**1, About sift**

Scale invariant feature transform (SIFT) is a computer vision algorithm used to detect and describe the local features in the image. It looks for the extreme points in the spatial scale, and extracts the position, scale and rotation invariants. This algorithm was published by David Lowe in 1999 and summarized in 2004.

The description and detection of local image features can help to identify objects. SIFT features are based on some local appearance interest points on the object, and have nothing to do with the size and rotation of the image. The tolerance of light, noise and some changes of micro viewing angle is also quite high. Based on these characteristics, they are highly significant and relatively easy to retrieve. In the feature database with a large number of parent numbers, it is easy to identify objects and there is little misidentification. Using SIFT features to describe the detection rate of some objects is also very high, and even more than three sift object features are enough to calculate the location and orientation. In today's computer hardware speed and small feature database conditions, identification speed can be close to real-time operation. SIFT features have a large amount of information and are suitable for fast and accurate matching in massive databases.

**2, sift algorithm principle**

The essence of SIFT algorithm is to find key points (feature points) in different scale space and calculate the direction of key points. The key points found by SIFT are those that are very prominent and will not change due to lighting, affine transformation, noise and other factors, such as corner points, edge points, bright spots in dark areas and dark spots in bright areas.

**1. characteristics**

1. SIFT feature is a local feature of image, which is invariant to rotation, scale scaling and brightness change, and stable to some extent to angle change, affine transformation and noise;
2. Good uniqueness and abundant information are suitable for fast and accurate matching in massive feature databases;
3. Multiplicity, even a few objects can produce a large number of SIFT eigenvectors;
4. High speed, the optimized SIFT matching algorithm can even meet the real-time requirements;
5. Extensibility, which can be easily combined with other forms of eigenvectors.

**2. Solvable problems**

The performance of image registration / target recognition and tracking is affected by the state of the target itself, the environment of the scene and the imaging characteristics of the imaging equipment. To some extent, SIFT algorithm can solve:

1. Rotation, scaling, translation (RST) of target
2. Image affine / projection transformation (viewpoint)
3. illumination
4. occlusion
5. clutter scene
6. Noise

**3. Algorithm decomposition**

Lowe decomposes SIFT algorithm into four steps as follows:

1. Extreme value detection in scale space  
For two-dimensional image, we build a DOG (difference of Gaussian) pyramid. The meaning of the DOG scale space is: it can be expressed by the Gauss function of a scale space and the convolution of the image.  
  
G is a Gaussian function with variable scale, \* \* \* I \* \* \* is the spatial coordinate, and Sigama is the scale.  
In order to determine the location of the feature points, we need to build a Gaussian pyramid.  
After getting the Gauss pyramid, we can get the Gauss difference DOC pyramid through two adjacent Gauss scale spaces. The formula of the whole process is as follows:  
  
After the establishment of the Gaussian difference pyramid, the feature points are many extreme points in the space of the DOG scale (I think that the location coordinates of the feature points can be obtained by calculating the partial derivative of the feature points). To find the extreme points, it is necessary to compare each point with 26 points in the neighborhood, including 8 points adjacent to each other on the same scale and 18 points adjacent to each other on the adjacent scale.

2. Determination of scale direction of feature points  
It is not enough to get the coordinates of the feature points. We must increase the direction scale information.  
·Using the finite difference method, the amplitude and phase of the image gradient in the range of taking the feature point as the center and taking 3 times of the radius of the West gamma as the radius are calculated.  
·Using histogram statistical method, the gradient direction and amplitude of all pixels in the neighborhood are obtained. The main direction of the feature points is the direction represented by the peak value of the histogram. If the main direction is determined, the SIFT algorithm can be rotation invariant.  
·The calculation formula of the main direction is: (Note: L represents the scale of the feature point)  
  
3. Generation of eigenvectors  
The feature vector is finally calculated by the neighborhood gradient information of the feature points.  
·First, rotate the coordinate axis position to the main direction of the feature point  
·Then, taking the feature point as the center, 16 points near the feature point are selected as seed points, and the gradients in 8 directions are calculated respectively  
·The final 128 dimensional vector is the feature vector  
4. Matching of feature points  
Usually, the nearest method is used, that is, to find the nearest feature point in another image, that is, the shortest Euclidean distance.  
The meaning of Euclidean distance is the distance between two points  
The details are as follows:

·For A feature point in the target image A, the Euclidean distance between that point and all feature points in the image B is obtained;  
·The order of the obtained Euclidean distance value is given;  
·Find out the feature points in the target graph B of the smallest and the next smallest (the second smallest) Euclidean distance value pair, and calculate the ratio of the two distances;  
·If the ratio is less than a certain threshold, the two points are matching points, otherwise they will not match.  
(Note: this threshold is an empirical threshold: it is the main reason for the influence of mismatch, and the threshold size generally selected in the experiment is about 0.6.)

**4. disadvantages**

SIFT has unparalleled advantages in image invariant feature extraction, but it is not perfect, and still exists:

1. Real time is not high.
2. Sometimes there are fewer feature points.
3. The feature points can not be extracted accurately for the target with smooth edge.