Image Feature Matching Based on Improved SIFT Algorithm

Jing Li

***Abstract*—**Image matching is a very important technology in the field of computer vision and image processing. SIFT algorithm can process feature matching issues between two images such as translation, rotation, scale change and illumination changes, and can have stable feature matching ability for perspective changes and affine changes to a certain extent. But SIFT feature matching also faces some problems, such as: more extracted image feature points, matching point redundancy, easiness to mismatch, large storage space and time-consuming matching etc. Therefore, this paper studies and improves SIFT algorithm to put forward an image feature matching by improved SIFT algorithm, and improves SIFT algorithm by combining the region extraction method to eliminate some unstable feature points beforehand, and extract the same target areas of two images before the matching, and then the area matching, and then match such region to reduce the SIFT feature points and increase the matching efficiency. Experiments show that such algorithm, based on the unchanged SIFT algorithm basic characteristics, has such advantages as large amount of matching points, no repeating point and higher matching efficiency, thus providing precise matching point for the image follow-up processing.

***Keywords*—**image matching, SIFT(scale invariant feature transform), feature points.

I. INTRODUCTION

Image matching is a kind of important image analysis and processing technology to determine one image area from another corresponding area taken from other sensors or find the corresponding relationship between images. When the images are acquired by different sensors or from different time and different viewpoints, the processing of image matching is usually necessary. The image matching also involves many related knowledge domains, such as image preprocessing, image sampling, image segmentation and feature extraction [1]. The image matching is the research basis in such fields as the image understanding and image restoration etc. Feature space refers to the information extracted from the image which is used to match, such as the image gray value, edge, outline, significant characteristics (such as the angular point, line intersection and high curvature point), statistical characteristics (such as moment invariants and center), high-rise structure description and syntactic description. Reasonably select matching features according to different images to improve the matching accuracy and reduce the matching complexity. SIFT operator is a very stable image feature descriptor. SIFT descriptor performance is better in the following 6 cases such as the change of illumination, image geometric relations, the resolution, image compression, rotation angle and fuzzy degree. This paper improves the extraction method of SIFT feature point, thus increasing the matching speed and the probability of correct matching [2], [3].

Image matching includes the following several aspects: feature space, search space, search strategy, the similarity measure and decision strategy. Nowadays, a lot of researches put forward all kinds of image matching methods to improve the accuracy, speed, versatility and anti-interference of the image matching. These methods are mostly focused on the matching method based on the image pixel values and matching method based on the image feature. With the clustering method, Stockman adopts the center of closed area as control point to match the image with geometrical transform relationship by using the line intersection and line endpoint matching [4]. Li combines the contour detection and gray scale local statistical features to extract feature points, and implements the initial matching and accurate matching of feature point to finally get the real matching results. Leese proposes MAD algorithm and Barnea proposes the sequential similarity detection algorithm (SSDA) which is helpful to improve the template matching speed. But these algorithms are susceptible to noise interference with low precision, and the matching result is not so good. SIFT algorithm was put forwarded by D.G.L owe in 1999. SIFT algorithm is a kind of algorithm to extract the local feature, seek local extreme point in scale space and extract the position, scale and rotation invariant. SIFT feature is the local feature of the image, which keeps invariance in terms of the rotation, scale zooming and brightness change, and also maintains a degree of stability in terms of perspective changes, affine transform and noise [5], [6].

This paper aims to effectively solve such issues of SIFT algorithm as too many extracted feature point and time-consuming matching by improving SIFT algorithm to effectively test and extract feature points in significant target area of the image. This paper first expounds and analyzes basic theory and key technology of the image matching preprocess, image feature extraction and matching, based on the above study, improves SIFT algorithm and also explains the basic idea and the algorithm flow of the algorithm including SIFT descriptor extraction steps, determination of key point position and scale, determination of key point direction, thus generating SIFT feature vector and extracting the target area in the image,

Jing Li is with the School of Intelligent Science and Information Engineering, Xi'an Peihua University, Xi’an 710125, Shaanxi, China.

and then re-using SIFT algorithm to implement the registration matching and at last analyzing the effectiveness of the algorithm in theory.

II. Scale Space Theory

Scale space theory is the basis for the detection of invariant feature. The Scale Invariant Feature Transform (SIFT) is an expression based on the area. As one important concept of scale space theory, the scale space is defined as $f_{out} = K * f_{in}$, for all input signal f_{in}, if the extreme value (order 1 differential zero crossing point) of the output f_{out} obtained by convolution of the input signal and the transform kernel K doesn't exceed the extreme value of original image, then we call K the scale space kernel and the convolution transform is the scale transform.

One nature of feature key point is to keep invariance for the scale change, so the feature point needed to be sought must be detected in different scales and can seek one certain stable feature in the scale space.

The only kernel function to transform into the scale space is Gaussian function. So one image's scale space image is defined as $L(x, y, \sigma)$, which is acquired by the convolution of the Gaussian function of variable scale $G(x, y, \sigma)$ and input image $I(x, y)$, namely:

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

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

In practice, in order to relatively efficiently calculate the key point position, it is recommended to use the difference of Gaussian function $D(x, y, \sigma)$. Its definition is as follows

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \tag{2}$$

From above formula, D is the difference of two adjacent scales, there exists one multiplied coefficient k difference for two adjacent scales [7].

Following Fig.1 is a set of Gaussian scale-space while scale factor is growing, Fig.2 is a set of difference-of-Gaussian scale-space.

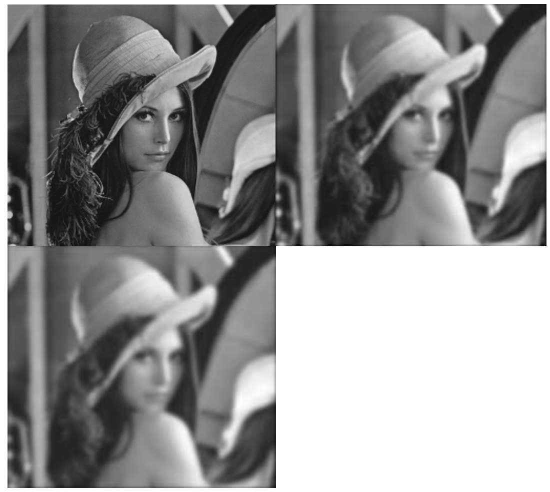


Fig. 1. A set of Gaussian scale-space while scale factor is growing

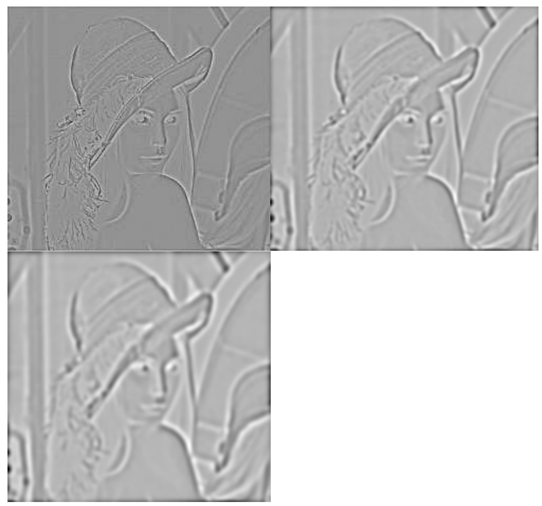


Fig. 2. A set of difference-of-Gaussian scale-space

III. 3 SIFT Feature Extraction

The generation of one image's SIFT feature vector mainly includes four steps: Gaussian scale space extremum detection, key point position and scale determination, key point direction determination and feature vector generation.

1) Extremum detection of Gaussian scale space

Gaussian kernel is the only transform kernel to implement the scale transform, which is used to realize the scale transform towards the original image to obtain the scale space expression sequence of the image under the multiple scales, and then implement the scale space feature extraction towards these sequences. The scale space expression of a two-dimensional image at different scales can be obtained by the convolution of the image and Gaussian kernel:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{3}$$

In which, $G(x, y, \sigma)$ is Gaussian kernel function.

$$G(x,y,\sigma)=\frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{4}$$

In which, (x,y) stands for pixel coordinates of the image point, $I(x,y)$ is the image data. σ is the scale space factor. $L(x,y,\sigma)$ represents the scale space of the image [8].

①Establish Gaussian pyramid

Gaussian pyramid is constructed by the convolution operation of the image $I(x,y)$ and Gaussian kernel function $G(x,y,\sigma)$ under different scale factors. The layer 1 of order 1 is the original image magnified twice, in the same order, the scale factor proportion of the adjacent two layers is k, and the scale factor of the layer 2 of order 1 is k. The layer 1 of order 2 is obtained by the sub-sampling of the middle scale image of the order 1, and its scale factor is k^2, then the scale factor of the layer 2 of order 2 is k times of the layer 1, namely k^3. The layer 1 of order 3 is obtained by the sub-sampling of the middle scale image of the order 2.

②Establish difference of Gaussian pyramid (DOG)

DoG operator is defined as the difference of two Gaussian kernels with different scales, is the normalization of LoG operator approximation. DoG pyramid is obtained by the subtraction of neighboring Gaussian image pyramids. Set k as the scale factor between two neighboring scales, then DoG operator is defined as follows:

$$\begin{aligned}D(x,y,\sigma)&=(G(x,y,k\sigma)-G(x,y,\sigma))*I(x,y)\\&=L(x,y,k\sigma)-L(x,y,\sigma)\end{aligned} \tag{5}$$

In order to get DoG extremum from the whole pyramid, smoothing images of layer $S+3$ should be generated in the Gaussian pyramid. The bottom layer of the next set of images is obtained by the downsampling with the factor of 2 implemented by the image in the former layer with the scale of 2σ, in which σ is the scale factor of the bottom layer image in the former layer [9].

2) Determine key position and scale

①Filter feature points of low-scale contrast

Define the offset of the candidate feature point x as Δx, and the contrast is the absolute value $|D(x)|$ of $D(x)$, and expand the scale space function according to the Taylor series:

$$D(x)=D+\frac{\partial D^T}{\partial x}\Delta x+\frac{1}{2}\Delta x^T\frac{\partial^2 D^T}{\partial x^2}\Delta x \tag{6}$$

In which, because x is the extremum point of DoG function, so $\frac{\partial D(x)}{\partial x}$, $\Delta x=-\frac{\partial^2 D^{-1}}{\partial x^2}-\frac{\partial D(x)}{\partial x}$, is got by the equation solution. The precise location and scale $\hat{x}$ of the final candidate point is obtained through times of iteration, and then $D(\hat{x})$ is solved after the substitution into the formula and its absolute value is $|D(\hat{x})|$. Set the contrast threshold value as T_c, the elimination formula of the low-scale contrast is:

$$\begin{cases}x\in X\mid D(\hat{x})|\geq T_c, x\in X_0\\ x\notin X\mid D(\hat{x})|< T_c, x\in X_0\end{cases} \tag{7}$$

$D(\hat{x})$ is very useful to eliminate the unstable feature points of the low-scale contrast. The extremum point of $D(\hat{x})<0.03$ is usually eliminated by regarding as the unstable feature points of the low-scale contrast. In this process, the precise location and scale of feature points is obtained [10].

②Filter edge feature points

For DoG function extremum point obtained at the edge, when compared to the point at the non-edge area, its principal curvature ratio is bigger; therefore, we can regard the point with bigger principal curvature ratio than a certain threshold point as the point at the edge and eliminate it. Hessian matrix is:

$$H=\begin{bmatrix}D_{xx} & D_{xx}\\ D_{xy} & D_{yy}\end{bmatrix} \tag{8}$$

In which, D_{xx},D_{xy},D_{yy} is the pixel difference of corresponding position of the candidate neighborhood. Preset α as the maximum feature value of H, β is the minimum feature value of H. Assume $\gamma=\alpha/\beta$, then the principal curvature ratio of $D(x)$ is proportional to γ.

Set α,β are respectively the maximum and minimum feature value of Hessian matrix H, and $\gamma=\frac{\alpha}{\beta}$, then :

$$\begin{aligned}tr(H)&=D_{xx}+D_{xy}=\alpha+\beta\\ Det(H)&=D_{xx}+D_{xy}-(D_{xy})^2=\alpha\beta\end{aligned} \tag{9}$$

$$\frac{tr(H)}{Det(H)}=\frac{(\alpha+\beta)^2}{\alpha\beta}=\frac{(\gamma\beta+\beta)^2}{\gamma\beta^2}=\frac{(\gamma+1)^2}{\gamma} \tag{10}$$

3) Determination of the key point direction

Use the distribution character of the feature point's neighborhood pixels in the gradient direction to specify the direction parameter for each feature point, so that the operator is endowed with the rotational invariance. The gradient value and direction of (x,y) are respectively:

$$\begin{cases}m(x,y)=\sqrt{(L(x+1,y)-L(x-1,y))^2+(L(x,y+1)-L(x,y-1))^2}\\ \theta(x,y)=\tan^{-1}\frac{((L(x,y+1)-L(x,y-1))}{(L(x+1,y)-L(x-1,y))}\end{cases} \tag{11}$$

In which, the scale used by L is the one of each key point. For each key point, the gradient distribution of the neighborhood pixel is summed up by histogram within the neighborhood window centered by the key point. Such histogram has 36 pillars with each column of 10° and 360 °in total. The peak value of the histogram in gradient direction represents the main direction of the neighborhood gradient of such feature point, namely, as the main direction of such feature point [11], [12].

4) Generate SIFT feature vector generation

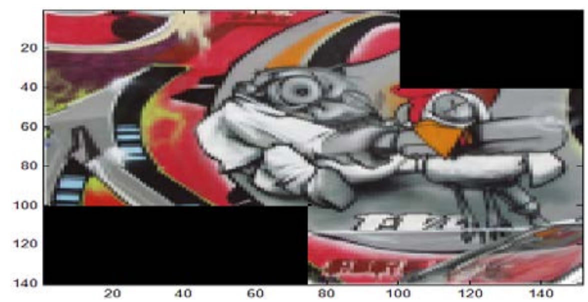
First rotate the coordinate axis to the direction of feature point. Take 8 x 8 window centering on the key points. Calculate the gradient value and direction of each pixel in the selected neighborhood area, and each square represents a pixel, and the arrow and length in the box respectively represent the gradient and direction size of such pixel. Then calculate the gradient direction histogram in 8 directions on each small piece of 4×4 and draw the accumulative value of each gradient direction to form a feature point, and each key point consists of four feature points of 2x2 with each feature point having vector information in eight directions. There are 4 such windows around the feature point to generate a 32-d vector. In actual calculation, in order to enhance the robustness of matching, 16 feature points of 4x4 are used to describe when calculating the feature description of each key point, and a key point is described with 128 data in total to finally generate 128 - d feature vector.

IV. Experimental Test and Analysis

This paper adopts Matlab 2014a software platform to test two images (reference image and the image to be matched), first detects the feature in scale space, determines the location of the feature points and the scale where feature points are located, and then uses the main direction of feature point neighborhood gradient as the direction feature of such feature point in order to realize the operator's independence on the scale and direction, and finally implements the weighted fusion of the reference image and the image to be matched. The following Fig.3 shows the matching result of SIFT feature vector, and Tab. 1 shows the time distribution comparison of each step.

(a) Rough matching results

(b) Matching results without error



(c) LM weighted fusion results

Fig. 3. SIFT feature vector matching

Tab.1. Time distribution comparison of each step by the algorithm

Initial test of extremum point in scale space	Precise positioning of key point	Direction allocation of key point	Generate feature descriptor	Total time
0.092	0.031	0.157	0.743	0.906

From the above Fig. 3, you can see that image fusion can eliminate the stitching line in the cracks and realize the smooth transition of the fusion to achieve better visual effect. After calculation of the transform matrix, on the basis of reference images, a projection process is implemented on images to be matched. Because front and back coordinate systems are different, it is necessary to implement the interpolation and re-sampling operation on the transform image's pixel coordinates with the purpose to have an integer coordinate. From Tab.1, it can be concluded that the phase to generate the feature descriptor accounts for the largest percentage. During the image feature extraction, the more feature points extracted, the longer the time consumed, therefore, the decrease of the extraction number of feature points will greatly improve the running speed of the algorithm. The proposed algorithm in this paper improves the matching speed and maintains the high accuracy of the original SIFT algorithm in the image matching at the same time, thus enhancing the matching speed and improving the robustness.

V. Conclusion

With the continuous development of computer technology, image feature point matching plays an important role in the field of computer vision and pattern recognition, at the same time, it is the basic issue in image analysis and processing, and such technology is widely used in such fields as the image stitching and industrial product testing. Because the traditional SIFT algorithm extracts more feature points, resulting in the extension of the identification time and the increase of error matching rate, this paper puts forward an improved SIFT feature point extraction method to make the detected extreme value point obtain the same observing effect as the one by human eyes and reduce the number of feature points at the same time, thus guaranteeing that the extracted target feature points are representative. The experiments show that this improved algorithm based on SIFT algorithm not only improves the efficiency of image matching, but also maintains high accuracy of image matching.

References

[1] Zaragoza Julio, Chin Tat–Jun, and Tran Quoc–Huy et al, “As-projective-as-possible image stitching with moving dlt”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1285–1298, 2014.

[2] Liu Dong, Wang Shitong, and Cao Pin et al, “Dark-field microscopic image stitching method for surface defects evaluation of large fine optics”, *Optics Express*, vol. 21, no. 5, pp. 5974–5987, 2013.

[3] Chia Wai Chong, Chew Li Wern, and Ang Li-Minn et al, “Low memory image stitching and compression for WMSN using strip-based processing”, *International Journal of Sensor Networks*, vol. 11, no. 1, pp. 22–32, 2012.

[4] Bergen Tobias, and Wittenberg Thomas, “Stitching and surface reconstruction from endoscopic image sequences: a review of applications and methods”, *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 1, pp. 304–321, 2016.

[5] Legesse F. B., Chernavskaia O., and Heuke S et al, “Seamless stitching of tile scan microscope images”, *Journal of Microscopy*, vol. 258, no. 3, pp. 223 –232, 2015.

[6] Huang Feng–Cheng, Huang Shi–Yu, Ker Ji–Wei et al, “High–performance sift hardware accelerator for real-time image feature extraction”, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 3, pp.340–351, 2012.

[7] Samsudin Salbiah, Adwan Somaya, and Arof H. et al, “Development of automated image stitching system for radiographic images”, *Journal of Digital Imaging*, vol. 26, no. 2, pp. 361 –370, 2013.

[8] Lv Xudong, and Wang Z. Jane, “Perceptual image hashing based on shape contexts and local feature points”, *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 1081–1093, 2012.

[9] Dellinger Flora, Delon Julie, and Gousseau Yann et al, “SAR-SIFT: A SIFT-Like algorithm for SAR images”, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 1, pp. 453–466, 2015.

[10] Aguilera Cristhian, Barrera Fernando, and Lumbreras Felipe et al, “Multispectral image feature points”, *Sensors*, vol. 12, no. 9, pp. 12661–12672, 2012.

[11] Sim Ngak–Leng, Kumar Prateek, and Hu Jing et al, “SIFT web server: predicting effects of amino acid substitutions on proteins”, *Nucleic Acids Research*, vol. 40, no. W1, pp. W452 –W457, 2012.

[12] Guan Juan, Wang Bo, and Bae Sung Chul et al, “Modular stitching to image single-molecule DNA transport”, *Journal of The American Chemical Society*, vol. 135, no. 16, pp. 6006–6009, 2013.