

A feature extraction and matching method in large parallax scenes

Guoqing Zhou^{a,b}, Jin Tian^{a,b}

^aCollege of Geomatics and Geoinformation, Guilin University of Technology, Guilin 541004, China

^bGuangxi Key Laboratory of Spatial Information and Geomatics, Guilin University of Technology, Guilin 541004, China

ABSTRACT

A feature point extraction and matching algorithm based on affine transformation space is proposed to address the shortcomings of existing feature extraction and matching algorithms in large view scenes with few effective matching points and slow matching speed. The algorithm first constructs the affine change space to simulate the viewpoint change and obtains the affine invariance; then avoids the feature point detection in the invalid region by dividing the valid region; in the feature description stage, the ORB algorithm is incorporated into the affine change space, while the gradient contrast information of multiple directions in the feature point sampling region is fused to obtain the final binary descriptor. Through experiments on large-view datasets and sequence images, it is demonstrated that the algorithm has better matching effect in large-view scenes, and also has more advantages in time efficiency.

Keywords: large view scene, image matching, affine transformation space, ORB algorithm

1. INTRODUCTION

As an important part of digital image processing technology, image matching has been widely used in the fields of autonomous robot driving [1], target tracking and recognition [2-3] and visual navigation [4-5]. Feature-based image matching algorithm is the mainstream direction of current image matching technology. Lowe proposed the SIFT (scale invariant feature transform) algorithm [6]. Bay et al reduced the dimensionality of the feature description operator to 64 dimensions and proposed the SURF (speeded up robust features) operator, which improves the matching efficiency [7]. Rub lee et al proposed the ORB (oriented FAST and rotated BRIEF) algorithm[8], which is one order of magnitude more efficient in feature point extraction than the SURF algorithm and two orders of magnitude more efficient than the SIFT algorithm. Although the ORB algorithm is faster in computation and can resist rotational transformation, its resistance to perspective transformation is poor. In view of this, Morel et al [9] proposed the ASIFT (Affine-SIFT) algorithm with anti-viewpoint transformation capability by simulating the viewpoint deformation using affine sampling. Based on the ASIF algorithm, the ASURF [10] (Affine-SURF) algorithm, the AORB [11] (Affine-ORB) method, and the AFREAK algorithm [12] (Affine-FREAK) utilize the SURF, ORB, and FREAK algorithm instead of SIFT algorithm, which effectively improves the computational speed of the algorithm. Cai et al proposed PSIFT (Perspective-SIFT) algorithm [13] based on ASIFT algorithm using the perspective transformation model instead of the affine transformation model, which improves the matching accuracy. However, the complexity of these two algorithms increases dramatically to ensure resistance to image rotation, image scale, and parallax changes in continuous images.

In this paper, a feature point detection and matching algorithm based on affine transformation space is proposed. At the stage of constructing affine transformation space, the valid regions in the image after perspective change are marked by tracking the four boundary corner points of the original image, thus avoiding feature detection in invalid regions. In the feature detection and description stage, the binary descriptor of the ORB algorithm is introduced into the affine transformation space, and a Gaussian pyramid is established to obtain the feature points with scale invariance. To improve the descriptive ability of the binary descriptor, the gradient contrast information in multiple directions in the neighborhood of feature points is added to the original grayscale contrast information to obtain the final binary descriptor instead of the original real-valued descriptor. Through experiments on the dataset of large-view scenes and sequence images, it is demonstrated that the algorithm can improve the time efficiency of the original algorithm while solving the matching problem in large-view scenes, and can guarantee the matching effect.

2. PRINCIPLE OF FEATURE MATCHING ALGORITHM BASED ON AFFINE TRANSFORMATION SPACE

2.1 Effective area identification

According to the affine transformation model and certain sampling range and sampling interval of tilt coefficient and longitude angle, the original image is perspective transformed to obtain the full perspective space, i.e., the affine transformation space, as shown in Figure 1a, which shows part of the image after perspective transformation. To avoid the computation of the detector in the invalid region, the four boundary corner points of the original image are tracked to construct the effective region mapping of the affine space, as shown in Fig. 2.1b.

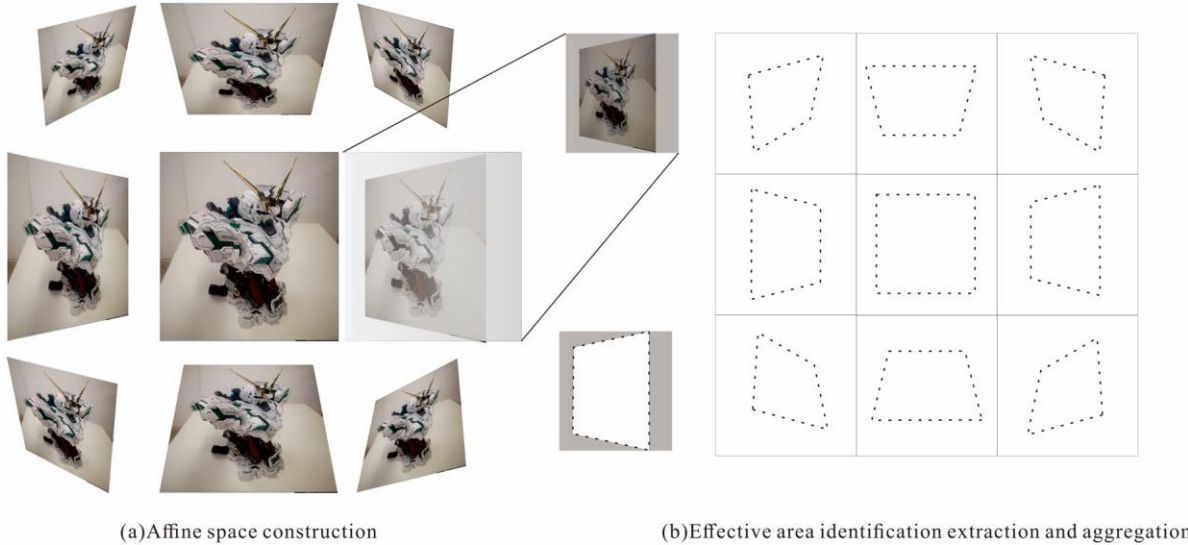


Figure 1 (a) affine space construction (b) Effective area identification extraction and aggregation

2.2 Feature point detection

This paper uses a modified FAST detector similar to the ORB algorithm for feature point detection in the feature point detection stage. FAST corner point is defined as: if any point to be detected is smaller or larger than the gray value of a certain number of sampled points in the neighborhood, it will be identified as a candidate corner point.

The feature points obtained by the FAST operator do not have scale information and orientation information, and the use scenario is limited. For scale invariance, this paper builds Gaussian pyramids by borrowing LOG detectors, DOG detectors and other multi-scale detectors. In order to achieve rotational invariance, the FAST corner points must be made directional, and the intensity center of the sampled region can be used as the principal direction of the feature points. Therefore, the detector in this paper obtains the feature point information with scale σ , direction θ , and fractional response R in the effective region of the affine transformation space.

2.3 Feature Point Description

(1) ORB algorithm descriptor

The ORB algorithm compares the feature points with the gray scale information of the sampled points in the surrounding neighborhood during feature description to obtain a set of binary strings as feature point descriptors.

$$\tau_1(p : x, y) = \begin{cases} 1, & \text{if } I(x) < I(y) \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

Where p is the current feature point, $I(x)$, $I(y)$ are the gray scale values corresponding to any sampled point pair x , y , respectively. For any feature point, N pairs of sampled points are obtained in its neighborhood, and an N -bit binary descriptor can be obtained in accordance with equation (1), as shown in equation (2), and N is 256 in the ORB algorithm.

$$des(p) = \sum_{i=1}^N 2^{i-1} \tau_1(p : x_i, y_i) \quad (2)$$

(2) Comparison of feature points and sample points

ORB descriptors use statistical learning to obtain sampled point pairs in order to improve the distinguishability of descriptors and reduce the correlation between descriptors. In order to improve the distinguishability and robustness of the descriptors, this paper adds the gray scale comparison information between the feature points and the sampled points based on the original descriptors.

$$P(I : p, x) = \begin{cases} 1, & \text{if } I(x) < I(y) \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

$$\tau_p(p : x, y) = \begin{cases} 1, & \text{if } P(I : p, x) < P(I : p, y) \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

Where denotes the gray scale comparison information between feature point p and sample point x. Then, using equation (3) to compare whether the gray scale comparison results between feature point p and sample point pair (x, y) are the same, if they are, the descriptor corresponds to position 1, otherwise it is set to 0. Then the binary descriptor part containing the gray scale comparison information between feature point and sample point can be expressed as (5)

$$des'(p) = \sum_{i=1}^N 2^{N+i-1} \tau_c(p : x_i, y_i) \quad (5)$$

(3) Descriptors for estimating gradient information

Due to the effect of viewpoint difference and illumination change, the gray scale value of the same local region in the image to be matched will change to a certain extent. In order to further reduce the sensitivity of the descriptor to illumination and viewpoint change, this paper introduces the gradient information into the binary descriptor.

In this paper, the gradient of the sampling point to each direction is first calculated, then the gradient difference in the corresponding direction between x, y at any sample point is compared to obtain the binary number as part of the descriptor.

$$\left\{ \begin{array}{l} P'(g_\theta; x) = \begin{cases} 1, & \text{if } g_\theta(x) > 0 \\ 0, & \text{otherwise} \end{cases} \\ \tau_{g_\theta}(p(x, y)) = \begin{cases} 1, & \text{if } P'(g_\theta, x) < P'(g_\theta, y) \\ 0, & \text{otherwise} \end{cases} \\ des^n(p) = \sum_{i=1}^N 2^{N+i-1} [\tau_{g_{0^\circ}}(p1) + \tau_{g_{90^\circ}}(p2) + \tau_{g_{180^\circ}}(p3) + \tau_{g_{270^\circ}}(p4)] \\ = \sum_{i=1}^N \sum_{j=1}^4 2^{(j+1)N+i-1} \tau_{g_{\theta_j}}(p_j) \end{array} \right. \quad (6)$$

The above feature information is integrated to obtain the final binary descriptor as shown in equation (7) :

$$\begin{aligned} Des(p) &= \sum_{i=1}^N 2^{i-1} \tau_1(p(x_i, y_i)) + \sum_{i=1}^N 2^{N+i-1} \tau_c(p(x_i, y_i)) + \sum_{i=1}^N \sum_{j=1}^4 2^{(j+1)N+i-1} \tau_{g_{\theta_j}}(p_j) \\ &= des(p) + des'(p) + des''(p) \end{aligned} \quad (7)$$

where, denotes the original ORB descriptor information, and represent the gray scale contrast information between feature points and sampled points and the gradient contrast information between sampled point pairs, respectively.

The detection and description of the above feature points are obtained in the affine transformation space constructed from the original image, and then all feature points need to be inverse-view transformed and remapped into the original single-view image, and finally the feature points and their corresponding binary descriptors based on the affine transformation space proposed in this paper can be obtained. The brief flow is shown in the figure 2.

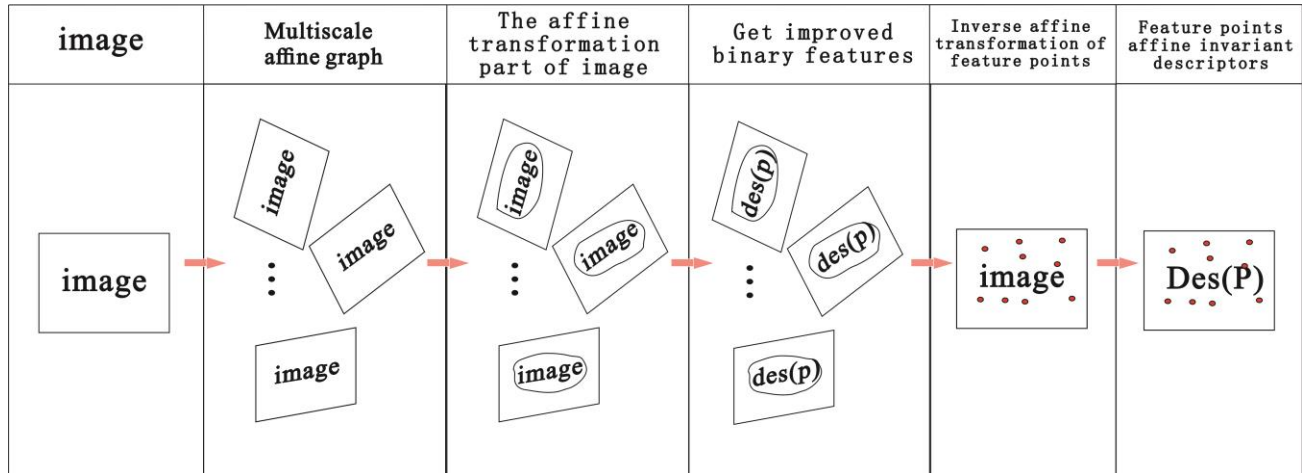


Figure 2 Brief flow of the algorithm in this paper

3. EXPERIMENTS AND ANALYSIS OF RESULTS

In this section, feature point matching experiments are first conducted in the Adam group, magazine group and cup group of the large view scene data dataset to demonstrate the effectiveness of the proposed feature point matching algorithm; then the gradient number comparison experiments are conducted to demonstrate the effectiveness of the gradient comparison information incorporated in the descriptors.

3.1 Experimental data set and environment

The image datasets Adam group, magazine group and cup group have large perspective variations. Figure 3 shows some of the images in the dataset, where the Adam group characterizes Absolute Tilt, and nine images are selected, which are positive angle of view images, left and right tilt 45°, 65°, 75°, 80° images. Figure 4 shows some of the images in the dataset, magazine group characterizes Transition Tilt, and 10 images are selected, which are 0° tilt and 10°, 20°, 30°, 40°, 50°, 60°, 70°, 80°, and 90° tilt angles. The data were obtained using the Wall group of images characterizing the viewpoint change in the Oxford set, as shown in Fig. 5, with the images from a to f plotted in order of viewpoint change enhancement.



Figure 3 the image datasets Adam group (Absolute Tilt)

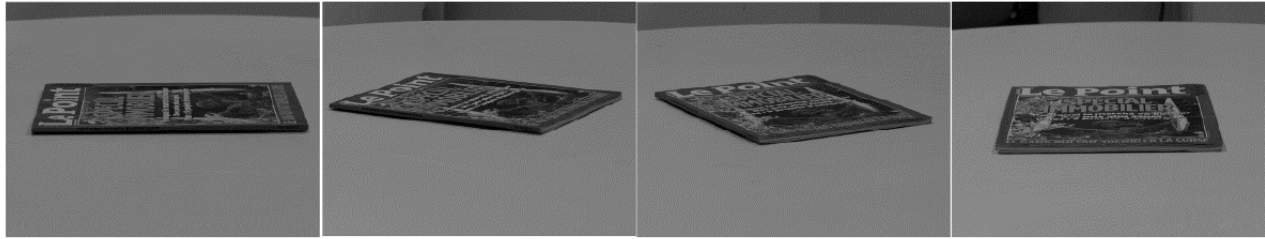


Figure 4 the image datasets magazine group (Transition Tilt)

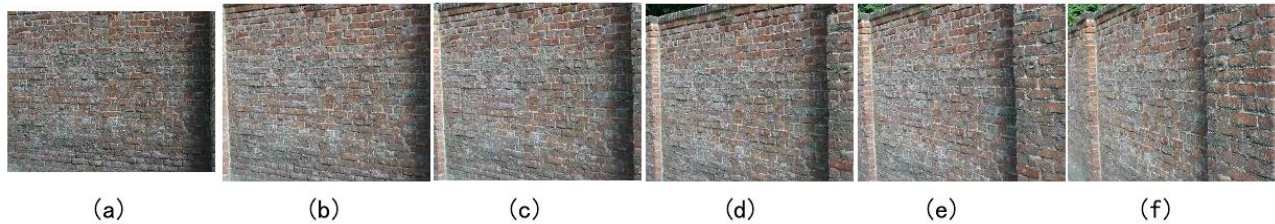


Figure 5 the image datasets Wall group

The experimental environment is VS2015 and OPENCV3.0, and the algorithm is implemented in C++. Since the feature matching algorithm proposed in this paper is based on Affine transform space, Gradient contrast information and ORB descriptors, it is uniformly identified as AGORB (Affine-Gradient-ORB) in the experiments.

3.2 Image feature extraction experiments with different viewpoint variations

Experiment 1 counts the number of valid matching point pairs between pairs of images with different viewpoint variations. As shown in Figure 6, the Adam group counted the number of valid matching pairs between the left (right) 5° view and the positive view, left (right) 65° view, left (right) 75° view, and left (right) 80° view images, respectively; the magazine group counted the number of valid matching pairs between the 10° to 90° view pairs, respectively, based on the 0° tilt view. The cup group counted, separately, the number of valid matching point pairs for each algorithm in image pairs 1-2, 1-3 ... 2-6.

As can be seen from the figure, the traditional SIFT and ORB descriptors are difficult to find a certain amount of effective matching point pairs in the image pairs with large changes in view angle. For example, in the Adam group, when the view angle of another image to be matched changes more than 75° (i.e., image pairs 45&75), SIFT and ORB have basically failed and it is difficult to find effective matching point pairs.

The algorithms of ASIFT and this paper, which introduce the affine transformation space, solve the problem of image matching under large viewing angle by simulating the change of viewing angle, and can still find a certain amount of effective matching point pairs even between pairs of images with large viewing angle, which is much better than the traditional SIFT and ORB algorithms.

Also from the comparison, it can be seen that compared with the ASIFT algorithm, the algorithm proposed in this paper incorporates the ORB algorithm into the affine transformation space, and in order to improve the description performance of the binary descriptors, the algorithm adds a certain amount of gradient contrast information to the original gray contrast information, which effectively improves the description ability of the feature descriptors and thus obtains more effective matching point pairs.

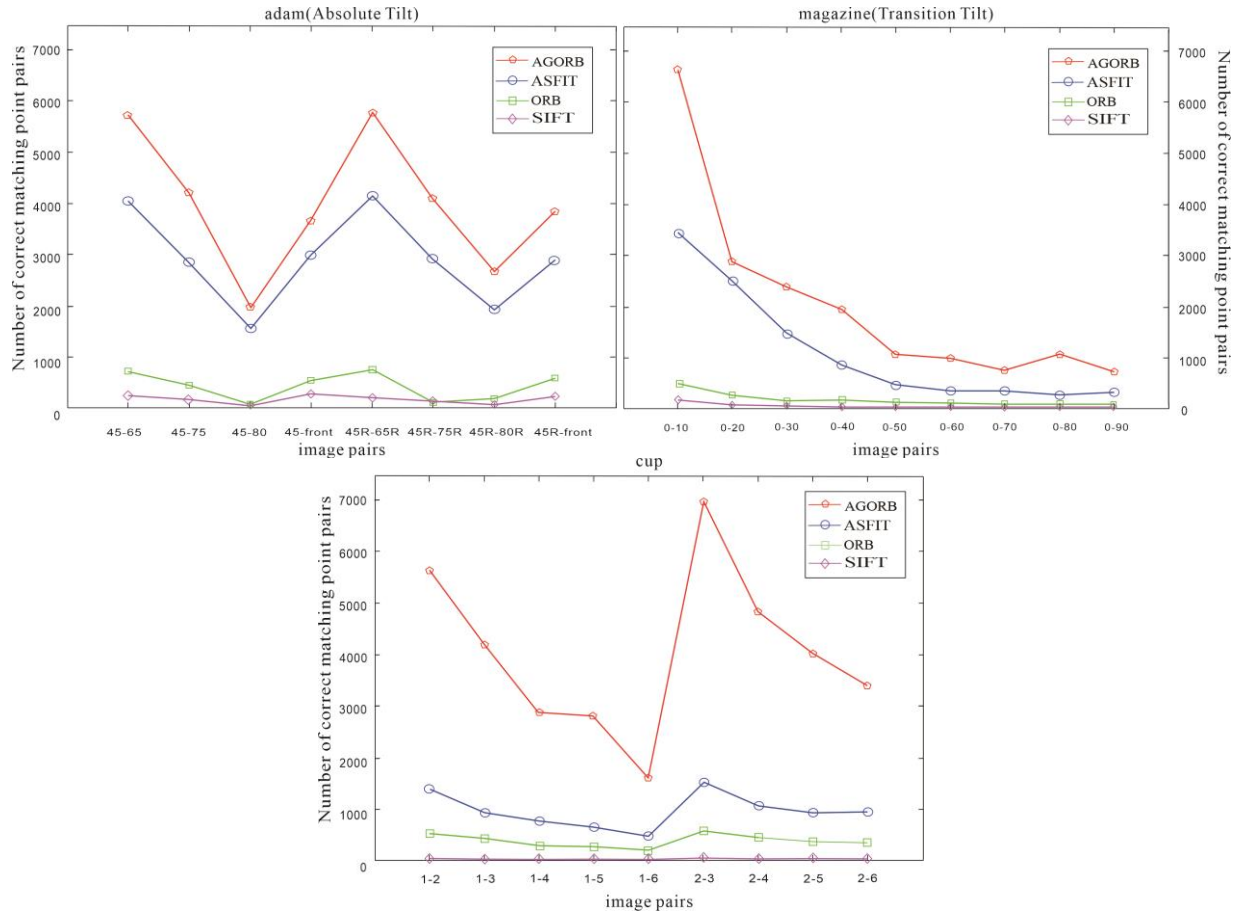


Figure 6 Effective matching point results for Adam, magazine and cup groups

Tables 1 and 2 statistically show the matching time between ASIFT and this paper's algorithm in the experiments. The comparison experiments on multiple image pairs show that the time used by the algorithm in this paper is 39.68%-44.65% less than the traditional ASIFT in the Adam group and 40.80%-59.62% less in the magazine group.

Table 1 Feature point detection time comparison (Adam)

| datasets | method | Image pairs | | | | | | | |
|-------------|--------|-------------|-------|-------|----------|---------|---------|---------|-----------|
| | | 45-65 | 45-75 | 45-80 | 45-front | 45R-65R | 45R-75R | 45R-80R | 45R-front |
| Adam | Ours | 14.78 | 14.44 | 14.21 | 14.94 | 14.41 | 14.55 | 15.10 | 14.97 |
| | ASIFT | 24.56 | 24.45 | 24.22 | 26.99 | 23.89 | 26.33 | 25.12 | 26.12 |
| Improve (%) | | 39.82 | 40.94 | 41.33 | 44.65 | 39.68 | 44.74 | 39.89 | 42.69 |

Table 2 Feature point detection time comparison (Magazine)

| datasets | method | Image pairs | | | | | | | | |
|-------------|--------|-------------|-------|-------|-------|-------|-------|-------|-------|-------|
| | | 0-10 | 0-20 | 0-30 | 0-40 | 0-50 | 0-60 | 0-70 | 0-80 | 0-90 |
| Magazine | Ours | 16.22 | 15.31 | 15.12 | 14.89 | 15.01 | 15.11 | 15.33 | 15.87 | 15.12 |
| | ASIFT | 27.44 | 28.78 | 26.89 | 28.99 | 33.34 | 31.21 | 26.76 | 32.13 | 37.44 |
| Improve (%) | | 40.89 | 46.80 | 43.77 | 48.64 | 54.98 | 51.59 | 42.71 | 50.61 | 59.62 |

This is due to the fact that although the algorithm in this paper also obtains the affine invariance by constructing the affine transformation space, it improves the execution efficiency of the algorithm by dividing the effective region when constructing the affine change space, which avoids the detection of feature points in the invalid region, and at the same time, this paper introduces the binary descriptor into the affine space, which greatly improves the speed of feature point description and matching.

3.3 Multi-gradient reference experiments

Experiment 2 was conducted to demonstrate the effect of the gradient contrast information introduced in the algorithm of this paper on the algorithm. In the experiments, AORB denotes the binary descriptor with no added gradient contrast information; AGORB1 denotes the binary descriptor with added gradient contrast information in two directions only; and AGORB denotes the binary descriptor with added gradient contrast information in four directions as proposed in this chapter.

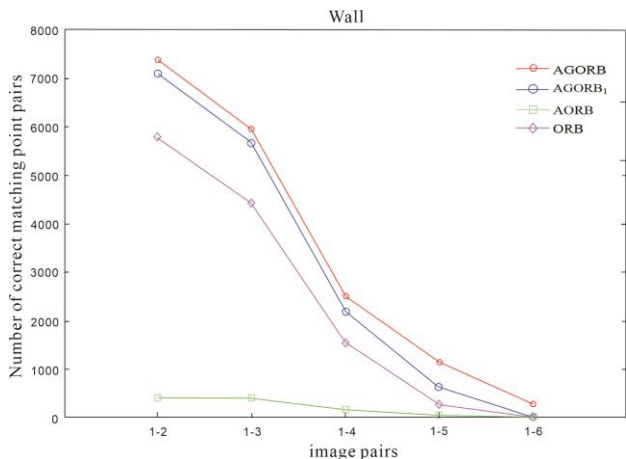


Figure 7 Number of correctly matched point pairs for ORB, AORB, AGORB1, and AGORB

As shown in Figure 7, the experiments measure the descriptor performance by counting the number of correctly matched point pairs in the same set of feature points. As can be seen from the figure, the resistance of the binary descriptor to viewpoint changes is significantly enhanced after the introduction of the affine transformation space. From the comparison of AORB, AGORB1 and AGORB, it can be seen that the number of correctly matched point pairs rises significantly after the introduction of the gradient contrast information, and the richer the gradient contrast information, the stronger the performance of the descriptor, and thus the greater the number of correctly matched point pairs. This experiment proves that the gradient contrast information can effectively reflect the local information of feature points, and the gradient features in multiple directions can effectively improve the matching performance and robustness of descriptors, and also proves the effectiveness of this algorithm. However, richer gradient comparison information also means a higher time complexity, which needs to be traded off in use on a case-by-case basis.

4. CONCLUSION

In this paper, the image feature matching algorithm for large view scenes is investigated. A feature point matching algorithm based on affine transformation space is proposed to improve the computational speed of the algorithm while ensuring the matching effect. The algorithm first constructs the affine transformation space to simulate the viewpoint change and obtain the affine invariance; then divides the valid region to avoid the feature point detection in the invalid region; finally, in the feature detection and description stage, the ORB binary descriptor is introduced into the affine transformation space, while adding certain gradient contrast information to obtain the final descriptor.

We compared the effective matching points of ORB algorithm, SIFT algorithm, ASIFT algorithm and the efficiency of this paper's algorithm and ASIFT algorithm in the experiments of continuous large parallax scenes and scenes with rapid change of viewpoint. The experimental results show that the algorithm of this paper has a higher number of correct matches with the addition of multi-directional neighborhood gradient comparison information of feature points, and the matching effect is better with the addition of more gradient information. Therefore, the experiments prove that (1) the algorithm in this paper has better matching effect in large view scenes. (2) the algorithm has advantages in time efficiency.

REFERENCES

- [1] Warrant E, Dacke M. Visual navigation in nocturnal insects [J]. *Physiology*, 2016, 31(3): 182-192.
- [2] Jia K, Chan T H, Zeng Z, et al. ROML: A robust feature correspondence approach for matching objects in a set of images [J]. *Int J of Computer Vision*, 2016, 117(2):173-197.
- [3] Chen Z X, He C, Liu C Y. Image saliency target detection based on global features and local features [J]. *Control and Decision*, 2016, 31(10): 1899-1902. (in Chinese)
- [4] Lu Y, Song D. Visual navigation using heterogeneous landmarks and unsupervised geometric constraints [J]. *IEEE Trans on Robotics*, 2015, 31(3): 1-14.
- [5] Xu Y X, Chen F. Scene matching algorithm based on CenSurE for SAR/INS integrated navigation system [J]. *Control and Decision*, 2011, 26(8): 1175-1180. (in Chinese)
- [6] Lowe D G. Distinctive image features from scale-invariant key points [J]. *Int J of computer vision*, 2004, 60(2): 91-110.
- [7] Bay H, Ess A, Tuytelaars T, et al. Speeded-up robust features (SURF) [J]. *Computer Vision and Image Understanding*, 2008, 110(3): 346-359.
- [8] Rublee E, Rabaud V, Konolige K, et al. ORB: An efficient alternative to SIFT or SURF[C]. *Int Conf on Computer Vision*. Barcelona: IEEE, 2011: 2564-2571.
- [9] Morel J M, Yu G. ASIFT: A new framework for full affine invariant image comparison [J]. *Siam J on Imaging Sciences*, 2009, 2(2): 438-469.
- [10] Su K X, Han G L, Sun H J. Anti-viewpoint changing image matching algorithm based on SURF [J]. *Chinese J of Liquid Crystals and Displays*, 2013, 28(4): 626-632. (in Chinese)
- [11] Hou Y, Zhou S L, Lei L, et al. Fast fully affine invariant image matching based on ORB [J]. *Computer Engineering and Science*, 2014, 36(2): 303-310. (in Chinese)
- [12] Fu C, Deng L, Lu G, et al. Improved image matching based on fast retina keypoint algorithm [J]. *Computer Engineering and Applications*, 2016, 52(19): 208-212. (in Chinese)
- [13] Cai G R, Jodoin P M, Li S Z, et al. Perspective-SIFT: An efficient tool for low-altitude remote sensing image registration [J]. *Signal Processing*, 2013, 93(11):3088-3110.