Research on Different Feature Matching Algorithms for Panoramic Image Stitching

Zhao Zhang

College of Physics, Xidian University, Xi'an, 710000, China.
21009100551@stu.xidian.edu.cn

Abstract. Panoramic image stitching technology has penetrated into every field of modern life. As an important part of the stitching process, image feature matching directly affects the quality and speed of the stitching. In this paper, photos taken in daily life are used for experiments, and the precision and computational efficiency of three different feature matching algorithms, Scale Invariant Feature Transform (SIFT), Speeded-Up Robust Features (SURF), Oriented Fast and Rotated BRIEF (ORB), under rotation, scaling, light intensity transformation and perspective transformation, are compared to explore their applicable scenarios. The experimental results show that SIFT is most appropriate for perspective transformation, but its running speed is so slow that it is only suitable for occasions where the real-time requirement is not high. SURF has the greatest stability when dealing with scale changes and different light intensities, while it operates far quicker than SIFT. ORB exhibits the best robustness in the case of rotation and runs the fastest in all cases, so it is most suitable for applications in real-time scenarios.

Keywords: Image stitching; computer vision; feature detection and matching.

1 Introduction

Since computer science and image processing have advanced so quickly, there's an increasing need for more accurate and stable panoramic image stitching, especially for image feature matching. Panoramic image stitching is a process that merges multiple images with overlapping areas taken at the same location but from different perspectives or light conditions to form a wide-view and high-resolution image. Modern image stitching technology can automatically and effectively process a large number of images and is widely applied into fields such as medicine, military and computer vision, significantly improving processing accuracy and efficiency. In the procedure of panoramic image stitching, the quality of it is largely determined by the quality of feature detection and feature matching. Therefore, it is vital to select a more robust and faster feature matching algorithm for image stitching. In general, image feature matching consists of three key steps: detection of key points, extraction of description vector and feature matching. Since the demand for precision, speed and robustness of feature matching is increasingly strict, people have done a lot of researches on the algorithms of feature matching. Nowadays, the most popular

Y. Wang (ed.), *Proceedings of the 2024 2nd International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI 2024)*, Advances in Computer Science Research 115,
https://doi.org/10.2991/978-94-6463-540-9_75

algorithms for feature matching in panoramic image stitching are Scale Invariant Feature Transform (SIFT)/Speeded-Up Robust Features (SURF)/Binary Robust Invariant Scalable Keypoints (BRISK)/Oriented Fast and Rotated BRIEF (ORB)/Fast Retina Keypoint (FREAK) and some others. And Random Sample Consensus (RANSAC) method is widely employed to eliminate the wrong matching point pairs, so as to optimize the matching rate of image matching. Although there have been many feature matching algorithms, the comparison between the matching effects of different algorithms in different scenarios is not in-depth.

The main task of this paper is to compare and summarize the advantages and disadvantages of different algorithms by using different feature matching techniques like SIFT, SURF and ORB. By matching the images under rotation, light intensity transformation and other conditions, the matching effect of each algorithm under different conditions is evaluated by using indicators such as matching point pairs and matching speed. From this, the applicable scenarios of different algorithms can be speculated, which is of great significance for researchers to select appropriate image feature matching algorithms in specific fields and scenarios, and can help them to quickly determine the method suitable for their research objects.

2 Related Work

The core issue of local image feature description is robustness. Many better feature matching algorithms have evolved since David Lowe proposed the Sift algorithm in 1999 [1]. This method is capable of withstanding some lighting and viewpoint modifications, and it possesses the qualities of scaling, rotation, and affine invariance. For instance, A.E. Abdel-Hakim and A.A. Farag introduced Colored SIFT (CSIFT), a SIFT descriptor with colour invariant characteristic [2]. Affine-SIFT (ASIFT) is a novel paradigm for totally invariant affine images comparison developed by Jean-Michel and Guoshen Yu [3]. In addition, to apply feature matching algorithm to different fields, researchers usually combine several algorithms with good properties to achieve better results. For example, a unique zero-watermarking technique according to SIFT and Bandelet and Discrete Cosine Transform (Bandelet-DCT) was presented by Yangxiu Fang et al. to address issues such patient privacy breaches and tampering with medical picture transmissions over the cloud [4]. Mohammad Manthouri et al. developed a computational intelligence technique for the detection of white blood cells using a mix of convolutional deep learning and SIFT [5], which has a high accuracy of segmentation.

Furthermore, because of the SIFT algorithm's inadequacies regarding its poor running speed, image feature matching cannot be applied to real-time scenarios like traffic or military ones. To solve this problem, Herbert Bay introduced the algorithm known as SURF [6], which is faster in calculation and comparison while ensuring robustness, repeatability and distinctiveness. SURF is also widely used. Showmik Setta et al proposed real-time face recognition with SURF [7] and Jiwei Fan et al applied it in order to propose a long-term visual tracking technique for Unmanned Aerial Vehicles (UAVs) [8]. As SURF algorithm is much faster than SIFT algorithm, these studies have well solved the difficulties of SIFT in dealing with real-time scenes, and will have great contributions to the field of information security and military. Furthermore, E Rublee et al came up with ORB algorithm [9] in2011, it is almost two orders of magnitude faster than SIFT and also noise-resistant and rotation invariant. Hence, it’s used to establish a feature-based monocular localization and mapping system called ORB-Simultaneous Localization and Mapping (ORB-SLAM) [10], which can be robust to severe motion clutter and includes full automatic initialization. The research can be of great significance for real-time localization and mapping.

3 Method

3.1 Stitching Process

With the development of computer vision and graphics, people have higher and higher requirements for the precision of image processing. As a kind of image processing technology commonly used in daily life, panoramic image stitching naturally needs to be improved and optimized. The main process of panoramic image stitching is shown in figure 1.

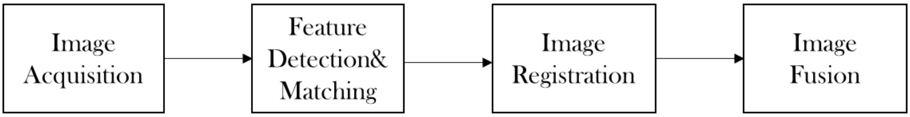

Fig1. Stitching process

The quality of image feature detection and matching straightly determines the quality of panoramic image stitching, so it is particularly important to choose a suitable feature detection and matching algorithm.

3.2 Data Collection

The raw data for this report comes from a set of original life photos. These photos were taken under different lighting conditions, different rotation angles and other factors that may affect the quality of the panoramic image stitching. The goal is to contrast the benefits and drawbacks of various algorithms in different scenarios from various perspectives and the key point is feature matching. The programme language and software used in this project are Python (version 3.12.1) and PyCharm (version 2023.3.1), and the feature matching algorithms performed and compared are SIFT, SURF and ORB.

3.3 Feature Matching

SIFT method. The fundamental idea behind SIFT is as follows: first, the image's scale-space representation is formed; next, the image's extreme value points are sought within the scale-space; and last the feature description vector is established based on the extreme value points [11]. When dealing with a two-dimensional image $I(x, y)$, the convolution of the picture with Gaussian function $G(x, y, \sigma)$ can yield the spatial scale representation $L(x, y, \sigma)$ at various scales

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{1}$$

Within the equation: $G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$ is a two-dimensional Gaussian function, σ is the standard deviation, point coordinates are represented by x and y. The picture to be matched is converted into grayscale image and convolved, and then the Difference-of-Gaussian (DOG) pyramid is established:

$$D(x, y, \sigma) = [G(x, y, k\sigma) - G(x, y, \sigma)] * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \tag{2}$$

Every single pixel in the Gaussian difference image is contrasted with a total of 26pixel points within the neighbourhood. If its greyscale value is the largest or smallest, this point is recorded as a candidate point of interest. Direction parameters for each feature point are specified using the gradient direction distribution of pixels in the neighborhood of the main point. Then the feature vector is generated and the similarity measure is performed at last. The brief flow of SIFT algorithm is shown in Table 1.

Table 1. SIFT algorithm procedure

Algorithm 1: SIFT method
Generate Gaussian Difference Pyramid
Scale space extremum detection
Specify direction parameters
Generate SIFT feature vector
Feature point matching

SURF method. Although SIFT has good robustness in various lighting circumstances, noise levels, scales as well as tiny rotational angles, it has the drawbacks of large computational data volume, high time complexity, and long algorithm runtime. Thus, the SURF was proposed. The overall idea and process of the SURF method are similar to SIFT, but different methods are used throughout the process. Scale-space theory remains the foundation of the SURF's feature point detection, but the biggest difference with SIFT algorithm is that SURF replaces second-order Gaussian filtering with box filtering approximation and uses integral images to accelerate convolution to improve computational speed [12].

ORB method. Although SURF speeds up feature matching, it is not enough to apply to real-time scenarios. Different from SIFT and SURF, ORB algorithm is built on the Features form Accelerated Segment Test (FAST) feature detection and Binary Robust Independent Elementary Features (BRIEF) feature descriptor.

FAST algorithm compares the greyscale pixel value in a neighbourhood with the greyscale value of the central point. If the greyscale value of n continuous pixels in a region is greater than that of the centre point plus t or smaller than that of the centre point minus t, then the point can be defined as a corner point. The FAST detector operates at very high speeds since it does not require sophisticated operations like gradient and scale.

The foundation of the BRIEF algorithm is the notion that image neighborhoods can be represented with a comparatively low degree of intensity contrast. The criterion τ for defining a picture neighbourhood *P* of $S \times S$ magnitude is defined as:

$$\tau(p; x, y) := \begin{cases} 1 & p(y) > p(x) \\ 0 & otherwise \end{cases} \tag{3}$$

Within the equation, $p(x)$ stands for the pixel grey value at $x = \begin{bmatrix} u \\ v \end{bmatrix}$ of the image neighbourhood P after Gaussian smoothing. n_d position pairs are selected to uniquely define the binary criterion, and the BRIEF feature descriptor is a n_d -dimensional binary bit string. In addition to saving storage space, the unique binary string representation of the BRIEF technique significantly cuts down on matching time. The ORB algorithm also takes into account the defects of FAST, which has no corner response function and no multi-scale features, and BRIEF, which is sensitive to noise and has no rotation invariance. By improving these problems, ORB has both high speed and good robustness.

In conclusion, the main parts of SIFT, SURF and ORB are compared in table 2.

Table 2. Comparison between SIFT/SURF/ORB

	SIFT method	SURF method	ORB method
Detect feature points	Different scale images convolve with Gaussian function	Using different box filters to convolve with the image	FAST algorithm
Orientation assignment	Using the gradient histogram in the neighbouring area	Calculate the Haar wavelet responses in x and y direction with in a circular of radius 6 s around the interest point	Using Intensity Centroid
Feature descriptor	In each sub-region, 8 gradient histograms are	Compute Haar wavelet responses at	BRIEF algorithm

	calculated. Thus totally 128 descriptor vectors	5×5 sample points. The responses and absolute values are summed up. Thus totally 64 descriptor vectors

RANSAC method. RANSAC algorithm is one of the most popular methods to determine the optimal homography matrix in order to remove the mismatching points and maximize the amount of data points meeting the matrix. The fundamental assumption of the RANSAC algorithm is that data is made up of "inliers" and "outliers". The "inlier" is the point that makes up the model parameters, and the "outlier" is the point that does not fit the model. And it makes the assumption that a program exists that can estimate a model that fits these points given a dataset that contains a minimal number of "inliers". The main procedure of RANSAC is briefly shown in Table 3.

Table 3. RANSAC algorithm procedure

Algorithm 1: RANSAC method
1. Select a threshold and the highest quantity of repetitions.
2. Loop for the highest quantity of repetitions:
i. Randomly select points from the dataset
ii. Compute the homography matrix from the sample points
iii. Apply the homography to all points in the other set of points
iv. Compute the distances between the actual points and the ones transformed by the computed homography
v. Determine the inliers where the distance is below the threshold chosen
vi. Update the best homography if the current one has more inliers
3. After the loop, the best homography matrix and the inliers are returned.

4 Experimental procedure

4.1 Image Selection

In this paper the raw data source is from a set of photos taken in quotidian life and the target object is expected to be detected from the features of the overall photo. The experiments used SIFT/SURF/ORB algorithm for comparison and applied RANSAC algorithm to remove points that are mismatched. The evaluation indexes are the quantity of points matched correctly, match rate and computation time. It should be noted that this experiment was carried out in four scenarios, namely rotation, scaling, different light intensity and different perspectives. In order to be closer to the actual application scenario, we put other objects around the object to be detected to ensure the validity of the experimental results. The image to be recognized is shown in figure 2.

Fig.2 Image to be matched.

4.2 Results

Matching results in the case of rotation Each pixel in the image will rotate at a corresponding angle around the rotation center when it is rotated. Additionally, pixels' gradient amplitude surrounding the original characteristic point and its direction information will change. The feature point's primary direction will rotate as well. In this experiment, the photo was rotated nearly 75degrees, and the rotated image was matched with the original object to be detected. Figure 3 and Table 4 display the output images and comparison results.

Table 4. The comparison results in the case of rotation.

Method	Match points	Correct matched points	Match rate	Computation time
SIFT	12829	5540	43.18%	43.6217s
SURF	2636	1137	43.13%	3.8034s
ORB	217	170	78.34%	0.8380s

It can be seen that SIFT, SURF and ORB algorithms all have good rotational invariance, which is consistent with the characteristics mentioned above. Although ORB merely obtain 217 match points, its match rate is as high as 78.34%, which is almost twice that of SIFT and SURF, which are 43.18% and 43.13% respectively. Meanwhile, the order in which the three algorithms execute is ORB, SURF and SIFT, which is also consistent with the theory. ORB can run for less than 1 second while SIFT needs more than 43 seconds to finish the task, which is about 52 times longer than ORB. This is mainly because of the fact that the feature point detection algorithm FAST and the feature description algorithm BREIF used by ORB are faster than that of the other two algorithms.

Fig. 3. Comparison of matching results of different algorithms under rotation. From top to bottom are: results of SIFT, SURF, and ORB

Matching results in the case of scaling Similarly, the image is scaled to 0.8 times the original size and the matching outcomes of diverse methods are exhibited in Figure 4 and Table 5

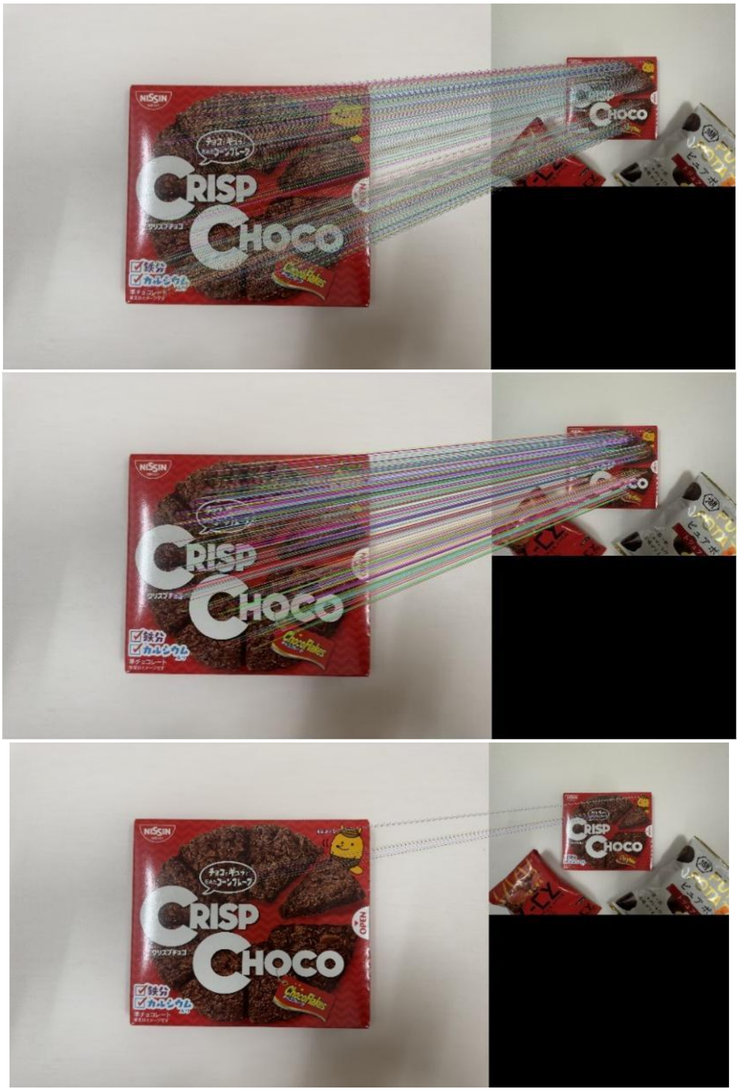

Fig. 4. Comparison of matching results of different algorithms under scaling. From top to bottom are: results of SIFT, SURF, and ORB

Table 5. The comparison results in the case of scaling.

Method	Match points	Correct matched points	Match rate	Computation time
SIFT	5755	1492	25.93%	36.6062s
SURF	1151	467	40.57%	2.6363s
ORB	147	41	27.89%	0.6662s

When the object to be matched is scaled, the accuracy of SIFT and ORB algorithms is not high, only 21% and 22% respectively, while the match rate of SURF is 40.57%.

The reason for this phenomenon is that SURF adopts box filtering approximation, which makes it more sensitive to image scaling. Meanwhile, SURF algorithm has a great improvement in speed compared with SIFT algorithm, and it is only about 3 seconds slower than ORB algorithm. This is due to the Fast Hessian used by SURF can calculate the space-scale faster than the Gaussian pyramid used by SIFT, while the integral image technology can quickly calculate the gradient and feature descriptors of the region around the key points.

Matching results in the case of different light intensity The main components of the image will appear in different gray scale regions depending on the lighting conditions, with some being more prominent than others due to occlusion, shadow, etc. In real life, light condition often has a great impact on panoramic image stitching and many other fields, so it is necessary to find the most robust algorithms under different lighting conditions.

Fig. 5. Comparison of matching results of different algorithms under different light intensity. From top to bottom are: results of SIFT, SURF, and ORB

Table 6. The comparison results in the case of different light intensity.

Method	Match points	Correct matched points	Match rate	Computation time
SIFT	4113	2328	56.60%	80.9563s
SURF	1410	1208	85.67%	3.5642s
ORB	183	120	65.57%	0.8299s

As Figure 5 and Table 6 demonstrate, the comparison of these three algorithms under different lighting intensity has similar results to the case of scaling, but the matching accuracy of each algorithm has been improved, especially the match rate of ORB is increased by more than 130%. This indicates that the robustness of the three algorithms under light intensity transformation is better than scaling. The SURF algorithm has a matching accuracy of up to 85.67% and maintains a fast speed, while SIFT algorithm, despite its promoted match rate, also runs much slower.

Matching results in the case of different perspectives Feature matching under different perspectives is also a vital part of panoramic image mosaic. These are the experimental results displayed (Figure 6 and Table 7).

Table 7. The comparison results in the case of different perspectives.

Method	Match points	Correct matched points	Match rate	Computation time
SIFT	8132	2233	27.46%	31.4119s
SURF	1861	316	16.98%	3.6420s
ORB	137	8	5.84%	0.8734s

Obviously, when processing images from different perspectives, SIFT algorithm has the slowest speed but the highest accuracy, which is 27.46%, while ORB algorithm, still fast but has the lowest accuracy, which is only 5.84%. This may be due to the image deformation when the perspective changes, which affects the performance of BRIEF algorithm.

Fig. 6. Comparison of matching results of different algorithms under different perspectives. From top to bottom are: results of SIFT, SURF, and ORB

5 Conclusion

The robustness of image feature matching algorithm under different transformations is an important direction of panoramic image stitching. In this paper, the match rate and running speed of SIFT, SURF as well as ORB algorithms under conditions such as rotation, scaling, light intensity transformation and perspective transformation are compared to determine the application scenarios applicable to different algorithms. Based on the experimental results, SIFT is best suited for fields with minimal real-time requirements and perspective transformation. SURF has the best stability when dealing with scale variations and varying light intensity, while it is much faster than SIFT. ORB has the fastest running speed, so it is suitable for the field with high real-time requirements and has the best robustness in the case of rotation. There are more than the above factors that affect the feature matching effect, and there may be more complex factors in practical applications. The comparison of feature matching algorithms under other complex cases is the main direction of further research.

References

1. Lowe D G. Object recognition from local scale-invariant features. Proceedings of the seventh IEEE international conference on computer vision. Ieee, 2: 1150-1157 (1999).

2. Abdel-Hakim A E, Farag A A. CSIFT: A SIFT descriptor with color invariant characteristics. 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06). Ieee, 2: 1978-1983 (2006).
3. Morel J M, Yu G. ASIFT: A new framework for fully affine invariant image comparison. SIAM journal on imaging sciences, 2(2): 438-469 (2009).
4. Fang Y, Liu J, Li J, et al. Robust zero-watermarking algorithm for medical images based on SIFT and Bandelet-DCT. Multimedia Tools and Applications, 81(12): 16863-16879 (2018).
5. Manthouri M, Aghajari Z, Safary S. Computational intelligence method for detection of white blood cells using hybrid of convolutional deep learning and SIFT. Computational and Mathematical Methods in Medicine, (2022).
6. Bay H, Tuytelaars T, Van Gool L. Surf: Speeded up robust features, Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I 9. Springer Berlin Heidelberg, 404-417 (2006).
7. Setta S, Sinha S, Mishra M, et al. Real-time facial recognition using SURF-FAST, Data Management, Analytics and Innovation: Proceedings of ICDMAI 2021, Volume 2. Springer Singapore, 505-522 (2021).
8. Fan J, Yang X, Lu R, et al. Long-term visual tracking algorithm for UAVs based on kernel correlation filtering and SURF features. The Visual Computer,39(1): 319-333 (2023).
9. Rublee E, Rabaud V, Konolige K, et al. ORB: An efficient alternative to SIFT or SURF, 2011 International conference on computer vision. Ieee: 2564-2571 (2021).
10. Mur-Artal R, Montiel J M M, Tardos J D. ORB-SLAM: a versatile and accurate monocular SLAM system. IEEE transactions on robotics, 31(5): 1147-1163 (2015).
11. Fu W., Qin C., Liu J., et al. Image target matching and localization based on SIFT algorithm. Chinese Journal of Scientific Instrument, 32(1): 163-169 (2011)
12. Zhang R., Zhang J., Yang C.. Research on image registration method based on SURF. Infrared and Laser Engineering, 38(1): 160-165 (2009).