A novel image matching algorithm based on PCA and SIFT

Wenju Li, Xinyuan Na, Pan Su, Yunfan Lu, Tianzhen Dong

School of Computer Science and Information Engineering, Shanghai Institute of Technology, 100 Haiquan Road, Fengxian District, Shanghai, China

Email: wjli@sit.edu.cn

Abstract: In order to solve the problem of low matching precision and slow matching speed of image matching algorithm based on classical SIFT, a new image matching algorithm based on PCA, SIFT and improved RANSAC is proposed. Firstly, the SIFT feature was extracted from images; Secondly, principal component analysis is used to reduce the dimension of SIFT feature descriptor, from 128 to 20; then, the EUCLIDEAN distance is used for feature matching; finally, an improved RANSAC algorithm is proposed to eliminate the mismatched feature points. Experimental results show that the proposed algorithm improves the accuracy and speed of image matching.

1. Introduction

Image matching [1] is a key technology in computer vision and image processing, which is widely used in the fields of object detection, object recognition, object tracking, three-dimensional reconstruction, industrial detection, and medical image analysis.

Image matching mainly has two types of methods based on region and feature [2]. Feature-based methods are currently the mainstream method of image matching and they have better robustness. Among the feature-based methods, the SIFT [3,4] algorithm has received widespread attention because of its invariance to scale changes, rotation transformations, and illumination changes. However, the classic SIFT algorithm has some shortcomings: for example, the number of feature descriptors is high, the algorithm calculation is large, it is difficult to meet the real-time requirements and the matching accuracy is low. For this reason, some improved SIFT algorithms have been proposed [5-9], these algorithms improve the matching speed or matching accuracy to a certain extent.

Aiming at the shortcomings of the traditional SIFT algorithm, this paper uses the PCA algorithm [10] to reduce the dimension of the SIFT descriptor to reduce the dimension of the description vector to improve the speed of feature matching and meet real-time performance; then use Euclidean distance for feature point matching; finally, the traditional RANSAC algorithm [11] is improved to reduce its excessive time consumption and remove mismatched point pairs.

2. Algorithms

2.1. SIFT feature extraction

The use of SIFT algorithm mainly includes the following steps: establishing the scale space of the image, generating difference of Gaussian (DOG), extreme point detection, precise locating of key points and direction assignment, and generating feature descriptors.
2.1.1. Image scale space establishment. The scale space $L(x, y, \sigma)$ of the two-dimensional image $I(x, y)$ can be obtained as follows:

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

Among them, $G(x, y, \sigma)$ is a Gaussian function, $x$ is the abscissa, $y$ is the ordinate, where $\sigma$ represents the scale factor, and $\ast$ represents the convolution symbol.

2.1.2. DOG pyramid generation. The DOG of the image is the difference between two adjacent scale images, which is expressed as follows:

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

Where $k$ is the scale parameter.

2.1.3. Extreme point detection. In the DOG scale space, each pixel is compared with 8 points of the same scale and 18 corresponding points in two adjacent scales. If the pixel is the maximum or minimum, then the point is an extreme point on the image.

2.1.4. Key point locating and direction assignment. The extreme points are accurately located by curve fitting, the edge response and the unstable points are eliminated through the Hessian matrix to obtain stable key points with sub-pixel accuracy.

Use the gradient direction of the neighborhood point adjacent to each key point to assign direction information to each key point, so that the feature descriptor has rotation invariance. The gradient value and direction information are as follows:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}$$  
$$\theta(x, y) = \tan^{-1}\frac{L(x+1, y) - L(x, y-1)}{L(x+1, y) - L(x, y-1)}$$

$L(x, y)$ is the scale space value of the key point; $m(x, y)$ is the key point gradient, and $\theta(x, y)$ is the key point direction ($^\circ$).

Draw the gradient direction of the neighboring pixels of the key point (one direction is taken every 45°, a total of 8 directions) into a histogram, and the peak direction of the histogram represents the direction of the key point.

2.1.5. SIFT feature descriptor generation. Take the key points as center, take a 16x16 neighborhood, divide the neighborhood into 4 x 4 sub-regions, then calculate the histogram of 8 directions for each sub-region. In this way, the feature vector of the SIFT feature descriptor of each key point is 128 dimensions. In order to remove the influence of illumination changes, the feature vector is normalized.

2.2. SIFT feature dimensionality reduction based on PCA

PCA is a widely used algorithm because of its good data dimensionality reduction performance. PCA transforms the original data into a set of linearly independent representations of each dimension through linear transformation, which can be used to extract the main feature components of the data.

The specific process of using the PCA algorithm to reduce the dimensionality of the traditional 128-dimensional SIFT feature descriptor is as follows:

- Set $n$ key points, and input the feature descriptors of these $n$ key points as samples. $X$ represents the sample matrix, then:

$$X = [x_1, x_2, ..., x_n]$$  

- Calculate the average feature vector $\bar{x}$ of $n$ samples:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$  

- Where $x_i$ represents the 128-dimensional feature descriptor of the $i$-th key point.
• Obtain the difference vector $d_i$ between the eigenvector of sample points and the average eigenvector:
  \[ d_i = x_i - \bar{x} \]  
  (8)

• Construct the covariance matrix C:
  \[ C = \frac{1}{n} \sum_{i=1}^{n} d_i d_i^T = \frac{1}{n} QQ^T, \quad Q = [d_1, d_2, ..., d_n] \]  
  (9)

• Find the 128 eigenvalues $\lambda_i$ and 128 eigenvectors $e_i$ of the covariance matrix.

• Arrange the 128 eigenvalues obtained in descending order $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_{128}$ and the corresponding eigenvectors $[e_1, e_2, ..., e_{128}]$.

• Select the eigenvectors of q largest eigenvalues as the direction of the principal component.

• Construct a $128 \times q$ projection matrix A, its columns are composed of q eigenvectors.

• Project the SIFT feature descriptor to obtain descriptors $y_1, y_2, ..., y_n$, namely:
  \[ y_i = x_i \times A \]  
  (10)

The size of the projection matrix A is $128 \times q$, and the size of $x_i$ is $1 \times 128$, so $x_i \times A$ is a matrix of size $1 \times q$, that is: each $y_i$ is a q-dimensional feature descriptor, so that the original 128-dimensional SIFT feature vector is reduced to a q-dimensional PCA-SIFT feature vector. q can be set according to experience or dynamically selected based on characteristic values. In this paper, q is 20.

2.3. Feature point matching

The manifestation of feature descriptor is feature vector. When the feature vectors of the two images are generated, the distance between the feature vectors of the feature points is generally used as the similarity measure of the feature points in the two images. The distance in this article uses Euclidean distance.

Suppose the two images to be matched are M and N. After the feature points are extracted through the above steps, the sets of feature points are: The set of feature points of image M is $F_M = \{f(m)_1, f(m)_2, ..., f(m)_Lm\}$, the feature point set of image N is $F_N = \{f(n)_1, f(n)_2, ..., f(n)_Ln\}$, where $L_m$ and $L_n$ are the number of feature points of images M and N, respectively. When the dimension of the SIFT feature vector is k, the distance is calculated as follows:

\[ d(F_M, F_N) = \sqrt{\sum_{i=1}^{k} (f(m)_i - f(n)_i)^2} \]  
(11)

And according to the distance ratio criterion for feature matching, that is, for a feature point, set the closest Euclidean distance in another image to $d_{min}$, and the next closest Euclidean distance to $d'_{min}$, and the ratio of $d_{min}$ to $d'_{min}$ is $r = d_{min}/d'_{min}$. If the obtained distance ratio is greater than the predetermined threshold $\epsilon$, the feature point is considered to be matched successfully, otherwise, it is considered that it does not match the feature point.

2.4. Elimination of Mismatched Feature Points Based on Improved RANSAC Algorithm

2.4.1. Introduction to RANSAC algorithm. The RANSAC algorithm is a robust model parameter estimation algorithm. It has strong errata for sample data sets and strong robustness. The basic idea is to use an iterative method to continuously extract sample sets from the data set, build a model in the data set containing outliers, and then extract the points in the dataset to test the model.

In the RANSAC algorithm, the points in the dataset that are suitable for the model are called interior points. Otherwise, they are called exterior points or wild points.

The traditional RANSAC algorithm is an iterative algorithm, the calculation process is as follows:

• For a dataset U containing N samples, first, according to the number of unknown parameters in the model to be calculated, the minimum number of data n required for each random sampling is
determined, and the sample set \( u \) is obtained and the data is guaranteed the total number of data \( N \) in the set \( U \) must be greater than the minimum number of data \( n \) per sampling, that is, \( N > n \) must be satisfied.

- After determining the minimum number of samples \( n \), randomly select \( n \) data from the data set \( U \), use these \( n \) data to estimate the model for the first time, which is the sample set \( u \), and obtain the model \( H \) that conforms to the sample set \( u \), \( H \) is the first estimation model.

- For the remaining data in the data set \( U \), sequentially calculate the error \( t \) between it and the model \( H \). The data point is an interior point if the error \( t \) is less than the error threshold \( \delta \), exterior point otherwise. All the interior points and the sample \( u \) are co-existed in the set \( U' \).

- Count the total number \( e \) of data elements in the set \( U' \). If \( e \) is greater than the threshold \( E \), the model \( H \) is judged to be the correct model, and the set \( U' \) is more accurate after removing a part of the outer points, use the set \( U' \) cycle steps (1)∼(3) to select the smallest sample set and estimate the model parameters.

- If the appropriate minimum sample has not been selected to estimate the model \( H \), when the sampling times exceeds the threshold \( L \), the proportion of interior points in the dataset \( U \) is too low, and the RANSAC algorithm ends in failure.

Whether the randomly selected data points are interior points have great influence on the sampling frequency of the RANSAC algorithm. The random sampling method in the RANSAC algorithm will cause high algorithm iterations, large amount of calculation, and long time-consuming.

2.4.2. Improvement of RANSAC algorithm. The interior points cannot be obtained at once during the random sampling process cause the problem that RANSAC algorithm has too many iterations and takes a long time. Therefore, this paper introduces a measurement factor \( \eta \) base on the RANSAC algorithm, and \( \eta \) can be used to measure whether a pair of SIFT feature points are interior points. The flow chart of the improved algorithm is shown in Figure 1:

Figure 1. Improved RANSAC algorithm flow chart
Let $\eta = \frac{1}{r_{d_{\text{min}}}}$, because the smaller the minimum Euclidean distance between matching point pairs, the two points are closest. The smaller $r$ also proves that the similarity of matching point pairs is higher. Therefore, the smaller the $r_{d_{\text{min}}}$, the higher the similarity of the two SIFT matching points; similarly, the larger the value of the measurement factor $\eta$, the greater the probability that the SIFT feature point pair is an interior point. According to this measurement standard, we can sort all SIFT feature matching points obtained from rough matching in descending order according to the measurement factor $\eta$. Therefore, when the RANSAC algorithm randomly samples the matching pairs in the sample set, the matching point pair with the higher value of the weighing factor $\eta$ is first selected as the sample, and the model is estimated to achieve the purpose of reducing the number of samples, reducing the amount of calculation and algorithm time consumption. In theory, the improved algorithm can maintain the robustness of the traditional RANSAC algorithm, while improving the efficiency of the RANSAC algorithm and reducing the time consumption of the algorithm.

3. Experimental results and analysis
The image data used in this experiment comes from the data actually taken by the NAO robot. Through the top camera with a resolution of 640×480, the image of the target to be recognized under normal, rotating, partial occlusion, multi-object and low-light conditions is obtained as test the input of the system. The hardware environment is a 64-bit operating system, Intel Core i5-4900 CPU, the main frequency is 3.3GHz, 6G memory, and the software environment is the MatlabR2017b software platform. In the scene of illumination change, rotation, partial occlusion and multiple targets, the traditional SIFT method and the algorithm in this paper are used to conduct experiments respectively. The matching effects of the two algorithms on the target image are shown in Figure 2 Figure 3, Figure 4, Figure 5 and Figure 6. As shown, the specific comparison result data are shown in Table 1.

(a) The classical SIFT algorithm     (b) The algorithm in this paper

Figure 2. Image matching in normal scenes

(a) The classical SIFT algorithm     (b) The algorithm in this paper

Figure 3. Image matching in object rotation scenes
(a) The classical SIFT algorithm (b) The algorithm in this paper
Figure 4. Image matching in partially occluded scene

(a) The classical SIFT algorithm (b) The algorithm in this paper
Figure 5. Image matching in multi-object scenes

(a) The classical SIFT algorithm (b) The algorithm in this paper
Figure 6. Image matching in low light scenes

Table 1. Comparison of matching results between our algorithm and traditional SIFT algorithm in multiple scenes

| Scenes              | Algorithm | Success matched feature points | Mismatched feature points | Accuracy(%) | Time(s) |
|---------------------|-----------|--------------------------------|--------------------------|-------------|---------|
| Normal              | SIFT      | 28                             | 7                        | 80          | 0.427   |
|                     | Our       | 24                             | 4                        | 85.71       | 0.202   |
| Object rotation     | SIFT      | 5                              | 2                        | 71.43       | 0.397   |
|                     | Our       | 5                              | 1                        | 83.33       | 0.192   |
| Partial occlusion   | SIFT      | 4                              | 4                        | 50          | 0.309   |
|                     | Our       | 4                              | 2                        | 66.67       | 0.135   |
| Multi-object        | SIFT      | 11                             | 5                        | 68.75       | 0.569   |
|                     | Our       | 10                             | 3                        | 76.92       | 0.306   |
| Low light           | SIFT      | 4                              | 6                        | 40          | 0.454   |
|                     | Our       | 3                              | 3                        | 50          | 0.223   |
It can be seen from Table 1:

- In terms of matching accuracy. No matter the image is under the condition of illumination change, rotation, partial occlusion and multiple objects, the algorithm in this paper has very stable matching performance.
- In terms of matching time. Compared with the traditional SIFT algorithm, the algorithm in this paper improves the object recognition speed by 50% on average, and improves the efficiency of matching.

Although the number of matching feature points in this algorithm is less than the traditional SIFT algorithm, it significantly reduces the number of mismatched feature points, and the matching accuracy is higher than that of the traditional SIFT algorithm. At the same time, the algorithm in this paper greatly reduces the matching time. Therefore, this algorithm not only maintains the stability and accuracy of the SIFT algorithm, but also reduces the matching time.

4. Conclusion
The descriptor vector generated by the traditional SIFT algorithm reaches 128 dimensions, which will cause the algorithm to consume a long time, and it is difficult to meet the requirements of real-time applications. In this paper, the PCA algorithm is used to reduce the dimension of the SIFT descriptor from 128 to 20, which reduces the computational complexity. Because the traditional RANSAC algorithm cannot guarantee that the samples are all interior points in random sampling of the sample points, which leads to a large number of iterations, a large amount of calculation and a long time-consuming, this paper proposes an improved RANSAC algorithm. The algorithm introduces a measurement factor $\eta$ that can reflect whether a pair of matching points are interior points. When selecting samples, the sample with a higher $\eta$ value is selected for model estimation, which can reduce the number of iterations of the RANSAC algorithm, thereby reducing time consumption. While maintaining the robustness of the original SIFT algorithm, the algorithm in this paper eliminates incorrect matching points, improves the recognition speed and improves the efficiency of the algorithm.

References
[1] Brown L G. A survey of image registration techniques[J]. ACM computing surveys(CSUR). 1992, 24(4): 325-376.
[2] LI Yufeng, LI Guangze& GU Shaohu, et al. Image mosaic algorithm based on area blocking and SIFT [J]. Optics and Precision Engineering,2016, 24 ( 5 ) : 1197-1205 ( in Chinese).
[3] Lowe D G.Distinctive image features from scale invariant key points[J].International Journal of ComputerVision, 2004, 60 (2) : 581-588.
[4] Hyun B J, Jae W J. Feature-point extraction based on an improved SIFT algorithm [C]. International Conference on Control & Automation and Systems. Jeju: IEEE, 2017: 345-350
[5] W.Ma.Remote Sensing Image Registration With Modified SIFT and Enhanced Feature Matching[C].IEEE Geoscience and Remote Sensing Letters.2017, Vol(14): 3-7
[6] Wu Lijan, WEN Jicheng&CHEN Zhicong, et al.Fast image match th an improved effi cient SIFT algorith[J]. Journal of Fuzhou University(Natural Science Edition),2017,45(6):801-809( in Chinese).
[7] FENG Wenbin, LIU Baohua. Research on image matching based on improved SIFT algorithm[J]. Computer Engineering and Applications,2018,54 (3): 200-205( in Chinese).
[8] SHAOJin G da,YANGShuai & CHENGLin. UAV Image Matching Algorithm Based on Improved SIFT Algorithm and Two-stage Feature Matching[J]. COMPUTER SCIENCE,2019,46(6):316-321
[9] CHENG De-qiang, LI Teng-teng & GUO Xin, et al. Improved SIFT neighborhood voting image matching algorithm[J]. COMPUTER ENGINEERING AND DESIGN, 2020, 41(1): 162-168 (in Chinese).

[10] Zhou Lifang, Li Weisheng, Huang Ying. Pattern recognition and engineering application[M]. Beijing: China Machine Press. 2013: 105-107. (in Chinese).

[11] R. Schnabel, R. Wahl, R. Klein. Efficient RANSAC for Point-Cloud Shape Detection. Computer Graphics forum[J], 2007, 26(2): 214-226