

Research on panoramic mosaic of multiple images

Haoxuan Liu

University of Science and Technology Beijing, Beijing, China

42021214@xs.ustb.edu.cn

Abstract. With the continuous improvement of the manufacturing process of mobile phone cameras, the demand for panoramic photos by the industry and the public is not limited to ordinary cylindrical panoramic images, so a series of accurate panoramic image stitching technologies have been derived. However, these techniques often fail to take into account the stitching effect and speed, and there is still a technical gap in current multi-image panoramic stitching techniques. This article mainly compares the accuracy and efficiency of various feature extraction algorithms to select the most suitable algorithm for obtaining feature point pairs and estimating the homography matrix. Finally, a dual band hybrid algorithm is used to distort and fuse multiple images into panoramic photos. The results show that the panoramic images generated with multiple images using the algorithm provided in this paper are very good, even though there is a little bit of light and dark interlacing and ghosting.

Keywords: multi-image, panorama mosaic, feature extraction, image blending.

1. Introduction

With the rapid development of computer graphics related technologies, the requirements for new technologies in various fields are also increasing. Panoramic stitching integrates image registration, image fusion and other technologies. It also involves optics, computer graphics, stereo vision and other scientific fields. Panoramic stitching is a hot research direction at present, and its development quality plays a vital role in medical, material, military, aerospace and other fields. At present, the panoramic image shooting function provided by mobile phones or cameras only generates a "flat view" of the scene, without stereoscopic and realistic sense, so the development of panoramic splicing technology will also bring benefits to photography. So far, many algorithms have been developed in the field of panoramic mosaic, but these algorithms cannot give consideration to both speed and accuracy, and there is also great room for improvement in image fusion. Therefore, the research and improvement of panoramic image mosaic technology is of both academic and practical significance.

The image mosaic requires numerous processing techniques, including boundary stitching, mixing, transform estimates across pictures, image warping into mosaic surfaces, and so forth [1]. Panoramic mosaic technology processes multiple images with overlapping parts, estimates the transformation matrix (a.k.a. Homography) of the coordinate system of two images through the coordinate matching of feature points between adjacent images, and then uses the transformation matrix to "distort" one image to another image, so that we can intuitively see that the field of vision of the entire image is broadened. Because the material pictures of the panorama are taken at the same position with different rotating states of the camera, it has a strong stereoscopic sense and excellent visual experience.

This paper mainly studies the selection and optimization of feature point matching and image fusion technology in panoramic mosaic algorithm, so as to realize a set of panoramic mosaic algorithm pipeline with high efficiency and accuracy, and highlight the problem of how to build a smoother and more realistic panoramic image with more material pictures. First, we will introduce the process of feature matching using SIFT. Next, we will use the Direct Linear Transform [2] algorithm to calculate the Homograph and use the Random Sample Consensus [3] algorithm to filter the interior points to optimize the calculation results of Homograph. After warping each image to the plane of the central image, select the appropriate fusion algorithm for image fusion to make the panorama look more natural.

2. Methods

2.1. Data collection

The experimental data is a group of pictures taken by the camera of the mobile phone around the horizontal line. In the process of obtaining the material, try to keep the camera position unchanged (the optical center of the camera corresponding to each picture almost coincides), only change the rotation angle, and shoot at the same time to ensure that the brightness and light line remain unchanged, otherwise it will lead to serious blurring effect of the mosaic panoramic image.

The process of obtaining data in this experiment is low cost and convenient, and the quality of the data obtained is relatively good due to the advantage of the advanced camera function of the current smart phone. Figure 1-3 are a part of the captured images.



Figure 1. Source Images (Stone, 13 in total).



Figure 2. Source Images (Sunshine, 5 in total).



Figure 3. Source Images (Different Brightness, 15 in total).

2.2. Algorithm

The majority of image stitching algorithms follow a similar process: after estimating the warping or transformations needed to align the overlapping images, the aligned images are composited into a single canvas [4]. More specifically, it can be divided into the following steps: feature extraction and matching, Homography computation and RANSAC optimization, Homography computation from each image to the central image, image mosaic and fusion. The specific process is shown in Figure 4.

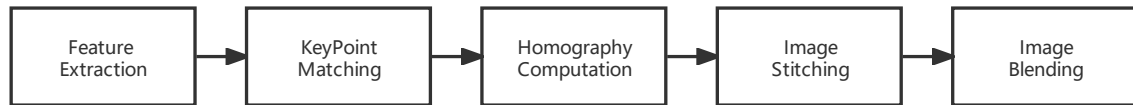


Figure 4. General process of panoramic splicing.

2.2.1. Feature extraction and matching

● SIFT Algorithm

SIFT is a feature extraction algorithm with scale invariance, and it can be concluded that SIFT algorithm has the following advantages: First off, despite changes in rotation, zoom, and illumination in the input photos, the inclusion of invariant features permits accurate matching of panoramic image sequences [5]. Table 1 shows the main steps of the SIFT algorithm.

SIFT has good stability and invariance, which means that it can ignore interference of affine transformation and noise. The use of SIFT algorithm in the multi-photo panoramic mosaic project also benefits from its distinguishing information quickly and accurately and generating a large number of eigenvectors. Figure 5 shows the generation process of SIFT descriptors.

Table 1. SIFT detector algorithm.

SIFT algorithm
1) Scale-space extremum detection: search for image positions on all scales. Identify potential points of interest that are invariant to scale and rotation by means of Gaussian differential functions.
2) Key point localization: At each candidate location, the location and scale are determined by fitting a fine-grained model. Key points are selected based on their stability. Low-contrast candidates and edge candidates are eliminated in the keypoint localization step.
3) Orientation determination: One or more orientations are assigned to each keypoint location based on the local gradient orientation of the image. All subsequent operations on the image data are transformed with respect to the orientation, scale and position of the keypoints, thus providing invariance with respect to these transformations.
4) Key point description: The gradients local to the image are measured at selected scales in the neighborhood around each key point. These gradients are transformed into a representation that allows for relatively large local shape distortions and illumination changes.

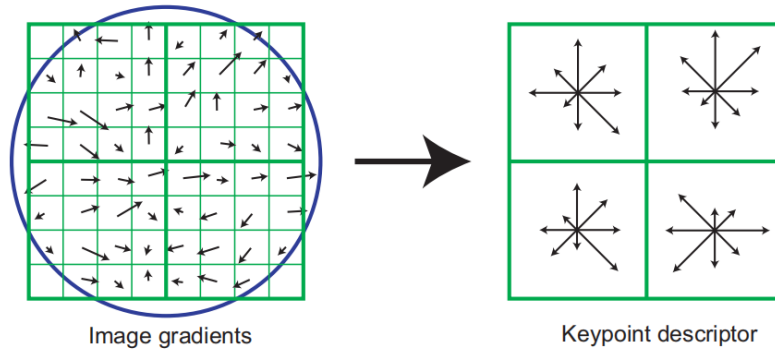


Figure 5. Keypoint descriptor created by gradient magnitude and orientation.

- KnnMatch Algorithm

This paper uses the knnMatch algorithm of $k=2$ as the descriptor matching algorithm. It finds the two feature points closed to the original image feature points, and only when the Euclidean distance between the two feature points is less than a certain value, will the matching be considered successful.

2.2.2. Homography computation and optimization

Although the cylindrical coordinate projection method is relatively easy to realize in the link of mosaic and image fusion, it is often only applicable to pictures that are far away from the camera and pictures that do not require strong perspective effect. Therefore, in this section, the method adopted by the project is to calculate the Homegraph matrix between adjacent pictures, and "twist" the pixel points in all pictures into the same plane through Direct Linear Transform coordinate transformation, as to form a panoramic view with strong perspective effect.

- Direct Linear Transform

Figure 6 describes the relationship between the coordinates of a point on the plane X_π on different projection planes. According to the relevant knowledge of Homegraph $X_1 \equiv HX_2$, we can get the expression with the shape of $Ah = 0$ through the transformation based on linear algebra. The A matrix is decomposed by SVD, and the column vector corresponding to the minimum singular value is obtained, which is the 9×1 column vector composed of the elements in the Homegraph matrix.

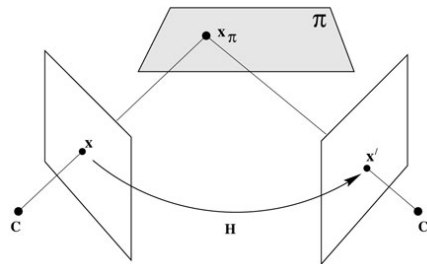


Figure 6. A homography H links all points x_π lying in plane π between two camera views.

- RANSAC algorithm

RANSAC is a frequently used algorithm for optimizing computer vision parameters. It can calculate the mathematical model parameters of data based on a set of sample data containing abnormal data, and obtain valid sample data. It is used to optimize the result of feature point matching in panoramic stitching, and to obtain the fundamental matrix in stereo vision. Table 2 shows the execution process of the RANSAC algorithm.

Table 2. RANSAC optimization algorithm.

RANSAC algorithm	
1)	Randomly select the number of sample points that can estimate the minimum number of points of the model (in the Homography Estimation algorithm, since the Homography matrix has 8 degrees of freedom, at least 4 pairs of feature points need to be selected here to generate the A matrix)
2)	Calculate the model parameters (elements of the Homography matrix)
3)	Bring all data into this model and count the number of inliers (inliers: use the Homograph matrix point to multiply the homogeneous coordinates of the point in the source image to obtain the estimated coordinates of the corresponding point in the target image, and then calculate the difference between the estimated coordinates and the accurate coordinates. If the difference is less than the threshold value, it means that the point is an interior point, otherwise it is an exterior point)
4)	Based on the number of interior points, maintain an optimal (maximum number of inliers) model (i.e., the Homegraph matrix). When the number of inliers in the new loop exceeds the currently maintained number of inliers, update the maximum value of inliers and the estimated Homegraph matrix.
5)	Repeat the above steps until the convergence (the number of inliers reaches the preset value) or the iteration reaches a certain number of times

2.2.3. Image stitching and blending

● Image stitching

One of these images must be transformed into a common field using the view transformation matrix H after the RANSAC model fitting comparison is finished. Give each of the photographs a different perspective in this case [6].

After computing the homography between each two adjacent images, it is necessary to warp each image onto the plane of the central image. The method used here is to dot multiply all the homography from the side image to the middle image to obtain the coordinate system conversion matrix from the side image to the central image, and then calculate the coordinates of each pixel point of the current image under the central image coordinate system through dot multiplication, as shown in Figure 7.

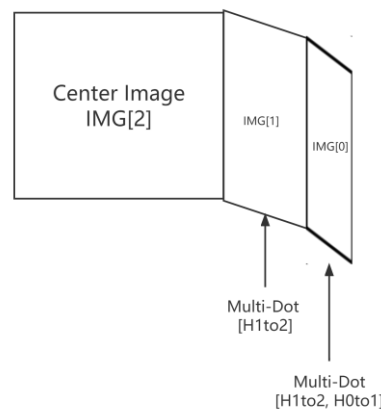


Figure 7. Warp each image into center image plane.

● Image blending

The image frames on the mosaic image are warped progressively and quickly in accordance with the camera's position. After the seam is found, due to factors such as image noise, illumination, exposure, and model matching error, direct image synthesis can produce significant edge traces at the stitching of

the overlapping regions of the image. Adaptive multi-band blending is used to produce stitching results that are aesthetically acceptable. The weight matrix and the image's Laplacian pyramids are kept and modified accordingly [7].

The idea of the multi-band blending algorithm is to directly decompose the two images to be spliced into Laplacian pyramids, and merge the latter half with the first half. In the project, because the material images are all taken at the same time and place, the difference between the images is not significant, so the algorithm used is two-band blending.

In this work, a common mean value fusion algorithm is also used for comparison. By comparing effect and time consumption of the resulting images generated by the two sets of methods, it is determined which image fusion algorithm is more appropriate to use. Figure 8 shows the basic principles of the two fusion algorithms.

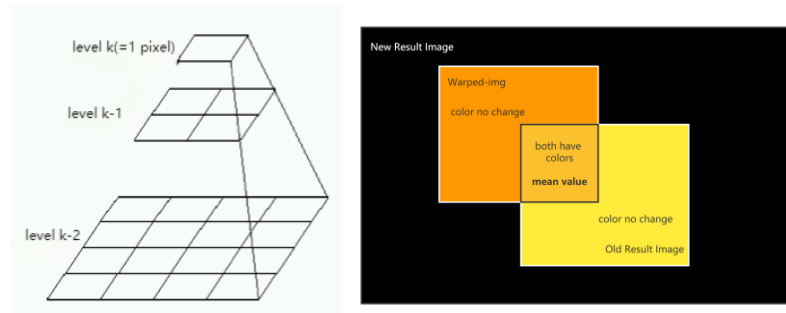


Figure 8. Two-band blending and uniform blending used in project.

3. Experiment

3.1. Platform introduction

In this report, the running environment of the project is Python 3.10.10 64-bit, and the external libraries are OpenCV-Python 4.7.0.22 and numpy 1.22.4. The IDE hosting the project is Visual Studio Code 1.71.2.

3.2. Results

3.2.1. Feature extraction and keypoint matching. The collection of image materials taken by mobile phones offers more detailed features that can be recognized by feature extraction algorithms due to the growing availability and sophistication of digital imaging technology (digital cameras, computers, and photo editing software), as well as the popularity of the Internet [8].

Table 3. The comparison results of image 1(Stone, 13 in total).

Detector	Key points	Match rate	Computation time(s)
SIFT	5237	36.10%	0.2834
ORB	6162	28.87%	0.3115
KAZE	4094	54.28%	1.0987
AKAZE	2544	57.57%	0.1611
BRISK	10973	26.59%	0.1685

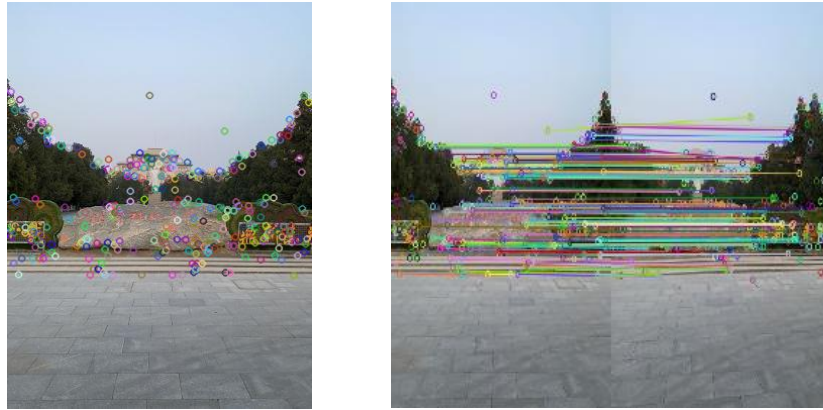


Figure 9. Feature Extraction and Keypoint Matching in image series 1.

Table 3 shows the number of feature points generated by different feature extraction and matching algorithms. The processing results from image series 1 (shown in Figure 9) indicate that the size relationship between the number of feature points generated by each image matching algorithm is: BRISK > ORB > SIFT > KAZE > AKAZE. The matching results are: AKAZE > KAZE > SIFT > ORB > BRISK. In terms of calculation time: AKAZE < BRISK < SIFT < ORB < KAZE, but the calculation time of these algorithms in this dataset is mostly less than 2 seconds, so there is little impact on the project in terms of speed.

From the above experimental data, it can be concluded that algorithms that consume more time have higher accuracy, but the calculation time is not directly related to the number of feature points generated. However, algorithms with a larger number of feature points generally have a lower matching degree (some of the generation of feature points is invalid). Based on the above analysis, this multi-image panoramic mosaic project selects the SIFT algorithm with balanced indicators as the feature extraction algorithm, and its corresponding image matching algorithm is the KnnMatch algorithm in the FLANN library, where $K = 2$.

3.2.2. Stitching. Figure 10 and Figure 11 are the results of direct stitching and the results obtained using two band blend (pixel RGB average fusion) processing, respectively. It can be clearly seen from the images that the results obtained using the direct stitching method have relatively obvious black seams, while using the two-band blend method can greatly eliminate the seams. The reason for this phenomenon is that the image fusion using pyramid decomposition is processed in different frequency bands, enabling a fusion effect that is closer to human visual characteristics [9].

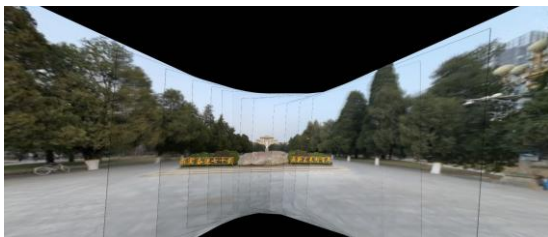


Figure 10. Results of direct splicing.

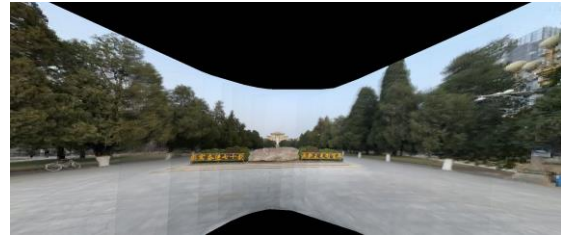


Figure 11. Results from using two-band blending.

Figure 12 shows the processing results of image materials taken under strong light intensity conditions. It can be seen that the multi-image stitching algorithm in this article has a relatively good processing effect for images with the same exposure (regardless of the intensity of light).

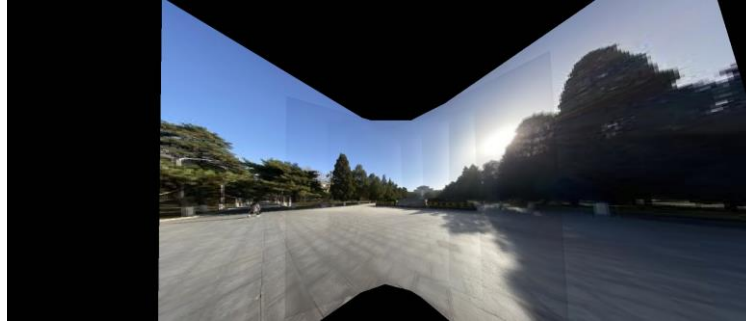


Figure 12. Splicing Results of Sunshine Datasets.

However, in Figure 13, we can see that there is a shading phenomenon between each material image. Although this is a small problem in image fusion, it also proves the correctness of the previous assumption: the camera used to take material images must have a consistent exposure. Overall, this issue has little impact on the multi-image mosaic algorithm in this article, and the overall effect is still acceptable.

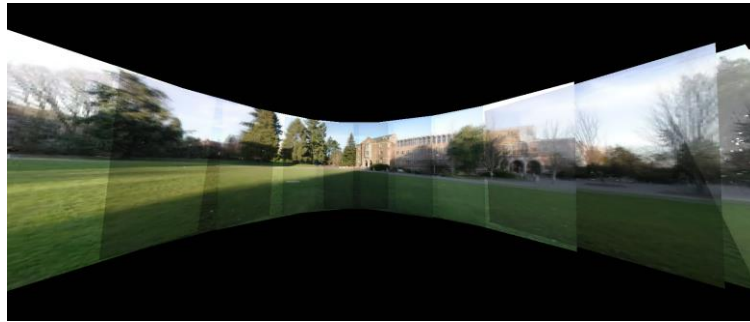


Figure 13. Splicing Results of Different Brightness Datasets.

In addition, we can see from the above panoramic mosaic results that there are still a small amount of ghosting and blurring between the various images, which is due to the slight displacement of the world coordinates of the camera used to collect the images. Images are related by a linear projective transformation known as a homography when the scene has a flat surface or when they were captured from the same angle [10]. Therefore, there are some ghosts in the panoramic image.

4. Conclusion

In this paper, we take the panoramic stitching of multiple images as the research goal, compare the differences between the effectiveness and efficiency of different algorithms in the feature extraction and matching, image fusion steps, and select the algorithm more suitable for doing panoramic stitching of multiple images. First, the SIFT algorithm is used to extract feature points between images and match them to generate feature point pairs. Then the RANSAC algorithm is used to reduce the estimation error of Homography. One image is selected as the central plane, and the other images are mapped to the plane of that central image. Finally, the two-band blending method is used to fuse the overlapping areas of these images. The results show that although fusion algorithms separate and fuse high-frequency and low-frequency information separately, shadow and ghosting effects may occur when there are objective interference factors such as significant changes in the brightness of the image, different camera exposures, and displacement of the device that captures the image. The relevant research results have certain reference value.

References

- [1] Ha, S. J., Koo, H. I. , Sang, H. L. , Cho, N. I. , & Kim, S. K. . Panorama mosaic optimization for mobile camera systems.2007 *IEEE Transactions on Consumer Electronics*, **53(4)**, 1217-1225.
- [2] Abdel-Aziz, Y. I., Karara, H. M. Direct linear transformation from comparator coordinates into object space in close-range photogrammetry. 1971 *American Society of Photogrammetry*, 1-10.
- [3] Fischler, M. A. , Bolles, R. C.. Random sample consensus. 1981 *Communications of the ACM*.
- [4] Zaragoza, J., Chin, T. , Tran, Q. , Brown, M. S. Suter, D.. As-projective-as-possible image stitching with moving dlt.2014 *IEEE Trans Pattern Anal Mach Intell*, **36(7)**, 1285-1298.
- [5] Brown, M. , & Lowe, D. G.. Automatic panoramic image stitching using invariant features. 2007, *International Journal of Computer Vision*, **74(1)**, 59-73.
- [6] Qi Fengshan & Jiang Tingyao. (2016). Improvement of QR code image angle point detection method based on Harris. 2016 *Software Guide* (**05**), 199-201.
- [7] Liu, X. , Yu, H. T. , & Chen, B. M. Adaptive Weight Multi-Band Blending Based Fast Aerial Image Stitching and Mapping. 2018 *15th International Conference on Control, Automation, Robotics and Vision*. 1-12.
- [8] Pan, X. , & Lyu, S. Region duplication detection using image feature matching. 2010 *IEEE Transactions on Information Forensics & Security*, **5(4)**, 857-867.
- [9] Liu Guixi,&Yang Wanhai Image fusion method and performance evaluation based on multi scale contrast tower, 2001 *Journal of Optics*, **21 (11)**, 7
- [10] Chiba, N. , Kano, H. , Higashihara, M. , Yasuda, M. , & Osumi, M. Feature-based image mosaicking, 1998 *Journal of Optics*, **19 (15)**.