Acta Technica *62* No. 1B/2017, 265–276

Image registration algorithm based on SIFT feature descriptor[1]

Changxing Geng[2], Peng Wang[2], Pengbo Wang[2]

Abstract. With the development of society, the single image has been unable to meet the needs of mankind. The new sensor allows the ability to capture images quickly. In order to meet this demand, the image mosaic technology is born. In this paper, a registration algorithm based on feature descriptors was proposed. The feature descriptor could fully reflect the shape and texture features of the image because the feature descriptor could quantify the local structural features of the image. The SIFT algorithm was used to extract image feature points and analyze the influence of different distance selection on the feature points matching in the similarity criterion. RANSAC algorithm was used to eliminate some misunderstanding matching, which improved the accuracy. The speed of pairing was analyzed by the parameter theory of Pyramid in SIFT algorithm.

Key words. SIFT, feature descriptor, image registration, algorithm.

1. Introduction

Through the human visual sensory system, the information described in an image can be analyzed by looking at an image. Human information is transported mainly through images because they can objectively display the real form of things, so that people can observe and understand [1]. The information expressed by voice and text is relatively simple, but the image is different, which contains many large amounts of information. The abundance of information makes it difficult for people to extract useful information from them. As the most important science in image information processing, digital image processing technology also has its branches [2], including image acquisition, registration, object detection, recognition and classification and so on. Image registration is the most basic task [3], and we only need to be responsible for that the same object or the same pixel image can match with others. No matter what the conditions are, we must obtain the same space in the final image [4].

[1]This work is supported by Jiangsu Province Natural Science Foundation for Young Scholars, No. BK20140325.

[2]Robotics and Microsystems Centre, Soochow University, Suzhou, 215021, China

Whether they are single sensors or multiple sensors, they have a certain relationship within the same data at different angles [5]. If the redundant information and complementary information can be used, a reliable basis can be obtained to improve the signal-to-noise ratio. Therefore, the integration of multiple resource information is the best way to get the information that we need [6].

2. State of the art

Image feature technology can be used in a variety of occasions requiring image processing. Foreign scholars have invested a lot of effort to study image processing techniques [7]. Beginning in the late 1970s, images began to be studied. Corner features were proposed by Moravec in 1977, but corner feature detection had many limitations [8]. For example, rotation invariance and noise sensitivity were not included. Harris and Stephens improved the Moravec's detector in 1988, which significantly improved detection rates and repetition rates with invariance to rotation and grayscale changes compared with the previous features [9]. In 1998, Lindeberg systematically proposed the scale space theory of signals, which realized scale invariant feature extraction. The method of region detection for maximum stable extremum was proposed by Matas in 2002, and affine invariant was obtained in strict sense [10]. In 2004, in order to solve many matching problems such as translation, rotation, affine transformation, perspective transformation, image features and so on, Lowe proposed scale invariant feature transformation (SIFT) algorithm, which was applied to a variety of situations.

In order to improve the speed of feature matching and the performance of matching, many scholars at home and abroad have done a great deal of improvement and research on the features according to the SIFT algorithm. In 2007, Tinne Tuytelaars improved the SIFT algorithm on the basis of gray information. Scholars from various countries have done a great deal of work and have reviewed the local feature detection operators [11]. After the proposal of n SIFT feature descriptor based o gradient image [12], SURF, GLOH and PCA-SIFT were improved based on SIFT feature descriptor [13], and was widely used in the follow-up.

2.1. Methodology

Image feature technology has been applied in many fields, such as image recognition, graph retrieval, image registration, image stitching, texture recognition and other fields. Its wide range of applications makes its research more thorough. Image registration means that no matter what conditions we are shooting, as long as we can be responsible for the same image or the same pixel, the same space can be obtained at last. Some local picture information obtained in the practical application is only by single sensor or multiple sensors in a same image or items, but the data is far more than the object or scene itself. In this paper, a registration algorithm based on feature descriptors is proposed. The feature descriptor can fully reflect the shape and texture features of the image, the reason is that the feature descriptor can quantify the local structural features of the image. In order to recognize objects, the first thing is to represent the image in a reasonable way. This is to make it easier for us to match, so that only the same target can be matched in the absence of conditions [14]. In addition, it is necessary to take into account factors such as time, resolution, light, posture, etc. Why the same object appears differently in different images because the various unstable factors affect the state of the target itself and the environment in which the scene is situated. But even with such uncertainty, people can be distinguished from different nationalities. When the characteristics of the information are judged, people can identify some objects through some local characteristics of the same object [15]. Local commonality allows us to use less resource to get the information that we need, so that the time and effort can be saved without tedious data analysis. Image local feature descriptor is one of the best matching and most widely studied algorithms. It not only has the characteristics of translation, scaling and rotation invariance, but also has good robustness to changes in illumination, affine and projection.

Local feature descriptor has the following characteristics: (1) local feature of image has stability and invariance to some extent; (2) The matching is fast and accurate with unique characteristics, and the information is rich; (3) it has a large quantity, and a few objects can produce a large number of eigenvectors; (4) it has the high speed of the optimized matching algorithm; (5) the scalability with other forms of feature vectors is very convenient. There are two steps of image registration by using SIFT algorithm: one is the extraction of SIFT feature points; another is the feature point matching.

Just like a light in a dark area and a black spot in a bright region, it still remains the same even when the light conditions change. In addition to the very stable feature points extracted by the SIFT registration algorithm and the corner and edge points in the image, there are some local extreme points. Usually, the transform relation of the image can be calculated by matching points. Typically, the matching points have the following characteristics. If the two images are registered in the same target area, the corresponding feature points and the corresponding relationship based on the SIFT feature points of the two images can be obtained. In the SIFT algorithm, the most important thing is to extract feature points. However, the premise of extraction is to build a multi-scale space with stable feature points to extract the invariant feature points of these scales. But in order to further accurately determine the location of the feature point, the interference of the unstable points should be eliminated, such as noise elimination and so on. These feature points are extracted for registration and the generation of SIFT feature descriptor at last. Through the main direction and auxiliary direction of feature points, the feature points can be rotationally invariant.

Usually it is arranged in the shape of Pyramid, and the resolution of each layer of the image in Pyramid is raised from top to bottom. In order to generate a space of Pyramid image, low resolution filtering and sampling are used for input brightness images, and multi-resolution processing mechanism is generally used. The bottom is high resolution image, and the top is low resolution image. The structure of the image gradually decreases from bottom to top and becomes smoother gradually. According to the spatial structure of Pyramid, the effect of noise on images can be reduced. The proposal of Pyramid provides strong evidence for our subsequent analysis. The scale of feature points is invariant because Gauss convolution kernel is used to establish scale space.

The two-dimensional Gauss function is

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

where (x, y) is the pixel coordinate position and σ is the scale factor. The larger the σ value, the more blurred the image. The image is represented with $I(x, y, \sigma)$. The image $I(x, y, \sigma)$ of different scales is obtained by calculating the product of image and two-dimensional Gauss function, and the formula is

$$L(x, y, \sigma) = G(x, y, \sigma) \times I(x, y, \sigma).$$

The Gauss difference scale space is $D(x, y, \sigma)$, which is obtained by different Gauss differential kernel product images:

$$D(x, y, \sigma) = G(x, y, k\sigma) - G(x, y, \sigma) \times I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma).$$

.

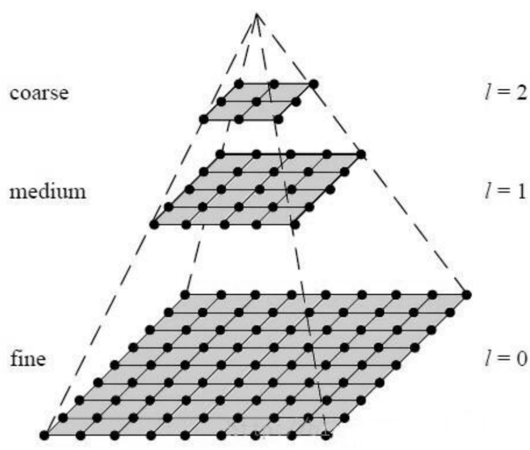

Fig. 1. Image pyramid

In order to detect the maximum and minimum points of $D(x, y, \sigma$, the key points are selected according to whether all of the sampling points can be greater than or less than the 26 adjacent points. So in order to judge, these 26 points are used to compare with each sampling point. Since DOG values are sensitive to noise and edge, the candidate feature points can be detected when the above method for detection is used: there are some low contrast points or some edge response points. These unstable candidate feature points need to be eliminated. The extremum point x_0 and the corresponding extreme value $D(x_0)$ are obtained by the derivation from the second-order Taylor expansion of the DOG function. If the value of $D(x_0)$ is not greater than the set threshold, the point is deleted.

In addition, it is necessary to take the edge response point into account and eliminate it. The method of trace and determinant ratio of Hessian matrix is used.

The Hessian matrix is defined as

$$H = \begin{bmatrix} D_{XX} & D_{XY} \\ D_{XY} & D_{YY} \end{bmatrix},$$

the trace of the matrix is

$$\mathrm{Tr}\,(H) = (D_{XY} + D_{YY})$$

and the determinant of the matrix is

$$\mathrm{Det}\,(H) = D_{XY} D_{YY} - (D_{XY})^2 .$$

If the value of $\mathrm{Tr}\,(H)^2/\mathrm{Det}\,(H)$ at the key point is not greater than the set threshold, the point is removed. The final threshold is set at 10, which is calculated by Lowe continuous experiments.

A feature descriptor with rotation invariant properties has been proposed by Schmid. But it is limited by its poor uniqueness and poor matching accuracy, and the descriptor does not take into account the direction. Therefore, the neighborhood pixel gradient direction is added into the SIFT algorithm. In order to improve the matching accuracy, it is guaranteed that the feature descriptor has rotation invariance. The following formulas represent the gradient direction and amplitude calculation of the feature points (x, y):

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

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

where, $L\,(x, y+1)\,, L\,(x, y-1)\,, L\,(x-1, y)\,, L\,(x+1, y)$ represent the gray values of the upper, lower, left, and right pixel points of the point M.

When the key point is the center, the histogram is used to calculate the gradient direction. According to the histogram of the gradient direction, the angle range is 0–360 degrees, and 36 columns with 10 degrees as a column can be obtained. The next thing is to find our peak representation, which represents the direction of our key points. But there are likely to be several main directions, at this time, it is necessary to judge. According to the robustness of the matching, our auxiliary direction needs to be judged whether it has 80 % energy of the main peak.

With the premise of not rotating, take the key point as the center and go to a 16×16 window. The main direction of the key is rotated around the coordinate axis. Then, the gradient direction and amplitude of each pixel are calculated with the 16×16 pixels as the center. After these pixels are computed, the Gauss weighting is also necessary. Through the pixels above, the window can divided into 4×4 windows. Histograms of gradient directions in 8 directions are computed on each small window. The cumulative values for each gradient direction are plotted. From the above division, the feature vector of 128 dimensions can be obtained, which consists of 4×4 16 seed points, each of which has vector information in 8 directions. Through such a joint domain method, we can effectively enhance our anti-noise performance and have a good fault tolerance for the positioning error.

The matching relation between images is set up by matching, so the image mosaic is finished. However, matching pairs is often found after extracting SIFT features, in which there are always incorrect matches. In order to reduce this situation, the matching must be focused on the extracted feature points. For the initial matching between two pairs of images, it is necessary to take the following two steps to match them.

1. Through the RANSACA algorithm, the error matching of initial matching points need to be eliminated to improve the accuracy. The initial matching points require that the nearest feature points are divided by the distance near them, and the ratio is the initial matching point that we need.

2. By referring to two original images, a feature point can be extracted from it. An efficient search of the nearest and sub adjacent points in a floating image is performed by using the BBF algorithm. But the initial matching must be judged by the Euclidean distance between the nearest feature points and the Euclidean distance between the sub adjacent feature points. The ratio is determined by the size of a specified proportional threshold.

The incorrect matches directly affect the parameters of the registration affine transformation, and then affect the image after stitching. The serious error matching results in the inaccurate transformation relation between the images. This situation occurs in the initial matching pair, which eventually leads to the poor accuracy of the stitching. The correct number of matches is obtained by using the RANSAC algorithm. After continuous screening, in order to be able to match more precise and let the final mosaic can have good accuracy, RANSAC algorithm is used to further screen in the correct matches. The original RANSAC algorithm is improved, and its improved algorithm steps are as follows.

1. Four pairs of matched pairs are extracted at random from the initial pair of points. The four pairs are set as initial interior points, and the transformation matrix H is calculated.

2. The distance between outside point of the set and the matching points after the transformation matrix is calculated. By setting a distance threshold T, the distance obtained can be judged. If it is greater than T, the remaining points are continued to be judged. If it is smaller than T, all previous points are added to the inner set.

3. The number of interior points under the transformation matrix H is counted.

4. After repeating the above three steps, one of the largest number of inner points is selected to compare with the threshold. If the quantity is greater than the threshold, the interior point is used as the initial value of the RANSAC. The change matrix H is re-calculated and the RANSAC is estimated. So set of interior points

under the new transform matrix H is the union of the original set of interior points and the new set of interior points.

5. The accurate matching point pairs are the all the feature points included in the best set with the largest number of interior points.

3. Result analysis and discussion

PC (Intel (R), Pentium (R), Dual, T2330@1.60GHZ, 8 G, memory, Windows, XP) were used as the adopted setting environment. The following points were needed to draw conclusions: the first step was to extract features from the SIFT and match the features. For the overlapping regions of the two images, feature extraction and matching were performed. The second step was screening, the initial screening was carried out. After screening, images were filtered by the improved RANSAC method for the second time to obtain accurate matching. The effects are compared in Figs. 2–7. Then images were fused by wavelet transform to obtain the final stitching image. The results show that in order to obtain more accurate stitching image, it is necessary to use the feature registration method described in this paper under the premise of improving the image registration parameters. In the process of matching the original image with the SIFT algorithm to register, it is necessary to adjust the value of the scale parameter. Table 1 shows the results for various values of threshold.

Table 1. Data results for various threshold values

Ratio threshold	Time (s)	Feature points extracted from reference images	Feature points extracted from floating images	Extraction duratio (s)	Matchpairs (unit)	Match time (s)	Total match time (s)
0.6	3.28	1783	1875	2.04	336	4.66	9.98
0.5	2.79	1783	1875	2.39	295	4.68	9.86
0.4	2.30	1783	1875	2.39	247	4.79	9.48

Fig. 4 shows the matching result of the two images with the proportional threshold of 0.6. Fig. 5 shows the matching result of the two images with the proportional threshold of 0.5. Fig. 6 shows the matching result of the two images with the proportional threshold of 0.4. The above experimental data shows that the total use time of the above three proportional thresholds was not much different. So the total match time did not have much influence on the proportional threshold. However, when the threshold was not used, the match pairs had obvious difference. Therefore, the appropriate proportion of the threshold was mainly reflected in the number of matching. When the threshold was set to 0.6, the number of incorrect matches was too high, but the total number of matches was higher. But when the threshold was proportional to 0.4, the number of matches was relatively small. Regardless of the

Fig. 2. Reference image

Fig. 3. Floating images

Fig. 4. Matching results with a threshold of 0.6

number of incorrect matches, it directly affected the final stitching effect. When the

Fig. 5. Matching results with a threshold of 0.5

Fig. 6. Matching results with a threshold of 0.4

Fig. 7. Stitched image using the method proposed in the paper

scale threshold was set to 0.5, the number of matches was big, and more accurate matches could be obtained. Therefore, the threshold in SIFT was selected as 0.5.

4. Conclusion

With the increasing demand for image acquisition in society, an accurate image registration method based on SIFT features was proposed in this paper by analyzing SIFT arithmetic extraction feature descriptor, so as to improve the registration accuracy, and the SIFT descriptor had a strong matching rate. However, in order to achieve more accurate of the matching number, the proportion of the threshold was adjusted, and the influence of matching threshold was analyzed according to the change of proportional threshold. In order to obtain seamless stitching images, the initial matching was obtained by using the Euclidean distance in the similarity criterion. The improved RANSAC method could further purify the matching pairs and obtain the desired mosaic images. The reason why the precisions of matching information obtained were different was that different thresholds were used with the RANSAC method to filter initial matches, although the different proportional thresholds had less impact on the total match time, they could directly affect the number of pairings, so as to affect the accuracy. So in the future, the SIFT feature points of a region can be calculated for corresponding registration, so as to improve our further registration effect through more direct and specific goals.

References

[1] W. Hu, L. Zhang, S. Liu, C. Shi: *An algorithm on registration of multi-view range images based on SIFT feature matching.* Journal of Computer-Aided Design & Computer Graphics *22* (2010), No. 04, 654–661.

[2] Y. Na, D. Wen: *An Effective Video Text Tracking Algorithm Based on SIFT Feature and Geometric Constraint.* Pacific-Rim Conference on Multimedia, 21–24 September 2010, Shanghai, China, Conference Paper LNCS, Springer, Berlin, Heidelberg *6297* (2010), 392–403.

[3] T. Y. Bai, X. B. Hou: *An improved image matching algorithm based on SIFT.* Transactions of Beijing Institute of Technology (2013), No. 06.

[4] J. Z. J. Zhao, L. J. Xue, G. Z. Men: *Optimization matching algorithm based on improved harris and SIFT.* International Conference on Machine Learning and Cybernetics, 11–14 July 2010, Qingdao, China, IEEE Conference Publications *1* (2010), 258–261.

[5] X. T. Wang, Y. Xu, F. Gao, J. Y. Bai: *An image matching algorithm based on SIFT and invariability of feature points set.* Applied Mechanics and Materials *121 to 126* (2011), Chapter No. 2, 701–704.

[6] D. Z. Cheng, L. I. Yan-Jun, Y. U. Rui-Xing: *Image matching method based on improved SIFT algorithm.* Computer Simulation (2011), No. 07.

[7] M. Y. Yin, F. Guan, P. Ding, Z. F. Liu: *Implementation of image matching algorithm based on SIFT features.* Applied Mechanics and Materials *602–605* (2014), Chapter No. 5, 3181–3184.

[8] H. Q. Zhang, L. G. Cao: *An image matching algorithm based on SUSAN-SIFT algorithm.* Applied Mechanics and Materials *325–326* (2013), Chapter No. 14, 1588–1592.

[9] B. Zitova, J. Flusser: *Image registration methods: A survey.* Image and Vision Computing *21* (2003), No. 11, 977–1000.

[10] F. Tian, Y. B. Yan: *A SIFT feature matching algorithm based on semi-variance function.* Advanced Materials Research *647*, (2013), Chapter No. 4, 896–900.

[11] A. Wang, D. Lu, Z. Wang, Z. Fang: *Research on non-rigid medical image registration algorithm based on SIFT feature extraction.* Journal of biomedical engineering *27* (2010), No. 4, 763–768,784.

[12] W. T. Wang: *Multi-sensor image registration algorithm based on SIFT points and canny edge features matching.* Computer Science *38* (2011), No. 07, 287–289.

[13] X. Dai, S. Khorram: *A feature-based image registration algorithm using improved chain-code representation combined with invariant moments.* IEEE Transactions on Geoscience and Remote Sensing *37* (1999), No. 5, 2351–2362.

[14] X. Zhang, Y. S. Zhang, H. Yao: *Image feature matching based on SIFT algorithm.* Applied Mechanics and Materials *644–650* (2014), Chapter No. 6, 4157–4161.

[15] Y. W. Wang, H. L. Yu: *Medical image feature matching based on wavelet transform and SIFT algorithm.* Applied Mechanics and Materials *65* (2011), 497–502.

Received June 29, 2017