An Improved SIFT Algorithm for Image Feature-Matching

Wenyu Chen, Yanli Zhao
School of Computer Science & Engineering
University of Electronic Science & Technology of China, UESTC
Chengdu, China
cwy@uestc.edu.cn, zhaoyanli2011@yahoo.cn

Wenzhi Xie, Nan Sang
School of Software
University of Electronic Science & Technology of China, UESTC
Chengdu, China

***Abstract*—SIFT algorithm has strong robustness and stability, which can be applied in many bad conditions to achieve high recognition rate. The 128-dimensional feature descriptor has good independence. However, it also contains redundant information, which makes the calculation amount of following matching increase. An improved SIFT algorithm is proposed to solve the disadvantages of large computation and not meet "real-time" of the original one, which reduces the dimension of the feature vector and reduces the mismatching of SIFT algorithm. It has been proved from the experiment that the improved algorithm can significantly increase running speed and guarantee the robustness. Besides, it can meet "real-time" requirements.**

Keywords-SIFT algorithm, robustness, real-time, dimension

I. INTRODUCTION

Researches on feature-based matching have been reported frequently. The matching process has been split into the following steps. Firstly, the features of each image are extracted. Secondly, the feature is matched to create a mapping transformation among each image. Finally, the matching image is attained. Feature profile, region, point and edge are common types of image features. Feature extraction is greatly related to image contents. Usually, the extraction of feature points is more feasible, since they remain unchanged regardless of illumination, translation, rotation, resolution and etc., which has been proved by several theses such as Hessian-Laplace, Harris-Laplace, SUSAN, DOG and Harris.

Lowe proposed SIFT algorithm [1] (Scale Invariant Feature Transform), which is a feature-describing method which has good robustness and scale invariance and has been widely used in image-matching, image stitching [2], classification of household goods, iris recognition [3] and other fields such as combines with other algorithm [4].

The 128-dimensional feature descriptor of SIFT is very descriptive with large amounts of feature points, but many feature points are not representative and very distinguishable for image detection. These are redundant information which not only reduces the speed of matching, but also affects the matching accuracy.

In this paper, an improved SIFT algorithm is developed to meet "real-time" requirements. In the feature extraction part, the number of dimensions of a feature descriptor is reduced from 128 to 48, omitting the determination of primary direction. The dimension performance also has been analyzed and the reasonableness of selected dimension has been verified. It is proved from the tests that the improved SIFT algorithm can significantly increase "real-time" and matching accuracy.

II. SIFT ALGORITHM

A. Establish a scale space

According to the theory of scale space, SIFT algorithm simulates multi-scale feature of image data. It is based on the theory that Gaussian kernel is the only linear transformation kernel. This had been proven by Lindeberg and Koenderink [5]. So, for a given image I(x, y), it's (Gaussian) scale-space L(x, y, D) can be defined as the convolution of scale variant Gaussian function G(x, y, D) and I(x, y)

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \qquad (1)$$

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

B. Detection extreme points in scale space

For each pixel in scale space, it is compared to 26 pixels which include surrounding 8 adjacent points and 18 points which are corresponding position of two images adjacent scale of up and down in pyramid. If the pixel value is greater or smaller than any of the 26 points, then it is the candidate feature point.

C. Determining direction of feature point

In the calculation, sampling in the neighborhood window centered at the feature points, using histogram to count neighborhood pixel gradient direction. Range of the gradient histogram is from 0 to 360^0, it is been divided into 36 directions averagely, and each column represents a direction in histogram. The peak of the histogram represents the primary direction of the neighborhood gradient which is set as the direction of the feature point.

Foundation items: Quantitative Analysis of the Motility of the Sub cellular Structures Based on Fluorescence Microscopy Image Sequences (61073120)

978-1-61284-774-0/11/$26.00 ©2011 IEEE

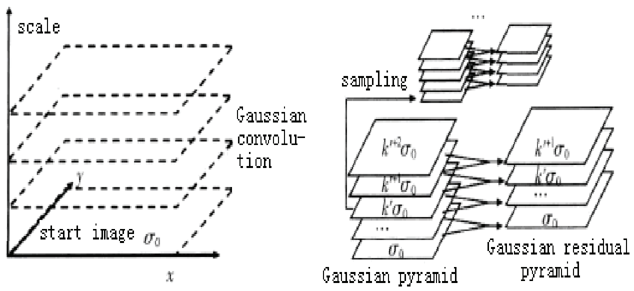
D. Generation of feature descriptor

Firstly, rotate the axis to direction of feature point and ensure rotation invariance. Secondly, select an 8*8 window centered rounds feature point. Then calculate the gradient histogram of 8 directions in each small 4*4 square and draw the accumulated value of each gradient direction, which can from a seed. Each seed has 8 directions vector information. Lowe suggested choosing 4 × 4 numbers of sub-regions which take key point as the center in the implementation. Thus a feature vector of 4*4*8 equal to128-dimonsion is constituted. United neighborhood orientation information enhances anti-noise ability, and provides good fault tolerance for those features-matching which contain error positioning.

III. IMPROVED SIFT ALGORITHM

A. Simplification the detection of extreme point in scale space

Scale space consists of a series of Gaussian convolutions. The bigger the scale factor is, the bigger the Gaussian convolution template is, and the longer the run-time is. In order to reduce calculated amount, the SIFT divides the design structure into Gaussian and Gaussian residual pyramid. The algorithm chooses the extreme point of the Gaussian residual pyramid in scale space, which is shown as Fig. 1.



(a)Scale space of image (b) Structure of pyramid scale space

Figure1. Scale space of image and SIFT pyramid based-on it

We can simplify the detection of SIFT feature points by changing the structure of pyramid scale space. We remove the top layer of each level, shown as r+2 layer in Fig.1. We do not extract feature points in the bottom layer of each level and just compare it to the regulated derivative $k\sigma_0$. Due to lack of top layer, the top two layers r+2 and r+1 of each level cannot compare extreme value. The simplified structure can compare the extreme values of last level's bottom layer and the second layer, which can substitute the pix comparison between the top two layers for the bottom layer of the last level, is gotten by sampling the top layer of this level. The simplified structure leads to eliminating one Gaussian convolution of each level.

It analyze the impact of Gaussian pyramid structure's change to the robustness of the SIFT algorithm. Sampling of the Gaussian pyramid structure do has an effect on invariance of original scale and displacement. We can compromise the gap of robustness by adding number of Gaussian residual pyramid's layers and levels.

B. Simplification of the SIFT feature descriptor

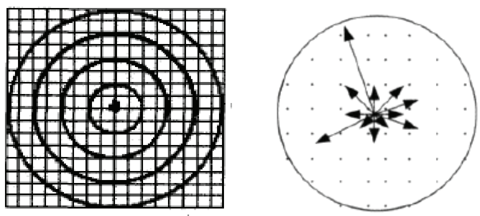
For the third step of SIFT algorithm, after the image rotation, the feature points around the sub-region will change, so the primary direction of each feature point need to be determined. Comparing with rectangular, circular has rotation invariance. If the feature descriptor has good anti-rotary capability, the determining of the direction of feature points can be omitted. Thus the improved algorithm use circle for feature descriptor, and replace rectangle region by diffused outwards concentric circle which take feature point as center.

Method of concentric feature descriptor is as follow. The circle region takes feature point as center with radius of 16. Thus it ensures that feature point neighborhood region and united Neighborhood orientation information are not changed. There are 5 steps.

a. Calculate the gradient of each sampling point in 16*16 circle neighborhood region.

b. For the selected feature point, extract four concentric circle sub-domains with radius of 2,4,6,8, as Fig.2 (a) shown.

c. Divide the circle into 12 equal pie-shaped sections and calculate histogram of 12 gradients in four concentric circles $D = (d_1, d_2, ..., d_{12})$.



(a) Concentric circle (b) 12-dimension vector

Figure2. Creating 12-dimension vector from the 4 concentric circle

d. Calculate accumulated value of 12 gradient in each concentric circle separately which is defined as d. Finally, a seed with 4*12 equal to 48-dimension is formed, as Fig.2 (b) shown.

e. Descending sorting of value d, then the ordered values of image do not change at any angle of rotation.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

Experiment environment: Windows XP2, Visual C++ 6.0, image processing software OpenCV2.0, simulation platform Matlab 7.0, 100 experimental images with the size of 320*240.

a. The first set of experiment

Compare of point detecting, matching and running speed, between the improved SIFT algorithm and the original one. Select the situations including zoom, rotation, light change, and partial occlusion with complex background. Some of the test results are showed as Fig.3. The left pictures show the results of using the original SIFT algorithm and the right one show the results of using the improved SIFT algorithm.

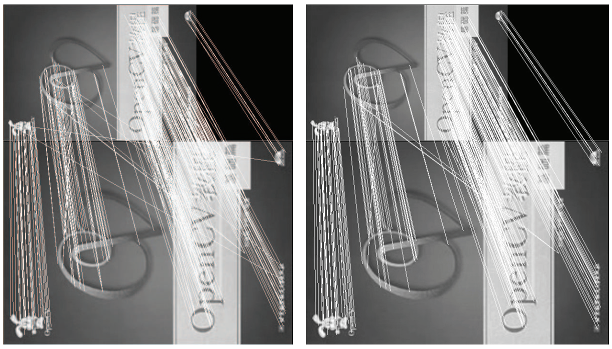


Figure3. (a) Zoom (*2)

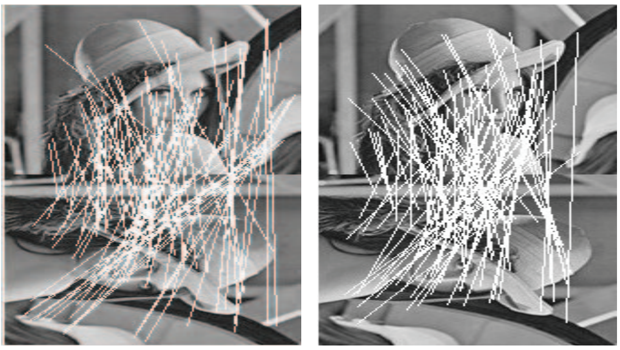


Figure3. (b) Rotation (90^{o})

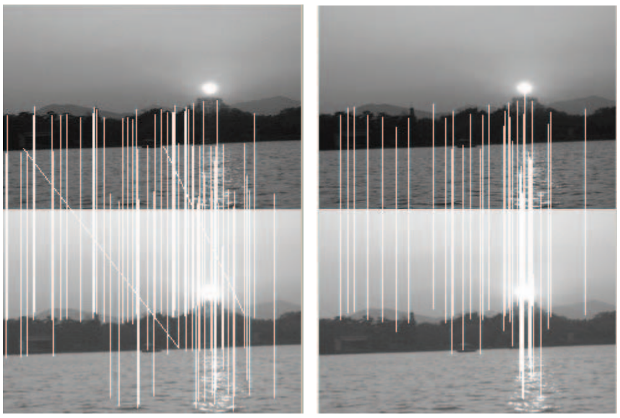


Figure3. (c) Illumination (*1.5)

Figure3. (d) Partial occlusion with complex background

Data of the experiments is shown as Tab. 1 and Tab. 2.

TABLE 1. Results of original SIFT algorithm

Case	Matching pairs	Original SIFT		
		Feature points	*Matching rate*	*Run time(ms)*
Zoom	121	135	91.29%	20.45
Rotation	384	848	92.97%	76.71
Light	96	260	63.87%	18.96
Partial occlusion	261	621	12.08%	76.61

TABLE 2. Results of improved SIFT algorithm

Case	Matching pairs	Improved SIFT		
		Feature points	*Matching rate*	*Run time(ms)*
Zoom	121	133	91.29%	20.45
Rotation	384	413	92.97%	76.71
Light	96	150	63.87%	18.96
Partial occlusion	261	2161	12.08%	76.61

We can get from Tab. 1 and Tab.2 that the improved algorithm brings more invalid values and lower matching rate, when processing images with complex background. It cannot match the original one, but in terms of zoom, it is obviously superior to the original one. In terms of rotation and light change, the running speed obviously increased without compromising matching rate. Totally, the improved algorithm largely shortcut the run time, is suitable. To the situation where little change of perspective and meet the "real-time" requirement.

b. The second set of experiment

Differences of the executing time between the improved SIFT Algorithm and the original one. SIFT image-matching has 5 steps, including generating scale space, detecting extreme points, determining direction of feature points, extracting feature descriptors and matching feature points. The improvements to the SIFT algorithm include canceling the determination of the feature descriptors and adding the 2-way matching. The run time differences of different step between the two algorithms are shown as Tab. 3.

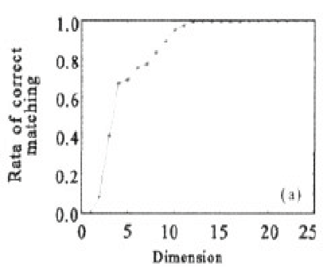
TABLE3. Two algorithms each step run time

algorithm / **step(ms)**	**SIFT**	**Improved SIFT(2-way matching)**	**Improved SIFT**
Scale space	10.11	10.11	10.11
Detection extreme points	21.02	21.02	21.02
Determination direction	9.11	0	0
Feature descriptor	80.5	39.83	39.83
Feature-matching	7.58	5.76	4.11
Total	128.32	76.71	75.06

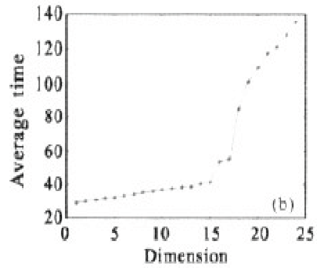
The data of Tab.3 also verifies that generation of feature descriptor is credited for 70% of the time the whole SIFT algorithm takes. Canceling the determination of the feature descriptors shortcuts the time that spent for the following procedures including generation of feature descriptor and image matching which lowers the run-time of the whole algorithm. Adding the 2-way matching makes it spend more 1.65ms, but also improves the matching accuracy rate.

c. The third set of experiment

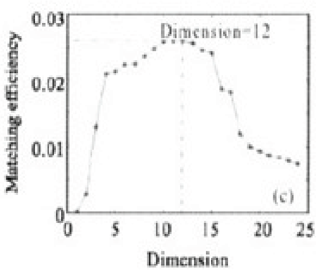
Determination of dimension in the improved SIFT Algorithm. In the construction of concentric circles feature descriptors, the number of concentric circles (d) and the dimension (n) that the concentric circles are split is the key to determine the dimension of the feature descriptor. The dimension of feature descriptor equals to d*n. The dimension of vector is not the bigger the better for calculation amount improves as the dimension rises. We get that 4 or 5 is the best through statistics of different kinds of images.



(a)Rate of correct matching



(b) Average time



(c)Matching efficiency

Figure4. Dimension assignment on matching task

Fig. 4(a) shows the relationship between dimension and matching rate. We can find that accuracy matching rate improves as the dimension (n) increases. But when the dimension (n) surpasses 12, there is no obvious improve of accuracy rate. Fig. 4(b) shows the relationship between dimension and counting time. Computing time improves as the dimension (n) increases. Especially when the dimension (n) surpasses 15, the time exponentially increases. The thesis determines reasonable dimension by matching rate and defines matching rate as the ratio of accuracy matching rate to computing time. And the maximum value is the best value of dimension. We can see from the Fig. 4(c) that we can get the best matching rate when the dimension (n) equals 12. So the dimension of the improved algorithm's feature descriptor is 4*12=48.

d. The fourth set of experiment

In order to remove those mismatching pairs, we add threshold to the Euclidean distance. Only the nearest points that are smaller than threshold are the matching pairs. Therefore, selecting reasonable threshold is especially important. In [6], Lowe recommends that setting threshold as 0.8. We test the threshold further in our algorithm and get the matching result in the Tab. 4.

TABLE4. Matching results of different threshold

Threshold	**0.2**	**0.3**	**04**	**0.45**	**0.5**	**0.55**	**0.6**	**0.65**	**07**	**08**
Pairs	112	108	90	88	80	72	70	61	54	45
Matching Pairs	14	21	34	49	56	59	51	40	27	7

This is maximum number of mismatching pairs when the threshold is about 0.55, which is gotten from the analysis of the Tab. 4. The farther the range of threshold is away from 0.55, the bigger the mismatching is.

V. CONCLUSIONS

The thesis mainly improves the image-matching based on SIFT algorithm. First of all, it points out the original SIFT algorithm's disadvantages, improves the algorithm's executing efficiency without compromising the original advantages. It improves the detecting of extreme point in scale space, generating of feature descriptor and successfully reduces the dimension of descriptor from 128 to 48. The executing efficiency improve 2/3 coming up to the speed of SURF Feature Descriptor algorithm based on integral[7], and overcomes the disadvantage that the SURF cannot fit for situation that gray scale and image perspective change.

REFERENCES

[1] LOWE D G. Distinctive image features from scale-invariant key points [J].International Journal of Computer Vision, 2004，60(2): 91-110.

[2] Yang Zhanlong, Guo Baolong. Image mosaic based on SIFT: International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2008[C]. Nanjing, China, 2008:1422-1425

[3] Belcher C., Du Yingzi. Region-based SIFT approach to iris recognition [J].Optics and Lasers in Engineering ,2009,47:139-147

[4] Zhou Huiyu, Yuan Yuan, Shi Chunmei. Object tracking using SIFT features and mean shift[J].Computer Vision and Image Understanding，2009, I 13:345-352

[5] J. Wood. Invariant pattern recognition: Areview. Pattern Recognition, 1996, Vo1.29(1):1-17.

[6] David G. Lowe. Distinctive image features from seale-invariant keypoints, International Journal of Computer Vision, 2004, 60(2):91-110P

[7]Herbert Features Bay,Andreas Ess,Tinne Tuytelaars,Luc Van Gool. Speeded-Up Robust (SURF) [J].Computer Vision and Image Understanding,2008,110(3):346-359