

Large-parallax Image Alignment Algorithm based on Sub-plane Segmentation

Haoming(Hammond) Liu
Class of 2023 - Computer Science
DURF Project Mentor: Prof. Xianbin Gu

Abstract

An alignment algorithm is proposed for stitching large-parallax images. The algorithm segments the regions that could reflect real situations of sub-planes. Experimental results demonstrate that the proposed algorithm provides accurate stitching results for images with large parallax, and the effect of stitching are quantitatively equivalent to the existing mainstream method.

1. Introduction

Image stitching is an important technique among diverse computer vision applications which aligns multiple images captured from different viewing positions onto a common coordinate domain to generate a wide-angle and high-resolution image. Currently, this kind of technique has been widely implemented in numerous fields, such as remote sensing images, medical images, aerial images, and virtual reality.

When the images are taken within the same plane or by stable rotation, the parallax of the images is relatively small, which can be easily stitched by simple computation and transformation. However, the parallax of input images in real life usually lies in a greater scale due to the influence of equipment and human. With different depths of field, the parallax of the input images would be inevitably enlarged and therefore influence the effect of stitching. Hence, the current researches on image stitching committed to alleviating the artifacts caused by parallax.

As a research field with decades of history, researchers around the world have gradually developed a systematic frame to stitch images with greater parallax. In fact, most of the conventional image stitching methods follow similar procedures [3]. Feature points are first detected from a pair of input images, and their correspondence matches are found between the images. Then parametric image warping models are estimated by using the detected feature matches, which warp a target image onto a reference image domain. Finally, we composite an output stitched image by determining the pixel values in the overlapped areas between the

warped target image and the reference image.

Since images with greater parallax contain three-dimensional spatial information in a two-dimensional plane, they always have projection distortion after certain transformation. Therefore, how to align images with greater parallax is the biggest challenge of image mosaic research.

To align the images with parallax, researchers have proposed many excellent models. Gao and Kim proposed a model that divided the image into two main planes, namely the ground plane and the distant plane, and it used the dual homography transformation to project the corresponding areas adaptively according to the positions of feature points [2]. Their model performs quite well on the datasets that comply with its segmentation conditions, yet for other images or images with much larger parallax, the accuracy of alignment would be insufficient. Zaragoza *et al.* proposed As-Projective-As-Possible (APAP) approach [5]. It divided an input image into regular grid cells and estimated an optimal homography for each cell by moving direct linear transformation (MDLT), which assigned more weights to the feature points spatially closer to the target cell while computing alignment error. Chang *et al.* applied the homography transformation to the overlapping regions and applied the similarity transformation to the non-overlapping regions respectively, which solved the problem of projection distortion in non-overlapping areas to some extent [1]. Yet, the alignment of huge-parallax images still relied on APAP algorithm. Lin *et al.* proposed the applied Adaptive As-Natural-As-Possible (AANAP) method [4]. It applied projection, similarity, and affine transformation to different regions in the images respectively, which could deal with the projection distortion effectively and make the stitched images much more in line with the visual aesthetics. AANAP approach also borrowed the idea of APAP algorithm in alignment, using the MDLT method to compute local projection and segment projection. Recently, Lee and Sim proposed an image stitching algorithm that was robust to large parallax [3]. The algorithm is based on the novel concept of warping residuals and uses superpixels to segment and wrap different regions adaptively.

2. Large-parallax Image Alignment Algorithm based on Sub-plane Segmentation

2.1. Hypothesis & Experiment

As it is mentioned in the introduction section, images with parallax contains three-dimensional spatial information in a two-dimensional plane, which has increased the difficulty of alignment. Gao and Kim's algorithm divides the images into ground plane and distant plane, yet this method cannot apply to all types of data [2]. APAP approach divides the input images into $C_1 \times C_2$ (100×100) grid cells and project all cells adaptively [5]. Up till now, APAP approach is still one of the best image stitching algorithms for large-parallax images. However, the high time complexity of APAP algorithm usually requires much more time to align the images. Furthermore, with the grid cells being too small, the accumulation of projection distortion sometimes leads to unnatural image distortion.

It needs to be emphasized that either thousands of grid cells or the simple hypothesis of the ground plane and the distant plane may not reflect the sub-plane segmentation in real life. Hence, this project expects to find an approach to segment the regions that can better represent real sub-plane situations. In general, the alignment of images relies on the computation for feature points, so the distribution of feature points might be correlate with the situations of sub-planes.

Here are two general descriptions on the distribution of feature points:

- (a) If a set of pixels in a continuous region belong to the same plane, then the feature points that lie in this area also belong to this plane.
- (b) The matched feature points have dense inliers and sparse outliers.

Given (a) and (b), we can raise the ensuing hypothesis:

- (c) If an image contains multiple sub-planes, then the feature points inside these sub-planes have a denser and closer distribution.



Figure 1. Aligning with all feature points (global homography)

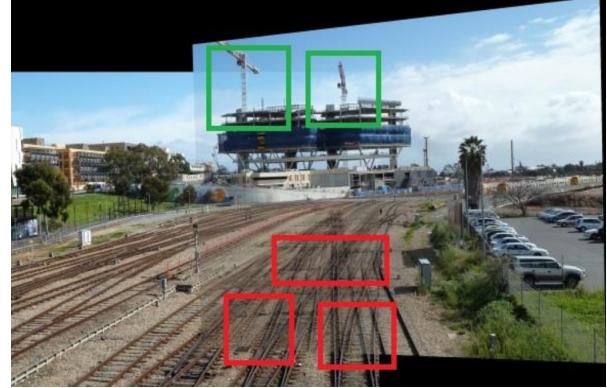


Figure 2. Aligning with the feature points that lie above the image

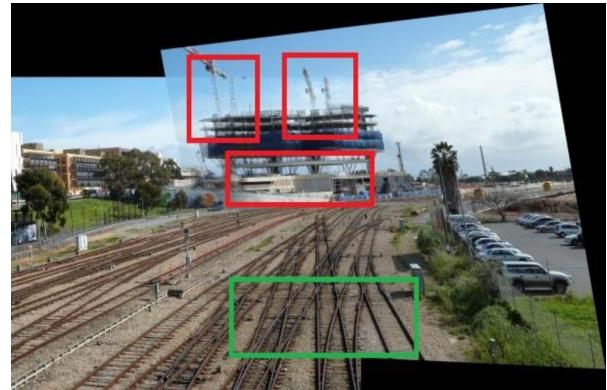


Figure 3. Aligning with the feature points that lie below the image

In the process of computing global homography, all the feature points (inliers) contribute to the parameters for projection, which has made a “compromise” between multiple sub-planes. When we align the images with the feature points inside a single sub-plane, the alignment accuracy of a specific sub-plane should be satisfying. Hence, local projection transformation is usually more accurate than global projection while aligning the images, and the segmentation of local regions, namely sub-planes, are be the cornerstone of accurate alignment.

2.2. Algorithm

2.2.1 Feature Extraction

Mathematically speaking, it is possible to represent high dimensional data in a feature space whose dimensions are less than that of the original data space. From the perspective of images, feature extraction is to extract the key and easily distinguishable information from the image. Currently, SIFT is a commonly used feature extraction algorithm for image stitching, so all the experiments proposed in this report use the SIFT algorithm in vlfeat library¹ to

¹<https://www.vlfeat.org/>

perform feature extraction. Due to limited mathematical knowledge, the implementation of the SIFT algorithm is not included as part of this project.

2.2.2 Eliminate mismatches (outliers) with RANSAC

Random sample consensus (RANSAC) is an iterative method to estimate parameters of a mathematical model from a set of observed data that contains outliers, so it is used as an outlier detection method. To eliminate the outliers, RANSAC algorithm finds an optimal homography 3×3 matrix H, where H needs to satisfy the most matching feature points of the matrix. In general, h_{33} is set as 1 to normalize the matrix, then the homography matrix H should have 8 unknown parameters, which means at least 8 linear equations are needed to solve. Given the point position information, a set of matching points can get two linear equations, so at least 4 sets of feature matching point pairs are needed to compute the global homography matrix H.

The procedures of the RANSAC algorithm:

- Randomly select 4 samples from the matched points (samples shouldn't be collinear), then compute the transformation matrix H, which is denoted as model M.
- Given model M, calculate the projection error of all other matches. If it is less than the set threshold, then add the sample data to the inliers set I.
- If the number of elements in the current inliers set I is greater than the optimal interior point set I_{best} , update I_{best} and the iteration number k.²
- If the number of iterations is greater than k, then exit; otherwise, increase the number of iterations by 1, repeat the steps above (The maximum of iteration is set as 1000).

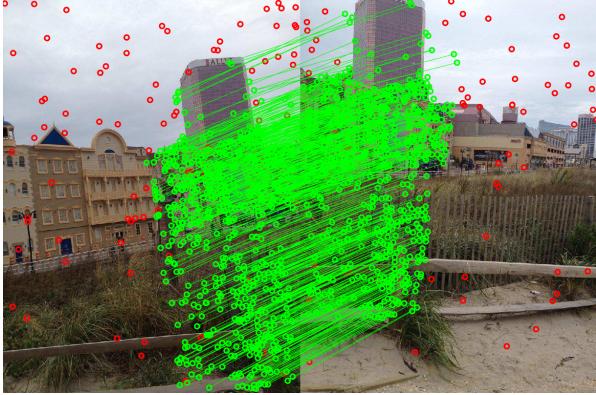


Figure 4. Inliers found by RANSAC algorithm

²p is the confidence level, usually 0.995; ω is the proportion of inliers; m is the minimum number of samples, namely 4; $k = \frac{\log(1-p)}{\log(1-\omega^m)}$.

2.2.3 Cluster the feature points with K-means

K-means algorithm is a relatively simple unsupervised learning method. It combines the points that are close in the sample and distinguishes them from the points that are far away. While processing feature points, K-means algorithm can classify densely distributed feature points into the same category, which meets the algorithm requirements of this project. In this step, we initialize a new canvas with the same size as the overlapping area, and cluster the feature points (inliers) in this area into K categories.

The procedures of the K-means algorithm:

- Randomly set K sample points as the initial centers.
- For each sample point, calculate the Euclidean distance between the point and all cluster centers, and assign it to the cluster center with the smallest distance.
- Re-compute the centroids of the clusters and define them as the new cluster centers.
- Repeat the above steps until the result converges.

Let $\{x_1, x_2, \dots, x_n\}$ denote the set of inliers, given the number of clusters K ($K \leq n$), we can divide the inliers into K categories $S = \{S_1, S_2, \dots, S_K\}$ that satisfy $\arg \min \sum_{i=1}^K \sum_{x_j \in S_i} \|x_j - \mu_i\|^2$, where μ_i denote the center of S_i .



Figure 5. Result of K-means clustering

2.2.4 Segment regions by clustering the pixels

In the previous step, we clustered the inliers into a certain number of categories, yet it is not enough for the segmentation of the overlapping area. To circumvent the problem that there no feature points in some regions, we adopt the principle of proximity to cluster the pixels.

Firstly, we delineate the rectangular boundary of a single cluster according to the extreme points in four directions.

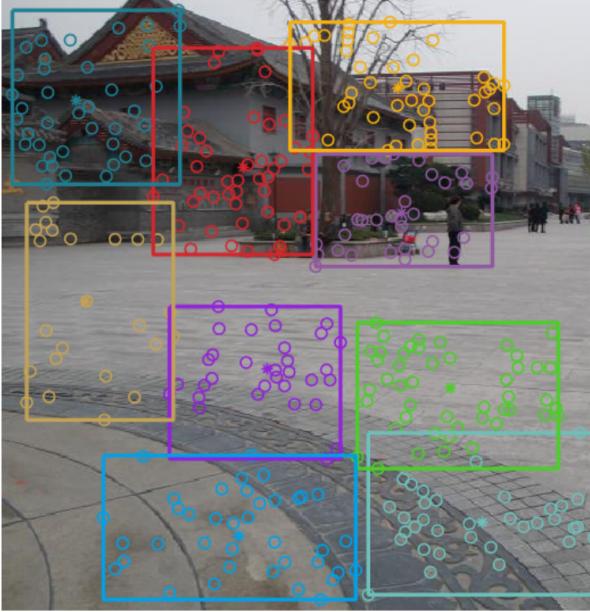


Figure 6. Rectangular boundaries of the clusters

Based on the boundaries of each cluster, the geometric centers of these rectangular regions are set as the cluster centers for pixel clustering. Here, we still use Euclidean distance as a measure of similarity. In this way, we finish the segmentation of regions, which is able to reflect the situation of real sub-planes with a properly chosen K.

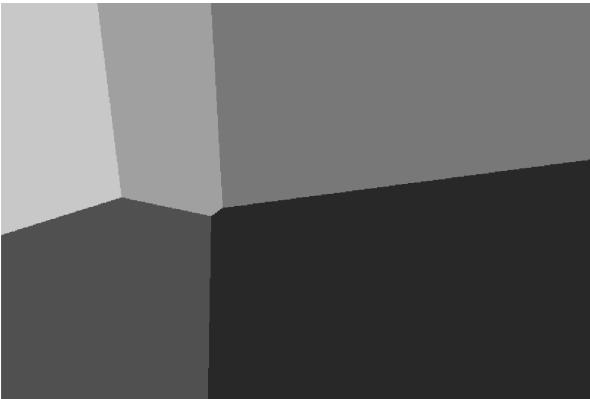


Figure 7. Result of sub-plane segmentation

2.2.5 Region completion & Local projection

After the segmentation, we can calculate the local homographies of different regions by RANSAC and draw the warped images by projecting local regions respectively. However, with different projection parameter H, applying projection transformation directly may sometimes lead to the existence of unfilled gaps in warped images.



Figure 8. Unfilled gap in warped images

To get well-filled images, we complete the polygon regions into rectangle regions, which could fill the gaps and reduce the runtime for processing the segmented areas. As for the non-overlapping area, we still apply the principle of proximity to cluster the pixels. With the blending algorithms, we can blend the warped images and get the stitched image.



Figure 9. Stitched result [Railtracks; Mine; K=14]

3. Experiment

All the datasets ($N = 37$) were collected from the work of Li *et al.*, which contains a collection of test datasets used by many other stitching approaches (e.g., APAP, SPHP, REW, etc.) [6]. The parameters of APAP were set as the default data in the open source code ($C_1 = C_2 = 100$, $\gamma = 0.1$, $\sigma = 8.5$). My approach set $K = 6$.

Table 1. Comparison of SSIM

Dataset	APAP	Mine
Boardingbridge	0.9645	0.9616
Building 1	0.8937	0.8925
Building 2	0.9339	0.9599
Cabin	0.8645	0.9172
Cabinet	0.9421	0.9144
Campus	0.9734	0.9735
Campussquare	0.9622	0.9613
Chessgirl	0.8660	0.8601
Corner	0.8626	0.8554
Desk	0.8822	0.8940
Door	0.9025	0.8865
Fence	0.9335	0.9301
Footpath	0.9367	0.9328
Foundation	0.9401	0.9317
Four	0.9193	0.9172
Garden	0.9507	0.9510
Guardbar	0.6831	0.7169
Gym	0.8798	0.8726
Intersection	0.8954	0.8913
Lawn	0.9538	0.9451
Maneki-neko	0.8202	0.8203
Memorial_hall	0.9758	0.9787
Park	0.9798	0.9814
Parksquare	0.9502	0.9553
Racetracks	0.8676	0.8549
Railtrack	0.7804	0.7832
Railtracks	0.9221	0.9395
Riverbank	0.9678	0.9644
Roof	0.9610	0.9534
Sculpture	0.8565	0.8562
Shelf	0.9496	0.9339
Temple	0.9225	0.9336
Theater	0.9448	0.9613
Tower	0.9246	0.9453
Wall	0.9100	0.9120
Window	0.9484	0.9250
Worktable	0.9255	0.9206
Average	0.9121	0.9131

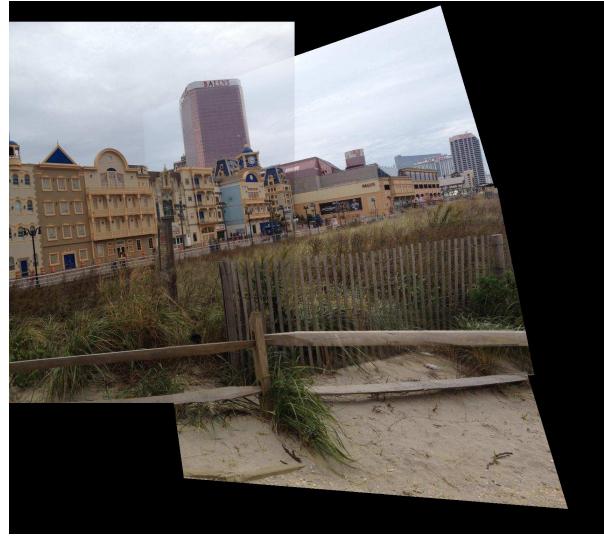


Figure 10. Stitched result (a) [Fence; Mine; SSIM=0.9301]

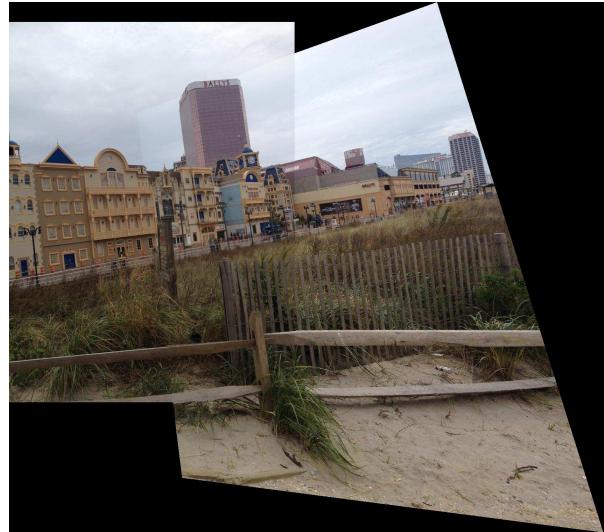


Figure 11. Stitched result (b) [Fence; APAP; SSIM=0.9335]



Figure 12. Stitched result (c) [Theater; Mine; SSIM=0.9613]



Figure 13. Stitched result (d) [Theater; APAP; SSIM=0.9448]

4. Conclusion & Future work

The accuracy of the proposed method is roughly the same as APAP method. Furthermore, APAP algorithm needs to adjust two relatively complex parameters (sigma and gamma), whereas the sub-plane segmentation method in this report only needs to adjust the number of regions (K). Thus, the proposed method is more applicable to some extent. Since the proposed method is essentially to segment the image based on the distribution of feature points. It will inevitably affect the accuracy of alignment and stitching.

Also, the proposed algorithm can be optimized in two different ways. In the procedure shown in section 2.2.5, it sometimes causes artifacts in the overlapping areas because the polygon regions are completed to the rectangular regions directly, so how to complete the raw segmented regions still worth further discussion. Besides, a hierarchical clustering might be helpful to develop an algorithm that can choose K automatically. In this way, we could obtain the better results without adjusting the parameter manually.

References

- [1] C.-H. Chang, Y. Sato, and Y.-Y. Chuang. Shape-preserving half-projective warps for image stitching. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3254–3261, 2014.
- [2] J. Gao, S. Kim, and M. Brown. Constructing image panoramas using dual-homography warping. pages 49 – 56, 07 2011.
- [3] K.-Y. Lee and J.-Y. Sim. Warping residual based image stitching for large parallax. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
- [4] C. Lin, S. U. Pankanti, K. N. Ramamurthy, and A. Y. Aravkin. Adaptive as-natural-as-possible image stitching. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1155–1163, 2015.
- [5] J. Zaragoza, T. Chin, Q. Tran, M. S. Brown, and D. Suter. As-projective-as-possible image stitching with moving dlt. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 36(7):1285–1298, dec 2014.
- [6] F. Zhang and F. Liu. Parallax-tolerant image stitching. pages 3262–3269, 06 2014.