An Improved Harris-SIFT Algorithm for Image Matching

Yu Cao(✉), Bo Pang, Xin Liu, and Yan-li Shi

School of Harbin University of Science and Technology, Harbin 150080, China
759734958@qq.com

Abstract. In view of the feature points extracted by the SIFT algorithm can not fully represent the structure of the object and the computational complexity is high, an improved Harris-SIFT image matching algorithm is proposed. Firstly, the feature points of the image are extracted by Harris corner detection operator. Then, the feature points are described by using the 28 dimension increasing homocentric square window. Euclidean distance is used as the similarity measure function in the matching process. Finally, simulation results show the validity of the improved algorithm, providing a new thought for the research into the image matching.

Keywords: Improved Harris-SIFT algorithm · Corner detection
Homocentric square window · Image matching

1 Introduction

Image matching is a kind of algorithm for finding similar image block, which is mainly based on the similarity and conformity analysis of image content, characteristics, grey degree, etc. [1]. Image matching technology has been applied to 3D Reconstruction [2], target tracking, remote sensing data analysis and many other fields [3, 4]. But it is still a problem to find one with high real-time and high precision from so many existing algorithms at present.

In recent years, there have been numerous image matching algorithms, including Moravec detection operator, Harris detection operator [4] and SUSAN detection operator [5]. SIFT (scale invariant feature transform) algorithm, PCA-SIFT algorithm [6], SVD matching method and Integration image method are new algorithms in modern society. In 2004, Lowe presented SIFT algorithm based on Local invariant descriptors. It has been broadly applied in many scenes with rotating zoom, partial occlusion, scale invariant, etc. [7]. But the feature points extracted by the SIFT operator do not fully represent the actual structure of the object. Beyond that, SIFT descriptors are quite complicated in calculation [8] and weak in real-time performance. On the basis of the above disadvantages, SIFT algorithm is not applicable to the higher requirements [8, 9]. The angle point features extracted by Harris operator is a good indication of the physical characteristics of the object, which is a stable angle point extraction algorithm [10].

This paper discusses the problem of the poor real-time performance of SIFT algorithm, and improves the descriptors of SIFT. In addition, this paper presents an

G. Sun and S. Liu (Eds.): ADHIP 2017, LNICST 219, pp. 56–64, 2018.
https://doi.org/10.1007/978-3-319-73317-3_8

improved Harris SIFT algorithm in combination with the Harris point detection operators. The simulation results demonstrate the effectiveness of the proposed algorithm.

1.1 Feature Point Detection Algorithm

The basic idea of SIFT algorithm: firstly, the scale space of image is constructed, and the key points are detected by extreme value detection in scale space, and the points with low contrast points and unstable edges are removed. Then the main direction of the key point is determined in scale space, and finally describe the key points, and make the descriptors unique.

1.2 Extreme Value Detection in Scale Space

Scale space is a theory which is simulating multi-scale characteristics to analog image data. SIFT algorithm finds the key points in different scale spaces. And the only linear kernel of scaling transformation is the Gauss convolution kernel. Therefore, the scale space of an image $L(x, y, \sigma)$ is defined as the convolution of a variable scale Gauss's function with the original image. e.g. (1):

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

where: σ is scale parameter.

In order to obtain more stable image features, difference of Gaussians (DoG) is proposed, e.g. (2):

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \tag{2}$$

where: k is related to the number of layers per dimension s in the scale space, $k = 2^{\frac{1}{s}}$.

In order to obtain the extreme points in the scale space. Each pixel needs to be compared with the 26 points, including the pixel scale, the same scale and the upper and lower scales. If it is the maximum or minimum in its adjacent points, the point is considered to be the extreme point in the scale space.

The key point which is obtained by the method of above-mentioned is the extreme points of discrete space, is not really extreme point. So the DOG function should be fitted of a curve, and then repeatedly interpolated to get continuous space extreme points. And the number of iterations or beyond the image boundary points are removed to obtain accurate positioning. At the same time, the low contrast points and edge unstable points are removed, so that the noise immunity can be enhanced.

1.3 Allocation of Key Points

In order to make the descriptor have rotation invariance, the directional parameters of each key point are obtained according to the gradient direction distribution feature of the neighborhood pixels of feature points. The formula of calculating the modulus $m(x, y)$ and direction $\theta(x, y)$ of the gradient which belongs to the feature points, (3) and

(4). Histograms are used to represent the gradient size and direction of pixels, and the peak direction of the histogram is the main direction of the key points. Using a column at 10 degrees, 0°–360° is divided into 36 columns. We give 8 columns, one of which is the main direction and the other is the auxiliary direction.

$$m(x, y) = ((L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2)^{\frac{1}{2}} \tag{3}$$

$$\theta(x, y) = \tan^{-1} \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \tag{4}$$

1.4 Key Point Descriptor Generation

First, the coordinate axis is rotated to be consistent with the main direction of the feature points, so as to ensure rotation invariance. Neighborhood statistics range of SIFT for each of the key points is 16 × 16 pixel window, and divide it into 4 × 4 sub regions, each sub region containing 4 × 4 pixels, and then calculate the gradient of 8 directions in each sub area of the histogram; last sort vector information on the 8 the direction of each sub region in the vector sorted form a 4 × 4 × 8 = 128 dimensional feature descriptor. Such a feature descriptor has the invariance of scale change, geometric deformation and illumination change.

The 128 dimensional descriptor of this algorithm makes the computing complexity and the real-time performance poor, and the detected feature points can not show the physical structure of objects.

2 Improved Harris-SIFT Algorithm

2.1 Detection of Feature Points

The principle of Harris corner detection is: taking a small window to move toward any direction in infinitesimal displacement by centering on target pixel; only when the center of this window is corner that the gray feature values of window change in all directions. According to the variation degrees of gray feature values in each direction, the feature and location information of corners in image can be determined. The gray degree variation is presented by the analytical formula as follows:

$$E(x, y) = \sum w_{x,y}(I_{x+u, y+v} - I_{x,y})^2 = [u \quad v] M \begin{bmatrix} u \\ - \\ v \end{bmatrix} \tag{5}$$

In this formula, u is the displacement of small window centering on (x, y) in the direction of X; v is the displacement in direction Y; $w_{x,y}$ is Gauss window function; I is function of image gray scale.

Calculating the value of matrix M:

$$M = \begin{bmatrix} I_x^2 & I_xI_y \\ I_xI_y & I_y^2 \end{bmatrix} \tag{6}$$

where I_x and I_y respectively are the gradient values of image pixels in horizontal as well as vertical directions.

Calculating the interest values of each pixel corresponding to original image, which is R value:

$$R = Det(M) - kTr^2(M) \tag{7}$$

where k is experience value taken 0.04–0.06. When the R value in certain point is greater than the given threshold value T, this point is deemed as corner. The threshold value T in this paper is 0.05.

2.2 Description of Feature Points

This paper will adopt neighborhood information to improve descriptor because of its strong anti-noise ability. The layer-by-layer increasing homocentric square window is applied in this paper to describe the neighborhood information of key points. This method divides the neighborhood around key point into 4 regions so as to establish 28 dimension feature point descriptor. The layer-by-layer increasing homocentric square window indicates that, the pixel closer to key point will generate stronger influence on ultimate feature description.

The specific construction method is: taking the window in key point closer to 20×20 pixel as the neighborhood scope of this key point for statistics, and regarding the 4 adjacent neighborhoods as first group; then expanding 2 pixels in each direction of first group and taking this neighborhood circle as second group; expanding 3 pixels in each direction of second group to obtain the corresponding neighborhood circle; expanding 3 pixels in each direction of third group to obtain the corresponding neighborhood circle as shown in Fig. 1. Finally, calculating the gradient accumulated values in 8 directions of pixel in each group. The feature vectors of 1–4 dimensions belong to first group, just as the style a shown in the diagram. The feature vectors in second group are adopted as 5–12 dimensions. I this ways, the feature vectors of 28 dimensions are acquired; such arrangement presents that, the pixel closer to key point will generate stronger influence on feature description. The grayscale accumulated values of pixels within each region is computed and normalized to obtain the feature descriptor with illumination invariants.

The grayscale accumulated values of pixels within each region is computed and normalized to obtain the feature descriptor with illumination invariants. The following formula is acquired by normalizing grayscale accumulated values:

$$\overline{f_i} = f_i \bigg/ \sqrt{\sum_{i=1}^{4} f_i}, \ i = (1, 2, 3, 4) \tag{8}$$

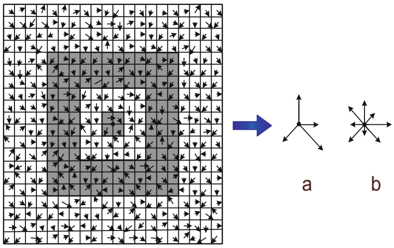


Fig. 1. Schematic diagram of improved feature descriptor

$$\bar{f_i} = f_i \Big/ \sqrt{\sum_{i=1}^{8} f_i}, \; i = (1, 2, 3, \cdots 8) \tag{9}$$

where formula (8) is to normalize the grayscale accumulated value of vectors in first group; formula (9) is acquired by normalizing the grayscale accumulated value of vectors in last three groups. Thus, the feature vector of 28 dimension is presented as follows:

$$F_i = (\overline{f_{i1}}, \overline{f_{i2}}, \cdots \overline{f_{i28}}) \tag{10}$$

The improved algorithm descriptor is reduced from 128 dimension to 28 dimension; meanwhile, the contained neighborhood is changed from 16 × 16 into 20 × 20 with more neighborhood information, which not only reduced calculated amount, but also avoided the loss caused by decrease of seed points.

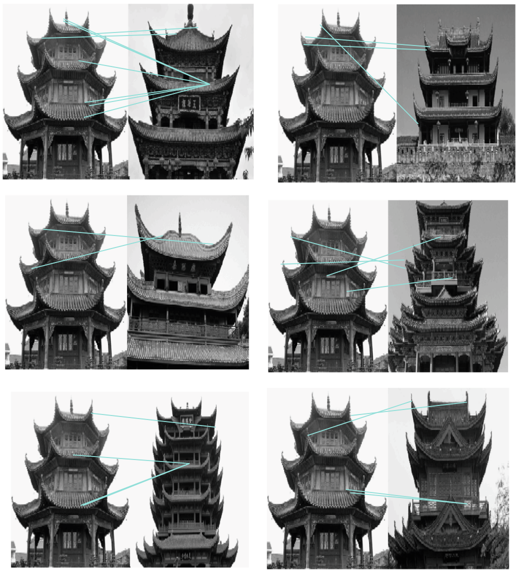
2.3 SIFT Feature Vector Matching

The Euclidean distance is adopted as similarity metric function to carry our feature vector matching; k-d tree is utilized to search so as to look for the nearest-neighbor and next nearest neighbor feature points corresponding to each feature points. In these two feature points, if the distance ratio through dividing nearest neighbor by next nearest neighbor is smaller than certain given proportion threshold value, then such pair of matching point is accepted.

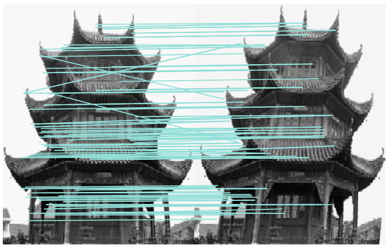
3 Experimental Result and Analysis

This paper firstly uses MATLAB programming to realize image matching and the shot pictures for test.

Firstly, the matching of similar image pairs in classical SIFT is research so as to look for the relationship between matching number of image pairs with scale variation of the same article and matching number of similar image pairs; so a threshold value of whether being the image matching number of the same article. The setting of such threshold helps prevent the wrong matching of similar image. The following is the result diagram of matching similar images and the same article in rotation transformation.



(a) Matching of similar image pairs



(b) Image matching of the same article in rotation transformation

Fig. 2. Matching test of similar image pair

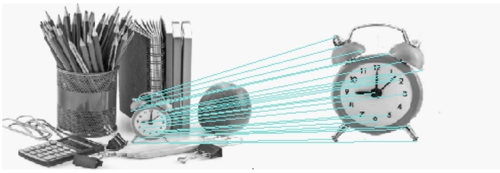
According to a group of test image above, it can be concluded that, the average matching number in the matching of similar image pair in Fig. 2(a) is 5, among which the maximum value 9 and the matching number in the same article is 125. Based on the test of abovementioned ten groups of similar image pairs, the following table is obtained:

According to the results in Table 1, it is concluded that, the more the feature points of articles detected, the more the wrong matching points of image will be. Based on the statistical data of Table 1, this paper sets wrong matching threshold value M = 15. The image with matching point pairs less than 15 will be output as wrong matching.

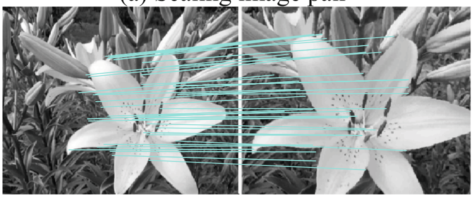
Table 1. Statistics of similar image pair matching test

Test group	1	2	3	4	5	6	7	8	9	10
Mean value	5	2	8	4	6	5	7	4	6	3
Maximum value	9	6	14	8	11	12	14	9	12	7
Matching number of same object	125	78	143	112	119	98	127	103	136	63

Secondly, the matching verification of improved Harris-SIFT algorithm proposed in this paper is implemented to analyze and compare it with original SIFT algorithm. Figures 3 and 4 respectively are the experimental results of two image pairs; Table 2 is experimental result statistics. It is obvious that, the feature points extracted by improved Harris-SIFT algorithm are corners, which reduced the extraction time of surplus feature points. In addition, the introduction of layer-by-layer increasing homocentric square window greatly reduced computational complexity, enhanced timeliness of original SIFT algorithm, and guaranteed matching correctness.



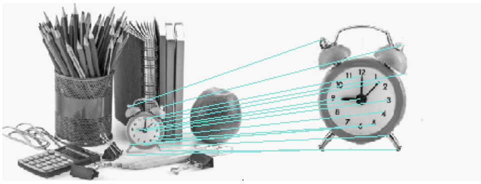
(a) Scaling image pair



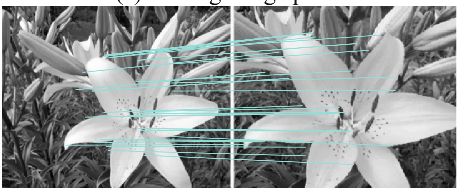
(b) Image pair in rotation compression

Fig. 3. Matching of original SIFT algorithm

(1) According to above experimental results, this paper solved three problems of original SIFT algorithm.
(2) By means of multiple matching experiments of similar images, this paper obtained the wrong matching value M = 15; the setting of threshold value can effectively prevent wrong matching of different articles.
(3) This paper adopted Harris corner detection algorithm in the detection stage of feature point, which eliminated many not obvious feature points, reduced computational complexity and enhanced correct matching rate; besides, the corners obtained can better embody article characteristics.
(4) In the stage of feature description, this paper adopted 28 dimension layer-by-layer increasing homocentric square window, which greatly shortened computation time and enhanced matching timeliness.



(a) Scaling image pair



(b) Image pair in rotation compression

Fig. 4. Improved Harris-SIFT algorithm matching

Table 2. Statistics of matching results

Image pair	a			b		
Reduced parameter	Matching number	Matching time/s	Correct matching rate/%	Matching number	Matching time/s	Correct matching rate/%
Original SIFT algorithm	30	0.098	90.0	63	0.134	92.1
Improved Harris-SIFT algorithm	22	0.054	95.5	46	0.085	95.7

4 Conclusion

The SIFT algorithm has the advantages of good scale, rotation, angle and light invariance, which is widely used in image matching. This paper presents an improved Harris SIFT algorithm based on the Harris angle point detection algorithm. The algorithm uses the Harris operator to detect angle points, then improves the descriptor for the SIFT operator. This algorithm describes the character points in a 28-dimensional incremented rectangular-ambulatory-plane descriptor, and finally uses European distance as the measure function to match. Experimental results show that the feature points extracted by the improved algorithm can be a very good reflect the structure of the object, and greatly reduce the matching time, improves the accuracy of matching as well.

Funding Project. This paper is supported by the project of young creative talents training program of Heilongjiang undergraduate higher education institution (UNPYSCT-2015039).

References

1. Zhou, R., Dexing, D., Han, J.: Fingerprint identification using SIFT-based minutia descriptors and improved all descriptor-pair matching. Sensors **13**(3), 3142–3156 (2013)
2. Guo, Y., Sohel, F., Bennamoun, M., et al.: An accurate and robust range image registration algorithm for 3D object modeling. IEEE Trans. Multimedia **16**(5), 1377–1390 (2014)
3. Chen, Y., Shang, L.: Improved SIFT image registration algorithm on characteristic statistical distributions and consistency constraint. Optik-Int. J. Light Electr. Opt. **127**(2), 900–911 (2016)
4. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: speeded up robust features. Comput. Vis. Image Underst. **110**(3), 346–359 (2008)
5. Ke, Y., Sukthankar, R.: PCA-SIFT: a more distinctive representation for local image descriptors. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, no. 2, pp. 506–513 (2004)
6. Lowe, D.G.: Object recognition from local scale-invariant features. In: Proceedings of International Conference on Computer Vision, pp. 1150–1157 (1999)
7. Gal, R., Cohen-Or, D.: Salient geometric features for partial shape matching and similarity. ACM (2006)
8. Wang, Y., Hu, J., Han, F.: Enhanced gradient-based algorithm for the estimation of fingerprint orientation fields. Elsevier Science Inc. (2007)
9. Maintz, J.B.A., van den Elsen, P.A., Viergever, M.A.: Evaluation of ridge seeking operators for multimodality medical image matching. IEEE Trans. Pattern Anal. Mach. Intell. **18**(4), 353–365 (2008)
10. Er-Sen, L.I., Zhang, B.M., Liu, J.Z., et al.: The application of SIFT feature matching method in the automatic relative orientation. Sci. Surv. Mapp. **33**(5), 15–16 (2008)
11. Tian, F., Yan, Y.B.: A SIFT feature matching algorithm based on semi-variance function. Adv. Mater. Res. **647**, 896–900 (2013)
12. Zhao, J., Xue, L.J., Men, G.Z.: Optimization matching algorithm based on improved Harris and SIFT. In: International Conference on Machine Learning and Cybernetics, pp. 258–261. IEEE (2010)