shape of the image. Some works afterwards continue to add constraints to further optimize registration accuracy. Lin et al. proposed adding the photometric term that combines the superior performance of dense photometric registration with the efficiency of mesh transformation [10]. They also proposed to add the line structure constraint to ensure the linearity of the transformation and the accuracy of the low-texture region.

Post-Processing. Post-processing methods can be generally divided into seam-based methods and complex fusion methods. The purpose of this kind of methods is to further process the registration image with a low registration accuracy to obtain a visually reasonable stitching result. The Seam-driven method firstly searches for the stitching line and then evaluate the seam, and iterates over to obtain the final stitched image [13]. Herrmann et al. proposed a Multi-registration method [14], which pre-calculates multiple candidate matches, and then uses each candidate's optimal part to fuse the final image. At the same time, they also propose an object-centered method [15]. The image area is divided, and the object term is added in the calculation of the seam by using the machine learning method. These two methods effectively improve the naturalness in the result image. However, the post-processing methods have the disadvantages of low computational

efficiency, and such methods do not consider the registration accuracy of the image, and can only act on the image stitching field and cannot be extended.

In this paper, a coarse-to-fine stitching scheme is proposed. The modified APAP is used as coarse registration and improved CPW is utilized as fine registration. In the coarse registration part, the feature point pairs are firstly grouped according to the plane which they are located in, and then each feature point pair is given a weight based on the group and the depth information. After that, the local homography matrix is calculated for image transformation. In the fine registration part, we add epipolar constraints on the basis of the traditional data term, similarity term, and photometric term to ensure image registration. Finally, we use several data for experimental verification. Through qualitative analysis and quantitative experiments, it can be seen that the proposed method is better than other methods.

The rest of this paper is organized as follows: Section 2 introduces the main content and implementation flow of our proposed method; Section 3 carries out experimental verification; Section 4 draws conclusions.

2 Proposed Method

2.1 Depth-Based APAP Warp

APAP is a mesh-based image stitching algorithm [4]. Firstly, the image is divided into uniform meshes, and then the transform matrix is calculated separately for each mesh according to the feature point distribution. The two images are denoted as I and I' respectively, and the corresponding points in the two images are denoted as $\mathrm{p} = [\mathrm{x\ y\ 1}]^T$ and $p' = [x'\ y'\ 1]^T$ (homogeneous coordinates). The homography transformation matrix is used to represent the relationship between the two images:

$$p' = \mathrm{Hp} \tag{1}$$

where H is a 3×3 matrix. The Eq. (1) can be rewritten as:

$$\begin{bmatrix} 0_{1\times3} & -p^T & -y'p^T \\ p^T & 0_{1\times3} & -x'p^T \end{bmatrix} \mathrm{h} = a_i h = 0_{2\times1} \tag{2}$$

where $\mathrm{h} = [h_1\ h_2\ h_3]^T$, and h_j is the j-th row of H. Then the DLT (Direct Linear Transformation) method is used to compute H:

$$\hat{h} = \underset{h}{\operatorname{argmin}} \sum_{i=1}^{N} \|a_i h\|^2 = \underset{h}{\operatorname{argmin}} \|Ah\|^2 \tag{3}$$

The Eq. (3) can be estimated by the least significant right singular vector of A.

The above is the method of global homography. The APAP method is proposed based on the global method. Firstly, the image is divided into uniform meshes, and then the transformation matrix is calculated separately for each vertices. All the feature point information is used in the calculation, but different points are given different weights according to the distance between points and mesh.

$$\hat{h}^* = \underset{h}{\operatorname{argmin}} \sum_{i=1}^{N} \|w_i^* a_i h\|^2 \tag{4}$$

where w_i^* denotes the matrix of weights, and can be represent as:

$$w_i^* = \max\left(\exp\left(-\frac{\|p_* - p_i\|^2}{\sigma^2}\right), \gamma\right) \tag{5}$$

where σ, γ mean adjustable parameters of the method, which usually be 8-12 and 0.0015-0.1 respectively.

The APAP method can achieve a good registration in most cases. However, when the camera is close to the scene and the images have large parallax, there will be a significant ghosting, as shown in Figure 1.

(a) (b)

Figure 1：The result of APAP (a), and proposed method (b).

Since the points in the same plane have the same transformation matrix, we propose an improved APAP method, which utilizes the depth information and grouping feature points to obtain a better registration result on the scene with large parallax and close distance.

The feature point pairs are detected firstly, and the RANSAC method is used to calculate the largest point pair that matches the model each time, then record them as a group, and the above steps are iterated until the remaining points cannot be calculated. Using this method, matching point pairs can be divided into groups according to the RANSAC method. The result of the grouping is shown in Figure 2.

(a) Closer group (b) Farther group

Figure 2: Red points indicate feature points. Different planes are clearly distinguished.

Then the average parallax of each group is estimated by the following equation:

$$\Delta p_j = \sum_{i=1}^{n} (x_i - x_i')/n \tag{6}$$

where j denotes the group number, and n is the number of points in each group.

Since the distribution of points within the group may be separated, the group which contains the closet point is assigned the biggest weight. And the other groups are sorted based on the difference between the average parallax value of the other groups and the closet group. The weight matrix is redesigned as:

$$w_i^* = \max(\exp\left(-\frac{\omega_j \|p_* - p_i\|^2}{\sigma^2}\right), \gamma) \tag{7}$$

where i, j denotes the mesh and the group number respectively, ω_j represents the weights depending on the sorting information, and 2,4,6⋯⋯ will be a good value according to experience. And then, the transformation matrix for each mesh can be solved by Eq. (4).

2.2 CPW with Epipolar Constraint

The result of the APAP is used as the coarse registration of the Content-preserving warp to further optimize the image registration accuracy. V denotes the coordinates of the vertices' points. And CPW is used to get the optimized vertices position $\hat{V}$.

Four constraints, including data term [9], similarity term [9], photometric term [10], and epipolar term are designed to form the objective function.

Data Term. The changes of the positions of the feature points are limited as small as possible. The pre-registered image and the target image are denoted as I_s, I_t respectively, and the matching point pairs of the two images are denoted as p, p'. Firstly, the feature points are represented by the interpolation of the four mesh vertices containing the points:

$$p = V_p w_p \tag{8}$$

where $V_p = \left[V_p^1\ V_p^2\ V_p^3\ V_p^4\right]$ denote four vertices which contain the point, and $w_p = \left[w_p^1\ w_p^2\ w_p^3\ w_p^4\right]^T$ denote the interpolation coefficient. This interpolation coefficient would not change after CPW. Therefore, Date term constraint is designed as:

$$E_D(\hat{V}) = \sum_{i=1}^{N_D} \left\| \hat{V}_{pi} w_{pi} - p_i' \right\|^2 \tag{9}$$

where p_i, p_i' denotes the i-th pair of matching points, and N_D means the number of matching points.

Similarity Term. The Similarity term is designed to reduce the deformation after the CPW. The same method in is used to calculate similarity term. The meshes are divided into triangles, and two of the vertices of the triangle are used to establish a local coordinate system to represent the third vertex:

$$V_1 = V_2 + u(V_3 - V_2) + vR_{90}(V_3 - V_2), R_{90} = \begin{bmatrix} 0 & 1 \\ -1 & 1 \end{bmatrix} \tag{10}$$

where V_1, V_2, V_3 denote three vertices of triangle, and u, v mean the calculated local coordinate. u, v should be changed as small as possible to make the image transformation more likely to the similarity transformation. Therefore, the similarity term can be represented as:

$$E_S(\hat{V}) = \sum_{i=1}^{N_S} \left\| \hat{V}_1^i - (\hat{V}_2^i + u(\hat{V}_3^i - \hat{V}_2^i) + vR_{90}(\hat{V}_3^i - \hat{V}_2^i)) \right\|^2 \tag{11}$$

where N_S means the number of triangles.

Photometric Term. Similar to [10], we use Photometric term to further constrain the global transformation of the image. The photometric error is defined as:

$$\left\| I_{tar}(q + \tau(q)) - I_{sou}(q) \right\|^2 \tag{12}$$

where $I_{tar}(q) I_{sou}(q)$ denote the intensity of image at q, respectively. And $\tau(q)$ represents a small displacement. The above equation can be simplified by first-order Taylor formula:

$$E_P(\hat{V}) = \sum_{i=1}^{N_P} \| I_{tar}(q) + \nabla I_{tar}(q)\tau(q) - I_{sou}(q) \|^2 \tag{13}$$

where N_P denotes the points, which are calculated (sample points in overlapping region as described in [10]). $\nabla I_{tar}(q)$ means the gradient at q, and is formulated as the interpolation of four mesh vertices as in data term.

Besides above three constraints, we add epipolar constraint to further improve sititching accuracy.

Epipolar Term. The epipolar constraint is added to further optimize the registration result. During the image transformation, the polar-pole correspondence of the image should remain unchanged. The $p^T F p' = 0$ relationship should be exhausted in the transformation. Therefore, the epipolar term is designed as:

$$E_E(\hat{V}) = \sum_{i=1}^{N_E} \| p^T F p' \| \tag{14}$$

where N_E denotes the number of points which are computed. Since p^T is known, p' can be represented as an interpolation of the vertices of the mesh, so that the above formula can be simplified as:

$$E_E(\hat{V}) = \sum_{i=1}^{N_E} \left\| b_i \hat{V}_{pi} w_{pi} \right\| \tag{15}$$

where $b_i = p^T F$, and p_i denotes the i-th points.

Optimization. The above terms are combined, and then the final objective function can be written as:

$$\mathrm{E}(\hat{V}) = \lambda_D E_D(\hat{V}) + \lambda_S E_S(\hat{V}) + \lambda_P E_P(\hat{V}) + \lambda_E E_E(\hat{V}) \tag{16}$$

where $\lambda_D, \lambda_S, \lambda_P, \lambda_E$ denote the coefficient of each constraint. According to [10], $\lambda_S = 0.2 \sim 0.5, \lambda_E = 0.1 \sim 0.5$, and the other two are set to 1. Then the constraints are superimposed and the optimal vertex position of the mesh can be obtained.

3 Experiments and Discussion

Images from [4] and captured by ourselves are used in our experiments. For the fairness of experiments, the same feature points are used for all methods. Some experimental results are shown in Figure 3. It can be seen from the figure that in the close-range large parallax scene, our method performs better than other method.

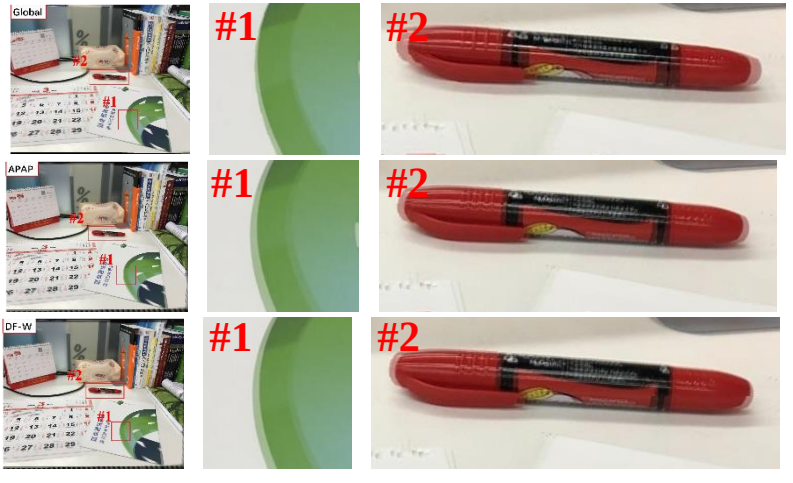

Figure 3: Qualitative comparison. List of method: Global homography, APAP, DF-W, Our method.

We selected the images in [4] and the images taken by ourselves are used to evaluate the results quantitatively. Figure 4 shows an overview of dataset. The first row is the images from [4], and the second row is the images of indoor and close-range scenes, and the third row contains the images of industrial parts.

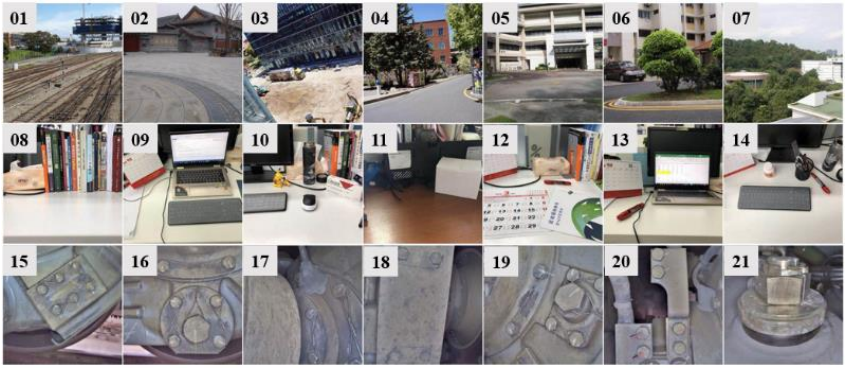

Figure 4: Images for quantitative evaluation. 01-07 are from [4], and 08-21 are ours.

For the lack of ground truth, a formula needs to be established to quantitatively compare the stitching results. RMSE error is chosen to analyse the stitching accuracy of various methods [5] (Since we use the different corresponding feature points, the values of APAP are different from that in [5]). Specifically, we compute the RMSE over a local w × w window for all pixels in the overlapping region.

$$\mathrm{RMSE}(I_s, I_t) = \sqrt{\frac{1}{N}\sum_{\Omega}(1 - NCC(p_s, p_t))^2} \tag{17}$$

where Ω represents the selected window size, N means the number of points in overlapping region, I_s, I_t, p_s, p_t represent the two images and points on them respectively.

Table 1 shows the results for various methods(Global [1], APAP [4], DF-W (Dual-feature warping) [5], D-APAP (depth-based APAP)). It can be seen that our method has better results in most scenes than other methods. However in some scenes, such as No.14 in the experiment, DF-W method gets a better result than our method. The reason is that there are many regions with low texture, and it can be hard to extract enough feature points there. But DF-W method use point features and line features, so it can get enough information in those low-texture regions, so the transform matrices are more accurate.

NO.	Global	APAP	DF-W	D-APAP	Ours
1	13.27	11.00	11.25	10.68	**10.14**
2	8.55	3.12	3.20	3.12	**2.99**
3	9.76	5.96	6.21	5.89	**5.85**
4	14.25	12.74	14.5	**12.45**	12.68
5	9.91	4.46	**3.98**	4.45	4.22
6	5.67	4.7	4.89	**4.44**	4.58
7	7.19	3.06	4.8	2.98	**3.01**
8	8.26	6.78	6.32	5.65	**5.36**
9	6.78	5.84	5.66	5.50	**5.44**
10	6.29	4.72	4.32	4.13	**4.1**
11	6.34	4.67	5.10	4.51	**4.43**
12	6.91	5.6	3.89	3.20	**3.2**
13	12.45	9.27	9.32	9.94	**9.22**
14	8.66	7.28	**6.32**	6.94	6.88
15	8.65	3.56	3.12	2.56	**2.44**
16	4.58	2.26	2.05	1.88	**1.96**
17	2.64	1.98	1.78	1.85	**1.76**
18	4.23	2.8	2.65	2.45	**2.41**
19	2.21	1.36	1.48	1.35	**1.35**
20	4.32	2.65	**1.98**	2.36	2.15
21	5.32	2.64	2.32	1.56	**1.29**

Table 1: RMSE results of different methods.

4 Conclusion

An image stitching method combining depth information and mesh optimization is proposed. Through the coarse-to-fine stitching scheme, an accurate registration is achieved, and high-quality stitching result images are obtained. The core of our method is the redesign of the weight coefficients in the APAP method. For most scenarios, our method can achieve precise stitching.

However, for the regions with few feature points, the method needs to be improved. In the future, other feature detection techniques can be combined with our method (such as point-line feature combination) to get more accurate results.

References

[1] M. Brown and D. G. Lowe. Automatic Panoramic Image Stitching using Invariant Features. *IJCV*, 74(1):59-73, 2007.

[2] J. Gao, S. J. Kim and M. S. Brown, Constructing Image Panoramas using Dual-Homography warping, *CVPR*, 45-53, 2011.

[3] Lin W, Liu S, Matsushita Y, et al. Smoothly varying Affine Stitching. *CVPR*, 345-352, 2011.

[4] Zaragoza J, Chin T J, Tran Q, et al. As-Projective-As-Possible Image Stitching with Moving DLT. *PAMI*, 36(7):1285-1298, 2014.

[5] Li S, Yuan L, Sun J, et al. Dual-Feature Warping-Based Motion Model Estimation, *ICCV*, 4283-4291, 2015.

[6] Lin C, Pankanti S U, Ramamurthy K N, et al. Adaptive As-Natural-As-Possible Image Stitching. *CVPR*, 1155-1163, 2015.

[7] Chang C, Sato Y, Chuang Y, et al. Shape-Preserving Half-Projective Warps for Image Stitching. *CVPR*, 3254-3261, 2014.

[8] Chen Y, Chuang Y. Natural Image Stitching with the Global Similarity Prior. *ECCV*, 186-201, 2016.

[9] Liu F, Gleicher M, Jin H, et al. Content-preserving Warps for 3D Video Stabilization. *CGIT*, 28(3), 2009.

[10] Lin K, Jiang N, Liu S. Direct Photometric Alignment by Mesh Deformation. *CVPR*, 2701-2709, 2017.

[11] Gao J, Li Y, Chin T, et al. Seam-Driven Image Stitching. *Eurographics*, 45-48, 2013.

[12] Lin K, Jiang N, Cheong L F, et al. SEAGULL: Seam-Guided Local Alignment for Parallax-Tolerant Image Stitching. *ECCV*, 370-385, 2016.

[13] Zhang F, Liu F. Parallax-tolerant Image Stitching. *CVPR*, 3262-3269, 2014.

[14] Herrmann Charles, Wang Chen, Strong Bowen Richard, Keyder Emil. Robust Image Stitching with Multiple Registrations. *ECCV*, 53-69, 2018.

[15] Herrmann Charles, Wang Chen, Strong Bowen Richard, Keyder Emil, Zabih Ramin. Object-Centered Image Stitching. *ECCV*, 846-861, 2018.

[16] Xiang T, Xia G, Bai X. Image Stitching by Line-Guided Local Warping with Global Similarity Constraint. *Pattern Recognition*, 83:481-497, 2018.