Portrait matching based on sift features

Xingxing Li¹, Jun Luo², Chao Duan³, Panpan Yin⁴, Yan Zhi⁵

Department of electronics and information engineering, Guangzhou College of Technology and Business, Foshan, 528100, China.

¹lixingxing@163.com, ²luojun@126.com, ³MDL0857a@126.com,
⁴abc0604@126.com, ⁵sparks666888@163.com

Abstract. Image matching refers to identifying homonymous points between two or more images by a certain matching algorithm. For example, in two-dimensional image matching, by comparing the correlation coefficients of the windows of the same size in the target image and the search image, the central point of the window corresponding to the maximum correlation coefficient in the search area is taken as the key point. In this paper, SIFT operator is used for image matching. The experimental results show that the method can effectively match the target.

1. Introduction

Image matching is an important research content in computer vision and image processing. Image matching includes template matching, histogram matching, shape matching and other matching algorithms. Template matching is to compare the target image with the source image to determine whether there is the same or similar region in the source image. If the region exists, it can also determine its location and extract the region. The main limitation of template matching is that it can only move in parallel. For example, the target matches in the original image rotates or changes in size, and the algorithm is invalid. Histogram matching is based on the statistical histogram of image color features. However, because the histogram cannot reflect the position information of color, it is possible that the content of two images is completely different and the histogram is similar. Therefore, simple color histogram matching may cause matching errors. Scale-invariant feature transformation (SIFT) is a computer vision algorithm. It is used to detect and describe local features in images. It searches for extreme points in spatial scale and extracts their position, scale and rotation invariants. This algorithm was published by David Lowe in 1999 and summarized perfectly in 2004. SIFT operator is invariant for image scaling, translation and rotation, and is partially invariant for illumination change and affine. It is the most appropriate for image matching.

2. SIFT principle

Here is to introduce the principle of SIFT and some steps.

2.1. Image scale space

Before understanding image feature matching, it needs to be clear that why two photos can match well is because of the high similarity of feature points. In order to find image feature points, first of all, we must know a concept—"image scale space". In ordinary life, when a photograph is viewed with the human eye, the image will become blurred gradually with the increase of the observation distance. When
a computer "looks" at a photograph, it will observe the photograph from different "scales". The larger
the scale it observes, the more blurred the image is.

2.2. Multi-resolution image pyramid
The pyramid representation of images is usually used in the multi-scale representation of early images. Image pyramids are a group of results obtained from the same image at different resolutions. Generally, the generation process includes two steps: smoothing the original image first, and then de-sampling the processed image. After down-sampling, a series of shrinking images are obtained. Obviously, in a traditional pyramid, the image of each layer is as half long and half high as that of the previous layer. Although the multi-resolution image pyramid is easy to generate, its essence is to down-sample, and the local features of the image are difficult to maintain, that is, the scale invariance of the features cannot be maintained.

2.3. Gaussian scale space
The image blurring degree is used to simulate the imaging process of the object on the retina from far to near. The closer the object is, the bigger the size of the image and the blurred the image is. This is the Gaussian scale space. The blurred image with different parameters (resolution unchanged) is another form of scale space.

The convolution operation of image and Gauss function can blur the image. Different "Gauss Kernel" can be used to get different blurred images. The Gaussian scale space of an image can be obtained by its and different Gaussian convolutions:

\[ L(x, y, \sigma) = G(x, y, \sigma) \ast I(x, y) \] (1)

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

Among them, \( L(x, y) \) is the Gauss scale space \( G(x, y) \) of the image, which is the Gauss kernel function. It is called scale space factor. It is the standard deviation of Gauss normal distribution, which reflects the degree of image blurring. The larger the value is, the more blurred the image is, the larger the corresponding scale is. \( L(x, y) \) represents the Gaussian scale space of the image.

3. DoG Spatial Extremum Detection
In order to find the extremum points in scale space, each pixel should be compared with all the adjacent points in its image domain (the same scale space) and scale domain (the adjacent scale space). When it is larger than (or smaller than) all the adjacent points, the change point is the extremum point. As shown in the figure, the detection points in the middle should be compared with 8 pixels in the 3 3 neighborhood of the image and 18 pixels in the 3 3 field in the upper and lower layers of the image, totaling 26 pixels.

3.1. Delete bad extreme points (feature points)
By comparing the detected DoG local extremum points with the real discrete space search results, since the discrete space is the result of continuous space sampling, the extremum points found in the discrete space are not necessarily the real extremum points, so we should try to eliminate the points that do not meet the conditions. Extremum points can be found by curve fitting of DoG function in scale space. The essence of this step is to remove points with very asymmetric local curvature of DoG.

4. Finding the Main Direction of Characteristic Points
After the above steps, we have found the feature points in different scales. In order to achieve image rotation invariance, we need to assign the direction of the feature points. The orientation parameters are determined by the gradient distribution characteristics of the neighborhood pixels of feature points, and then the stable orientation of the local structure of key points is obtained by using the gradient histogram of the image.
When the feature points are found, the scale of the feature points can be obtained, and the scale image \( L(x, y, \sigma) \) where the feature points are located can also be obtained. Calculating the magnitude and angle of the region image centered on the feature point and radius \( 3 \times 1.5 \sigma \), the modulus \( m(x, y) \) and direction \( \theta(x, y) \) of the gradient of each point \( L(x, y) \) can be obtained by the following formula.

\[
m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}
\]

\[
\theta(x, y) = \arctan \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)}
\]

After calculating the gradient direction, the histogram is used to calculate the gradient direction and magnitude of the pixels in the neighborhood of the feature points. The transverse axis of the histogram of gradient direction is the angle of gradient direction (the range of gradient direction is 0 to 360 degrees, the histogram has 10 bin per 36 degrees, or 8 bin per 45 degrees). The longitudinal axis is the accumulation of gradient magnitude corresponding to gradient direction, and the peak value of the histogram is the main direction of the feature points.

Lowe pointed out in his paper that after getting the main direction of feature points, three information \((x, y, \sigma, \theta)\) can be obtained for each feature point, i.e. location, scale and direction. Thus, a SIFT feature area can be determined. A SIFT feature area is represented by three values. The Center represents the position of the feature point, the radius represents the scale of the key point, and the arrow represents the main direction. Key points with multiple directions can be duplicated into multiple copies, and then the direction values are assigned to the duplicated feature points. A feature point produces multiple feature points with equal coordinates and scales, but different directions.

5. Experiment

5.1. Experimental process

5.1.1. Image preprocessing: identifying source and target images. Choose a life photo as the source image, as the left image in Figure 1, and intercept a female head as the original target image, as the right image in Figure 1. Then, a series of operations such as clockwise rotation of 90 degrees, 165 degrees and affine transformation are performed on the original target image as the target image, as shown in Figure 1.

![Figure 1](image.png)

**Figure 1.** The left two images are obtained by rotating different angles of the original target image. The rotating angles are clockwise 90 degrees and 165 degrees, respectively. The right two images are obtained by affine transformation of the original target image.
5.1.2. Experimental results of image matching using SIFT operator

![Image 1](image1.jpg)  
Figure 2. Result display of original target image

![Image 2](image2.jpg)  
Figure 3. Image Matching Display Map with Different Rotation Angles

![Image 3](image3.jpg)  
Figure 4. Image Matching Display Graphs with Different Affine Transformations
6. Conclusion
SIFT feature is a very stable local feature because of its invariance to rotation, scaling, brightness and so on. It plays an important role in image processing and computer vision. It is also very complex. In this paper, a method of extracting invariant features from images is used to reliably match different views of objects or scenes. These features are invariant and rotational to image scale, and show robust matching across a large number of affine distortions, in changing 3D viewpoints, adding noise, and changing lighting. These features are very unique, in a sense, a single feature can correctly match the features of large databases with high probability, from many images.

7. References
[1] Chen K, Song X, Liu J. Pedestrian tracking framework based on deep convolution network and SIFT [J]. Scientia Sinica (Informationis), 2018.
[2] Zhang H, Han J, Jia H, et al. Features Extraction and Matching of Binocular Image Based on SIFT Algorithm [C]// International Conference on Intelligent Transportation. IEEE Computer Society, 2018.
[3] Zhang W, Liu D, Xiong Z, et al. SIFT-based adaptive prediction structure for light field compression[C]// 2017 IEEE Visual Communications and Image Processing (VCIP). IEEE, 2017.
[4] Lowe D G. Object recognition from local scale-invariant features [C]. international conference on computer vision, 1999: 1150 - 1157.
[5] Lowe D G. Distinctive Image Features from Scale-Invariant Keypoints [J]. International Journal of Computer Vision, 2004, 60(2): 91 - 110.