Applied Mechanics and Materials
ISSN: 1662-7482, Vols. 644-650, pp 4157-4161
doi:10.4028/www.scientific.net/AMM.644-650.4157

Submitted: 2014-07-20
Accepted: 2014-07-21
Online: 2014-09-22

Image Feature Matching Based on SIFT Algorithm

ZHANG Xin[1, a], ZHANG Ya Sheng[1,b] ,Yao Hong[1,c]

[1] Equipment Academy of PLA, Beijing, China, 101416

[a]zhangxin081404@163.com

[b]lizhizys@263.net,

[c]momoridy@163.com

Keywords: SIFT, Image matching, Scale invariance

Abstract. In the process of image matching, it is involved such as image rotation, scale zooming, brightness change and other problems. In order to improve the precision of image matching, image matching algorithm based on SIFT feature point is proposed. First, the method of generating scale space is introduced. Then, the scale and position of feature points are determined through three dimension quadratic function and feature vectors are determined through gradient distribution characteristic of neighborhood pixels of feature points. Finally, feature matching is completed based on the Euclidean distance. The experiment result indicates that using SIFT feature point can achieve image matching effectively.

Introduction

In the fields of digital image processing, pattern recognition and computer vision, the image matching is a very important technology which develops quickly in recent years and is applied widely on medical diagnosis, remote sensing image matching, object pose determination, image fusion and other aspect[1]. Image matching is a process of feature extraction and matching whose target is two or more images coming from the same object in different times and imaging conditions, namely a kind of space transformation relationship is searched to make two correspondence points in two images represented the same object point identical in space position[2].

When mankind recognizes an object, no matter the object is distant or near, it can be correctly identified. That is called “scale invariance”. Similarly, when the object rotates, we also can identify it correctly. That is called “rotation invariance”. So, how to make the machine have the ability like human? This is what image local invariance need to solve. In recent years, an algorithm based on local invariant description operator is developing rapidly in image matching field because of high stability and high matching degree of local feature.

SIFT (Scale Invariant Feature Transform) algorithm was proposed by David G Lowe[3] in 1999 and improved in 2004 which is a extraction method of local invariant feature point. SIFT algorithm could handle two images matching when their translation, rotation scale and light conditions are changed. It has stable feature matching performance towards visual angle transformation and affine transformation.

SIFT feature vector has some performance as follows[4]:

1. SIFT feature is the local feature of the image, which keeps invariance towards rotation, scale zooming and luminance transformation and keeps stability towards visual angle transformation, affine transformation and noise;
2. Uniqueness and rich information;
3. Even if a handful of rich texture object also can produce a large number of SIFT feature vectors;
4. Scalable, it can be easily combined with other forms of feature vectors.

SIFT algorithm is divided into the following three steps[5]: First, gaussian pyramid and difference pyramid are built in turn, extremum is calculated in the latter. So that we can get feature points, at the same time their coordinates and scale factors are recorded; Second, more accurate coordinates and

scale factors are positioned mainly by using the method of curved surface fitting. Meanwhile, direction factor is identified as key points of the gradient direction of neighborhood, which gives the characteristics of rotation maintain stability. Finally, through the feature vector of Euclidean distance, image matching result is given.

Basic Principle of SIFT Algorithm and Implementation Steps

(1) Scale space extrema detection

Scale space theory is the basis for the detection of invariant features. Witkin proposed the theory of scale space in 1983, in which the problem of one-dimensional signal smoothing processing is discussed. Koenderink in 1984 expanded the theory to the two-dimensional image and proved the gaussian convolution kernel was the only transformation kernel for scale transformation.

A two-dimensional image scale space at different scales can be obtained by the convolution of image and gaussian kernel.

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

where $G(x,y,\sigma)$ is a gaussian kernel function.

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

where (x,y) is the image pixel coordinates, $I(x,y)$ is image data. σ is scale space factor, which is gaussian normal distribution of the variance and reflects the image smooth degree. Its value is smaller, image smooth degree is smaller, scale is also smaller. $L(x,y,\sigma)$ represents the scale space of image.

To efficiently detect the stable feature points in scale space, Low used the difference gaussian DoG extremum in scale space as judgment. DoG operator is defined as the difference of two different scale gaussian kernels and is approximation of the normalization of LoG operator. K is the ratio between two adjacent scale factors and the DoG operator is defined as follows:

$$\begin{aligned} D(x,y,\sigma)&=(G(x,y,k\sigma)-G(x,y,\sigma))*I(x,y)\\ &=L(x,y,k\sigma)-L(x,y,\sigma) \end{aligned}. \tag{3}$$

$D(x,y,\sigma)$ is structured as shown in Fig.1. The left is gaussian images and the right is DoG.

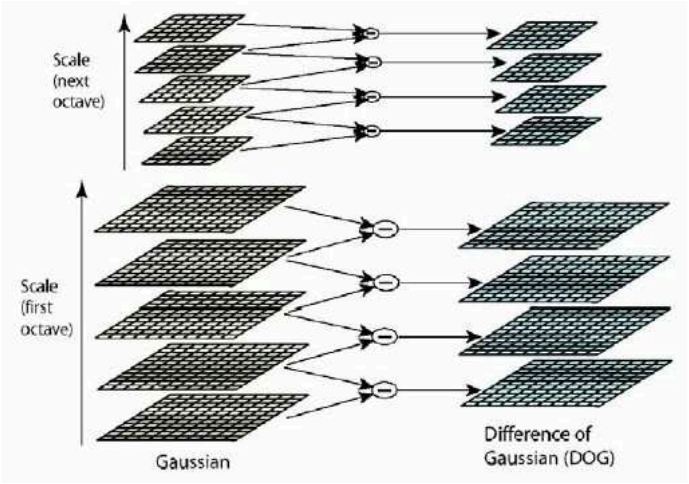

Fig.1 Gaussian and DoG

The gaussian image pyramid is divided into groups, each group is divided into layers. A set of layers differ scales, scales between adjacent layers differ a scale factor k. In S scale interval changes scale factor , such as double, k is should be $2^{\frac{1}{s}}$. In order to get DoG extremum in the all pyramids, S+3 layers of gaussian smoothing images should be generated in the gaussian pyramid. The bottom layer of next group images is obtained through last group image whose scale is 2σ for down sampling with factor 2, where σ is scale factor of most bottom image in last group. DoG pyramid is obtained by the neighboring gaussian image pyramid subtracting.

Each gaussian image σ is:

$$\sigma(o,s)=\sigma_0 2^{\frac{o+s}{S}}. \quad (4)$$

where σ_0 is basis scale factor; o , s respectively represent image group coordinates and group inner coordinates, o∈o_{min}+[0，…O-1]，s∈[0，…S-1]; o_{min} is the first pyramid group coordinate, usually o_{min} is 0 or 1, when set to 1, the image doubles before calculating gaussian scale space.

After the pyramids are structured, DoG local extremum is starting testing. Each pixel need to be compared with the same scale of neighborhood around eight pixels and adjacent scale corresponding location of the neighborhood around 9 x 2 pixels totally 26 pixels. Only when the DoG value is more or less than the 26 pixels, that point is as extreme value point and stored for subsequent calculations. As is shown in Fig.2.

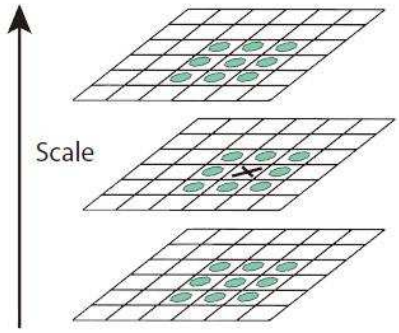

Fig.2 DoG scale space local extremum detection

(2) Identify position and scale of the key point

The next step is precisely positioning for the key point. Firstly, the image is shown as curved surface in the scale space, according to the each pixel gradient calculation of surrounding the feature points, using three dimensional quadratic function to fit and calculate. The coordinates of the feature points and scale factor are obtained. This step has very high accuracy which can achieve sub-pixel level. Secondly, through extrema detection we get many candidate feature points which need to be filtrated through two steps: First, because the difference pyramid edge response is relatively intense and the edge corresponding points are not stable, which should be filtered; Second, in order to ensure stable matching and strengthen antinoise performance, the low contrast points need to be screened out.

(3) Determination the key point orientation

By determining the orientation of the key point, it can make the feature points with rotation invariance. The key point orientation is determined according to the gradient distribution characteristics of the neighborhood pixels. For each gaussian image, each point $L(x,y)$ gradient module $m(x,y)$ and orientation $\theta(x,y)$ can be calculated as follows.

$$m(x,y)=\sqrt{\left[L(x+1,y)-L(x-1,y)\right]^2+\left[L(x,y+1)-L(x,y-1)\right]^2}. \quad (5)$$

$$\theta(x,y)=\tan^{-1}\frac{L(x,y+1)-L(x,y-1)}{L(x+1,y)-L(x-1,y)}. \quad (6)$$

(4) Generation feature vector

First, in order to ensure the rotation invariance, the coordinate axes are rotated to the orientation of the key point. Then to the key point as the center, an 8×8 window is taken, as is shown in Fig.3. Every compartment represents a pixel of the key point neighborhood in scale space. The arrow direction represents the gradient orientation of the pixel and the arrow length represents the gradient modulus value. The circle in Fig.3 represents the scope of gaussian weighting (closer to the key point pixel gradient orientation , more contribution). Then on each 4×4 small piece, gradient orientation histogram is calculated in eight directions. Drawing the accumulative value of each gradient orientation can form a seed point as shown in Fig.3. A key point consists of 2×2, amount to 4 seed

points, each seed point has eight direction vectors. This kind of idea of neighborhood direction information coalition enhances the ability of resisting noises, at the same time for feature matching with position errors it provides good fault tolerance.

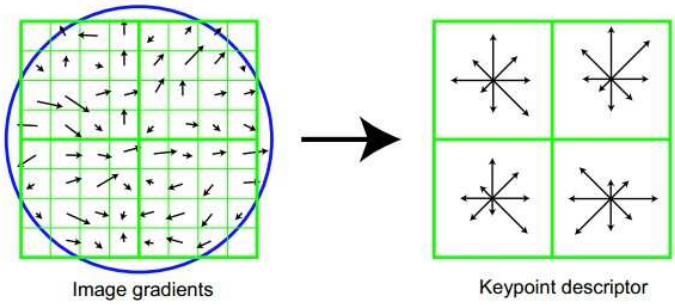

Fig.3 Structure of feature vector of feature point

In the process of actual calculation, in order to enhance robustness of matching, Lowe advices for each key point using 4×4 (16) seed points to describe. So a key point can produce 128 data and finally form a 128 dimensional SIFT feature vector. The SIFT feature vector has eliminated the influence of geometric deformation of scale transformation, rotation and so on. Sequentially, normalizing the feature vector can further eliminate the influence of illumination change.

(5) Feature matching

After generating SIFT feature vectors for two images, how to determine the suitability is the last key step of the SIFT matching algorithm. Here the Euclidean distance of feature vectors of key points is selected as the judge of similarity.

Judgement steps are follows. First, take one image as standard and take out a feature point from it, then search feature points of another image and record two points with nearest Euclidean distance. Second, the Euclidean distances from last step are compared. If the value of the smallest distance dividing second-smallest distance is less than the threshold, two feature points match successfully, otherwise it will not be accepted.

The flow chart is shown in Fig.4.

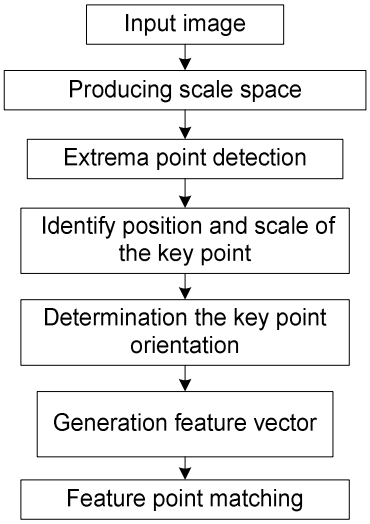

Fig.4 Implementation Steps of SIFT

Experiment Result

To the same object in different angle and distance imaging, the size of each image is 381×510 pixels. 488 feature points are extracted in Fig.5 (a) and 394 feature points are extracted in Fig.5 (b). After calculating and matching, there are 107 feature points matching successfully, the matching result is shown in Fig.6. The matching effect is very good and verifies the stability of the SIFT method.

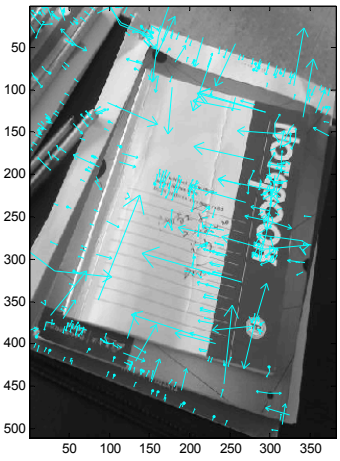

Fig.5 (a) Image 1

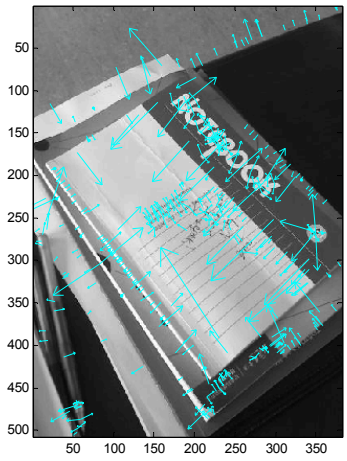

Fig.5 (b) Image 2

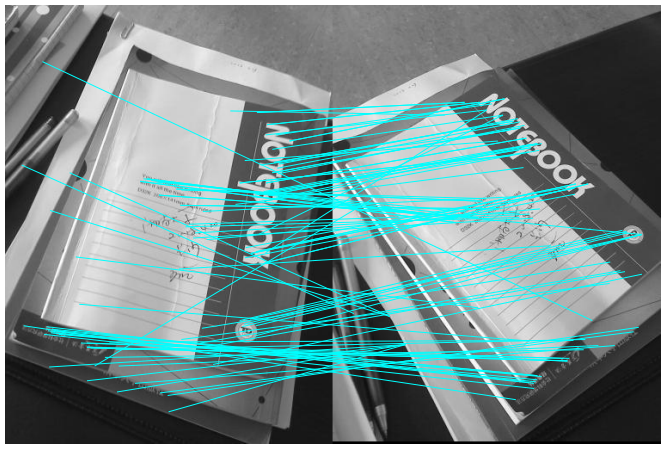

Fig.6 Matching result

Summary

Through a brief introduction and analysis to the principle of SIFT and programming implementation, experiment result indicates that SIFT algorithm, as a kind of image local feature extraction and description method, realizes image matching steadily when there are scale, rotation, illumination and perspective changes. Meanwhile, the result shows that SIFT has a high correct matching rate and can meet the demand of the target recognition and image processing.

References

[1] Ni Guoqiang, Liu Qiong. Analysis and prospect of multi-source image registration techniques [J].Opto-Electronic Engineering，2004,31(9):1-6

[2] Ji Li'e, Shi Jisheng. Image matching algorithm based on SIFT [J]. China Radar, 2013,2:48-50.

[3] Lowe,D. Distinctive image feature from scale-invariant keypoints [J]. International Journal of Computer Vision, 2004,60(2):91-110.

[4] Zhao Xiaochuan. MATLAB image processing [M]. Beijing:Beihang university press,2014.1.

[5] Ge mengyao. Video sequence image processing and intelligent matching system.[D]. Beijing:Beijing University of Posts and Telecommunications，2012,12:28-33.