doi:10.21311/001.39.11.09

Local Feature Point Extraction Method Based on SIFT

Ye Chen, Xiaoqun Qin*, Haibo Gao, Kuangfeng Ning
School of Information Science and Engineering, Hunan International Economics University, Changsha 410205, Hunan, China
**Corerresponding author (Email: 104956905@qq.com)*

Abstract
Feature extraction uses the calculation of the image features as the intermediate result to make decisions on the local information contents of the image. To extract the image local feature points is to extract the feature points which include effective information from the image and then describe these feature points. Moreover, this process is required to have certain robustness in the viewpoint change, grey-level change, noises and other factors. This paper optimizes the structure process and feature descriptor of scale space to overcome the defects of the optimization space in the structure and generation feature descriptor of the scale space of the conventional SIFT (Scale-Invariant Feature Transform) algorithm and proposes an image feature point extraction method based on improved SIFT algorithm. It abstracts the edge information of the shape of the image objects into the feature graph constituted by the relative orientations of the feature points and the edges. Based on the feature graph, the object shape recognition algorithm with invariance in translation, scale and rotation can be designed so as to improve the precision and accuracy of the traditional SIFT algorithm.

Key words: Scale-Invariant Feature Transform, Feature Point Extraction, Grey-Level Change.

1. INTRODUCTION

Image feature refers to the significant basic feature or characteristic in an image. Only these significant features can reflect the image contents and they are also the key elements of the image. Generally, the image features include three types: point, line and plane. To analyze the image is to recognize its objects. Segment the objects, extract the features of the objects to be described and recognize the objects according to these features. From this perspective, the key issue of image feature extraction is the image object recognition. Scale invariant feature transform (SIFT) is one of the most brisk algorithms in the existing image feature extraction field. SIFT algorithm is an algorithm to extract local features. It searches extreme point in the scale space and extracts position, scale and rotation invariants. SIFT feature has excellent invariance in image scaling, translation and rotation and it has certain robustness on the lighting changes and affine transformation or three-dimensional projection(Xiaorong and Yingyong et al., 2015; Peizhi, 2015).

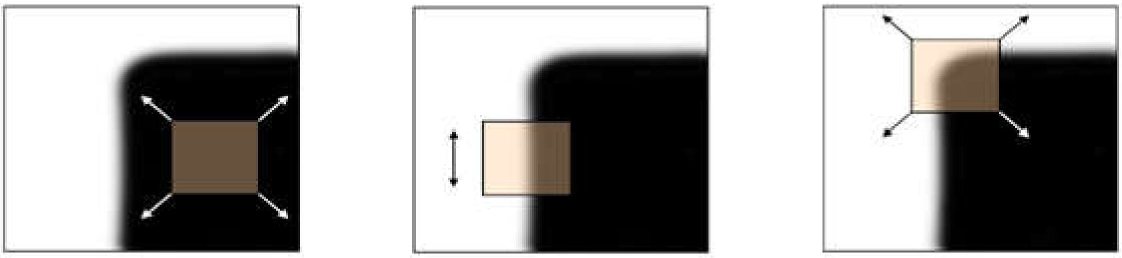
During the image feature extraction, although the image has been preprocessed and the image quality has been enhanced, the preprocessing algorithm is different in adaptability and effectability for each feature point and new noises are introduced, making the refined image has numerous false feature points, therefore, after the features are extracted, it recognizes and filters the feature points as accurately as possible and preserves the true feature points according to the correlation between the features. In as early as 1983, Witkin has proposed the concept of scale space and pointed out that to detect the positions of the feature points with scale invariance can be obtained by searching the stable points of every scale in the scale space. Afterwards, Lindeberger has found out that on the premise of a reasonable assumption, the only possible scale-space kernel is Gaussian kernel, namely Gaussian function(Yu et al., 2015; Stefan and Stanciu et al., 2015). SIFT algorithm is proposed by David Lowe based on the image local feature descriptor with the image translation, rotation and scaling invariance in the scale space by summarizing the feature detection methods based on the existing invariant technology in 1999 and it has been developed and improved in an in-depth manner in 2004. Due to the invariance of SIFT feature, SIFT algorithm has been widely applied in the field of image matching. However, the conventional SIFT algorithm has certain limitations in coping with the practical problems, therefore, the research progress of image feature point extraction has always attracted the attention of the researchers(Robert E. and Jacques-Donald et al., 2015).

This paper firstly introduces the image feature extraction, local feature point extraction and description process. Then, it analyzes the image feature extraction procedures and steps in SIFT and on this basis, it proposes the method of this paper to make the local image features extracted maintain invariant in brightness change, scaling and rotation and maintain stable in affine transformation and viewpoint change. Finally, it is the simulation experiment and analysis.

2. IMAGE FEATURE EXTRACTION

Feature extraction is a primary operation of image processing. In other words, it is the first processing on an image. Feature extraction uses computer to extract image information and determines whether the points of each image belong to an image feature. The result of feature extraction is to divide the points of the image into different sub-sets, which usually belong to isolated points, continuous curves or continuous regions. According to the image to be recognized, generate a group of original features through calculation. There are many original features in the image to be processed. In another word, the original sample is in a high-dimensional space and the feature description in the high-dimensional space can be described with the features in low-dimensional space through mapping or transformation methods. This process is called feature extraction. It checks every pixel to determine whether this pixel represents a feature. As a premise operation of feature extraction, the input image will be smoothened in the scale space. After that, one or several features of the image will be calculated through local derivative operation(Amin and Hamid, 2015; Gholam and Davar, 2015).

Image feature extraction can be divided into 2 kinds: global features, including image texture feature, color feature, shape feature and so on, and local features such as the extraction of local information of the image, including the extraction of point feature. The extraction of global feature is based on the global information of the image. The global feature has a rapid calculation speed, but it is not distinctive enough, therefore, the extraction methods of local features come into being(Jonathan and Golden G, 2015). There are 3 kinds of local image features, flat region, edge and corner, which form the feature template, as shown in Fig.1.



(a) Flat region: no change in all directions (b) Edge: no change alone the edge directions (c) Corner: significant change in all directions

Figure 1. Classification of image features

In the broad sense, image feature point is the interesting point to be matched by certain way in two images. And the extraction of feature point is the most important and the most difficult link in the image matching based on local feature point. Generally, the descriptor of extraction algorithm shall be unique, profound in information, applicable for the matching of plenty of feature data, highly accurate in matching manner and difficult to generate repetitive matching points. Local feature point extraction and description procedures are indicated in Fig.2.

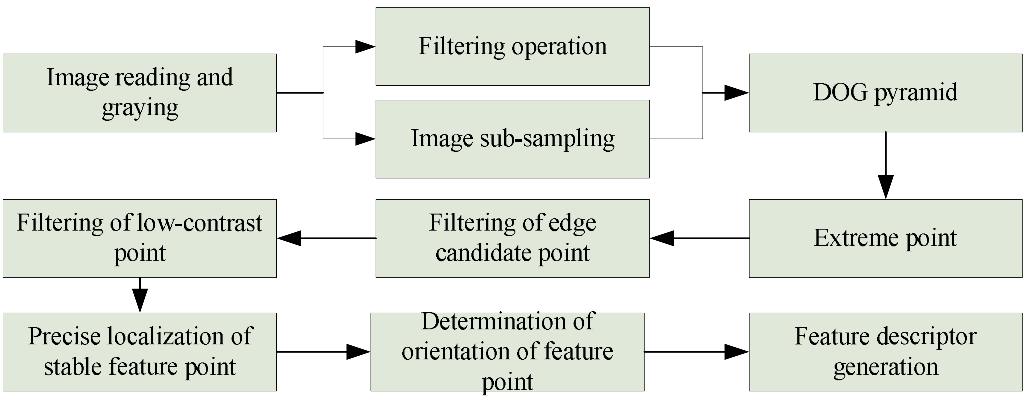


Figure 2. Feature point extraction and description procedures

3. IMAGE FEATURE EXTRACTION BASED ON SIFT ALGORITHM

The image feature extraction based on SIFT can be further divided into five parts: scale-space structure and its description, key-point localization, orientation assignment, key-point descriptor generation, comparing

Euclidean distance of the descriptors for matching.

3.1. Scale-space Structure and Its Description

Scale-space theory performs scale transform on the original image, obtains the scale-space representation sequence of the image in multi-scale, conducts principal contour extraction based on scale space on these sequences and takes the principal contour extracted as a kind of feature vector to achieve the feature extraction in different resolutions. The production of scale space is aimed to simulate the multi-scale features of image data. Because Gaussian convolution kernel is the only linear convolution kernel to realize scale transform, Gaussian function is used here to perform the image convolution(Wassim and Guillaume, 2015). For example, the scale space $L(x, y, \sigma)$ of a two-dimensional image $I(x, y)$ is shown by Formula (1).

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

In Formula (1), $\otimes$ represents the convolution operation of the functions; $I(x, y)$ is the input image and $G(x, y, \sigma)$ is the two-dimensional Gaussian function, as indicated by the following formula.

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

In Formula (2), σ is the scale factor. The scale space $L(x, y, \sigma)$ of the image is produced by the convolution between the Gaussian kernel $G(x, y, \sigma)$ of the factors σ with different sizes and the image $I(x, y)$ and it is represented by the Gaussian pyramid of the image, as shown by Fig.3, which are the first order and second order of the Gaussian pyramid formed by an image.

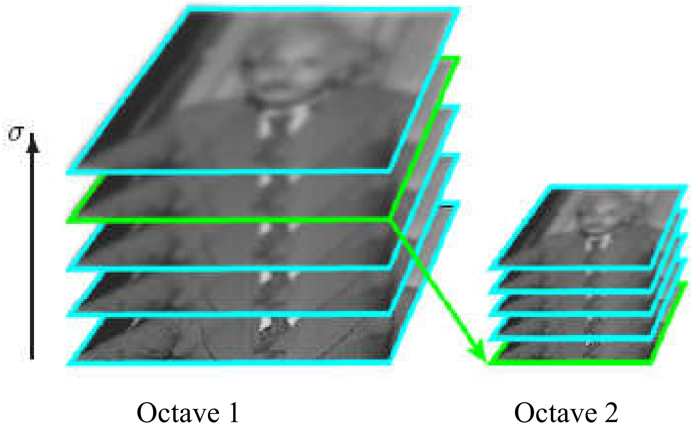


Figure 3. Gaussian pyramid

In practical applications, in order to calculate the position of the key point highly efficiently, it is suggested to use difference of Gaussian $D(x, y, \sigma)$, which is defined as follows.

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \tag{3}$$

As indicated in the above formula, D is the difference of two neighborhood scales (one of the two scales differs by a multiplication coefficient k from the other). Besides, in order to detect the stable key point in the scale space, difference of Gaussian (DOG) is required to construct. For instance, Fig.4 is the scale-space image cluster of DOG.

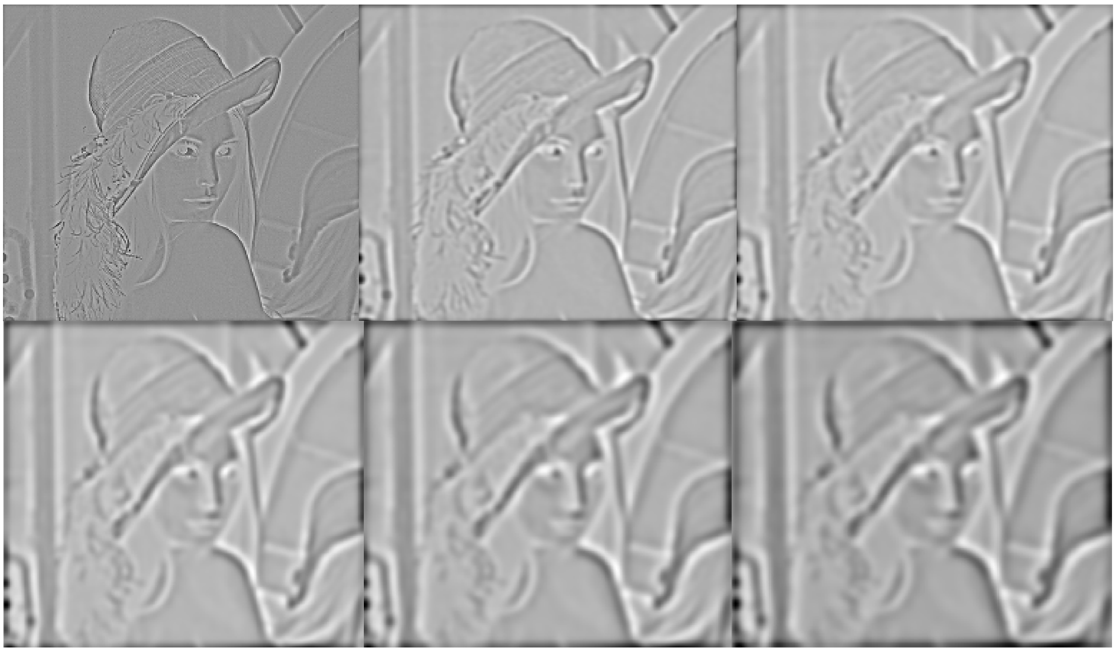


Figure 4. Scale space of difference of gaussian

Gaussian pyramid is the representation model of the image sequence structure. It includes n levels and there are m layers in every level. The first layer of each level is obtained by resampling the uppermost layer of last level. The scale factor between the images at each layer differs by k times from each other in the pyramid of the same level. If the scale factor of the first-layer image is 2, the scale factors at the second- and third-layer are $k\sigma$ and $k^2\sigma$ respectively and the rest can be deduced in the same manner(Yu and Wang, 2015).

Build Gaussian pyramid and DOG pyramid in order to make sure that the image can detect the extreme point in both the scale space and the two-dimensional image space. Then, conduct extremum detection within DOG pyramid so as to preliminarily confirm the position and scale of the feature point. If this pixel point has the maximum or minimum value among all the comparison points, then this point is one of the extreme points of the scale space. Fig.5 shows the principle of such scale-space local extremum detection.

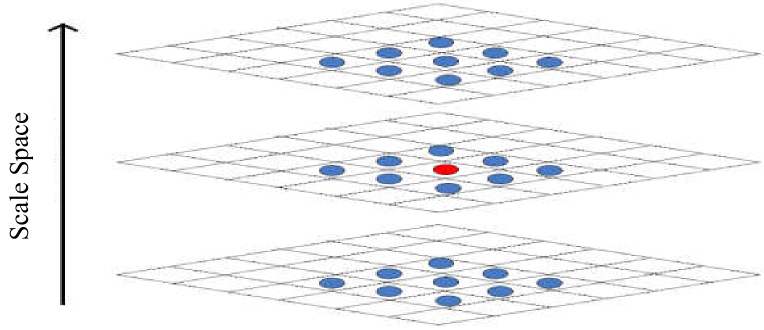


Figure 5. Local extremum detection in DOG scale space

As indicated in Fig.5, compare the red pixel detection point in the middle with 26 pixels points, 8 neighborhood pixels points with the same scale and the corresponding 9×2 pixel points in the upper and lower neighborhood scales. Besides, two important parameters are required to be confirmed in building the scale space. They can be described as scale-space sampling frequency and spatial-doprincipal sampling frequency. Scale-space sampling frequency is shown as all the DOGs of each DOG cluster. Since the maximum scale in each DOG cluster is 2 times of the minimum scale, the more DOGs are in this range, the higher the sampling frequency will be. This frequency affects the result of the feature extraction. And the other parameter is spatial-doprincipal sampling frequency, a numerical value represented as σ . As the convolution between the image and the Gaussian function can be seen as the spatial filtering, then σ is greatly related to the cut-off frequency of the filter. The bigger σ is, the smaller the cut-off frequency will be and the smaller the visible sampling frequency will be(Ladislav and Pavel, 2015).

3.2. Key-point Localization

Perform three-dimensional Taylor series expansion on DOG within the pixel points, localize the extremum position to sub-pixel level and obtain the fitting function of the key point. The expansion formula is shown as follows.

$$D(x) = D + \frac{\partial D^T}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 D}{\partial X^2} X \tag{4}$$

It can be seen that all the partial derivatives can be approximated with the difference of the pixel value. Among them, D and the partial derivative of D are the values calculated in the expansion, after localizing the key point, its DOG function can be obtained from the expansion of DOG in its adjacent pixel point. The function value $D(\vec{x})$ of DOG can be used to remove the unstable key points with low contrast. The lower the value is, the less stable key points shall be ignored, as indicated by the following formula.

$$D(\vec{x}) = D + \frac{1}{2} \frac{\partial D^T}{\partial \vec{x}} \vec{x} \tag{5}$$

The extremume of a Gaussian difference operator with a bad definition has a big principal curvature in the horizontal edge and a small principal curvature in the vertical edge. Only the principal curvatures of the key point are required to calculate so that it can be decided whether it is located in the edge and it shall be eliminated. The principal curvatures are calculated through the 2×2 Hessian matrix H .

The two principal curvatures of this point are proportional to the two feature values of Hessian matrix. In this sense, we can only consider the bigger feature value and we can confirm whether this point is located in the edge(Xiongwen and Qingzhang et al., 2105).

3.3. Orientation Assignment

Give an orientation assignment, include the orientation into the descriptor feature of the key point and then this key point has rotation invariance. The gradient of the pixel point is will be used in the description of the orientation. The model and orientation of the gradient can be defined and calculated with the following pixel difference method.

$$\alpha(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \tag{6}$$

$$\beta(x, y) = \tan^{-1}(L(x, y+1) - L(x, y-1)) / (L(x+1, y) - L(x-1, y)) \tag{7}$$

The orientation of the key point can be confirmed by the gradient orientations of all points within the neighborhood with the key point as the center. In practical calculation, such statistics can be confirmed through gradient orientation histogram. If a feature point has an auxiliary orientation, another new feature point shall be built. This new feature point has the same coordinate as but a different orientation from the original feature. In other words, some feature points with the same coordinates but different orientations will appear(Huanqiang and Kai-Kuang et al., 2015).

3.4. Key-point Descriptor Generation

In order to make the key point not change with the changes of lighting and angels, a descriptor is required to build for each key point and the descriptor shall be highly unique in order to improve the matching probability of the feature point. When building the descriptor, rotate the coordinates of the orientation of the descriptor to the orientation of the key point so that it is ensured that the key point has rotation invariance.

To guarantee the rotation invariance, firstly rotate the coordinates to the orientation of the key point and then take a 16×16 window with the key point as the center, the gree grid in the middle of the left of Fig.6 is the position of the current key point, each grid represents a pixel of the neighborhood of the key point in its scale space, the direction of the arrow is the gradient orientation of the pixel, the length of the arrow is the size of the module value of the gradient and the blue circle is the range of Gaussian weight (the gradient orientation information of the neighborhood pixel closer to the key point makes a bigger contribution). Next, divide the entire region into 16 4×4 small regions; calculate the gradient orientation histograms in 8 orientations of each 4×4 regions; calculate the accumulated value of each gradient orientation and obtain a seed point, as shown in Fig.6. In Fig.6, a key point is consisted of 4 seed points and each seed point has the vector information in 8 orientations. There are 16 seed points in total, the federation of the neighborhood orientation information provides excellent fault tolerance for the feature matching with location error. Finally, normalize the feature vector and remove the effect caused by lighting(Yifang, 2016).

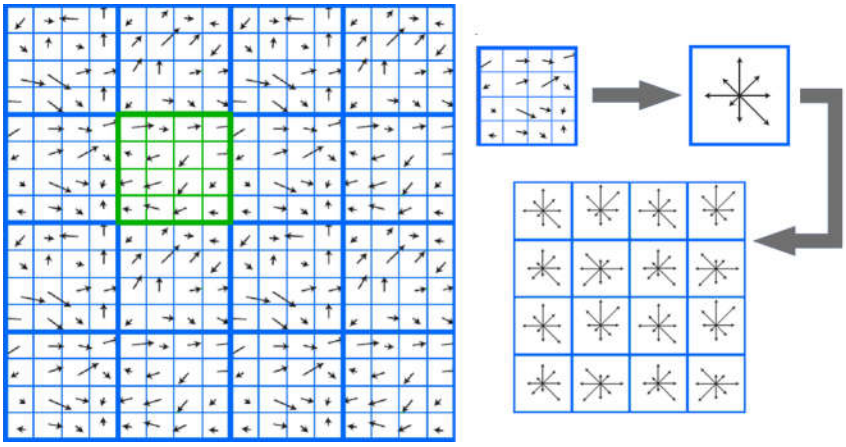


Figure 6. Key-point descriptor generation

We have made some improvements to SIFT algorithm. After rotating the coordinates to the orientation of the key point, make a circular window with a radius of 16 pixels and with the key point as the center in order to confirm the neighborhood range of the key point. Then divide the neighborhood region into 16 sub-regions by the concentric circles with the radiuses of 1-16 pixels and with the key point as the center. Mark the accumulated value of the gray-levels of each pixel within 16 sub-regions as f_i and normalize f_i, namely

$$\overline{f_i} = \frac{f_i}{\sqrt{\sum_{n=1}^{16} f_n}}, i = 1, 2, \ldots, 16 \tag{8}$$

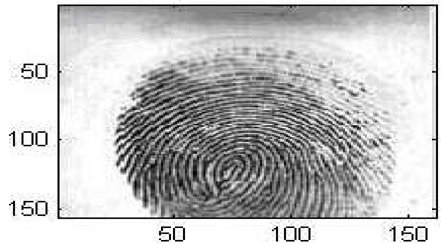
Use first-order numerical difference to calculate the difference value $d\overline{f_i}$ of $\overline{f_i}$, the accumulated value of the gray-levels of these 16 sub-regions, as indicated by Formula (9).

$$d\overline{f_i} = \begin{cases} \left|\overline{f_i} - \overline{f_{i+1}}\right| & i = 1 \\ \left|2f_i - f_{i-1} - f_{i+1}\right| & 1 < i < 16 \\ \left|\overline{f_i} - \overline{f_{i-1}}\right| & i = 16 \end{cases} \tag{9}$$

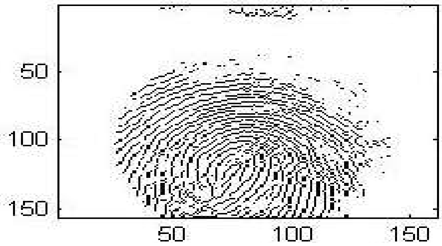
Form the descriptor of the feature point with and normalized grey-level accumulated value $\overline{f_i}$ and the difference value $d\overline{f_i}$. In other words, a simplified 32-dimensional descriptor $F = (\overline{f_1}, \overline{f_2}, \cdots \overline{f_{16}}, d\overline{f_1}, d\overline{f_2}, \cdots, d\overline{f_{16}})$ will be obtained. SIFT features are all concentrated in SIFT vector, namely above the feature descriptor. As a representation of the local image features, it determines the result of various subsequent processing methods based on features(Sana and Alireza et al., 2012).

4. EXPERIMENT AND ANALYSIS

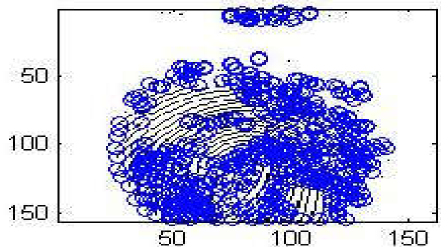
In order to verify the effectiveness of SIFT image feature point extraction algorithm of this paper, it has made the experiment and tested its performance. The CPU of the simulation computer is dual-core 1.6G, the computer memory is 4G and the simulation platform is Matlab2012. The extraction of image feature points will directly affect the matching result. In the image input, the image feature point obtained will have the problems of crease, blur and uneven gray-level due to the influence of various objective factors, therefore, the feature point extraction algorithm shall recognize and remove the false feature points as accurately as possible according to the correlation of the image features and preserve the true feature points. This paper uses 3 real images and extracts the texture features with SIFT algorithm to effectively obtain the image feature point information. Fig.6-Fig.8 has shown the results of different fingerprint texture feature points by the method of this paper.



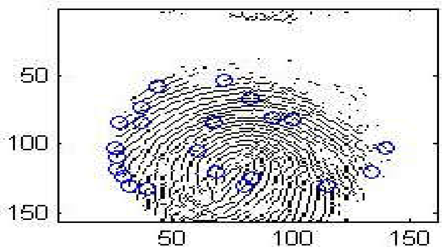
(a) Original image



(b) Refined fingerprint image

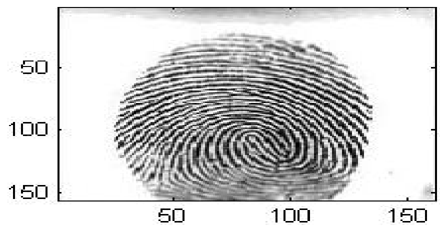


(c) Refined fingerprint image

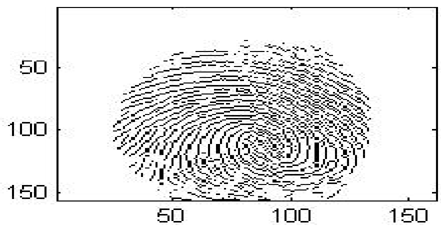


(d) Refined purified fingerprint image feature point

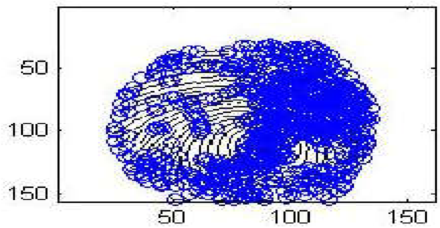
Figure 7. Fingerprint feature point extraction result of loop-shape



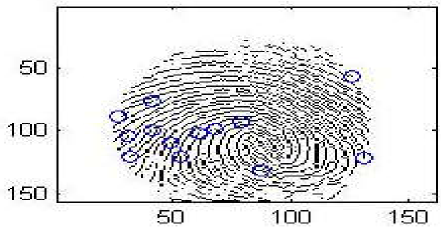
(a) Original image



(b) Refined fingerprint image

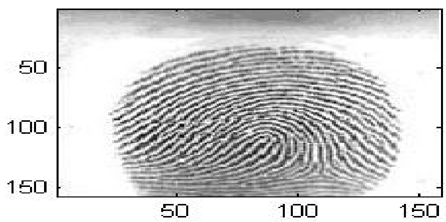


(c) Refined fingerprint image

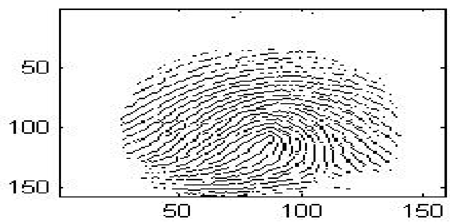


(d) Refined purified fingerprint image feature point

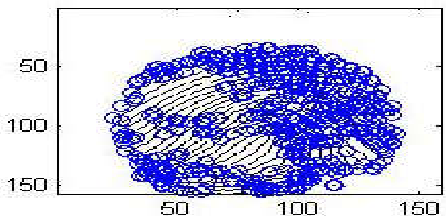
Figure 8. Fingerprint feature point extraction result of whorl-shape

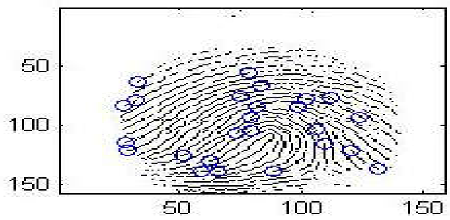


(a) Original image



(b) Refined fingerprint image





(c) Refined fingerprint image (d) Refined purified fingerprint image feature point

Figure 9. Fingerprint feature point extraction result of arch-shape

In the above experiments, the fingerprint feature purification operation mainly removes the redundant fingerprint feature points. The false features usually have the following characteristics. Most of them are located in the image edge; the false feature points within the image are quite close and two or more false features exist in the small region at the same time. The high discrimination of SIFT feature point makes the feature point matching more simple and effective. The feature point extraction algorithm based on SIFT in the experiment is effective given its experimental results. It can be seen from the experiment result that the threshold of DOG function value $D(\vec{x})$ of the key feature point is 0.035. In other words, when $|D(\vec{x})| < 0.035$, all the feature points will be removed completely. The number of filtered matching points will reduce, therefore, while preserving a certain number of matching points and ensuring the matching accuracy, remove some points with bad matching quality on this basis. The experiment shows that when the spatial-domain sampling frequency σ varies within the range of 1.5-1.6 will have a better matching result. The registration ratio is 57.7%~80.2% and the matching efficiency is 51.5%-83%. It can be seen from the experiment result that the majority of the details will be extracted accurately and the fingerprint image feature points shall be purified to greatly reduce the number of the fingerprint image feature points.

5. CONCLUSIONS

The basic idea of local image feature point extraction is to extract the feature points which contain sufficient effective information from the image. SIFT operator has strong robustness in brightness change, scaling, rotation, noises and affine transformation; however, it can be seen from the in-depth analysis of SIFT algorithm that SIFT operator has some weaknesses in the feature extraction and matching. Therefore, this paper has proposed an improved SIFT algorithm, the key-point feature which obtains has invariance in translation, scaling and rotation changes. Moreover, the feature this algorithm extracts is very strong in robustness. The experiment in the final part has fully verified the said strengths of the algorithm of this paper.

ACKNOWLEDGEMENTS

This work was supported by the Hunan Provincial Department of Education Science Research Project (No.16C0903) and the Hunan Provincial Education Science five-year plan funded project(No:XJK014BGD046).

REFERENCES

Amin Sedaghat, Hamid Ebadi (2015) "Distinctive Order Based Self-Similarity Descriptor for Multi-Sensor Remote Sensing Image Matching", *ISPRS Journal of Photogrammetry and Remote Sensing*, 108(10), pp.62-71.

Gholam Ali Montazer, Davar Giveki (2015) "Content Based Image Retrieval System Using Clustered Scale Invariant Feature Transforms", *Optik-International Journal for Light and Electron Optics*, 126(18), pp.1695-1699.

Huanqiang Zeng, Kai-Kuang Ma, Chen Wang, Canhui Cai (2015) "SIFT-Flow-Based Color Correction for Multi-View Video", *Signal Processing: Image Communication*, 36(8), pp.53-62.

Jonathan Grier, Golden G. Richard III. (2015) "Rapid Forensic Imaging of Large Disks with Sifting Collectors", *Digital Investigation*, 14(1), pp.S34-S44.

Ladislav Lenc, Pavel Král (2015) "Automatic Face Recognition System Based on The SIFT Features", *Computers & Electrical Engineering*, 46(8), pp.256-272.

Peizhi Chen (2015) "Fusing Two- And Three-Dimensional Feature to Correlate Volumetric Images", *Optik-International Journal for Light and Electron Optics*, 126(24), pp.5787-5789.

Robert E. Smith, Jacques-Donald Tournier, Fernando Calamante, Alan Connelly (2015) "SIFT2: Enabling Dense Quantitative Assessment of Brain White Matter Connectivity Using Streamlines Tractography", *NeuroImage*, 119(1), pp.338-351.

Sana Sadeghi, Alireza Rezvanian, Ebrahim Kamrani (2012) "An efficient method for impulse noise reduction from images using fuzzy cellular automata", *AEU - International Journal of Electronics and Communications*, 66(9), pp.772-779.

Stefan G. Stanciu, Denis E. Tranca, Dinu Coltuc (2015) "Contrast Enhancement Influences The Detection of Gradient Based Local Invariant Features and The Matching of Their Descriptors", *Journal of Visual Communication and Image Representation*, 32(10), pp.246-256.

Wassim Bouachir, Guillaume-Alexandre Bilodeau (2015) "Collaborative Part-Based Tracking Using Salient Local Predictors", *Computer Vision and Image Understanding*, 137(8), pp.88-101.

Xiaorong Gao, Yingyong Wu, Kai Yang, Jinlong Li (2015) "Vehicle Bottom Anomaly Detection Algorithm Based on SIFT", *Optik-International Journal for Light and Electron Optics*, 126(23), pp.3562-3566.

Xiongwen Cui, Qingzhang Wu, Jin Zhou (2015) "Online Fragments-Based Scale Invariant Electro-Optic Tracking with SIFT", *Optik - International Journal for Light and Electron Optics*, 126(18), pp.1720-1725.

Yifang Wang (2016) "Motion Blurred Image Restoration Based on Improved Genetic Algorithm", *Rev. Téc. Ing. Univ. Zulia.*, 39(5), pp.231-237.

Yu Liu, Zengfu Wang (2015) "Dense SIFT for Ghost-Free Multi-Exposure Fusion", *Journal of Visual Communication and Image Representation*, 31(8), pp.208-224.

Yu Wang, Xiaojuan Ban, Jie Chen, Bo Hu, Xing Yang (2015) "License Plate Recognition Based on SIFT Feature", *Optik-International Journal for Light and Electron Optics*, 126(21), pp.2895-2901.