Panoramic image mosaic method based on image segmentation and Improved SIFT algorithm

Lei Zhuang, Jiyian Yu, Yang Song*
Mechanical Engineering, NJUST, Nan Jing, Jiangsu Province, 210000, China
*Email: SongY126@njust.edu.cn.

Abstract. Aiming at the problem of large amount of calculation in extracting image feature points in panoramic image mosaic by SIFT algorithm, a panoramic image mosaic algorithm based on image segmentation and Improved SIFT is proposed in this paper. The algorithm fully considers the characteristics of panoramic image stitching. Firstly, the stitched image is divided into blocks, and the maximum overlapping block of image pairs is extracted by using mutual information. The SIFT key points are extracted by SIFT algorithm, and the dog is filtered before the spatial extreme value detection of SIFT algorithm to eliminate the feature points with small intensity value; When establishing the feature descriptor, the 128 dimension of the original algorithm is reduced to 64 dimensions to reduce the amount of calculation. In the feature point registration process, the feature descriptor is reduced to 32 dimensions, the feature point pairs are roughly extracted by the optimal node first BBF algorithm, and the feature point pairs are registered and screened by RANSAC; Finally, the image transformation matrix is obtained to realize panoramic image mosaic. The experimental results show that the proposed algorithm not only ensures the panoramic mosaic effect, but also extracts the feature points in 11% of the time of the traditional SIFT algorithm, and the feature point registration speed is 27.17% of the traditional SIFT algorithm.

1. Introduction
Panoramic images are spliced from actual scene images, which can display a wide range of information. At present, panoramic images have been widely used in remote sensing image processing, military applications, medical image analysis, machine vision and other fields. Two or more images taken from different positions and angles are spliced into a panoramic image with wide viewing angle, which has the characteristics of large field of view and high resolution [1]. Image registration is a key technology in panoramic image Mosaic and it is a process of finding geometric relations between different images [2]. Image registration methods are mainly divided into region-based and feature-based registration methods [3]. The image registration method mainly uses the feature-based registration method [4], which uses the feature information in the image to establish the geometric relationship between the image and the image, and has feature invariance under the conditions of image scale transformation, translation, rotation and so on.

SIFT first proposed SIFT algorithm in 1999 and optimized in 2004. As the algorithm has good robustness under the conditions of image affine transformation, perspective change, noise and illumination, it has gradually become a panoramic image Mosaic based on feature points Research hotspot of algorithm. Li Yufeng et al. [1] proposed an image Mosaic algorithm based on region segmentation and SIFT features, which only divided the image into 4 equal parts and extracted feature points of similar regions using SIFT algorithm. Li Jia et al. [5] proposed a video panorama Mosaic
method based on image block matching, which divides images into blocks according to the number of processors, uses multi-thread to match and search sub-images, and establishes the relationship between the sub-images. The speed of this method is too dependent on the performance of hardware. Zhao Yan et al. [6] divided the overlapping area into 128×128 image blocks and only selected the two image blocks with large standard deviation and distributed in different positions to detect their SIFT feature points. This method needs to adjust the size of sub-image blocks according to different images and has poor adaptability. Li Liang et al. [7] reduced the DIMENSION of SIFT feature descriptor from 128 to 64, and changed the direction of feature descriptor to 0, 1/4π, 1/2π and 3/4π, but did not include the direction between 3/4π and 2π. Li Dan et al. [8] reduced the dimension of the feature vector by reducing the size of the window centered on the feature point, from the original 128 dimensions to 24 dimensions. In this method, the reduction of the window centered on the feature point is easy to make the feature point descriptor affected by illumination and Angle of view changes.

Based on the above deficiencies, this paper proposes a panoramic image Mosaic method based on image segmentation and improved SIFT, which extracts the overlapping areas of images and reduces the feature descriptor from 128 dimensions to 32 dimensions. This method can obtain the panoramic image quickly and stably while retaining the CORE SIFT algorithm.

2. The general process of stitching

SIFT has a large amount of calculation when establishing feature point descriptor, and the real-time performance of panoramic image mosaic is not strong. According to the mosaic characteristics of panoramic images, only the feature points of overlapping parts are effective. Therefore, the images to be spliced are processed in blocks to extract the overlapping area of the image, which will greatly reduce the running time of the algorithm.

2.1. Regional block

The two blocking methods in Figure 1, 2 are selected to extract the feature points of four sub images A, B, C and D. It is concluded that the blocking strategy reduces the time used to search the feature points of similar areas by 1 / 3 compared with the traditional SIFT algorithm, makes full use of the mosaic characteristics of panoramic images, eliminates a large number of invalid feature points, and fully retains the feature points of overlapping areas.

2.2. Analysis of regional similarity

In this paper, the image after blocking is analyzed with mutual information as a similarity measure. Mutual information measures the similarity of two random variables, which is commonly expressed by joint entropy and individual entropy. H(a) and H(b) represent the entropy of image a and image B respectively, and H(a, b) represents the joint entropy of image a and image B. Then the joint entropy H(a, b) and individual entropy H(a) of the image are expressed as:

\[ H(A) = - \sum_a P_A(a) \log P_A(a) \]  
\[ H(A, B) = - \sum_{a,b} P_{AB}(a,b) \log P_{AB}(a,b) \]

Where, \( P_A(a) \) represents the probability distribution of A, \( P_{AB}(a, b) \) represents the joint probability distribution of A and B.

Then the mutual information of image A and image B can be expressed as:
\[ I(A,B) = H(A) + H(B) - H(A,B) \]  
\[ I(A,B) = \sum_{a,b} P_{AB}(a,b) \log \frac{P_{AB}(a,b)}{P_A(a)P_B(b)} \]  

In practical problems, the entropy values of different images vary greatly. NMI is used as a similarity measure. The normalized mutual information can be expressed by the following formula:

\[ \text{NMI} = 2 \frac{I(A,B)}{H(A) + H(B)} \]  

When the two images are completely consistent, the normalized mutual information is 1; When the two images are completely inconsistent, the normalized mutual information is 0.

3. Improved SIFT algorithm

The steps of generating image feature point feature descriptor by Improved SIFT algorithm include: 1) improved scale space extreme value detection; 2) Determination of the principal direction of feature points; 3) Improved feature point feature descriptor.

3.1. Improved scale space extreme value detection

The Gaussian pyramid of the image is used to realize the scale space of the image, establish the dog of the image, set the threshold, propose the points with low intensity in the image, and finally find the local extreme points in the dog scale space.

The image scale space \( L(x,y,\sigma) \) is defined as the convolution of the two-dimensional Gaussian \( G(x,y,\sigma) \) function with the original image \( I(x,y) \), \( \sigma \) is the image scale parameter. As follows:

\[ L(x,y,\sigma) = G(x,y,\sigma) \ast I(x,y) \]  
\[ G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \]  

Subtract two Gaussian images of the same group of adjacent scales to obtain the Gaussian difference scale space of the image, namely:

\[ D(x,y,\sigma) = (G(x,y,k\sigma) - G(x,y,\sigma)) \ast I(x,y) = L(x,y,k\sigma) - L(x,y,\sigma) \]  

Remove the points with small intensity value in Gaussian difference scale space, namely:

\[ |D(x,y,\sigma)| < K, \ K \text{ is the threshold} \]

If the above conditions are true, the pixel is updated to 0. As shown in Figure 4, the detection point is compared with 26 points in the field. If the detection point is a local extreme point, the point is tentatively regarded as a feature point. Finally, feature points with poor stability and feature points under edge effect are eliminated to reduce the error of image matching.

3.2. Determination of principal direction of feature points

In order to realize the rotation of feature points without deformation, the gradient of the neighborhood of feature points is used to describe the reference direction of feature points.

Figure 3. Dog pyramid  
Figure 4. Extreme point detection
\begin{equation}
m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2}
\end{equation}

\begin{equation}
\theta(x,y) = \tan^{-1} \frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)}
\end{equation}

Calculate the gradient amplitude and direction of each pixel in the neighborhood of feature points, and use histogram to represent the gradient and direction of pixels in the neighborhood of feature points. The gradient histogram divides 360° into 36 histograms on average, and each histogram represents the range of 10°. The direction with the largest histogram peak is selected as the main direction of the feature point, and the direction with the histogram peak greater than 80% of the maximum peak is reserved as the auxiliary direction of the feature point.

### 3.3. Determination of principal direction of feature points

SIFT algorithm takes a lot of time to establish the feature descriptor of feature points. In this paper, the 128 dimensional feature vector is simplified to 64 as the feature vector, and the four directions of 0, 1/2π, π, and 3/2π are selected to describe the feature points. The four directions of 1/4π, 3/4π, 5/4π, and 7/4π in SIFT algorithm can be projected in the four directions of 0, 1/2π, π, and 3/2π, which can not only reduce the dimension of feature vector, but also represent the characteristics of feature points in 360° direction. As shown in the figure 5:

![Figure 5. Feature descriptor dimensionality reduction graph](image)

The description of the sub region in the neighborhood of each feature point is changed from 8 directions to 4 directions. This method not only reduces the running time of the algorithm, but also retains the rotation invariance of the feature points, which can better realize the mosaic of panoramic images.

### 4. Feature point matching and registration

In this paper, an algorithm combining average dimensionality reduction with BBF (best bin first) is proposed. The feature points are roughly matched by this algorithm, and then the transformation matrix of the image to be spliced is obtained by RANSAC algorithm, so as to obtain the panoramic image through geometric transformation.

#### 4.1. The average dimensionality reduction method and BBF algorithm in this paper

BBF algorithm is an improvement on the k-dimension tree nearest neighbor search algorithm. It is based on KNN algorithm, and is mainly used for the search of key data in multi-dimensional space. However, the dimension of the feature descriptor of the Improved SIFT algorithm in this paper is as high as 64 dimensions. When processing high-dimensional data, the performance of BBF algorithm will be seriously degraded due to backtracking search, so this paper first reduces the dimension of the feature descriptor.

In this paper, an average dimensionality reduction method is adopted. Let the feature point descriptor be represented by matrix A of M * n (n is an even number). M is the number of feature points and N is the dimension of feature points. In this paper, n = 64, and each feature point has 64 dimensions. After the dimensionality reduction of the feature descriptor, the BBF algorithm is used to roughly match the feature points of the stitched image, eliminate the wrong matching points and improve the registration accuracy.
The average of odd and even dimensions of each feature point is obtained as the new first dimensional feature descriptor of the feature point, that is, the feature descriptor after dimension reduction can be expressed as:

\[ A_{m^*k} = \frac{A_{m^*i} + A_{m^*(i+1)}}{2}, \quad i \text{ is } 1, 3, 5 \ldots n-1, k = (i+1)/2 \]  \hspace{1cm} (11)

4.2. RANSAC obtains the transformation matrix

In this paper, RANSAC algorithm is used to further optimize the filtered feature point pairs. Through this algorithm, the transformation matrix of the feature point pairs of the two images can be obtained, so as to transform and splice the images with the transformation matrix to obtain the panoramic image.

Randomly select three pairs of feature points \( x \) of two images to be spliced \( X_1=(x_1, y_1)^T \), \( X_2=(x_2, y_2)^T \), \( X_3=(x_3, y_3)^T \) and \( X'_1=(x'_1, y'_1)^T \), \( X'_2=(x'_2, y'_2)^T \), \( X'_3=(x'_3, y'_3)^T \), then the feature point pair meets:

\[
\begin{bmatrix}
    x_1 \\
    y_1 \\
    x_2 \\
    y_2 \\
    x_3 \\
    y_3
\end{bmatrix} = H
\begin{bmatrix}
    x'_1 \\
    y'_1 \\
    x'_2 \\
    y'_2 \\
    x'_3 \\
    y'_3
\end{bmatrix}
\]

\[ H = \begin{bmatrix}
    a_{11} & b_{12} & c_{13} \\
    a_{21} & b_{22} & c_{23}
\end{bmatrix} \]  \hspace{1cm} (12)

Where, \( H \) is the affine transformation matrix of the image, \( a_{11}, a_{21}, b_{12}, b_{22} \) is the image rotation and scale transformation parameter, \( c_{13}, c_{23} \) is the offset parameter in the horizontal and vertical directions of the image.

5. Result and analysis

5.1. Feasibility verification of image segmentation

In this paper, two images to be spliced are selected for experimental analysis. The original size of the images is 574 * 324. According to the blocking method proposed in this paper, the spliced image is processed in blocks. The blocking results are shown in Fig. 4. The blocks a, B, C and D of the image are extracted for verification test.

![Figure 6. Image to be spliced 1](image1.png)

![Figure 7. Image to be spliced 2](image2.png)
Table 1. NMI values between block images

|   | A’    | B’    | C’    | D’    |
|---|-------|-------|-------|-------|
| A | 0.170500 | 0.093356696 | 0.082555360 | 0.07765363 |
| B | 0.3124302 | 0.1160261831 | 0.128396 | 0.097717 |
| C | 0.11396 | 0.0522 | 0.137880508 | 0.08615464410 |
| D | 0.10215183 | 0.069312 | 0.108013 | 0.098617 |

Mutual information is used to measure the similarity of images. The greater the normalized mutual information NMI value, the higher the similarity of images. As can be seen from table 1, block B in Fig. 6 and block A’ in Fig. 7 have the largest NMI and have the highest similarity. The panoramic image is spliced by extracting the feature points of the two blocks.

5.2. Data dimensionality reduction efficiency analysis

This algorithm reduces the feature descriptor dimension of the traditional SIFT algorithm, tests 13 images, and obtains the time distribution of this algorithm and the traditional algorithm in Fig. 8.

For a single image, the time used by this algorithm to extract feature points is much less than that of the traditional SIFT algorithm, which is about 11% of that of the traditional algorithm.

Figure 8. Time curve used to extract feature points

5.3. Registration efficiency comparison

In order to further reduce the time of panoramic image mosaic, the feature descriptor is reduced to 32 dimensions in the image registration stage, and 5 images to be spliced in different scenes are selected for comparative analysis.

Table 2. Comparison of two algorithms

| Picture number | Number of feature points to be spliced | Matching time | Number of correct matches |
|----------------|---------------------------------------|---------------|--------------------------|
|                | Figure 1 to be spliced | Figure 2 to be spliced | Paper algorithm (reduced to 32 dimensions) | Traditional SIFT algorithm | Paper algorithm 128 → 32 | Traditional SIFT algorithm |
| 1              | 1006 | 959 | 0.089 | 0.313 | 258 | 300 |
| 2              | 757 | 1136 | 0.077 | 0.285 | 98 | 122 |
| 3              | 763 | 958 | 0.067 | 0.242 | 128 | 143 |
| 4              | 997 | 1225 | 0.105 | 0.388 | 79 | 88 |
| 5              | 2101 | 1967 | 0.370 | 1.442 | 40 | 60 |

It can be seen from table 2 that the feature point pairs successfully registered by the dimensionality reduction algorithm in this paper are about 82.46% of that of the traditional algorithm, but the registration time is only 27.17% of that of the traditional algorithm, which greatly reduces the time cost. The registration results of the traditional SIFT algorithm are shown in Figure 9, and a total of 51
pairs of feature points are successfully registered. The registration results of the algorithm in this paper are shown in Figure 11, and a total of 42 pairs of feature points are successfully registered.

5.4. Panoramic image mosaic effect
Two images with the same scene are experimentally analyzed by using this algorithm and the traditional SIFT algorithm, and figures 11 and 12 are obtained. There is no obvious difference in the image quality of the two algorithms, and there is an obvious transition stage at the splicing.

6. Conclusion
Aiming at the problems of large amount of calculation in extracting image feature points and high dimension of feature point registration in panoramic image mosaic, this paper proposes an improved panoramic image mosaic algorithm, which blocks the image to be spliced, extracts the feature points of the image block with the largest coincidence degree, reduces the dimension of the feature points before registration. This method avoids the problems of slow speed and high dimension of the traditional SIFT algorithm, and reduces the time of the algorithm. Experimental results show that the proposed algorithm reduces the registration time by about 73% on the premise that the feature point pair of successful registration is 82.46% of that of the traditional algorithm. At the same time, there is no obvious difference between the panoramic image stitched by this algorithm and the panoramic image stitched by traditional SIFT algorithm.

References
[1] Li, Y.F., Li G.Z., Gu, S.H. (2016) Image mosaic algorithm based on region segmentation and scale invariant feature transformation. J. Optics and Precision Engineering., 5: 1197-1205.
[2] Xiao, M., Bao, Y.L., Yan, Z.X. (2015) Overview of image registration methods based on point features. J. Acta Armamentarii., S2.
[3] Barbara, Z., Jan F. (2003) Image registration methods: A Survey. J. Image and Vision Computing., 11: 977-1000.
[4] Lowe, D.G. (2004) Distinctive Image Features from Scale-Invariant Keypoints. J. International Journal of Computer Vision., 60(2): 91---110.
[5] Li, J., Duan, P., Zhang, C. (2018) Video panoramic mosaic method under image block matching. J. Journal of applied basic and Engineering Sciences., 26(4): 697-708.
[6] Zhao, Y., Chen, Y., Wang, S.G. (2017) Fast sift image mosaic combined with projection error correction. J. Optical precision engineering., 25(6): 1645-1651.

[7] Li, L., Gu, G.H., Qian, W.X. (2013) Infrared sequence image mosaic based on feature point and Poisson fusion. J. Infrared and laser engineering., 42(9): 2584-2588.

[8] Li, D., Sun, H.T, Wang, H.L. (2015) An Improved SIFT image stereo matching algorithm. J. Journal of Southwest Jiao Tong University., 50(03): 490-496.