

# Optimization of Image Matching Algorithm with Scale Invariant Features Transform

Li Guo<sup>1</sup>, Yuqin Yao<sup>1,a</sup>

<sup>1</sup>Chengdu University of Information Technology, Chengdu, 610225, China

<sup>a</sup> email: yyq@cuit.edu.cn

**Keywords:** Image matching, DOG pyramid, SIFT feature matching algorithm, Scale space

**Abstract.** Image exists in many key points, these points to the change of brightness the phenomenon such as scale; scale has the characteristics of stable constant. SIFT feature matching algorithm is based on the key points to complete the image matching. It has the very strong matching ability, as the image illumination change between, rotation, displacement, affine transformation and perspective transformation, and so on and so forth, can still complete accurate image matching.

## Research overview

A fundamental problem in image processing is the image matching problem, which reflected from many practical problems in many fields, such as mixing information from many sensors; comparing the differences of the images at different times and in different conditions<sup>[1]</sup>. Collect the 3D information through changes in the imaging system and the object scene; identify the target in the image; retrieval of multimedia database and so on.

In short, image matching is aligned different images from the same scene, that is to say finding the mapping relationship of the points between the different images, or establishing relations with features which we are interested in.

Correlation of matching test process is determined by the similarity; the searching space is a set of image transform, it's used to calibrate the image: searching strategy provides ideas which how to find the next transform in the space and how to test and search the best transform is in this space: refers to the extraction of feature space get the matching data information from the image<sup>[2]</sup>. The feature space is data information extracted from image and it's being used to do matching work.

The final matching accuracy of image matching will be influenced by the similarity measurement, searching space, searching strategy and feature space, all the image matching methods can be seen as the assembling of these options.

## SIFT feature matching algorithm

Scale space theory is first used in the field of computer vision; it is mainly used to simulate the multi-scale characteristics of image information<sup>[3]</sup>. The only linear kernel scaling is Gauss convolution kernel, which has been proved. Therefore, when making scale transformation to the original image, it can be accomplished by using the Gauss kernel, so as to get the sequences describing the scale space within the multi-scale image. We use formula (2.1) to define two-dimensional Gauss kernel, which  $(x, y)$  is the space coordinates and also the location of the pixels in the image,  $\sigma$  is the scale space factor.

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (2.1)$$

In different scale, after the convolution between 2D image  $I(x, y)$  and Gauss kernel  $G(x, y, \sigma)$ , we can get the description in scale space  $L(x, y, \sigma)$ , the formula is as follow<sup>[4]</sup>:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (2.2)$$

It can be derived by using the properties of Gauss function:

$$I_{p+1} = G(\sigma_p) * I_p = G(x, y, \sigma_p) * G(x, y, k^p \sigma) * I = G(x, y, k^{p+1} \sigma) * I \quad (2.3)$$

so

$$G(x, y, k^{p+1} \sigma) = G(x, y, \sigma_p) * G(x, y, k^p \sigma)$$

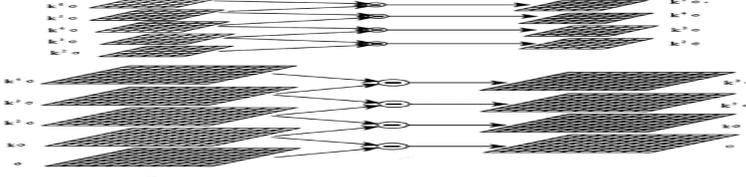


Figure 2.1 DOG Pyramid

So, if image  $I$ ,  $\sigma$  and hierarchy number  $s$  are all known, we can get an order.  $\sigma$  is scale factor which shows the degree of smoothing of an image, and the smaller the value is, the smaller the image is smoothed, and vice versa. The large scale corresponds to the overall characteristics of the image, and the small scale corresponds to the detail characteristics of the image<sup>[5]</sup>. The key to establish the scale space is to choose proper smoothing coefficient. Here, we mainly establish Gauss Pyramid and DOG Pyramid, and then detect the extreme point in the DOG Pyramid, so as to ensure the location of the feature point and it's scale.

In order to get the stable feature point in different scale space, we make convolution between image  $I(x, y)$  and Gaussian kernel  $G(x, y, \sigma)$  within different scale factor, to constitute Gaussian Pyramid. DOG Pyramid is the difference of the adjoining scale space function, describes as  $D(x, y, \sigma)$ , shows as formula (2.4):

$$D(x, y, \sigma) = G(x, y, k\sigma) - G(x, y, \sigma) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (2.4)$$

DOG Pyramid comes from the difference between the adjoining scale space function, the first floor of scale space factor in DOG Pyramid and Gaussian Pyramid is the same.

In order to detect the extreme point in DOG space, we need to compare every pixels in DOG space with and it's adjoining 8 pixels, 9 adjoining pixels in upper layer, middle layer, lower layer, all are 26 pixels, only in this way can ensure the part extreme point in both scale space and 2D space. However, the local extremum points obtained by the above methods must be validated to get the accurate feature points, this is because the DOG value is more sensitive to noise and edge. We can ensure the accurate feature point and scale by fitting 3D quadratic function, and in order to improve the stability and accuracy, we can ignore the extreme point of low contrast and unstable edge response point, so as to enhance the anti-interference ability. The two time Taylor expansion of  $D(x, y, \sigma)$  is as follow:

$$D(x) = D + \frac{\partial D^T}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 D}{\partial x^2} x \quad (2.5)$$

The extreme point's location  $\hat{x}$  is determined by the function and its inverse. When  $x \rightarrow 0$ , make derivation with the formula. When samples with different near points,  $3 \times 3$  linear system can solve the problem with lowest cost. If the offset is higher than 0.5 and it's the same in every direction that means this point is closer with other sample point. The extreme point may change, and one point may be replacing by another point, and then can be obtained formula (2.6):

$$D(\hat{x}) = D + \frac{1}{2} \frac{\partial D^T}{\partial x} \hat{x} \quad (2.6)$$

Formula (2.6) will finally get one value  $|D(\hat{x})|$ , the point can be seen as the point with lower

contrast if the value is smaller than 0.03, and we can ignore it. Only by ignoring points with lower contrast can't obtain the stable feature point, DOG scale space exists large edge response, these location can have little capacity of noise, so we have to remove these points. Here, we use  $r$  to ensure a threshold to remove the edge response points. A bad definition of Gauss differential

operator extremum exists large curvature across the edge, and the curvature at the vertical edge of the directions is small. The main curvature can be calculated by 2×2 Hessian matrix:

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad (2.7)$$

The main curvature of D is proportionate to the characteristic value of H, assume  $\alpha$  is the largest characteristic value,  $\beta$  is the smallest characteristic value. Let  $\alpha = r\beta$ , then the formula:

$$\frac{Tr(H)^2}{Det(H)} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(r\beta + \beta)^2}{r\beta^2} = \frac{(r+1)^2}{r} \quad (2.8)$$

$$\frac{Tr(H)^2}{Det(H)} < \frac{(r+1)^2}{r} \quad (2.9)$$

When  $r=1$ ,  $(r+1)^2/r$  get the smallest value, with the increasing of  $r$ , therefore, in order to detect the main curvature whether in threshold  $r$ , we only need to detect the points satisfy formula (2.9). If the point is lower than the threshold, that means this point is useful, otherwise the point should be ignored. In Lowe algorithm,  $r = 10$ .

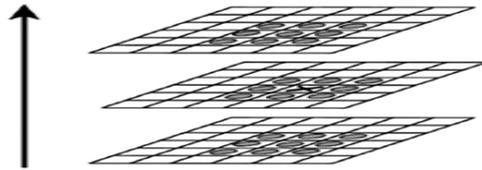


Figure 2.2 scale space extrema detection

### Image mosaic experiment of large field of view based on SIFT

In our everyday life, we usually encounter this situation: we want to put all the scene we love in one image, but the screen of the camera or mobile phone is not enough, if we focus on the scene, the whole image will be vague, at last we have to take a picture without the whole scene. Fortunately, image stitching technology can help solve this problem. In this experiment, two groups of picture are taken as experiment object, the key is that there must exist overlapping area between the two images, as shown below:



Figure 3.1 original image a



Figure 3.2 original image b

We made experiments on the original images,  $distRatio=0.5$ , the key points of SIFT in  $a$  and  $b$  are 1055 and 1492. After calculation, there are 195 pairs of key point in the two images.

Feature points matching connection diagram:

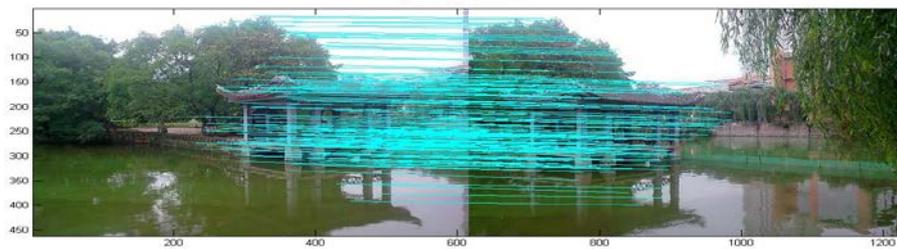


Figure 3.3 feature points matching connection diagram

a. The result of basic scene image mosaic:



Figure 3.4 the result of basic scene image mosaic

b. The experimental results of image gray stitching:

After make gray processing on original image *b*, with the same steps as above, figure 3.5 is the result of image mosaic after gray processing.



Figure 3.5 the experimental results of image gray stitching

Through the image stitching test, we verify the performance of SIFT feature matching algorithm, it's also stable when the gray level changes. In addition, the uniqueness is good, the amount of information is huge, it's also satisfy the huge data base, it can achieve matching work fast and accurately. There are other advantages with SIFT: multi quantity, small number of objects can also bring large number of SIFT feature vectors; high speed, after optimizing with SIFT, it can satisfy the real-time requirement; extendibility, it's very easy to combine with other feature vectors of comes from other algorithm.

## Conclusion

This paper has a brief study and application on SIFT algorithm. The amount of information of image matching is large, and the imaging conditions and scene is complex. At present, there still exists some problems on how to make the image feature corresponding reliable with the existing image processing technology. Due to the method used in this experiment is the hotspot in digital image processing field, the relevant literatures are limited, with great difficulty and limited time, there still exists many problems to be solved.

## References

- [1] David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints[J]. International Journal of Computer Vision . 2004 (2)
- [2] Stephen M. Smith,J. Michael Brady. SUSAN—A New Approach to Low Level Image Processing[J]. International Journal of Computer Vision . 1997 (1)
- [3] Tony Lindeberg. Feature Detection with Automatic Scale Selection[J]. International Journal of Computer Vision . 1998 (2)
- [4] Stephen M. Smith,J. Michael Brady. SUSAN—A New Approach to Low Level Image Processing[J]. International Journal of Computer Vision . 1997 (1)
- [5] Barnea DI,Silverman HF.A class of algorithms for fast digital image registration. IEEE Transactions on Computers . 1972