

## Vectorization Method on the Color Cloud Image

Xumin Liu<sup>1</sup>, Cong Zhang<sup>1</sup>, Sen Na<sup>1</sup> and Weixiang Xu<sup>2</sup>

<sup>1</sup>*College of Information Engineering  
Capital Normal University, Beijing 100048, P. R. China*

<sup>2</sup>*College of Traffic and Transportation  
Beijing Jiaotong University, Beijing 100044, P.R. China  
liuxumin@126.com*

### Abstract

*The vectorization, which is an interdisciplinary combining subjects of the computer vision, computer image processing, computer graphics and artificial intelligence, is an important issue in the digital image processing. This paper makes a study of the vectorization method on the satellite color cloud image, Specific techniques such as cloud image layering, threshold selection, edge detection and tracking, feature point extraction, curve fitting are involved. This paper proposes a color layering algorithm whose parameter is the difference between luminance and saturability by analyzing the color features of the satellite cloud pictures. In this algorithm method, we first generate the reference image and then histogram equalize the reference image, afterwards do de-noises process to the above histogram and smooth it for the threshold choice. After analyzing the uncertainty of satellite cloud image edge, this paper puts forward an edge detection algorithm based on adaptive threshold. In order to modify the shortcomings of edge data of raster and edge data of chain code, this paper puts forward the corner detection method based on edge data chain. Also this paper develops a prototype system of automatic vectorization on satellite cloud images by curve fitting the feature points and gets the final vector graphics. Experiments show that the system has the basic functions of automatic extraction, recognition and vectorization for the information of satellite cloud clouds, and carry out the comprehensive verification and analysis for the theory and algorithm by using the real satellite cloud image data.*

**Keywords:** *Satellite cloud image; Vectorization; Image segmentation; Edge detection; Corner detection*

### 1. Introduction

The satellite observation on earth has become an indispensable source of information. The scope of the observation includes meteorology, remote sensing, aviation, agriculture, environment, construction, electricity, water conservancy, transportation and many other fields, the satellite plays an important role in the sustainable development of human society, global disaster prevention and other aspects. Satellite cloud pictures are images of the cloud cover on the earth surface and the features of the ground surface taken by the meteorological satellite while overlooking the earth.

Satellite cloud image vectorization technology has become the main study direction of satellite image information processing in the world today. The satellite data is playing an increasingly significant role in the global weather forecasting, climate monitoring, disaster prevention and mitigation, climate change and other areas. Satellite cloud image vectorization technique can greatly improve the degree of accuracy and service efficiency in monitoring and predicting disastrous weather including typhoon, rainstorm, fog, dust and so on.

However, presently the main approach of satellite cloud image analysis is an artificial one, by which the observer makes a subjective judgment after the visual analysis. This

method is not only easily influenced by the various factors of the observer, but also go against the fully extraction of the information of the satellite cloud images, thus, hindering the development of weather forecast with the scientific automation and quantification.

In recent years, many foreign scholars devoted to the study of image vectorization method, and worked out a lot of good methods for vectorization. For example, Ailing De *etc.*, put forward “SOM neural network to realize the adaptive vectorization algorithm” [1]. Pascal Barla *etc.*, put forward “Vectorization method of photos” [2]. Guillaume Lobet *etc.* put forward “split-and-combine vectorization new method” [3]. Zhang Xianquan *etc.*, put forward “Binary image boundary vectorization algorithm based on control points” [4]. At the same time, the study of image vectorization has also rapidly developed. Xie Guofu *etc.*, put forward “Precise automatic image vectorization method based on the layering diffusion curve” [5]. Dong Wei *etc.*, put forward “Image vectorization method based on mathematical morphology” [6]. Zhao Jiaojiao *etc.* put forward “Application of blue noise sampling method for image vectorization” [7]. Gao Wei *etc.*, do research on “Compiler optimization for SIMD automatic vectorization” [8].

Although much progress has been made on the study of image vectorization, but there are few vectorization methods based on the meteorological satellite cloud images of meteorological fields, and among which most analyze and extract useful image information using the image processing methods after analyzing a large number of satellite cloud images and the traditional methods. Therefore, reasonable processing, analysis and application of the satellite cloud image information have become the main study direction of satellite cloud image processing.

This paper synthetically analyzes the existent image vectorization theory, strategy and method according to the characteristic properties of satellite cloud image, discusses the characteristic of cloud image region of satellite cloud image, points out the two attributes of the cloud image, puts forward the image layering method for the cloud image, analyzes the uncertainty of the edge in the cloud image, puts forward the cloud edge detection based on adaptive threshold, combining with the macroscopic property of the image, researches the advantages and disadvantages of the edge data of raster and edge data of chain code, puts forward the corner detection method based on edge data chain, combining theory with practice, and puts forward a kind of vectorization algorithm based on satellite cloud image, develops a automatical vectorization prototype system of satellite cloud images.

## **2. Color Layering of the Satellite Cloud Image**

### **2.1. Color Layering**

Color layering [9] can also be called the image decomposition, which is used for image recognition field. It refers to the layer processing of the image based on color information. The expecting result of the extracted target area is that it contains a series image element of the same property. The data gotten by color layering has regularity in a certain sense, thus simplifying the subsequent processing.

There are mainly two methods for color layering-threshold and clustering. In the satellite cloud image, the universal similarity between the cloud regions makes, it difficult to distinguish the boundary of the clouds. In practice, there are always fusions and deformations of the clouds. And a commonly used approach to weather analysis is the water vapor distribution diagram which is similar to the contour plot. So this paper selects the method based on threshold to analyze the color layering.

In the satellite cloud image, the system mainly pays attention to the area of the cloud image. But there is no appropriate parameter for the cloud image in the existing standard color space for choosing the reference data of the color processing. Therefore, the first problem to be solved is the parameter selection in view of the cloud image.

## 2.2. Parameters selection in view of the satellite cloud image

Color space is to describe color using mathematical method. Historically there are various color space resulted in different applications. RGB is a color space built by the theory of the three-primary colors—red(R), green (G), and blue (B) based on human vision. It holds that the proper mixture of the three colors would generate the perception of all colors of the electromagnetic spectrum. HSV, however, is a color space based on human visual perception characteristics, of which Chroma (H) represents different colors, saturability(S) represents the depth of the color, and luminance (V) represents the lightness of the color. YUV is the color space used by PAL and SECAM, of which luminance(Y) represents grey value of image, hue (U, V) describes the properties of the image color and saturability. The conversion among a variety of different color spaces is needed according to practical applications.

The conversion formula from RGB space to HSB space is:

$$h = \begin{cases} 0^\circ & \text{if } \max = \min \\ 60^\circ \times \frac{g - b}{\max - \min} + 0^\circ & \text{if } \max = r \text{ and } g \geq b \\ 60^\circ \times \frac{g - b}{\max - \min} + 360^\circ & \text{if } \max = r \text{ and } g < b \\ 60^\circ \times \frac{b - r}{\max - \min} + 120^\circ & \text{if } \max = g \\ 60^\circ \times \frac{r - g}{\max - \min} + 240^\circ & \text{if } \max = b \end{cases} \quad (2.1)$$

$$s = \begin{cases} 0 & \text{if } l = 0 \text{ or } \max = \min \\ \frac{\max - \min}{\max + \min} = \frac{\max - \min}{2l} & \text{if } 0 < l \leq \frac{1}{2} \\ \frac{\max - \min}{2 - (\max + \min)} = \frac{\max - \min}{2 - 2l} & \text{if } l > \frac{1}{2} \end{cases} \quad (2.2)$$

$$l = \frac{1}{2}(\max + \min) \quad (2.3)$$

The conversion formula from RGB space to YUV space is:

$$Y = 0.30R + 0.59G + 0.11B \quad (2.4)$$

$$U = 0.493(B - Y) \quad (2.5)$$

$$V = 0.877(R - Y) \quad (2.6)$$

### 2.2.1. The Generation Algorithm of the Reference Image for the Cloud Image

In view of the cloud image, the present standard color space is not able to well-perform the satellite cloud image features. By observing the satellite cloud image, we can see that the generally cloud images appear to be white and the thin cloud images grey white. As a result, this paper considers using luminance image as reference images of clouds image. But it also contains the brightness of the land information, so it also need to combine other images as reference images. Through further observation, it can be found that the cloud image has two characteristics: high brightness, low saturability. In view of the cloud, the system uses the difference between luminance and saturability as the parameter. As shown in formula 2.7:

$$Y(\text{parameter}) = V(\text{luminance}) - S(\text{saturability}) \quad (2.7)$$

And then the generation algorithm of the reference image is proposed, and the detailed steps of the reference image generation algorithm are as follows:

- (1) Select each point in sequence in the image as the operating point.
- (2) Get the brightness parameter of the point according to the formula 2.4.
- (3) Get the saturability parameter of the point according to the formula 2.2.
- (4) Calculate to get the parameter of the cloud characteristic according to the formula 2.7.
- (5) Repeat these steps until all the points in the image are operated.

### **2.3. Optimization for the Feature Image of Satellite Cloud Image**

#### **2.3.1. Histogram Equalization for the Feature Image**

We can discover from the system generated reference cloud feature image that the light area of the cloud feature image was not all the cloud images. The system considers that the high-light area of the cloud feature image can be used as the cloud image and the middle-light area of the cloud feature image may not necessarily be the cloud image. So the system has reserved the high-light area and weakened the middle-light area, it is called the optimization for the satellite cloud feature image. The main idea of the algorithm is to make the image information distribute evenly between 0 and 255 –that is to stretch the primary data horizontally to get an ideal cloud feature image.

The specific steps of histogram equalization algorithm for feature image are as follows:

- (1) Count the number of every value in the image.
- (2) Find the maximum and minimum value.
- (3) Calculate the difference value between maximum and minimum.
- (4) Calculate the offset between each value and the minimum.
- (5) Calculate amplification coefficient, that is 255/difference.
- (6) The result of the offset multiplied by the amplification coefficient as the transformation applied to each point in the image.

#### **2.3.2. Denoising Processing for the Feature Image**

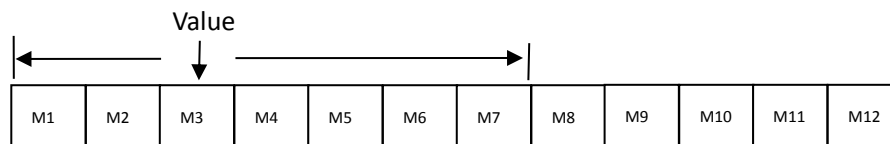
After obtaining the cloud feature image of histogram equalization, we can find that the unduly weak false area in the image was well-decayed and the area including larger amount of information is fully expressed, so it can be considered that the obvious feature of cloud images is obtained. It's important to note that the elements whose data is zero should be ignored, that is to say the histogram can't include the elements whose data is zero. Because it can be found in the actual processing that the region whose data is zero is generally the meaningless background area and occupies a large part of the image.

Because of all kinds of noise in satellite cloud image, the system only counts the value which is easily identified from macroscopic view, to eliminate the noise effects on extremum selection. Because the part which is easily identified includes points occupying a large part of the image. The system counts all identifiable target points and then takes the average value of the numerical domain as a threshold. By looking for the maximum and minimum greater than the average value, the noise is excluded.

## 2.4. Threshold Selection

There are two ways to choose the threshold-the semi interactive mode and fully automatic mode and this paper adopts the latter. In the selection of the threshold, the system usually hopes that the desired target can be separated and the undesired part can be hidden by the selected threshold. Therefore, the histogram of the image has a good reference function. The main idea of the automatic threshold [10] selection is to select the macro curvature maximum point in a certain sense in the histogram. In practice, the system selects the local minimum value as the curvature maximum point. Because of the strong jitter of the histogram, the selection of local minima becomes meaningless. Therefore, the histogram should be smoothed before selecting the local minimum to eliminate the jitter on macro. Here, the system uses four-neighborhood smoothing method to smooth the histogram.

Four-neighborhood smoothing algorithm is that selecting separately four unit length forward and backward for each point in the histogram, constituting a total of nine data sets and then calculating the average value of the data set as the value of the point. The points which exceed the boundaries of two sides are rejected.



**Figure 1. The Diagram of Four-NeighborhooD Smoothing Algorithm**

The system uses the point in position 3 in Figure 1 as an example and selects forward four unit length's area. Because the starting position is in position -1, while the histogram of the system starts from position 1, so only selecting forward two points in position 1 and 2 to add to the data set, without considering the excess part. The histogram of the system ends up with position 255 and when the system selects backward four unit length's area, the final position is in position 7 which does not exceed the range, so selecting backward four points from position 4 to position 7. Together with the selected data point, the system selects seven data points in all from position 1 to position 7 to calculate. Calculate the average value of the seven points in the data set and then assign the average value to the point in position 3 in the new histogram.

After smoothing, the original histogram becomes smoother than ever thus, curbing the jittery effect and makes it feasible and useful to select the local minimum as the curvature maximum in the macro. The minimum of the array is obtained by means of statistics and iteration to update the known minimum in a set of data, which is the usual means by which system selects the minimum. However this method is not appropriate at this time. The curves in the graph are mainly composed of two parts which are increasing part and decreasing part. The main idea of the selection of local minimum is to exclude the increasing and decreasing parts and the remaining part is the expected local minimum. The judgment is made by the minimum position of the sequence, whether it is an increasing sequence or a decreasing sequence. If a set of data sequences whose minimum is located at the starting position, they are increasing sequences and if a set of data sequences whose minimum is located at the ending position, they are decreasing sequences.

The system uses a fixed length window to detect the local minimum in the histogram, and the system selects 15 as the width of the window. When a window is selected, the minimum position of the data set is obtained by the window firstly, and then to see if the position is at the beginning or in the end of the window. If so, jump straight to the next window (using the ending position of the present window as the starting position of the next window). If it is not, then the system will find a local minimum. After the system

recorded the minimum position, the position of the minimum is used as the starting position of the next window to start detecting the next window, until the local minimum of the histogram is detected.

### **3. Automatic Vectorization of the Satellite Cloud Image**

The edge detection of the image [11] is an important method of image processing which can make the image contour more prominent and it is a kind of important region processing technology. Because the feature points to be extracted are located in the edge of the image, so edge detection technique is used to obtain the edge of the image based on which the feature points of the image can be determined.

#### **3.1. Edge Detection based on Binary**

Now edge detection technology has been more mature, and the key in the technology is to determine the detection criteria. The simple edge detection algorithm is to judge whether the current point is at the edge. That is whether there exist target points and background points in the eight-neighborhood at the same time. The four-neighborhood  $N4(r, c)$  of pixel  $(r, c)$  includes pixels  $(r-1, c)$ ,  $(r+1, c)$ ,  $(r, c-1)$ , and  $(r, c+1)$ , which are often referred to as its north, south, west, and east neighbors, respectively. The eight-neighborhood  $N8(r, c)$  of pixel  $(r, c)$  includes each pixel of the four-neighborhood plus the diagonal neighbor pixels  $(r-1,c)$ ,  $(r+1,c)$ ,  $(r,c-1)$ , and  $(r,c+1)$ , which can be referred to as its northwest, northeast, southwest, and southeast neighbors, respectively. The main idea of the algorithm is that making statistics on the value of the eight-neighborhood of the current target pixels whose value is one, if there is more than one point in the eight-neighborhood different from the current point, then the current point is at the edge.

The edge line could be detected by this method, but there exists redundant edge points. After analyzing all kinds of possible situations, this paper presents another algorithm, of which the speed is two times faster than ever and the algorithm can effectively avoid the occurrence of the above mentioned problems.

However, for satellite cloud image, the edge detection based on information can't get a clear definition of the edge, so the edge detection based on threshold is proposed in this paper.

#### **3.2. The Edge Detection based on the Threshold**

The basic rule of the edge detection algorithm based on the threshold is that if a certain pixel (including itself) whose surrounding not only has a value bigger than the threshold but also has a value smaller than or equal to the threshold, it is considered that the pixel is the edge. The element edges produced by this algorithm are all closed, and the background edges can be well detected. In the actual process, the definition of a certain point surrounding can be four-neighborhoods or eight-neighborhoods. Four-neighborhood computes fast and gets the sparse edges which are the key edges and the eight-neighborhood computes slow, but the gotten edges is much thicker. In this paper, we select the four-neighborhoods as the definition of the point surrounding.

The specific steps the of edge detection algorithm based on the threshold are as follows:

- (1) Search the first point whose value is above the given threshold.
- (2) Detect if there is a point in the four-neighborhood whose value is under the given threshold.
- (3) If the step (2) is set up, the point is marked on the new image, else carry out the step (4).

(4) Search the next point whose value is above the given threshold, repeat the step (2) until the point whose value is above the given threshold could not be found.

### 3.3. Freeman Chain Code

Freeman chain [12] is an encoding representation method of boundary, using the boundary direction as the basis for encoding, for the simplified boundary description and it generally describes points set of the boundary. The purpose of the freeman chain code is to pick up the edges one by one, and then store them in order in some arrays whose structure is similar to chain. When the system needs to access a certain edge, it can easily get the point data of the edge in order. First, the system needs to select the edges one by one, so as to turn the information attached to the grating image into some recognizable elements. In this paper, the contour tracing algorithm is used to complete the operation.

The characteristics of the contour tracking is that it can only act on one edge at a time and the tracking edge must be continuous and uninterrupted. The main idea of the contour tracking algorithm [13-14] is to look for the next edge point according to the closure and continuity of the edge based on the current edge points. The eight-neighborhood used this time is different from before and it needs to label the eight-neighborhood respectively according to the order of the clockwise or counterclockwise.

The disadvantage of the algorithm is that it can only detect the edge of one connected graph at a time. But in practice the contour boundary of the actual image is the close and disjoint 2D curve, and there are many disjoint contour lines on the same level layers. In this paper, we propose a modified algorithm to solve the multi object problem.

The specific steps of the modified algorithm are as follows:

- (1) Scan the whole image from the beginning line by line and search the first target boundary point as the starting point.
- (2) Search the next point of the starting point according to the eight-neighborhood search rule, take the adjacent point as a current point and continue to search
- (3) That the search returns to the starting point indicates the searching through a closed 2D curve is completed. Record the starting point and boundary chain code of the contour line, remove this contour line from the image, repeat steps (1) and search the next contour line until search through the contour lines of the whole graph are finished.

Since the edge line generated previously is ideal, generally the direction of the search rules is clockwise or counter clockwise, but a lot of redundant search is found after analyzing the composition of the actual edge line. In the actual edge, the main object is straight line and, curve and there are few inflection points. In the clockwise search process, the search order is first inflection point, then curve, and finally straight line. Therefore, we change the search order to straight line first, then curve, and finally inflection point.

The system has gotten the raster pattern which recorded the edge information of which the position belonging to the edge in the raster pattern is identified by nonzero value. When the system carries on contour tracking for the edge of the raster pattern, it gets the position information of edge points in order. When the track for one edge is finished, the edge will be wiped off, so that it will not affect the next edge tracking. In practice, the edge erasing is performed simultaneously with the contour tracking, that is, when tracking a point, the label for the point is set to zero. The system picks out the starting point of the edge based that this point is identified as the edge and does not belong to the previous detected edge. When the searching for the whole image is finished, the original image recording the edges is completely deleted. The system finally gets the sequences of edges.

### 3.4. Corner Point

Usually, the points gotten by edge detection is plenty, it would require too much computing to curve fitting all the points representing the edges. In order to reduce the amount of computation, it is a must to sample the boundary and the feature points [15] on the boundary are selected to represent the boundary.

#### 3.4.1. Traditional Corner Detection

In the digital image edge curve, the corner point exists where the edge direction is changed sharply, so that the feature point can be judged by calculating the curvature of each point. In this paper, we reference the SUSAN algorithm whose central idea is the using of a circular template to match the target pixels, and calculate the number of target pixels in the template, according to the number to determine the type of target point.

SUSAN operator method is simple and effective. But the prototype of the algorithm is based on a single target image and sparse target image, so for the image which needs to consider the order of the corner points and carry on curve fitting, the algorithm can't play a role perfectly. After the edge chain code from the contour lines of the image is obtained, the method based on boundary chain code is chosen to extract the feature points of the boundary.

#### 3.4.2. Corner Detection based on Edge Data Chain

The biggest difference between the corner detection based on edge data chain [16] and the grating corner detection is to make statistics on the number of target points in the neighborhood. The edge chain only includes the position information of the edge, all the information of the target points is filtered, and thus it is difficult to make statistics on the points that do not exist. So the system selects translation machine to solve practical problems.

As shown in Figure 2, it is a circle in a matrix among which X represents certain point, 0 represents the region outside the neighborhood of the point, 1 represents the region inside the neighborhood of the point, 2 represents the edge of the neighborhood of the point.

0	0	2	2	2	0	0
0	2	2	1	2	2	0
2	2	1	1	1	2	2
2	1	1	X	1	1	2
2	2	1	1	1	2	2
0	2	2	1	2	2	0
0	0	2	2	2	0	0

**Figure 2. The Structure Diagram of SUSAN Operator**

In Figure 2, there are a total of 37 points, including 24 points in the edge of the area. According to the idea of circular area statistic algorithm, if the X point on the edge, the edge is a straight line through X, and it must have the intersection with the boundary of the region. Through the analysis, it can be found that the distance from edge to the center point is about two or three pixels, mapping to the edge data chain, the area which is about two or three pixels apart from the center point must have the intersection of area edge and the target image edge. Therefore, the system only needs to detect two to four points to determine the intersection of the two region edges with the target edge. These two points

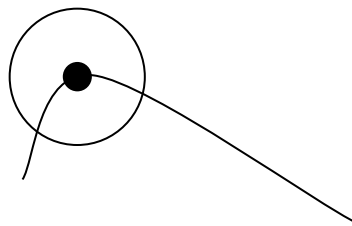


will always meet certain vector relationships relative to the X. In order to facilitate the analysis, specify the coordinate of x is (0,0), unit distance is 1, the right and the up direction is positive and vector diagram of SUSAN operator is generated, as shown in Figure 3.

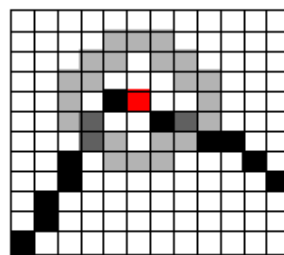
-3, 3	-2, 3	-1, 3	0, 3	1, 3	2, 3	3, 3
-3, 2	-2, 2	-1, 2	0, 2	1, 2	2, 2	3, 2
-3, 1	-2, 1	-1, 1	0, 1	1, 1	2, 1	3, 1
-3, 0	-2, 0	-1, 0	0, 0	1, 0	2, 0	3, 0
-3, -1	-2, -1	-1, -1	0, -1	1, -1	2, -1	3, -1
-3, -2	-2, -2	-1, -2	0, -2	1, -2	2, -2	3, -2
-3, -3	-2, -3	-1, -3	0, -3	1, -3	2, -3	3, -3

**Figure 3. The Vector Diagram of SUSAN Operator**

When the corner detection is performed on certain edge in the image, the SUSAN operator method shown in Figure 2 is applied to the point to be detected on the edge. You can get two reference points on the dark edge. As shown in Figure 4, it is a schematic diagram for the idea of the corner detection among which the black dot at the center represents the edge points to be detected.



**Figure 4. The Schematic Diagram of Corner Detection**

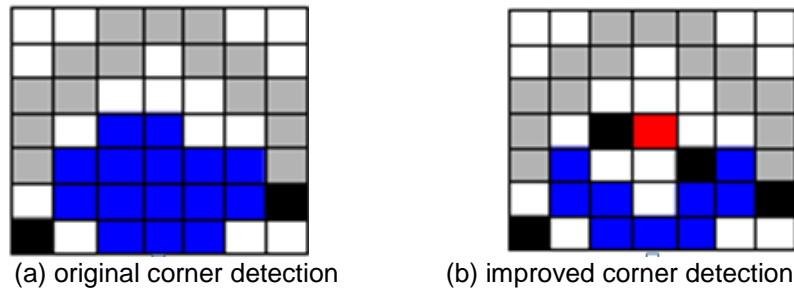


**Figure 5. The Detection Region Image after Grating Amplification**

Figure 5 shows the detection region image after grating amplification among which red dot represents the point to be detected, the light gray part represents the edge region of the SUSAN operator, and the black part is the edge of the image, and the dark gray part is the intersection of the SUSAN operator edge with the image edge which is the overlapping part.

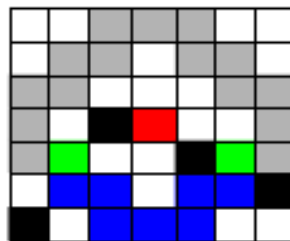
In the original corner detection method, the area of the region that is surrounded by the image edge and the edge of the SUSAN operator template should be gotten that is the area of the blue region in Figure 6(a). However, because the processed image only includes the edge information, there is no target point's information and background point's information, it is very difficult to make statistics on Figure 6(a). Therefore, the method

making statistics on the edge SUSAN operator shown in Figure 6(b) is proposed. In Figure 6(b), the blue area to be counted is less than 6(a). When the statistical method based on image position implements, it can reference the contour tracking algorithm, but this kind of method is too slow. Therefore, the paper proposes a translation mechanism based on the look-up table.



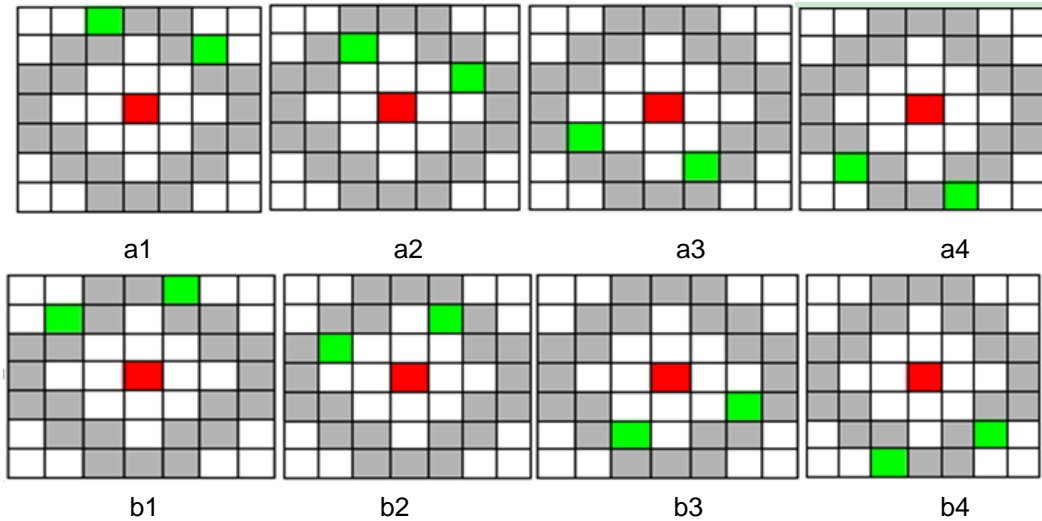
**Figure 6. The Diagram of Detection Region**

The translation mechanism select the intersection of two edges as the parameters, the principle of selection is to give priority to the selection of the intersection close to the target point on the edge of the edge chain, as the green dot shown in Figure 7. Combining the Figure 3, we can get the coordinates of the two green points relative to the red dot at the center in Figure 7. The left is (-2,-1), the right is (2,-1). In the process of searching for a reasonable solution, this paper proposes a method that using the vector difference of two intersections as reference. It is possible that the result is the same when use the negative vector difference and the positive vector difference respectively, so the method of using the absolute value of the vector difference as a reference is proposed.



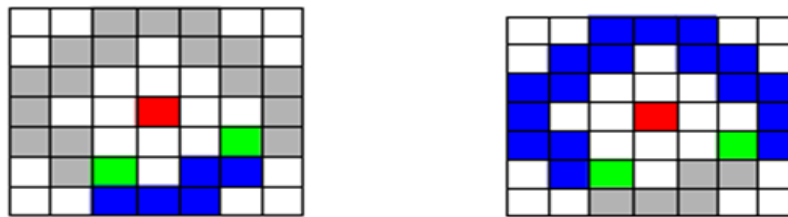
**Figure 7. The Diagram of Detection Structure**

Take coordinate (3, 1) as example, the layout may appear shown in Figure 8. Using coordinate (3,-1) and coordinate (-3, 1) to search, we can get the four cases of the group a. Using coordinate (3, 1) and coordinate (-3,-1) to search, we can get the four cases of the group b. Comparing the four cases in group a with group b, it can be seen that the actual real distribution only has two cases which is represented by a1 and b1 respectively. The remaining six cases can be obtained by the means of symmetry and rotation of a1 and a2, and its structure is consistent with a1 and a2. For example, b2 is obtained by the horizontal flip of a2 on the vertical axis.



**Figure 8. The Different Cases with the Same Vector Difference**

Because the shorter edges are more meaningful when judge the corner point, so when make statistics on the results, use the shorter edge as a result, as shown in Figure 9(a), and not use the longer edge as a result, as shown in Figure 9(b). As a result, the data of Table 1 is obtained by taking account of all the cases.



(a) Using the Shorter Edge as the Result      (b) Using the Longer Edge as the Result

**Figure 9. The Selection of Edge**

**Table 1. The All Results**

	0	1	2	3	4	5	6
0	0	1	2, 4	5	6, 8	9	10, 12
1	1	2	3	4, 6	7	8, 10	11
2	2, 4	3	4	5	6, 8, 12	9, 11	12
3	5	4, 6	5	6, 10	7, 9	10	
4	6, 8	7	6, 8, 12	7, 9	8, 12		
5	9	8, 10	9, 11	10			
6	10, 12	11	12				

Table 1 lists all the possible results for the reference values where the row represents the transverse absolute value for the vector reference, the column represents the longitudinal absolute value for the vector reference, the content of the table indicates the number of the points on the edge when the absolute value of the vector difference of the two points' intersection is relative to the value of the table. That is the blue part shown in

Figure 9. The practical significance is the needed number of steps from one intersection to another. The oblique line represents the meaningless or inexistent cases. Because the parameter is the absolute value, the vector difference can be expressed four cases (X, Y), (-X, Y), (-X, -Y), (X, -Y). Because of the symmetry of the operator template in the horizontal and vertical, the above four cases can be classified as one condition.

Observing Table 1, two important characteristics will be found, the first one is that the table is symmetric in the diagonal direction, the second is that there are multiple values in the table. Aiming at the first characteristic, the table is simplified as shown in Table 2.

**Table 2. The Simplified Results**

	0	1	2	3	4	5	6
0	0	1	2, 4	5	6, 8	9	10, 12
1		2	3	4, 6	7	8, 10	11
2			4	5	6, 8, 12	9, 11	12
3				6, 10	7, 9	10	
4					8, 12		

Table 2 removes the repetitive part for symmetry in Table 1, gets a valid data area similar to the lower triangle where the data is not repeatable. For multi-valued case, a separate discussion is conducted in process. So at first, these multiple values are individually numbered so as to make it convenient to search when it is needed. System chooses to take the negative value of the order of appearance as the label and uses the specific negative value to label the inexistent and meaningless value.

**Table 3. The Translation Table**

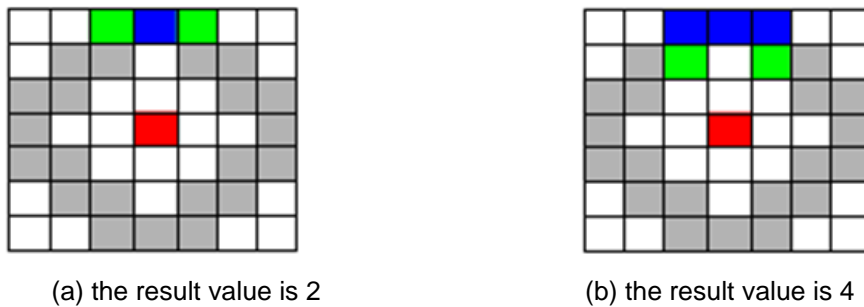
	0	1	2	3	4	5	6
0	0	1	-1	5	-2	9	-3
1	-99	2	3	-4	7	-5	11
2	-99	-99	4	5	-6	-7	12
3	-99	-99	-99	-8	-9	10	-99
4	-99	-99	-99	-99	-10	-99	-99

In Table 3, -99 represents the inexistent and meaningless value, from -1 to -10 represents multiple values, so further analysis is needed. The positive value indicates the number of points on the neighborhood edge which is in the target area and they are directly available.

**Table 4. The Translation Result with Multiple Values**

-1	-2	-3	-4	-5	-6	-7	-8	-8	-10
2	6	10	4	8	6	9	6	7	8
4	8	12	6	10	8	11	10	9	12
					12				

Table 4 lists the labels that need to be further discussed and the possible results. Since the data is not repeated, it is necessary to consider the twenty-one cases separately. In the process of the specific segmentation from the label to the translation result, the final result will be gotten by combining the unique characteristics of all kinds of situations. That is to say, when it is unable to give a unified judgment standard, do independent judgment for each label. Take the -1 label as example to see the practical and realized case of the system. The absolute value of the vector difference corresponding to the -1 label is (2, 0), and the results obtained respectively are 2 and 4, as shown in Figure 10(a) and 10(b).



**Figure 10. The Different Cases with the Same Vector Difference**

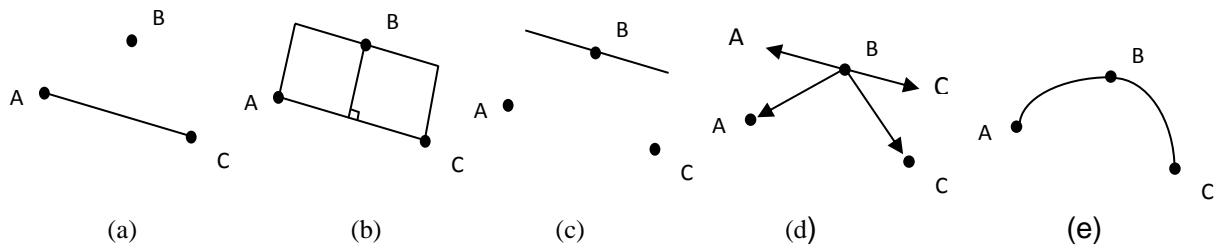
Observing the difference between 10(a) and 10(b), the coordinates of the two intersections in figure 10(a) are (-1, 3) and (1, 3), the coordinates of the two intersections in figure 10(b) are (-1, 2) and (1, 2). From this, it can be concluded that the difference between (a) and (b) in Figure 10 is that the position of the vertical coordinate of intersections is different. Because there exists swap mapping of the absolute value of the vector difference, we need to consider the case that figure 10 rotates 90 degree. That is the coordinate of the intersection are (3, 1), (3,-1) and (2, 1), (2, -1) respectively.

The SUSAN algorithm detects the number of targets in the neighborhood, but the number of the adjacent edges in the target area is detected there, which is the length of the arc. If the intersection between the edge and the field is A and B, then the SUSAN algorithm detected the area of sector AXB, and the length of the arc AB is detected by the system. After the system analyses the experiment, it is found that choosing the arc as judge basis is much simpler and more efficient.

### 3.5. Curve Fitting Based on Feature Points

The boundary fitting of the feature points could be carried out after obtaining the edge feature points. Curve fitting is the process of constructing a curve that has the best fit to a series of data points, possibly subject to constraints. Curve fitting can involve either interpolation, where an exact fit to the data is required, or smoothing, in which a “smooth” function is constructed that approximately fits the data.

The state-of-the-art methods of curve fitting are based on B-spline. Considering that there is a wide variety of B-spline curve fitting methods and these methods are so complex, the system doesn't use them. The simplest method to fitting data is linear interpolation. The result obtained by direct interpolation is a straight line. However, the fact is that the feature points in images are not in a straight line. So we choose one more point to control curvature. The idea of the three-point-control method is that we choose three feature points A, B, C, in order, connect the point A and the point B, and bend the line AB based on the point B. In this way, we can generate the curve ABC.



**Figure 11. The Algorithm of the Three-Point-Control Method**

The algorithm of the three-point-control method is illustrated in Figure 11. Firstly, we get the gradient and the length of the line between the start point A and the endpoint C, which is illustrated in Figure 11(a). Then we translate the line AC to the point B along the direction of the vertical of AC, which is illustrated in Figure 11(b). In this way, we can get the parallel line of AC through B and the length of the parallel line is equal to the length of AC, which is illustrated in Figure 11(c). Finally, we can get the vector  $\overrightarrow{BA}$ ,  $\overrightarrow{BA}$ ,  $\overrightarrow{BC}$ ,  $\overrightarrow{BC}$ , which is illustrated in Figure 11(d). The arc line between the point A and the point B can be easily got from B to A, where  $V=B+K*+(1-K)*$ . Thus, we achieve the curve connection of the data. This algorithm is simple and can be easily implemented.

#### 4. Design and Implementation of the System

In this paper, we analyze and study the characteristics of satellite cloud image, and propose a series of algorithms, such as image preprocessing, color layering and vectorization processing based on satellite cloud image. A satellite cloud image vectorization system is developed on the matlab platform, and a comprehensive experiment is carried out by using multiple satellite images.

The conversion from raster image to vectorization image and the recovery from vectorization image to edge information are realized. The system has a small memory space, a hierarchical structure and independent element structure, a more flexible and more convenient features compared with the raster image.

##### 4.1. The Main Functions of the Vectorization Prototype System for Satellite Images are as Follows

- (1) Image layering of satellite cloud image.
- (2) Image preprocessing of satellite cloud image.
- (3) Threshold selection of satellite cloud image.
- (4) Edge extraction of satellite cloud image.
- (5) Corner detection of satellite cloud image.
- (6) Curve fitting of satellite cloud image.

Image vector system flow chart shown in Figure 12:

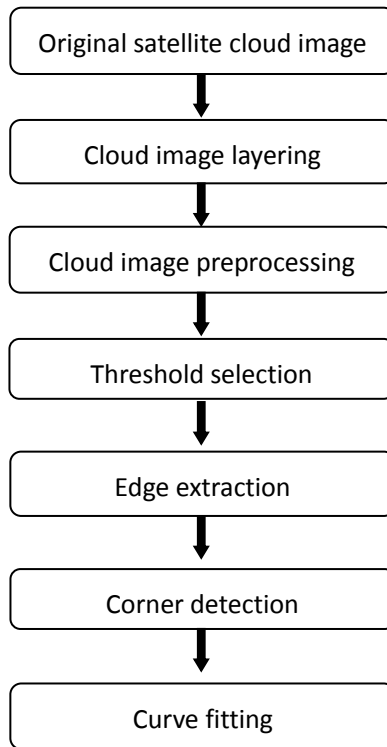


Figure 12. The System Flow Diagram

#### 4.2. Software Components of the System

The system is divided into three fundamental functional modules from the composition of components-the preprocessing module, color layering module and automatic vectorization module.

Preprocessing module is mainly denoising module, considering the vectorization effect, the denoising part is placed in the middle of the two color layering modules. Color layering module is divided into image layering module and threshold selection module. Automatic vectorization module is divided into edge extraction module, corner detection module, and curve fitting module.

The main interface of the system is shown in Figure 13:

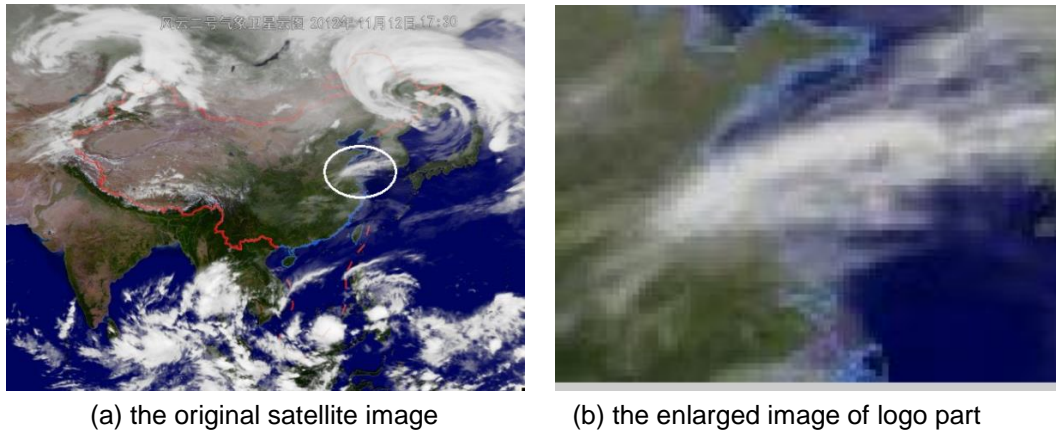


Figure 13. The Main Interface of the System

### 4.3. The Experimental Results of the System

We use a variety of satellite cloud images as the experimental data to test the functions realized in this system. Following is the satellite image with vectorization effect.

Figure 14 is the original satellite image, where 14(b) is the enlarged image of white circle denoted part in 14(a).



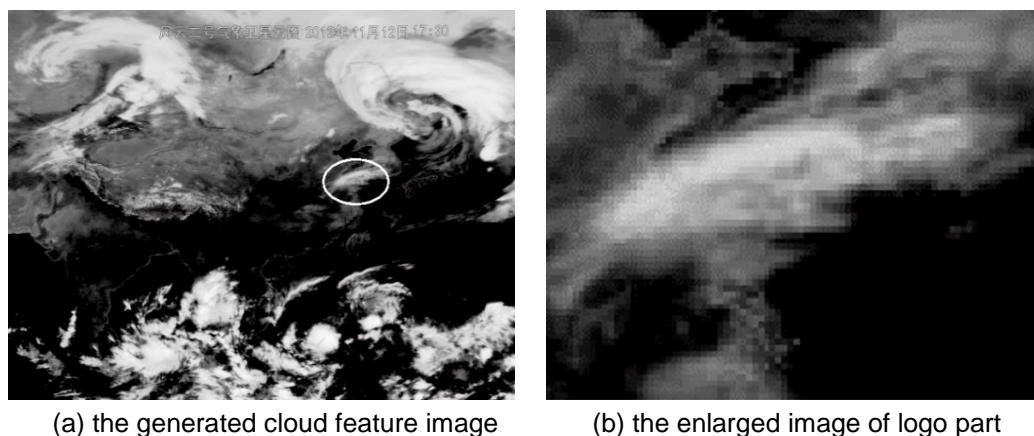
**Figure 14. The Original Image**

It can be seen from Figure 14 that the cloud graphics and the edge of the image is fuzzy and the cloud layer is not clear. First of all, we do color layering for the original image using the color layering algorithm which takes the difference value between brightness and saturation as a reference value. Color layering is the process of processing the image according to the color information, extracting the meaningful target area which mainly includes the cloud image segmentation, image preprocessing (denoising) and threshold selection.

In regard to the extraction method of cloud image, after analyzing several commonly used color spaces and satellite cloud images, it can be seen that the cloud images generally appear to be white, and thin clouds appear to be gray white. From this, we can consider the using of the brightness image as a reference image of the cloud image.

However, after a careful study of the luminance image, we find that luminance image is cannot well-express the cloud information, because it also contains the brightness information of the land. Therefore, in view of the cloud characteristics of high brightness and low saturability, the system takes the difference of brightness and saturability as a parameter.

Figure 15 is the generated cloud feature image which is taken as the reference image, where 15(b) is the enlarged image of white circle denoted part in 15(a).

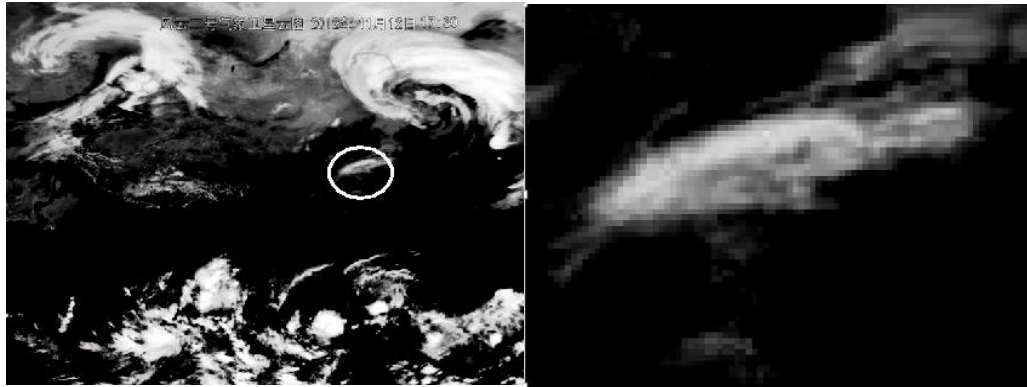


**Figure 15. The Generated Cloud Feature Image**



It can be seen from Figure 15 that the bright part of the directly generated cloud feature image is not all cloud images. The system considers that the high-light part can be treated as cloud image and the middle-light part may not necessarily be the case. So a processing method called the histogram equalization of the feature image, which is to retain the high-light part and weaken the middle-light part, is proposed.

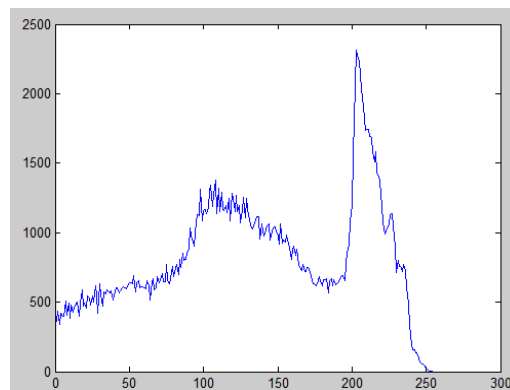
Figure 16 is the histogram, equalized cloud feature image, where 16(b) is the enlarged image of white circle denoted part in 16(a).



(a) Histogram Equalized Cloud Feature Image (b) The Enlarged Image of Logo part

**Figure 16. The Cloud Feature Image after Histogram Equalization**

We can see from Figure 16 that the excessively weak misjudgment area was well decayed and the area containing a large amount of information was fully shown, so it can be considered that the obvious features cloud image is obtained. Figure 17 is the original histogram, where horizontal ordinate represents the value in the image and longitudinal coordinate represents the number of the value in the image.

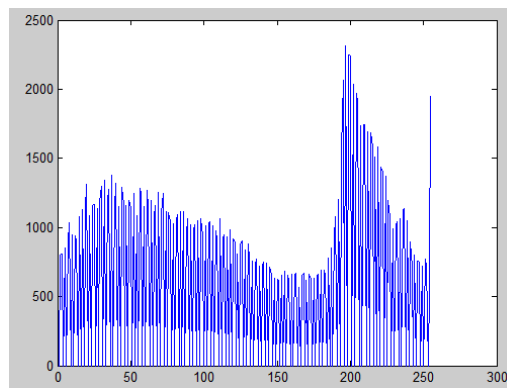


**Figure 17. The Original Histogram**

The total amount of effective points in the statistical graph is 211112, so the obtained threshold is  $211112/255$  and the result is 827.8902, so the effective maximum that is the first one found from the right of the graph whose value is bigger than the threshold value is obtained, it's 229. And the effective minimum that is the first one found from the left of the graph whose value is higher the threshold value is also obtained, it's 86. We need to find the maximum and minimum of the histogram to histogram equalize it.

In order to eliminate noise effects on extreme value selection, the system only makes statistics on extremum which is easily identified in macro. The points of the easily identified parts occupy a large proportion of image area. The system makes statistics on all the target points and takes their average value as threshold, thus eliminating noise.

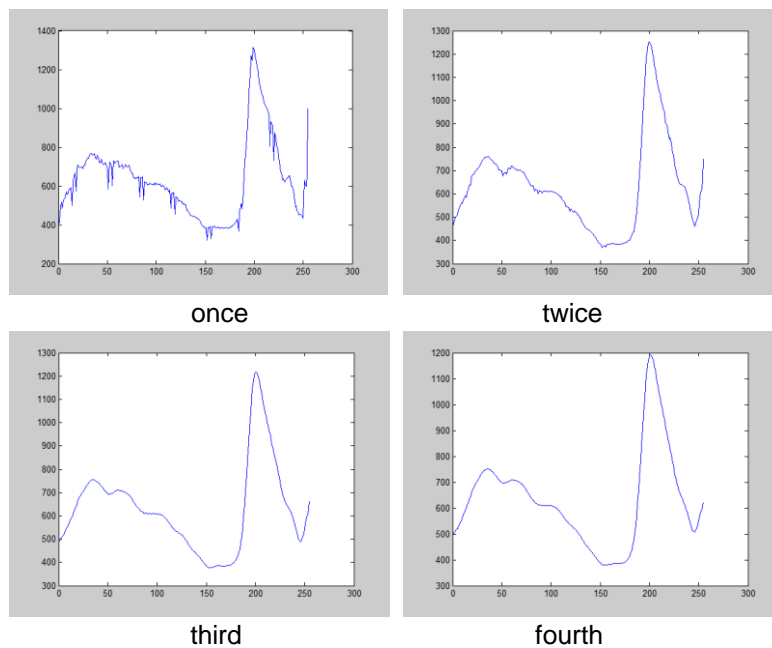
Figure 18 is an equalization histogram.

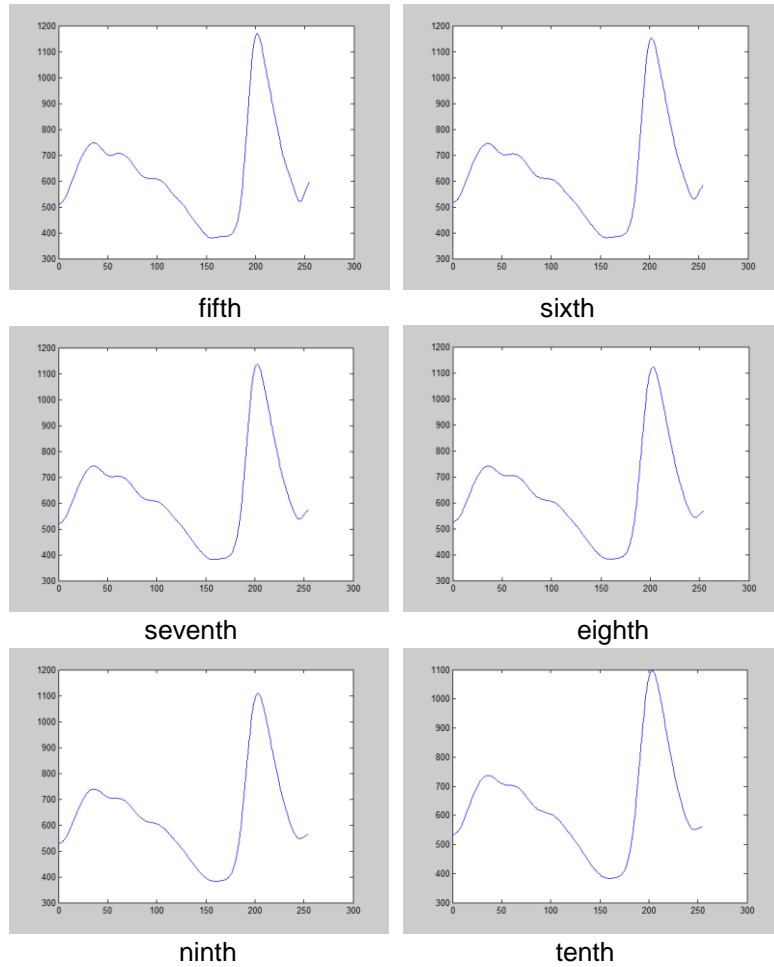


**Figure 18. The Equalization Histogram**

It can be seen that Figure 18 is obtained by the stretching the remaining part after removing the extreme values of both ends in Figure 17, thus effectively eliminating the influence of noise. The threshold selection is to complete the segmentation for different image blocks. For the optimized cloud images as shown in Figure 18, take 180 as the segmentation point, the two part of the region can be seen clearly. The central idea of selecting automatic threshold is to select the macroscopic curvature maximum point in the histogram in a certain sense. In the implementation of the system, the system selects the local minimum as the maximum curvature point. Due to strong jitter in histogram, we need to smooth the histogram to eliminate jitter in macro before the selecting of the local minimum.

Figure 19 is histogram after multi iterations and smooth operation.





**Figure 19. The Histogram after Multi Iterations and Smooth Operation**

From Figure 19, it can be seen that after smoothing for many times, the impact of the jitter is well controlled. Finally five minimums, 55, 99, 154, 78,245, are got. The system do color layering for the image according to the gotten minimums.

After finishing color layering the image, feature extraction for the image is needed. Feature extraction is to extract the features from the image to be used for fitting, it includes three parts-edge extraction, freeman chain code and corner detection.

Edge detection is used to obtain the edge of the image, but the edge detection based only on information cannot provide a clear definition of the edge, so the edge extraction based on the threshold is proposed.

In this paper, the adaptive threshold is obtained by generating the histogram and calculating.

The pixel is the edge if there are both values that bigger than the threshold and the ones smaller than or equal to the threshold around certain pixel (including itself).

Figure 20 is a superimposed image after the edge extraction using 61 and 184 as threshold respectively, where 20(b) is the enlarged image of logo part in 20(a).

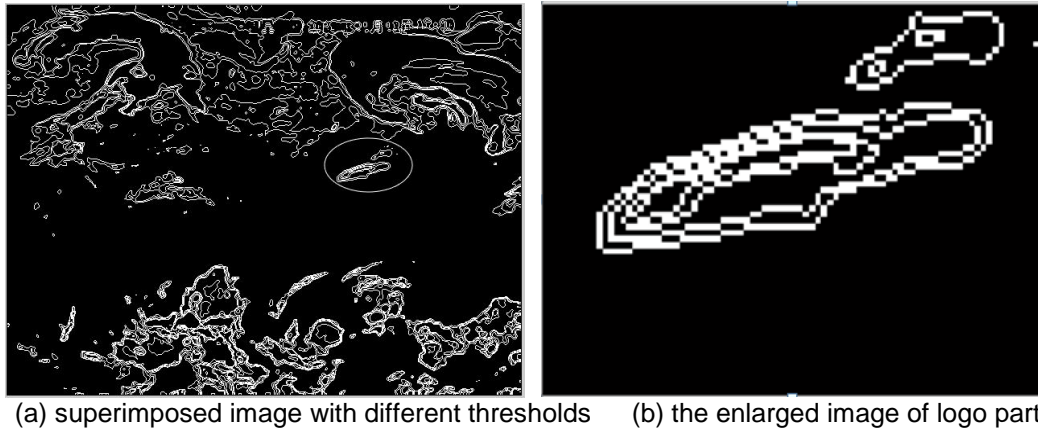


Figure 20. The Superimposed Image after the Edge Extraction

Compared with Figure 16, it can be seen that the expected edge is detected. But the points gotten by edge detection is vast, it will cost a lot of computing to use all the points to do edge fitting. In order to reduce the amount of computation, the boundary is sampled and the feature points on the boundary are selected to represent the boundary. Traditional corner detection method cannot meet the requirements, so we use the corner detection method which based on edge data chain. Figure 21 shows the corner points using 10 as the test value, where 21(b) is the enlarged image of logo part in 21(a).

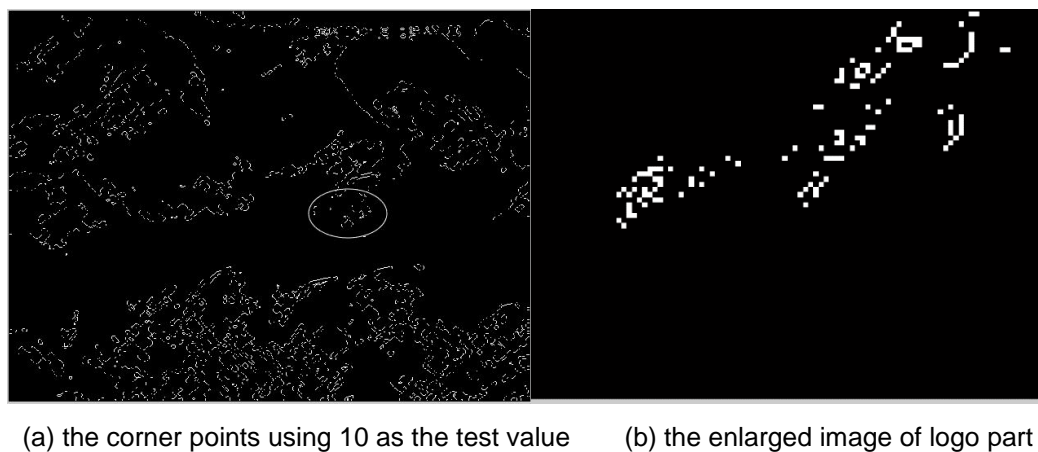
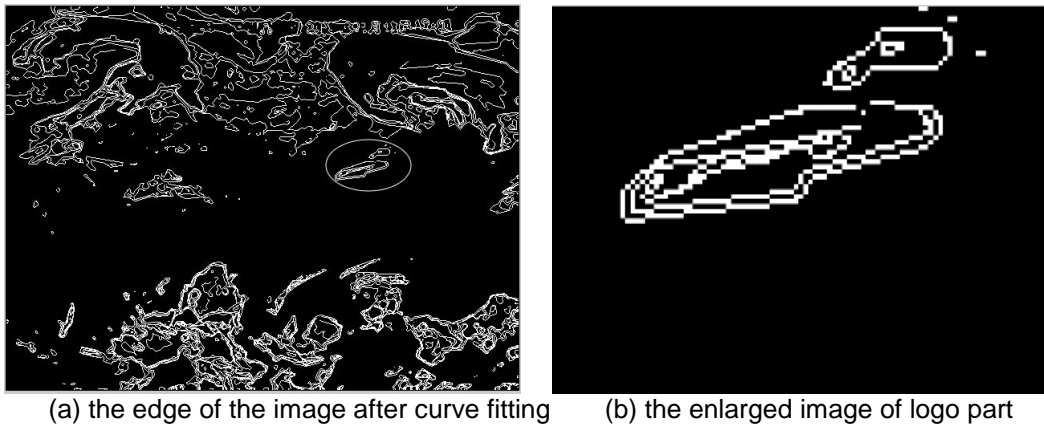


Figure 21. The Corner Points

From Figure 21 it can be seen that the feature points on the boundary are extracted very well. After obtaining of the feature points, we can carry out the boundary fitting of the feature points. In this paper, the three-point-control algorithm is used for the curve fitting. Figure 22 is the edge of the image after curve fitting, where 22(b) is the enlarged image of logo part in 22(a).



**Figure 22. The Edge of the Image after Curve Fitting**

After comparing with Figure 20, it can be seen that the application of three-point-control algorithm on curve fitting is very good and the boundary of the image is well fitted.

At this point, the system has completed the processing for satellite cloud image, which makes the boundary and hierarchical information of cloud layer in cloud picture clear and intuitive, so that the satellite cloud images changed from the qualitative subjective fuzzy identification into quantitative objective clear and accurate identification.

## 5. Conclusion and Prospect

This paper studies the implementation process of thought way and technology for satellite cloud image vectorization. According to the characteristics of satellite color image, using image processing technology, to study from aspects of the satellite cloud image, the image layer, threshold selection, edge detection and tracking, feature point extraction, curve fitting and so on.

Put forward an image layering algorithm which using the difference between brightness and saturation as reference images to realize the image layering for cloud image. Analyze the uncertainty of the edge of the satellite cloud image and an edge detection algorithm based on adaptive threshold is proposed.

Studying the advantages and disadvantages about the grating edge data and chain edge data and put forward the corner detection method based on edge data link. A vectorization algorithm based on satellite color image is proposed, after color layering, edge extraction, feature extraction and curve fitting, generate the final vector graphics. Developed automatic vector system about satellite color images, and the experimental results were carried out by using real satellite color image data. The feasibility of the method and the validity of the algorithm in this paper were verified. According to the extractive characteristics of color image and interested cloud area, the structure and evolution trend of color cloud picture can be analyzed, and provide effective basis for meteorological workers.

## Acknowledgment

This research was supported by the National Natural Science Foundation of China General Projects Grant No. 61272029.

## References

- [1] A. De and C. Guo, "An image segmentation method based on the fusion of vector quantization and edge detection with applications to medical image processing", *International Journal of Machine Learning and Cybernetics*, vol. 5, no. 4, (2014), pp. 543-551.
- [2] P. Barla and A. Bousseau, "Gradient Art: Creation and Vectorization. Image and Video-Based Artistic Stylisation", vol. 42, (2013), pp. 149-166.

- [3] G. Lobet and X. Draye, "Novel scanning procedure enabling the vectorization of entire rhizotron-grown root systems", *Plant Methods*, vol. 9, no. 1, (2013).
- [4] Z. Xianquan, T. Zhenjun and W. Jianjun, "Binary image boundary vectorization algorithm based on control points", *Computer Science*, vol. 1, (2009), pp. 169-170.
- [5] X. Guofu and S. Xin, "Hierarchical Diffusion Curves for Accurate Automatic Image Vectorization", *ACM Transactions on Graphics*, vol. 33, no. 6, (2014).
- [6] D. Wei and L. Yong, "Image vectorization based on mathematical morphology in geographic information system", *Applied Mechanics and Materials*, vol. 532, (2014), pp. 231-234.
- [7] Z. Jiaojiao and F. Jie, "Image vectorization using blue-noise sampling", *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 8664, (2013).
- [8] G. Wei, Z. Rong Cai, H. Lin, P. Jian Min and D. Rui, "Research on SIMD auto-vectorization compiling optimization", *Computer Software*, vol. 26, no. 6, (2015), pp. 1265-1284.
- [9] J. B. Mena and J. A. Malpica, "An automatic method for road extraction in rural and semi-urban areas starting from high resolution satellite imagery", *Pattern Recognition Letter*, vol. 26, no. 9, (2005), pp. 1201-1220.
- [10] C. Armenkis and F. Leduc, "A comparative analysis of scanned maps and imagery for mapping applications", *Photogrammetry&Remote sensing*, vol. 57, (2003), pp. 304-314.
- [11] A. Kosehan and M. Abidi, "Detection and Classification of Edges in Color Images", *Signal Processing Magazine, Special Issue on Color Image Processing*, vol. 22, no. 1, (2005), pp. 64-73.
- [12] S. Jingrong and X. Luping, "An improved graphics vectorization method", *Computer Engineering and Applications*, vol. 1, (2004), pp. 88-90.
- [13] R. Mingwu, Y. Jingyu and S. Han, "Tracing boundary contours in a binary image", *Image and vision computing*, vol. 20, no. 2, (2002), pp. 125-131.
- [14] C. Bin and Z. Shusheng, "Research Summary of Engineering Graphics Vector recognition technology", *Journal of Engineering Graphics*, vol. 6, (2008), pp. 7-11.
- [15] Y. Surong, Z. Guilin and Xucongfu, "A new method of bitmap vectorization", *Computer Engineering and Applications*, vol. 14, (2005), pp. 85-87.
- [16] Z. Xiaoli, W. Min and H. Xinhan, "An effective corner detection method based on Freeman", *Journal of electronic measurement and instrument*, (1999), pp. 14-18.

## Authors



**Liu Xumin**, was born in Liaoning, China in 1956. She took the B.S degree in computer application and M.E. degree in computer application from the Dalian University of Technology, in 1982 and 1993, respectively. She received the Ph.D. degree in School of Computer and Information Technology from the Beijing Jiaotong University in 2008. Currently, she is a professor at College of information Engineering, Capital Normal University. Her research interests include graphics and image processing and data mining.



**Zhang Cong**, was born in Shijiazhuang, Hebei in 1992. She received the B.S degree in Tangshan College, Hebei, in 2014. Her research interests include digital image processing, data mining and image processing.



**Na Sen**, was born in Nei Menggu, in 1991. He received the B. S degree in College of information Engineering, Capital Normal University, Beijing, in 2013. His research interest is digital image processing.



**Xu Weixiang**, is presently a Professor of School of Traffic and Transportation at Beijing Jiaotong University of China. He obtained his M.S degree in computer science from Dalian University of Technology of China in 1993 and PhD degree in System Engineering from Northern Jiaotong University of China in 2000. His current research interests focus on data mining, intelligent information processing and management, the analysis and integration for Transport systems, and their joint applications in engineering design, Rail and urban rail transit for knowledge discovery and management purpose.

