Proceedings of the 3rd International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/6/20230635

Research of feature matching algorithm on panoramic mosaic

Ruyan Guo [1,4,†] Baiyi Xiao[2,†] and Shilun Zhang[3,†]

[1]School of Softwares, Heilongjian University, Harbin, China
[2]School of Information Science and Technology, Guangdong University of foreign studies, Guangzhou, China
[3]School of Foreign Languages, Southeast University, Nanjing, China

[4]20196135@s.hlju.edu.cn

[†]These authors contributed equally.

Abstract. Focusing on the application problem according to feature matching and panoramic image stitching, comparing direct stitching and smoothing after feature matching algorithms. Through time performance, synthesis accuracy, and synthesis efficiency, comprehensively compared their effects of synthesizing panoramic images to find a better solution for panoramic splicing. In the study, there are three groups of image pairs with parallax and brightness contrast, then use SIFT algorithm to conduct feature point and matching feature point. The RANSAC algorithm obtains the homograph matrix and completes the panoramic image direct splicing and smoothing test through the matrix transformation. Because of the luminance difference, the panoramic picture after direct splicing will show more obvious light and dark. Though the smoothing treatment could be performed rigidly in the light and dark difference, the deformation, and heavy shadow are easier produced when the disparity and brightness difference are large.

Keywords: feature matching, panoramic mosaic, SIFT, feature point extraction.

1. Introduction

With the progress of computer and computer graphics, and the growing demand for the contrast-wise fusion of images, also the requirements for computer graphics detection are gradually diverse and complex, and unavoidable, the need for computer image detection is coming specific and detailed. Panoramic splicing is a technology based on image stitching, which is one of the nowadays hot topics today. Panoramic splicing is a technique to produce a full and clear panoramic picture. It is based on the continuous photos taken in a specific environment, through picture splicing technology to process and compare the photos and other preliminary preparation. Finally, filter and match the feature points [1]. This technology could enable the generation of wide-angle 360 panoramic images. Not only bring people a more comfortable and special visual experience in the living field. But also plays an important part in education, medical care, the military, and many other fields. What is most common in nowadays picture splicing technology is the extracted feature by SIFT / SURF / OBR / AKAZE and some others, realizing alignment through RANSAC and some others, and alignment image splicing

DOI: 10.54254/2755-2721/6/20230635

based on image fusion or seamless cloning algorithm [2][3]. The image splicing can also be divided into two parts, one is similar pixel images the other is similar features.

Present panoramic picture splicing technology has vigorous development space. It combined technology and knowledge in multiple fields, including optics, computer science, computer vision, computer graphics, and so on. Its fast development could also bring many fields explored together [4]. Also, provide significant technology support for artificial intelligence and other fields.

This paper mainly focuses on matching algorithms and mosaic and fusion of the images. Introduce some algorithms for point feature detection, such as SIFT、RANSAC. Secondly introduce the main algorithm we had realized, and how to realize image splicing and optimize image splicing. Finally, SIFT feature matching experiment and image Mosaic and fusion are carried out and produce a summary report.

2. Method and technology

2.1. Data collection

Figure 1. Ceiling 001 (left) and 002 (right).

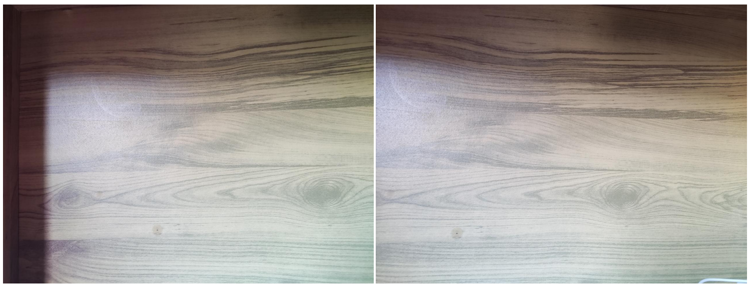

Figure 2. Wood grain 001 (left) and 002 (right).

Figure 3. Room 001 (left) and 002 (right).

With the development of science and technology, the growing demand for artificial intelligence in computer vision is increasing, people are devoted to the study of feature matching and panoramic Mosaic of VR images in computer vision, this algorithms for feature detection require convenient and efficient performance, and the accurate which feature matching needs to reach.

The main data is from the mobile phone camera, which changes its view to take a picture of an object. As well as the main advantage of mobile phones, it’s convenient, fast cost and low cost also cost camera distortion is large parallax, and brightness is obvious. The following figures (Figure 1-Figure 3) are the selected data sets.

2.2. Algorithm

The majority of Image Mosaic is divided into two parts: To complete docking and fusion, the first stage is feature point matching, which establishes the positional relationship between pictures. The second step is transforming all image projections to the same coordinate system [5]. SIFT feature detection and matching algorithm, RANSAC algorithm for deleting mismatched point pairs, picture stitching algorithm, and gap smoothing algorithm are the key algorithms used in this paper [6].

2.2.1. Feature matching. Feature point extraction and feature point matching are the two primary steps in image feature matching. With the SIFT algorithm, key points are extracted from the image as features; Feature matching is the calculation of Euclidean distance between feature vectors.

- Feature point extraction

SIFT Algorithm (Scale-Invariant Feature Transform) The fundamental objective of the SIFT technique is to locate key points in different scale areas and establish the direction of those key points. The image's important details may be recognized and is a local feature descriptor. The source picture and the target image are separately detected by feature points and described by feature points to generate the feature point set of the target. The key points are then compared, and the corresponding points are corrected. To achieve feature matching, there are main procedures in table 1.

Table 1. SIFT algorithm procedure.

SIFT algorithm
1) Extract key points, that is, some points that will not disappear due to illumination, scale, and rotation, such as corner points, edge points, bright spots in dark areas, and dark spots in light areas. This step is to search for image positions on all scale Spaces， Potential key points with invariant scale and rotation are identified by the Gaussian function, a series of reduced-size images were obtained by dropping the sampling, and then the extreme values of DOG space were detected to remove some edge response points;
2) Locate the key points and determine the feature direction, The choice of key points depends on the degree of stability. Gradient amplitude and gradient direction based on local image, assign one or more directions to each key point location, and all posterior operations on image data are transformed concerning the orientation, scale, and position of key points, to provide invariance to these transformations;
3) Through the feature vectors of each key point, the pairwise comparison is carried out to find out several matching pairs of feature points, and the corresponding relationship is established.

- Feature point matching

Because the data in the SIFT feature vector are floating-point numbers, the similarity of the key-point descriptors is calculated using the Euclidean distance. The two feature points are considered to have successfully matched when the distance is below a predetermined threshold.

Key point descriptors in image 1 of the data set：$R_i = (r_{i1}, r_{i1}, \cdots, r_{i128})$;

Key point descriptors in image 2 of the data set：$S_i = (s_{i1}, s_{i1}, \cdots, s_{i128})$;

Similarity measure of any two descriptors： $d(R_i, S_i) = \sqrt{\sum_{j=1}^{128} \left(r_{ij} - s_{ij}\right)^2}$

We will get matching pairs of feature points in the two images when the feature points have been matched. But, some of the feature point pairs we get will be mismatched. Matching point pairs must be eliminated, RANSAC is frequently used to eliminate incorrect matching point pairs, and the position relationship is improved to ensure that feature point matching point pairs are accurate.

The RANSAC algorithm is a straightforward and efficient algorithm used in estimating models to reduce the effect of noise (Table 2). Given a set of n data points, assuming that the majority of the points can be generated by a model and that at least M points can be used to fit the model parameters, the parameters can then be fitted repeatedly in the following ways. n iterations later, the largest model was selected as the fitting result.

Table 2. RANSAC algorithm procedure.

RANSAC algorithm
1) Select n data points randomly from them;
2) Estimate parameter x to calculate the transformation matrix;
3) Use this data point to fit a model;
4) The remaining data points' distance from the model should be calculated. An outlier point is one when the distance is greater than the cutoff. An intra-office point is one where the value does not exceed the threshold. Consequently, identify the model's corresponding intra-office point value.

2.2.2. Image stitching and smoothing

- Image stitching process

Once the RANSAC model fitting comparison is complete, the view transformation matrix H is given, and one of these images must be converted into a common field. In this instance, give one of the images a perspective modification [7]. For the most part, perspective transformations can be combined as rotation, scaling, translation, or clipping. Create a new image by merging two existing ones. The detailed information is listed in Table 3.

Table 3. Image stitching procedure.

Image stitching
1) Image display function;
2) Read the input image;
3) Detect SIFT key feature points of images A and B, and calculate feature descriptors;
• Build SIFT generator;
• Detect SIFT feature points and calculate descriptors;
• Converts the result to an array;
• Return the set of feature points and remember the corresponding description feature.
4) Set up matcher;
5) Use KNN to detect SIFT feature matching pairs from A and B images, K=2;
6) When the matching point pairs after screening are larger than 4, the view transformation matrix is calculated, and H is the view transformation matrix of 3×3;
7) Match all the feature points of the two images and return the matching result;
8) If the return result is empty, it proves that there is no feature point matching the result and exits the program;
9) Otherwise, the matching result is extracted;
10) Transform the view Angle of image A, and the result is the transformed image;
11) Pass image B to the left end of the resulting image.

DOI: 10.54254/2755-2721/6/20230635

According to the transformation matrix between images, the corresponding images can be transformed to determine the overlapping area between images, and the fused image can be mapped to a new blank image to form a Mosaic. However, due to the brightness difference between the input images, the mosaic image will appear with obvious bright and dark changes, so based on the mosaic, the Alpha channel is used for feather processing to smooth the image gap.

- Gap smoothing process

Alpha channels are used to describe an image's transparency and semi-transparency (table 4). Red, Green, Blue, and Alpha are represented by the color space RGBA [8-9]. Alpha rendering, Alpha composition, and this crucial alpha number can be created using numerical computation. Utilize the alpha factor. The value at the center pixel is 1, decreasing linearly with the boundary pixel becoming 0 following. When at least two photos are utilized in the output and it depends on the image, the alpha value is used to calculate the color at one of the pixels.

$$I = \alpha F + (1 - \alpha)B \tag{1}$$

The actual pixel value of the alpha template is normalized from [0,255] to [0,1]. From the above equation we know:

1. When A =0, the output pixel value belongs to the background.
2. When A =1, the output pixel value belongs to the foreground.
3. When 0<a<1, the output pixel value is a mixture of the front and background, why the fusion effect is more natural, usually the pixel value on the boundary of the alpha template is between 0 and 1.

Table 4. Gap smoothing procedure.

Gap smoothing
1) Separate the channels; 2) When the matching point pairs after screening are larger than 10, calculate the perspective transformation matrix; Take one of the images as the foreground and remove the alpha channel part in this image; 3) Obtain the alpha channel of the image, namely the alpha mask; 4) Normalize the value of alpha between 0 and 1 as the weighting coefficient; 5) The foreground and background are weighted, and the weighting coefficient of each pixel is the value of the pixel corresponding to the alpha mask; 6) The foreground part is 1, and the background part is 0; 7) Start overlapping the leftmost end; 8) Overlap the rightmost column.

3. Experiment

3.1. Platform introduction

The platform configuration parameters used in this report are the compiler and version vscode 1.70.1, using OpenCV in the python language as a framework, creating a virtual environment using Anaconda, and installing Python 3.10.0 in the virtual environment [10]. The code has been uploaded to the Code Cloud (gitee): https://gitee.com/susica/image-stitching.git)

3.2. Splicing results

The feature matching and image splicing algorithms were studied and optimized in this project. To avoid errors, this project repeated multiple comparisons, and different matching feature point pairs, matching results, and splicing fusion results were collected. All the algorithms were developed in the python environment.

3.2.1. Direct splicing results. The results of two different perspectives of the data set are shown in Figures 4-6.

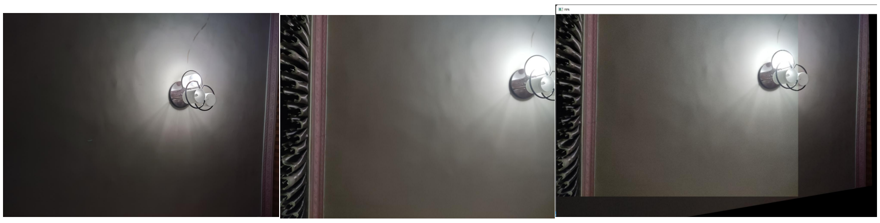

Figure 4. (a) Ceiling image 001 (left) and (b) 002 (middle) (c) directly spliced ceiling image (right).

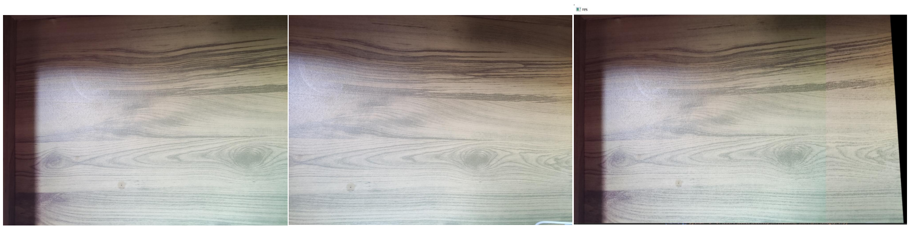

Figure 5. (a) Wood grain image 001 (left) and (b) 002 (middle) (c) directly stitched wood grain image (right).

Figure 6. (a) Room image 001 (left) and (b) 002 (middle) (c) Room image directly stitched together (right).

The above three sets of data sets show the two original images to be stitched and the images formed after direct splicing processing. It can be seen that the results of direct splicing are relatively accurate. However, due to the difference in the brightness of the camera, there is an obvious gap in all three images, especially the ceiling result in Figure 4. Because the brightness difference between the two images (ceiling images 001 and 002) is the largest, and the brightness difference after direct splicing is also the largest.

3.2.2. Smooth processing result. Direct splicing results and overlap at the leftmost end, and overlap by column, are shown in Figure 7-9.

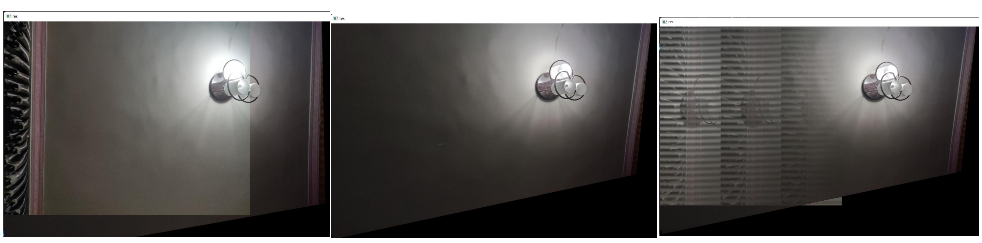

Figure 7. (a) ceiling splicing (left) and (b) overlap left most (middle) (c) overlap by column (right).

Figure 8. (a) Direct stitching (left) and (b) overlap the most left end (middle) (c) overlap by column (right).

Figure 9. (a) Room direct splicing (left) and (b) overlap far most left (middle) (c) overlap by column (right).

The above three data sets show the results of direct splicing processing and smoothing processing (overlapping the leftmost end and overlapping by column). It can be seen that the bright and dark area is significantly eliminated when smoothing processing, but the deformation and double shadow will still occur because the matrix is a single stress matrix rather than a global single stress matrix.

3.2.3. Fusion time comparison. During the result processing, the direct splicing times and smoothing times were recorded for the three group images, as shown in Table 5.

Table 5. Compares the demonstration images at the splicing and fusion time.

	Direct splicing time (the time is the average value)	Smooth processing time
Ceiling image	0.41s	7.90s
Wood pattern image	0.49s	8.49s
Room image	0.46s	9.46s

3.3. Results and discussions

3.3.1. Experimental results and causes analysis. We verified three points through experimental images and performance tables:

- In the image splicing, the SIFT algorithm is better than the ORB algorithm. Although the ORB algorithm is faster, image splicing mainly requires the stability of the rigid body transformation and the reduction of mismatching;
- Images were smoothed better than direct splicing, with higher accuracy but faster direct splicing in splicing time than smoothing. The high accuracy is due to the Alpha channel processing between the image channels, which makes the superposition and fusion of the two maps smoother and smoother.
- When smoothing, the leftmost method is prone to deformation, but overlap by column is prone to residual shadow. This is because, in the image mapping global single stress transformation, the image often has a different depth, of the different depth of characteristic points (ceiling curtain in

figure 7), because the object in the image cannot meet the same plane, the calculated matrix error will be larger, along with a single stress transformation cannot fully describe the transformation relationship between the images.

3.3.2. Experimental comparison and analysis. In this experiment, the performance of the three groups of data was analyzed, mainly through the key points and matching time, to obtain the performance comparison diagram (Table 6).

Table 6. Compares the performance of each splicing for the demonstration images.

	001 Figure Key Points	002 Figure Key Points	Match key points	Direct splicing time	Smooth processing time	Matching rate	Fusion efficiency
Ceiling image	220	1126	220	0.41s	7.90s	19.5%	2.47%
Wood pattern image	892	614	614	0.49s	8.49s	68.8%	8.10%
Room image	510	808	510	0.46s	9.46s	63.1%	6.67%

Combined with chart comparison, each group of experiments' smooth processing time is greater than the direct splicing time, in addition, on the matching rate, the ceiling image disparity and brightness are large, so the matching rate (matching rate in this table calculation formula: matching rate = two key points more matching points), the same fusion efficiency (fusion efficiency in this table calculation formula: fusion efficiency = weighted fusion time matching rate) is the lowest, the fusion image is worse than other groups. The wood grain and room image matching rate reach more than 60%, so the fusion efficiency is also high. The room images have a larger parallax than the wood image, but less luminance than the ceiling image, so the fusion efficiency is in the middle. The wood grain images had less parallax; thus the matching rate and fusion efficiency were the highest among the three experiments.

This leads to the following conclusion:

- Parallax and brightness are large, so that the key point matching rate and fusion efficiency are low, resulting in the largest fusion image and visual image comparison;
- The disparity is large but the brightness is consistent, the key points are identified, and the matching rate is high, so the fusion efficiency is relatively good;
- Lax small brightness is consistent and is optimal in both matching rate and fusion efficiency.

4. Conclusion

In this project, for feature matching and image splicing, first use SIFT to match key points, using the RANSAC algorithm to obtain a single stress matrix, through the matrix transformation to determine the overlap area between images, but found in the splicing of the images due to the different brightness information between the two images will appear obvious bright gap. Then, Alpha channel normalization was performed to make the image fusion smoother. In the above process, the more efficient algorithm and splicing method were found by comparing the experimental data set. It is concluded that under the SIFT algorithm, the gaps and visual errors are almost not found for the fusion by using the processing method of gap smoothing of the images with small brightness differences in disparity.

References

[1] Huang Tianyi. (2021). Research and implementation of panoramic image splicing algorithm based on feature matching (master's thesis, Nanjing University of Posts and Telecommunications).

[2] Wang Bo, Guan Yonghong & Liu Yang. (2021). Research on indoor panoramic image registration algorithm based on improved SIFT. Electronic Design Engineering (06), 83-87.

DOI: 10.54254/2755-2721/6/20230635

doi:10.14022/j.issn1674-6236.2021.06.018.
[3] Chen Xiaodu. (2020). Research on panoramic image splicing algorithm based on improved AKAZE and RANSAC (Master's dissertation, Nanchang University).
[4] Yi Caiqin. (2020). Research on the Development of Panoramic Image Splicing Technology. Information and Computer (Theory edition) (14), 149-151.
[5] Bao Yiming, Chen Qi, Zhu Zhongjin, Chen Yi & Lu Wei. (2022). Research and application practice of VR panoramic technology. Radio and Television Network (05), 101-103.
[6] Zhao Wenbin & Zhang Yanning. (2006). Review of corner-point detection techniques. Computer Application Research (10), 17-19 + 38.
[7] Qi Fengshan & Jiang Tingyao. (2016). Improvement of QR code image angle point detection method based on Harris. Software Guide (05), 199-201.
[8] Liu Wenyu, Zhu Guangxi. (2000). Morphological skeleton method for corner point detection in binary images. Signal processing (03), 276-280.
[9] Liu Yawei, Li Jian for & Zhang Xiaohong. (2010). An LoG angle point detection based on the edge contour line. Computer Engineering and Application (10), 140-143.
[10] Qu Tianwei, An Bo & Chen Guilan. (2010). Application of the improved RANSAC algorithm in image registration. Computer Applications (07), 1849-1851 + 1872.