

IMAGE STITCHING BASED ON MULTI-SCALE MESHES

*Yixuan Li^{*1}, Haotian Zhao^{*2}, Nan Pu³, Qi Jia¹*

¹International School of Information Science Engineering, Dalian University of Technology

²School of Software Microelectronics, Peking University

³ School of Computer Science, Catholic University of Louvain

ABSTRACT

Generating high-quality stitched images with a natural structure is a challenging task in computer vision. Recent image stitching methods based on warps failed to suppress the distortion of the images. They often bend the salient lines in the image, which is inconsistent with human perception. In this paper, we succeed in proposing a novelty model called multi-perspective warps for natural image stitching which is related to the density of feature points in the images. With it we can get more precise matching results. Three new energy terms are developed to stitch quality to specify and balance the expected for aligning the vertices of the multi-scale mesh, which can constrain the transformation of the mesh. We also explore and introduce three feature point reconstruction algorithms to enrich the features in the images. Extensive experiments demonstrate that the proposed method outperforms most state-of-the-arts by effectively preserving the linear structure in the image and improving the robustness.

Index Terms— image stitching, multi-mesh, energy terms

1. INTRODUCTION

Image stitching is a process of combining multiple images with overlapping fields of view into a larger image with a wider field of view [1]. This technology can be of great use in areas as diverse as medical image processing [2], robotics [3] and daily photographic panoramas on smartphones [4]. However, it is still a challenging task to generate high-quality stitched images for state-of-the-arts as they are subject to various unpleasant effects such as artifacts and distortions. In early research, feature matching is the key point to get a more accurate global prior homography. For point features, SIFT [5] features including extracting and matching are widely used in many traditional algorithms [6] [7] [8]. Recent years, SURF features are often used due to its fast-running speed and pinpoint accuracy [9]. However, point features usually have spatially biased distributions that are insufficient for stable estimation of warp models in large parallax and/or low-texture situations where points are prone to mismatch [10]. To address this problem, Lin *et al.* [11] utilized both point and line features by different weights in



Fig. 1. Comparisons of stitching methods.

the objective function. Unfortunately, these methods match points and lines separately, thus the proposed method was not applicable to images with large parallax rendering resulting in unnatural artifacts in the stitched images. Liao *et al.* [12] employed the RANSAC strategy to refine point and line pairs by applying the homography between images. But their method still failed to resolve the conflict between distortion and linear structure. As shown in Fig. 1, the magnified overlapping area on the right exhibits artifacts on the spire and the doors. It is highly desirable to refine the mesh wrap and take the relevance of the lines and points into the consideration. Considering the mentioned problems, our solution for this dilemma is to propose a multi-scale mesh warp based on the point features. We convert the linear features into point features, and then the distribution of point features is used to construct a multi-scale mesh. The main contributions of this paper are: (1) To the best of our knowledge, we are the first to introduce multi-scale mesh into image stitching and provide accurate pre-alignment to significantly mitigate unnatural distortions. Since the grid is constructed taking into account the density of the features, it can provide finer stitching results. (2) We designed three corresponding energy function terms to achieve a refined warp, based on the multi-scale mesh. These energy terms are adapted to multi-scale mesh, so they can provide accurate pre-alignment while eliminating artifacts and distortion. (3) We applied three

^{*} Qi Jia is the corresponding author.

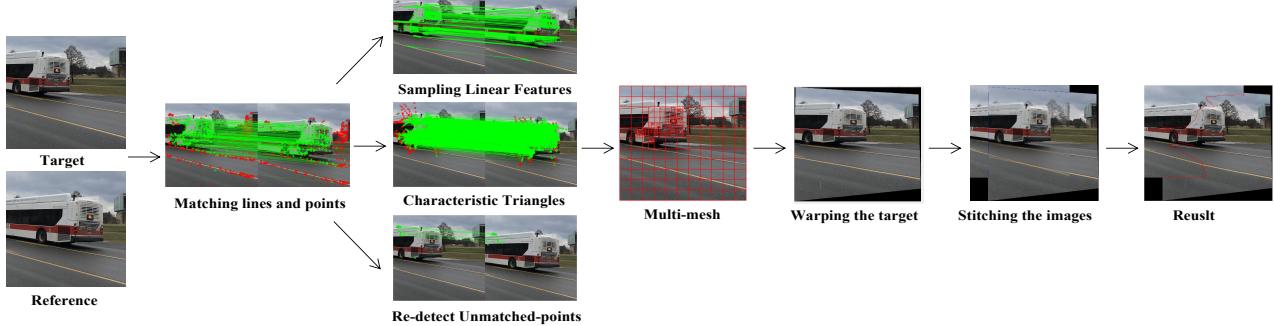


Fig. 2. Process of our model for image stitching base on multi-scale meshes

methods to take the expansion point features. (4) We compare the proposed approach with state-of-the-arts with prominent point-line structures.

2. METHODOLOGY

We denote I and I' as the target image and the reference image respectively. As shown in Fig.2, the process of our model for image stitching can be divided in 4 steps: (1) detecting the matching dual-feature (lines and points). (2) reconstruction the point features with the dual-feature we have acquired. (3) mesh distortion based on the density of the points. (4) image warping base on terms designed for the multi-mesh.

2.1. Reconstruction of Point Features

In this subsection, three methods are applied to reconstruct the feature points, which are demonstrated in Fig. 2. Since the multi-scale mesh is constructed by the density of the feature points, it is necessary to specify the texture of the image.

Sampling Linear Features. First, we use the robust line matching algorithm LSD [13] to detect the matching lines L_m . Then, we uniformly sample the matching lines to obtain equidistant pairs of points. Since these lines have already been matched, uniformly sampling an equal number of points on these lines ensures that these points will be matched. These col-linear points can represent the linear features of the image.

Re-detect Unmatched-points. First, we can obtain the local homography of the linear neighborhoods, which can be calculated by the matched points. Second, the unmatched points in the neighborhood can be mapped to the reference image. Third, we consider them as the re-matched points P_{rm} if the distance between mapped point and homologous point satisfies the constraint δ .

Characteristic Triangles. The triangle characteristic number is a special form of characteristic number [14]. According to the definition of characteristic number, we can use five points to construct a triangle and equal intersections on each edge. As shown in the upper Fig. 3, l_j and l'_j are the matching lines in the original images, $K_{l_j}^1$ and $K_{l_j}^2$ are two interest points

on line l_j and $K_{l_j}^1$ and $K_{l_j}^2$ are two interest points on line l'_j , which are considered as the endpoints or the intersections with other matching lines closest to the endpoints. P_1 is the base point while P_2 and P_3 are two of the randomly selected non-col-linear matching points. Above the five key points in a same neighborhood $K_{l_j}^1$, $K_{l_j}^2$, P_1 , P_2 , P_3 , any three points are not col-linear.

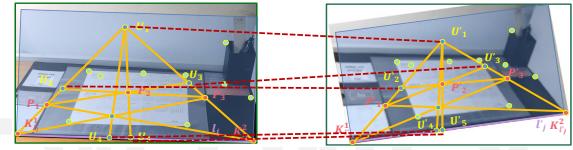


Fig. 3. Characteristic triangles.

We denote the line through two points P_i and P_j as $\overline{P_iP_j}$, the intersection of two line $\overline{P_iP_j}$ and $\overline{P_kP_m}$ as $\langle \overline{P_iP_j}, \overline{P_kP_m} \rangle$. We can obtain a set of intersections by five key points: $U_1 = \langle \overline{P_1K_l^1}, \overline{P_3K_l^2} \rangle$, $U_2 = \langle \overline{P_1K_l^1}, \overline{P_2P_3} \rangle$, $U_3 = \langle \overline{P_1P_2}, \overline{P_3K_l^2} \rangle$, $U_4 = \langle U_1 \langle \overline{P_1K_l^2}, \overline{P_3K_l^1}, l \rangle, U_5 = \langle \overline{U_1P_2}, l \rangle$ [15]. Thus, we have $\Delta U_1, K_l^1, K_l^2$ in the neighborhood of line l and $\Delta U'_1, K_l^1, K_l^2$ in the neighborhood of line l' . With it we are able to get five new pairs of feature points U_1, U_2, U_3, U_4, U_5 can be reconstructed by the characteristic triangle. It signify that we can generate $C_{N_j-1}^2$ characteristic triangles P_{rc} in this neighborhood.

2.2. Multi-Scale Mesh

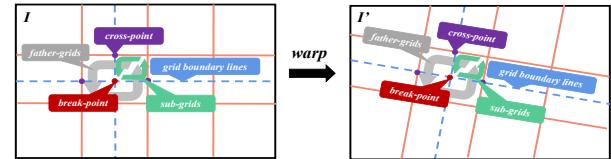


Fig. 4. Basic configuration for multi-scale mesh.

To begin with, we construct multi-scale mesh for each im-

age pair. The multi-scale mesh is constructed by the density distribution of matched points (including the point features and the points reconstruction). We perform fine meshing on areas with high feature point density. Thus, we set a threshold ξ as $\xi = \xi_{alpha} \times \rho$, where ρ denotes the average the density of matching point and ξ_{alpha} represents the coefficient.

When the number of matched points in the grid is greater than the threshold, we divide the grid into sub-grids. In a multi-scale mesh model, we call the original grid parent-grids. As shown in Fig. 4, it is noted that the demarcation point of the four sub-grids is called the break-point, and the demarcation point at the edge of the parent-grid is called the cross-point. In addition, the straight lines which are col-linear with all boundary segments of each grid are called the grid boundary line.

2.3. Energy Function Definition

In order to obtain warping results that are more adapted to multi-scale grids, we apply three new energy functions.

Break-point Alignment term. We define the set of break-points P_b in image I as $P_b = p_i^b \{i=1,2,\dots,P^b\}$, $P_i^b = (x_i^b, y_i^b)$, where P_i^b denotes the break-points in image I , which is settled by the property of multi-scale. We denote the warped break-point as \widehat{P}_i^b . We construct the break-point alignment term $E_b(\hat{V})$ based on the above inference

$$E_b(\hat{V}) = \sum_{i=1}^{P^b} \|\tau_{fa}(\widehat{P}_i^b) - \tau_*(\widehat{P}_i^b)\|^2 = W_b(\hat{V}), \quad (1)$$

where $\tau_{fa}(\widehat{P}_i^b)$ denotes the bi-linear interpolation calculated by the parent-grid and $\tau_*(\widehat{P}_i^b)$ denotes the bi-linear interpolation of itself, $W_b \in \mathbb{R}^{2P^b \times 2n}$.

Cross-point Alignment term. We define the set of cross-points P_c in image I as $P_c = p_i^c \{i=1,2,\dots,4P^b\}$, $P_i^c = (x_i^c, y_i^c)$, where P_i^c denotes the cross-points in image I , which is settled by the property of multi-scale. We denote the warped cross-point as \widehat{P}_i^c . It is convenient to observe that the number of the cross-points is four times the number of break-points. We can infer that the vector starts with a cross-point and ends with a break-point, which should be perpendicular to the line where this intersection is located. We construct the cross-point alignment term $E_c(\hat{V})$ based on the above inference

$$E_c(\hat{V}) = \sum_{i=1}^{P^b} \sum_{e(i)=1}^4 \|\tau_*(\widehat{P}_i^b) - \tau_*(\widehat{P}_{e(i)}^c) \cdot \vec{d}_p^e\|^2 = W_c(\hat{V}), \quad (2)$$

, where \vec{d}_p^e and $e(i)$ denote the direction vector of the line where the cross-point is located and the four cross-points of i th break-point partly, $W_c \in \mathbb{R}^{4P^b \times 2n}$.

Col-liner term. We define the col-liner term as $E_{colliner}(\hat{V}) = E_H(\hat{V}) + E_V(\hat{V})$, where $E_H(\hat{V})$ and $E_V(\hat{V})$ denote the horizontal and vertical col-liner term separately. Obviously,

each cross-point is the intersection of two grid boundary lines. Therefore, we define the set of grid boundary lines as $L_H = l_j^h \{j=1,2,\dots,L_H^M\}$, $L_V = l_k^v \{j=1,2,\dots,L_V^M\}$, where l_j^h represents the horizontal grid boundary lines and l_k^v represents the vertical grid boundary lines in image I . We sample the grid boundary lines with equal spacing. There are P_H^j points on each horizontal grid boundary lines l_j^h and P_V^k points on each vertical grid boundary lines l_k^v . These sampling points are represented as $P_H^j = p_f^{h,j} \{f=1,2,\dots,P_H^j\}$, $P_V^k = p_f^{v,k} \{f=1,2,\dots,P_V^k\}$. The vector starts at a sampling point located on the grid boundary lines and ends at the cross-points, hence it should be perpendicular to the normal vector of this line. Thus, we construct the horizontal col-liner $E_H(\hat{V})$ term and the vertical col-liner term $E_V(\hat{V})$ as

$$\begin{cases} E_H(\hat{V}) = W_{colliner}(\hat{V}) \\ = \sum_{i=1}^{P^b} \sum_{e(i)=1}^4 \sum_{f=1}^{P_H^j} \|\tau_*(\widehat{P}_f^{h,j}) - \tau_*(\widehat{P}_{e(i)}^c) \cdot \overrightarrow{N^{h,j}}\|^2 \\ E_V(\hat{V}) = W_{colliner}(\hat{V}) \\ = \sum_{i=1}^{P^b} \sum_{e(i)=1}^4 \sum_{f=1}^{P_V^k} \|\tau_*(\widehat{P}_f^{v,k}) - \tau_*(\widehat{P}_{e(i)}^c) \cdot \overrightarrow{N^{v,k}}\|^2 \end{cases} \quad (3)$$

, where $\overrightarrow{N^{h,j}}$ is the normal vector of l_j^h , $W_{colliner}(\hat{V}) \in \mathbb{R}^{2(\sum_{i=1}^{P^b} \sum_{e(i)=1}^4 \sum_{f=1}^{P_H^j}) \times 2n}$ and $\overrightarrow{N^{v,k}}$ is the normal vector of l_k^v , $W_{colliner}(\hat{V}) \in \mathbb{R}^{2(\sum_{i=1}^{P^b} \sum_{e(i)=1}^4 \sum_{f=1}^{P_V^k}) \times 2n}$.

3. EXPERIMENT

3.1. Experiments Settings

We demonstrate the effectiveness of the proposed method with an ablation study and with quantitative and qualitative comparisons to the state-of-the-art testing image pairs. Input images pairs remain unchanged in its original size and the size for each mesh grid is set as 75×75 . Our code is implemented in MATLAB and experiments are run on a desktop PC with Intel i7 2.8GHz CPU and 16GB RAM. More details are included in the supplementary file.

3.2. Comparison with the state of arts

Quantitative comparisons. We use images from the testing set of SPHP [8], APAP [10], AANAP [16], SPW [12], ELA [17], DHW [11], SVA [18] and GSP [19] datasets to make comparisons. We take RMSE as evaluation metrics to quantify the performance of models. As listed in Table 1, our method outperforms all other methods in these metrics on these datasets. In terms of speed, we can get stitching results in a few seconds, and compared with deep learning methods, the upfront training cost is greatly reduced.

Qualitative comparisons. A comprehensive visual comparison is demonstrated in Fig. 7. We found that existing

Datasets	SVA	DHW	CPW	APAP	GSP	SPW	Ours
APAP-railtracks [10]	7.30	14.12	6.77	4.66	4.58	3.76	3.17
DH-temple [8]	12.21	6.84	2.54	2.04	2.21	2.31	1.94
APAP-consite [10]	11.36	6.30	7.06	5.83	5.01	5.46	2.26
APAP-garden [10]	8.98	2.26	6.36	1.75	4.15	1.74	1.71
SVA-chessgirl [18]	20.78	4.08	9.45	4.21	2.88	1.88	1.83
SVA-rooftops [18]	4.11	4.95	3.54	2.82	3.05	3.74	2.73
AVG-RMSE	10.79	6.43	5.95	3.55	3.65	3.15	2.27

Table 1. RMSE on matched feature points

methods exhibit severe curvature and produce significant artifacts when stitching all three types of images. Our method, on the other hand, shows amazing results, with the stitching of leaves, person and buildings structures significantly better than other methods in terms of clarity and without significant distortion, perfectly keeping the line structure clear.

3.3. Ablation Study

Multi-scale mesh. We validate the effectiveness of it by comparing it with the ordinary mesh (equal-scale mesh) module (SPW), and add an energy function adapted to the mesh, and other parts remain unchanged. As shown in Fig. 6, our approach with multi-scale mesh produces clear stitching results.

Multi-mesh energy terms. The distribution of the lower-right image is chaotic, which did not introduce any energy terms. Cross-point terms are introduced into the lower-left image which maintains the stability of some cross-points, but the distribution of dense areas is still disordered. Break-point alignment term and cross-point alignment term are introduced into the image of the upper-right image, which has aligned most vertices but there's still some confusion. All of the terms are introduced into the the upper-left image, which has aligned all of the vertices fitly, which are shown as Fig.5.

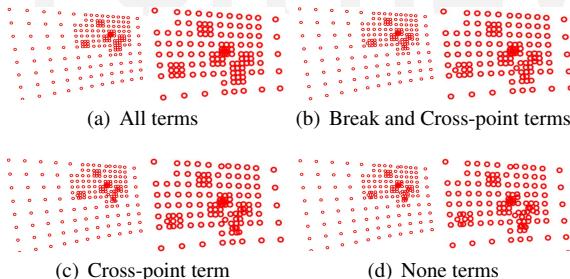


Fig. 5. The ablation study of new energy terms.

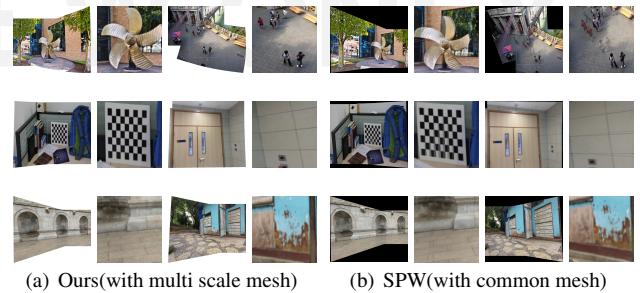


Fig. 6. Ablation study on multi-scale mesh stitching.



Fig. 7. Comparison of different stitching methods.

4. CONCLUSION

We propose a structure-preserving image stitching method based on the multi-scale mesh. We partition common mesh into sub-regions according to the density of feature points and match the images with three new energy terms which are designed for this novelty model. Furthermore, to have higher robustness and adaptability to this model, we applied three methods for feature point reconstruction. Experimental results demonstrate that the proposed method can accurately align overlapping and non-overlapping regions on challenging test images and has a significantly better performance.

5. REFERENCES

- [1] Richard Szeliski et al., “Image alignment and stitching: A tutorial,” *Foundations and Trends® in Computer Graphics and Vision*, vol. 2, no. 1, pp. 1–104, 2007.
- [2] Savita Singla and Reecha Sharma, “Medical image stitching using hybrid of sift & surf techniques,” *International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)*, vol. 3, no. 8, pp. 838–842, 2014.
- [3] Abhishek Kumar Dewangan, Rohit Raja, and Reetika Singh, “An implementation of multi sensor based mobile robot with image stitching application,” *Int J Comput Sci Mobile Comput*, vol. 3, no. 6, pp. 603–609, 2014.
- [4] Yingxin Xiong and Kari Pulli, “Sequential image stitching for mobile panoramas,” in *2009 7th International Conference on Information, Communications and Signal Processing (ICICS)*. IEEE, 2009, pp. 1–5.
- [5] David G Lowe, “Distinctive image features from scale-invariant keypoints,” *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [6] Weiqing Yan, Chunping Hou, Jianjun Lei, Yuming Fang, Zhouye Gu, and Nam Ling, “Stereoscopic image stitching based on a hybrid warping model,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 9, pp. 1934–1946, 2016.
- [7] Fan Zhang and Feng Liu, “Parallax-tolerant image stitching,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3262–3269.
- [8] Che-Han 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*, 2014, pp. 3254–3261.
- [9] Shuaicheng Liu, Lu Yuan, Ping Tan, and Jian Sun, “Bundled camera paths for video stabilization,” *ACM Transactions on Graphics (TOG)*, vol. 32, no. 4, pp. 1–10, 2013.
- [10] 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*, 2013, pp. 2339–2346.
- [11] Shiwei Li, Lu Yuan, Jian Sun, and Long Quan, “Dual-feature warping-based motion model estimation,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 4283–4291.
- [12] Tianli Liao and Nan Li, “Single-perspective warps in natural image stitching,” *IEEE transactions on image processing*, vol. 29, pp. 724–735, 2019.
- [13] Rafael Grompone Von Gioi, Jérémie Jakubowicz, Jean-Michel Morel, and Gregory Randall, “Lsd: A line segment detector,” *Image Processing On Line*, vol. 2, pp. 35–55, 2012.
- [14] Francis D Murnaghan, “On the unitary invariants of a square matrix,” *Proceedings of the National Academy of Sciences*, vol. 18, no. 2, pp. 185–189, 1932.
- [15] 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/CVF conference on computer vision and pattern recognition*, 2021, pp. 12186–12195.
- [16] Chung-Ching Lin, Sharathchandra U Pankanti, Karthikeyan Natesan Ramamurthy, and Aleksandr Y Aravkin, “Adaptive as-natural-as-possible image stitching,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1155–1163.
- [17] Jing Li, Zhengming Wang, Shiming Lai, Yongping Zhai, and Maojun Zhang, “Parallax-tolerant image stitching based on robust elastic warping,” *IEEE Transactions on multimedia*, vol. 20, no. 7, pp. 1672–1687, 2017.
- [18] Wen-Yan Lin, Siying Liu, Yasuyuki Matsushita, Tian-Tsong Ng, and Loong-Fah Cheong, “Smoothly varying affine stitching,” in *CVPR 2011*. IEEE, 2011, pp. 345–352.
- [19] Yu-Sheng Chen and Yung-Yu Chuang, “Natural image stitching with the global similarity prior,” in *European conference on computer vision*. Springer, 2016, pp. 186–201.