



Local Feature Acquisition Method of Multi-layer Vision Network Image Based on Virtual Reality

Jinzhu Liu^(✉) and Shuai Zheng

XiaMen HuaXia University, XiaMen 361024, China
liuwenping74@163.com

Abstract. As an important part of image features, local image features reflect the changes of local information in the image, which are not easily disturbed by various changes such as noise, illumination, scale, and rotation. Faced with this situation, this paper proposes a method for local feature acquisition of multi-layer visual network images based on virtual reality. Based on virtual reality technology, the multi-layer visual network image is reconstructed in layers, and the reconstructed images are preprocessed by histogram equalization and denoising. The SUSAN algorithm and the SIFT algorithm are combined to realize the acquisition of local image features. The results show that compared with the original SUSAN algorithm and the original SIFT algorithm, the average running time of the researched method is shorter and the overlap error is smaller, which indicates that the researched method has lower time complexity and higher acquisition accuracy.

Keywords: Virtual reality technology · Multi-layer visual network image · Layered reconfiguration · Pretreatment · SUSAN algorithm · SIFT algorithm · Image local feature acquisition

1 Introduction

How to obtain a more discriminative image local feature descriptor is an important hotspot in the field of computer vision. It plays an important role in image modeling and reconstruction. Image local feature descriptor has a good effect on maintaining the invariance of image rotation. The scale invariant feature transform descriptor proposed by Lowe can efficiently detect key points in images because of its scale invariance. It is the most widely used descriptor in image recognition. These feature descriptors have different design and implementation methods, but their purpose is to optimize the general performance of image local features. However, the advantages of human visual system for extracting image local features have not been taken into account in the implementation of existing algorithms. The purpose of computer vision is to extract meaningful descriptive information from images or image sequences. Traditional methods establish algorithm models with the help of geometry, physics, learning theory and statistical methods, sample and process visual information, and achieve some results, but they are still far from the cognitive ability of biological visual system [1]. Some researchers

have proposed an improved SUSAN corner detection algorithm to extract image features through corner detection. After Canny edge detection, SUSAN corner detection is performed on the detected edge pixels, and then Euclidean distance is used to make the detected corners more accurate. However, because the modified algorithm requires layered reconstruction of multi-layer vision network images, it may increase the time of image feature extraction [2]. Other researchers have done image feature matching by improving SIFT algorithm. By reducing the vector dimension, the selection range of the neighborhood around the feature points is reduced, the matching time of the image feature points is accelerated, and the matching error rate is reduced to a certain extent [3]. In order to solve this problem, this paper will combine SUSAN algorithm and sift algorithm to design a multi-layer visual network image local feature acquisition method based on virtual reality, in order to improve the accuracy of image local feature acquisition.

2 Research on Local Feature Acquisition of Multi-layer Visual Network Images

Local feature acquisition is usually the first step for many problems in computer vision and digital image processing, such as image classification, image retrieval, wide baseline matching, etc. the quality of feature extraction directly affects the final performance of the task [4]. Therefore, the local feature extraction method has important research value. However, the image often changes in scale, translation, rotation, illumination, angle of view and blur. Especially in practical application scenarios, the image will inevitably have large noise interference, complex background and large target attitude changes. This brings more challenges to the problem of image local feature acquisition. Therefore, the research on local feature acquisition method still has important theoretical significance and application value, which is worthy of researchers' continued attention.

2.1 Layered Reconstruction of Multi-layer Visual Network Images Based on Virtual Reality

Virtual reality, also known as spiritual environment technology, is a lifelike three-dimensional virtual reality environment built by people using computer technology, which provides users with a new technology that can independently interact with them through computer input devices [5]. Based on many characteristics of virtual reality technology and the rapid development of computer technology and virtual reality technology, virtual reality technology has made great achievements. While changing people's lifestyle, it is also considered as one of the major technologies that may cause great changes in the world in the 21st century.

At present, the interactivity in the virtual reality system is mainly realized through nine aspects. Conceptuality is the rationality of the existence of various things (real or imagined) in the virtual environment, and most other digital media also have this feature, such as animation, movies, etc. The rapid development of digital processing, display technology and media acquisition technology has led to the explosive growth of 3D image data. As a new type of multimedia image mode, 3D image mainly provides images with different viewing angles for the left eye and right eye of the human, so as to

use the human perception ability to create a three-dimensional perception scene for the viewer. Therefore, 3D images have attracted more and more attention in the multimedia field. It has an increasingly broad market demand [6]. However, with the continuous development of 3D images, there are still many technical problems that need to be solved urgently, among which the most urgent problem is the layered reconstruction of 3D images. The layered reconstruction of 3D images is directly related to the compression quality of the images. However, there are certain problems in the traditional methods of layered reconstruction of 3D images, which lead to a great decrease in the quality of images during compression.

For the above situation. In this chapter, virtual reality technology is used to realize layered reconstruction of 3D images: firstly, 3D images are layered based on virtual reality technology, and then layered reconstruction of 3D images is realized by triangulation algorithm. The specific steps of hierarchical reconfiguration are as follows:

- 1) Firstly, the virtual reality technology is used to layer the 3D image;
- 2) Set a point corresponding to the minimum value and the maximum value, and extract a total of four maximum values. The four maxima extracted represent the quadrangular coordinates of the corresponding containment rectangle of the scattered point set, and the circumscribed circle radius value and center coordinates of the containment rectangle are obtained to obtain the circumscribed circle coordinates of the containment rectangle;
- 3) Obtain the vertex coordinates of the circumscribed triangle corresponding to the circumscribed circle of the containing rectangle, and obtain the specific coordinates of the circumscribed triangle;
- 4) Taking the vertex coordinates of the circumscribed triangle as additional points, the three vertices are placed in the points group of the layered 3D image in a counter-clockwise direction, and the circumscribed triangle is placed in the triangle array of the layered 3D image;
- 5) Insert data points point by point: Initialize the triangulation and insert data points point by point. First, the relationship between the three sides a , b , c and the insertion point V is judged by the CCW judgment method.
- 6) Find the triangles that the insertion point V can insert until a triangle mesh is formed, so as to realize the hierarchical reconstruction of the 3D image.

2.2 Image Preprocessing

When an image is sent to the computer as input information for processing, there is often noise due to various reasons, which is not suitable for the recognition of the machine vision system. Generally, the image will be preprocessed before entering the vision system. The preprocessing process is not a complete denoising process. Its main purpose is to enhance the useful information of the image and eliminate the unnecessary information as much as possible. In fact, Image preprocessing is the enhancement of useful information.

Histogram Equalization

Histogram equalization is a kind of point processing. The gray value distribution of most images is uneven, and there is often a phenomenon of centralized gray value distribution. The uneven distribution of gray values is very unfavorable for image segmentation and image comparison. The result of image equalization is that the number of pixels of each gray level is basically the same, so that the upper limit value of the histogram obtained does not have much difference, showing a horizontal state [7]. Histogram equalization can be done in the following ways:

$$S_{D(x,y)} = N \cdot \frac{\sum_{a=1}^A h(a) \cdot D(x,y)}{A} \quad (1)$$

In the formula, $S_{D(x,y)}$ represents the image output after histogram equalization processing, A represents the number of pixels, N represents the number of gray levels, $D(x,y)$ represents the input image, and $h(a)$ represents the histogram of the input image.

This method is relatively simple, but after the histogram is extended, the new histogram is still unbalanced, so the new histogram should be modified.

Image Denoising

Image information inevitably carries noise in the process of transmission, which will affect image recognition, so the process of denoising in image processing is essential [8]. Gaussian noise is a kind of noise with normal distribution. Linear filter can select the desired frequency from many frequencies, and it is also suitable for removing the unwanted frequency from many frequencies. It has an obvious effect on Gaussian noise.

Mean Filter

The mean filter is a relatively simple linear filter, which belongs to the local spatial domain algorithm. The value of each pixel is replaced by the average pixel value of its surrounding points to achieve the purpose of smoothing the image [9]. The specific formula is as follows:

$$G(x,y) = \frac{\sum_{(x,y) \in F} H(x,y)}{n} \quad (2)$$

In the formula, $H(x,y)$ represents the original image, $G(x,y)$ represents the filtered image, F is a neighborhood set generated around (x,y) , and n is the number of all points in the neighborhood F . The size of F determines the value of n , and also determines the radius of the neighborhood. The mean filter will blur the image while eliminating noise, and the size of its radius will have a proportional impact on the blurriness of the image.

Gaussian Filter

The function of Gaussian filter is to eliminate Gaussian noise. It is a process of weighted average of image. For each pixel, the convolution template is used to weighted average the pixel gray values in the neighborhood to determine its value. Gaussian filtering can be realized by discrete window sliding window convolution or Fourier transform. The

former is commonly used, but with the increase of the amount of calculation, the latter should be considered when it causes great pressure on time and space.

The discrete window sliding window convolution is completed by a separable filter, which is the one-dimensional decomposition of multidimensional convolution. Generally, two-dimensional zero mean discrete Gaussian function is used for noise removal. Equation (3) is the specific formula of two-dimensional zero mean Gaussian function.

$$E(x, y) = e^{-\frac{x^2+y^2}{2d^2}} \quad (3)$$

In the formula, d is the standard deviation; $E(x, y)$ is the Gaussian function; (x, y) is the point coordinate.

Gaussian filter belongs to low-pass filter, which plays an important role in machine vision. Its wide application is inseparable from the characteristics of Gaussian function: two-dimensional Gaussian function has rotational symmetry; Gaussian function is a single valued function; The Fourier transform spectrum of Gaussian function is single lobe; The adjustability of Gaussian filter width makes a compromise between image features and abrupt variables; Gaussian function is separable.

2.3 Realization of Image Local Feature Acquisition

Local feature acquisition is a necessary step before local feature description. After the collected feature points or feature areas are described by description operators, corresponding processing is carried out according to the characteristics of different application algorithms. For example, in the image mosaic and matching algorithm, feature points or feature regions are matched [10].

There is no general or precise definition of image features so far. For different problems, different application objects and application scenarios, the definitions of features will be different to varying degrees. Simply put, a feature is the interesting part of a digital image, which is the basis of many image processing algorithms. Therefore, the success of an algorithm has a lot to do with the features it defines, selects, and uses.

Some basic concepts related to image features are edges, corners, regions, features, and ridges.

- Edge is the set of pixels that constitute the boundary (or edge) between two regions in the image. It can be summarized as line feature or point feature. It mostly exists in places where the gray information of the image changes dramatically. Generally, the shape of the edge may be arbitrary, and may even include crossed pixels. In the practical operation of feature detection, the edge is usually defined as a set of points with large gradient amplitude in the image, and the gradient angle of these points is the orthogonal direction of the edge. When describing edges, because discrete point sets are extracted, and "false edges" are often detected (shadows in images form edge like features under different illumination). Therefore, some algorithms often filter the points with high gradient amplitude (select some points and eliminate some points) and use a certain way to connect them to form a perfect description of edge features.
- A corner is a point feature in an image, which is different from the one-dimensional structure of the edge, and has a two-dimensional structure in the local corner. It

generally occurs where the direction of the boundary changes drastically, where the edges of multiple straight lines intersect, or where the grayscale of the image changes drastically. Therefore, several early corner detection algorithms usually first perform edge detection on the image, and then find the place where the edge direction changes abruptly, that is, the position of the corner. The subsequent corner detection algorithm removes the step of detecting the edge, but directly finds the point with large curvature value in the image gray gradient information, so as to obtain the corner information more accurately. Compared with other image features, corner points are easier to detect, and at the same time, it has good stability for various image transformations, noise interference, changes in external conditions and other factors.

- The regional feature is different from the corner, but it is also composed of point and line features, and the information contrast between the sub blocks of the image is high.
- The local feature points obtained from the local feature description reflect the local characteristics of the image. Local features are not easy to be affected by image rotation, scale, angle of view and other change factors, but also try to avoid the shortcomings of some global features, such as easy to be affected by factors such as target occlusion or complex background. Therefore, local features can better solve the problem of image recognition in the case of translation, scaling, angle change, rotation, scale transformation, noise or occlusion. At the same time, they are widely used in image matching, image mosaic, image registration and three-dimensional reconstruction, which is a hot issue in recent years.

The local feature points of an image are mainly divided into two categories: blobs and corners. Spots represent areas that are different from the surrounding color or grayscale; corners usually refer to the corners of objects in the image or the intersections between edge lines. Correspondingly, the local image feature description methods include blob detection and corner detection.

SUSAN Algorithm

The local gradient method is sensitive to the influence of noise and has a large amount of calculation, while SUSAN corner point is a morphological-based corner point feature acquisition method, which is directly calculated based on the image gray value. The method is simple and the calculation efficiency is high. The idea of the SUSAN corner point algorithm is to use a circular template with a fixed radius to slide on the image. The center pixel of the template is called the kernel. If the difference between the gray value of other points in the template and the gray value of the kernel is less than the threshold, it is considered that The point and the nucleus belong to a similar gray level, and the area composed of all pixels that satisfy this condition is called the nucleus value similarity area (USAN). In this way, the local neighborhood of the core pixel is divided into two parts, that is, the similar area and the dissimilar area, and the image texture structure of the local area is reflected by counting the size of the similar area of the core value, that is, the number of points that make up the USAN area. When the USAN area is large, it is generally a smooth area, and when the USAN area is small, it is generally a corner point.

SIFT Algorithm

SIFT algorithm has scale invariance and rotation invariance, and is widely used in image matching. The algorithm mainly includes four steps: scale space extreme value detection, key point location, direction assignment and key point description.

Step 1: scale space extreme value detection. In order to ensure the scale invariance, the concept of scale space is introduced. The scale space has a parameter that controls the change of scale, so as to construct different scale spaces. The scale space function $P(i, j, z)$ is obtained by convolution of the input image $Q(i, j)$ and the variable scale Gaussian kernel function $T(i, j, z)$, expressed as follows:

$$P(i, j, z) = Q(i, j) \otimes T(i, j, z) \quad (4)$$

In,

$$T(i, j, z) = \frac{1}{2\pi z^2} e^{-\frac{i^2+j^2}{2z^2}} \quad (5)$$

where \otimes represents the convolution operation, (i, j) represents the pixel coordinates in the image, and z is the scale factor.

The size of z determines the clarity of the image. The large-scale factor corresponds to the overall features of the image, and the small-scale factor corresponds to the detailed features of the image. Therefore, the large z value corresponds to the low resolution (blur), and the small z value corresponds to the high resolution (clarity). To establish the scale space, firstly, Gaussian pyramid is constructed, which is completed by fuzzy filtering and down sampling. The pyramid consists of multiple sets of images. Each group of images has multiple sub images with the same size. The first group is the original image, and the next group of images are obtained from the down sampling of the previous group of images. These images are arranged from bottom to top, from large to small, forming a Gaussian pyramid. The number of groups n of Gaussian pyramid image should not be too large, otherwise the image size on the top of the tower is very small, and the significance of feature point detection will be lost.

The Laplacian of Gaussian (LoG) takes the second-order derivation of the image to complete the detection of feature points at different scales, but there is a problem of low detection efficiency, so the SIFT algorithm uses the difference of Gaussian (DoG) image instead of the Laplacian of Gaussian image., and DoG has better stability and stronger anti-interference than LoG. The difference image can be obtained by subtracting two adjacent sub-images in each group of images, and the extreme value detection is performed on the basis of the difference image. The difference image is represented as

$$\begin{aligned} L(i, j, z) &= [T(i, j, kz) - T(i, j, z)] \otimes Q(i, j) \\ &= P(i, j, kz) - P(i, j, z) \end{aligned} \quad (6)$$

where k is the multiplicative factor of two adjacent scale spaces.

Step 2: Keypoint positioning. After the extreme points are detected in the scale space, they are used as keypoint candidates. For each candidate location keypoint, its stability will be evaluated to decide whether to retain the detection result. The detection of extreme points in the previous step is carried out in discrete space. In order to obtain

a more accurate extreme value position, the sampling points are fitted with a three-dimensional quadratic function, and the unstable and low-contrast key points at the edge are eliminated at the same time. So as to find the real feature points.

Step 3: key orientation matching. After the position information of feature points is determined, SIFT algorithm counts the gradient direction distribution of pixels in the neighborhood of each key point, so as to determine the direction for them, so that the final extracted features meet the rotation invariance. The gradient and amplitude of each feature point are solved as follows.

$$\phi(i, j) = \sqrt{\frac{[P(i+1, j) - P(i-1, j)]^2}{[P(i, j+1) - P(i, j-1)]^2}} \quad (7)$$

$$\varphi(i, j) = \arctan \frac{P(i+1, j) - P(i-1, j)}{P(i, j+1) - P(i, j-1)} \quad (8)$$

In the formula, $\phi(i, j)$ and $\varphi(i, j)$ represent the magnitude and direction of the gradient, respectively, and $P(i, j)$ is the value in the space of the scale where the feature point is located.

The gradient amplitude and direction of each pixel in the neighborhood of key points are counted, and then histogram statistics are carried out. The direction 3600 is equally divided into 8 sub intervals, and the sum of gradient amplitudes of all surrounding pixels falling into the corresponding interval is counted. The angle with the largest cumulative amplitude is taken as the main direction of the key point, and the angle with an amplitude greater than 80% of the maximum value is taken as the auxiliary direction of the key point.

Step 4: Description of key points. The SIFT feature descriptor is the result of statistics on the Gaussian image gradient in the neighborhood of the feature point. The algorithm first divides the neighborhood of the feature point into blocks, and counts the gradient histogram in each block to generate the feature vector descriptor. The description Symbol is an abstract representation and is unique. Take the main direction of the feature point as the direction axis and the feature point as the origin to establish a new coordinate. In the 16×16 neighborhood of each feature point, the SIFT algorithm calculates the gradient direction histogram for $4 \times 4 \times 4$ regions. There are 8 statistical directions, and a statistical interval is set every 45° , thereby forming For an 8-dimensional vector, 4×4 such vectors can be generated in a 16×16 neighborhood, thus finally forming a 128-dimensional feature vector in the SIFT feature extraction algorithm. Finally, in order to make the feature extraction algorithm have good illumination invariance, the generated feature vector descriptor is normalized, and the processing process is as follows.

$$v_j = \frac{l_j}{l_1 + l_2 + \dots + l_{128}} \quad (9)$$

where, $L = (l_1, l_2, \dots, l_j, \dots, l_{128})$ is the currently obtained 128 dimensional eigenvector descriptor, and $V = (v_1, v_2, \dots, v_j, \dots, v_{128})$ is the normalized eigenvector descriptor.

Image Local Feature Acquisition Based on SUSAN + SIFT

The SUSAN algorithm is combined with the SIFT algorithm to realize the acquisition of local image features. The specific process is shown in Fig. 1 below.

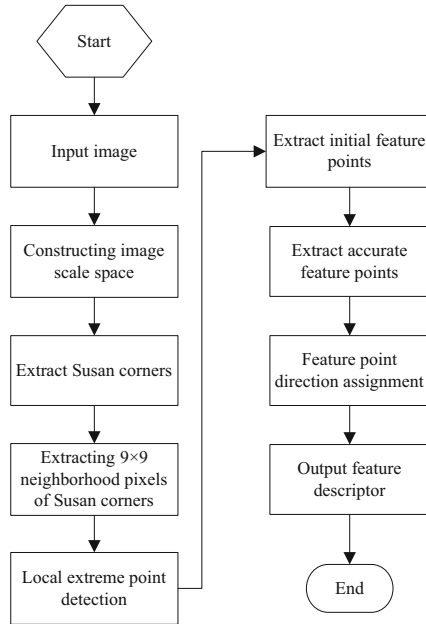


Fig. 1. The process of image local feature acquisition based on SUSAN + SIFT

The implementation process of the improved SIFT feature extraction algorithm is divided into the following five steps.

Step 1: SUSAN corner detection. The input original image is filtered using a 37-pixel circular template to count the area of the USAN area at all pixels. When the area is less than the threshold, it is determined as the initial corner point, and then the initial corner point is subjected to local non-polarization. A large value is suppressed to obtain SUSAN corner points, and the position and grayscale information of the corner points are saved.

Step 2: construct the scale space. The operation of constructing scale space is completely consistent with the original operation of SIFT algorithm. Through continuous downsampling of images and Gaussian filtering of different scales for each size of images, a Gaussian pyramid is formed, and then the images of adjacent two layers in each group in the Gaussian pyramid are differenced to form a Gaussian difference pyramid to complete the construction of scale space.

Step 3: Extremum detection in scale space. Map the positions of the detected SUSAN corners to the corresponding positions of each group and each layer (the mapping method is to downsample the SUSAN corners), and remove the edge SUSAN corners in each group of images, Then, each pixel in the 9×9 neighborhood of the corner point in each layer of the image is detected by the local maximum value of 26 neighborhoods. If the point is a local maximum value, it is determined as the initial feature point.

Step 4: Location of key points. First, through the fitting function, the position and scale information of the initial feature points are corrected, and the points with too low contrast are eliminated. Then, the edge response points are eliminated by judging the

relationship between the two eigenvalues of the Hessian matrix, and finally each feature point is calculated. Harris's corner response value CRF, eliminates the feature points less than 1/10 of the absolute value of the absolute value of all feature points, so as to obtain the final accurate feature points.

Step 5: determine the key direction. The main and secondary directions of feature points are obtained by statistical gradient direction histogram of pixels in the neighborhood of feature points.

Step 6: generation of feature descriptor. Feature descriptors are generated according to the formation process of feature descriptors, and normalized to generate feature descriptors with scale invariance, rotation invariance and illumination invariance.

3 Method Test

3.1 Sample Preparation

For the multi-layer visual network image local feature acquisition method based on virtual reality, when it has not been tested, it is only limited to the theoretical significance, and has no practical value for the actual scene. Therefore, it is necessary to carry out a series of algorithm tests, and compare and analyze with the original algorithm that has not been improved, so as to comprehensively and objectively evaluate the improved algorithm. In this paper, a large number of data tests are carried out for the multi-layer visual network image local feature acquisition method based on virtual reality. 20 images are selected as the original test set to test the performance of this method, the original SUSAN algorithm and the original SIFT algorithm, and the relevant experimental data are obtained. Thus, the experimental evaluation indexes are calculated, displayed and analyzed, and the final evaluation results are obtained. Some image samples are shown in Fig. 2 below.



Fig. 2. Partial image sample

3.2 Image Hierarchical Reconstruction Results

The images in the samples are reconstructed in layers with the help of 3ds Max software of virtual reality technology. Taking one of the samples as an example, the image layer reconstruction results are shown in Fig. 3 below.



Fig. 3. Image layered reconstruction results

3.3 Image Feature Collection Results

The method based on virtual reality is used to collect local features of 20 sample images. The results are shown in Table 1 below.

Table 1. Image feature collection results

Image samples	Number of eigenvectors
1	244
2	584
3	423
4	756
5	1244
6	357
7	844
8	654

(continued)

Table 1. (continued)

Image samples	Number of eigenvectors
9	578
10	424
11	365
12	772
13	201
14	585
15	1255
16	418
17	842
18	427
19	624
20	515

3.4 Evaluation Indicators

- (1) For feature extraction of such large-scale and complex images, the time complexity of the algorithm is the primary evaluation standard. If the time complexity is too high, the image processing time is too long, and the computational memory is also very large. Even if the performance of the algorithm in other aspects is very good, it is also very limited in application scenarios. Therefore, this paper first analyzes the time complexity of the method for the research method. The analysis method is: use the original SUSAN algorithm, the original SIFT algorithm and the research method to test 30 images in the image test set, count the average running time of the three algorithms, and then compare the time complexity of the method.
- (2) Use the overlap error between regions to measure the acquisition accuracy of the method. Let μ and η denote local features detected from two images with the same scene but with affine changes, O denote the homography matrix between the two images, and O_η denote the mapping of regions to regions through the homography matrix. In the image where it is located, ξ_μ represents the ellipse fitting region corresponding to the region, and the overlap error $\Phi(\mu, \eta)$ between the two regions is expressed as:

$$\Phi(\mu, \eta) = 1 - \frac{\xi_\mu \cap O_q}{\xi_\mu \cup O_q} \quad (10)$$

If the overlap error between two regional features is less than a set threshold, the two regional features are considered to be related.

3.5 Method Test Results

Among the 20 sample images, one image is randomly selected and compared with the original SUSAN algorithm, the original SIFT algorithm and the studied method. The effect of image local feature acquisition is shown in Fig. 4.

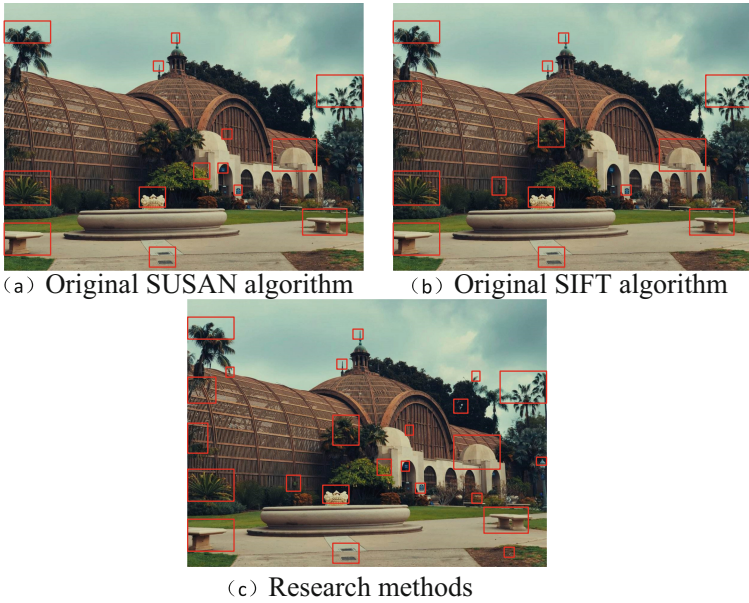


Fig. 4. Image local feature acquisition effect of different methods

It can be seen from Fig. 4 that the local feature points of the image that can be collected by the research method in this paper are far more than those of the two methods, which proves that the local feature collection effect of the image of the research method in this paper is better.

At the same time, the process of image local feature acquisition is recorded, and the overlapping error between the collected running time and the image area is compared. The test results are shown in Table 2 below.

Table 2. Method test results

Method	Average running time (s)	Overlap error (%)
Research methods	5.25	2.25
Original SUSAN algorithm	8.32	5.87
Original SIFT algorithm	7.31	5.66

As can be seen from Table 2, the running time of the original SUSAN algorithm is 8.32 s, the running time of the original SIFT algorithm is 7.31 s, while the running time of the method studied in this paper is 5.25 s, which is shorter. In terms of error, the overlap error of the original SUSAN algorithm is 5.87%, the overlap error of the original SIFT algorithm is 5.66%, while the overlap error of the research method in this paper is 2.25%, the overlap error is smaller, and the acquisition accuracy is higher.

4 Conclusion

This paper studies the local feature acquisition method of multi-layer visual network image based on virtual reality, and extracts the local feature of the reconstructed multi-layer visual network image. Combining the existing SUSAN algorithm with SIFT algorithm, the ability of image feature acquisition of multi-layer visual network is enhanced. The experimental test proves that the research method can effectively speed up the running time of image local feature acquisition, reduce the overlapping error between image regions, and improve the accuracy of acquisition.

However, due to time constraints, this paper did not consider the threshold problem in SUSAN algorithm when studying the new image local feature acquisition method, so we can continue to improve the design method in the next research to enhance the adaptability of the algorithm.

References

1. Liu, S., et al.: Human memory update strategy: a multi-layer template update mechanism for remote visual monitoring. *IEEE Trans. Multimedia* **23**, 2188–2198 (2021)
2. Zheng, H., Lin, Y.: An improved SUSAN corner detection algorithm. *Computer Knowledge and Technology, Academic Edition* **16**(22), 3 (2020)
3. Cheng, J., Zhang, J., Hu, J.: Improvement of SIFT algorithm in image feature matching environment. *J. Heilongjiang Univ. Sci. Technol.* **30**(4), 4 (2020)
4. Liu, S., Liu, D., Muhammad, K., Ding, W.: Effective template update mechanism in visual tracking with background clutter. *Neurocomputing* **458**, 615–625 (2021)
5. Shuai, L., Shuai, W., Xinyu, L., et al.: Fuzzy detection aided real-time and robust visual tracking under complex environments. *IEEE Trans. Fuzzy Syst.* **29**(1), 90–102 (2021)
6. Saad, E., Hirakawa, K.: Improved photometric acceptance testing in image feature extraction tasks. *J. Electron. Imaging* **29**(4), 1 (2020)
7. Ramkumar, B., Laber, R., Bojinov, H., et al.: GPU acceleration of the KAZE image feature extraction algorithm. *J. Real-Time Image Proc.* **17**(5), 1169–1182 (2020)
8. Classification of wood knots using artificial neural networks with texture and local feature-based image descriptors. *Holzforschung* **76**(1), 1–13 (2022)
9. Wang, Y., Song, X., Gong, G., et al.: A multi-scale feature extraction-based normalized attention neural network for image denoising. *Electronics* **10**(3), 319 (2021)
10. Wang, Y., Yang, Y., Zhang, P.: Gesture feature extraction and recognition based on image processing. *Traitement du Signal* **37**(5), 873–880 (2020)