Research of Meteorological Facsimile Map Vectorization Algorithm*

Feng Lingling, Qi Guoqing

Information Science and Technology Collage, Dalian Maritime University, Dalian, 116026, China.
flingling_1122@126.com, qgq@dlmu.edu.cn

Abstract - An algorithm for extracting information from meteorological facsimile maps and transform the map into vector format is proposed. The algorithm contains two parts. The first one is annotation information extraction method based on the regular pattern of position and angle, the second part is isobars vectorization based on tracking and filtering. As the annotations change with latitude line, relevant characters can be extracted on the basis of removing the base map and then thinning the isobars. After removing noise and connecting broken isobars, seamless and smooth isobars are extracted by means of α-β filtering. Then extracted isobars are approximated by polygonal lines to compress the data of isobars. Experiments show that the proposed algorithm can extract annotation and vectorize isobars.

Index Terms - Meteorological facsimile maps, annotation, isobars, vectorization.

I . Introduction

The meteorological facsimile map is a kind of weather map which contains meteorological information. Ships at sea can receive real-time meteorological facsimile maps from countries near the navigation area. By knowing about the whether and sea condition that has occurred and will take place and more weather variations in a large scale in time, mariners can make decisions to avoid disasters and ensure the safety of maritime activities. The meteorological facsimile map is broadcast in dot matrix data format, so it is not convenient to process and display relevant information together with other navigation information. In order to make it easy to store and process, especially to overlay the weather information on the ECDIS(Electronic Chart Display and Information System), it is necessary to extract the useful information on the map and convert the map from the original raster format to vector format. Some algorithms have been put forward to extract useful information (isobars, isotherms, etc.) on the meteorological facsimile maps[1-4]. By doing so, the interested information can be added to the electronic chart, making it easier for users to be aware of the weather and sea conditions on the route. In this paper the algorithm of meteorological facsimile map vectorization is investigated. For the convenience of extracting the dynamic information, the static information contained in the map(such as latitude and longitude lines and coastlines for a specific area) are eliminated first, then the algorithm for extracting annotations(including numbers, letters and other symbols) and vectorizing isobars are proposed. As the annotations of isobars are located along the tangential direction of latitude lines, they can be extracted by the direction information. After the extraction of annotations, the remaining information of the map including isobars is processed by thinning and converted to binary format. To connect and track the broken isobars the least square tracking filter(α-β filter) is applied. After the extraction of seamless and smooth isobars, they are approximated by polygonal lines and denoted by the coordinates position of the turning points. Experiments show that the method proposed in this paper can extract annotations and vectorize isobars successfully.

II . Annotations Extraction

The meteorological facsimile maps used in this paper are the Asian ground analysis chart monitored by Japan Meteorological Agency, which can be downloaded from the Internet[5] as shown in Fig.1. The geographical scope covered by this kind of meteorological facsimile maps is fixed, so the latitude and longitude lines and coastlines are same in maps issued at different time(i.e., they are static). The static components in the maps are called base map in this paper and can be removed from each real-time map for the convenience dynamic information extraction and data compression. Fig.2 shows the base map extracted from 40 maps downloaded at different time. After removing the base map, the information left in the map is the dynamic component which needs to be processed. Since the gray level values of the map pixels do not provide extra information, the map is converted to binary format image by OTSU algorithm[6].

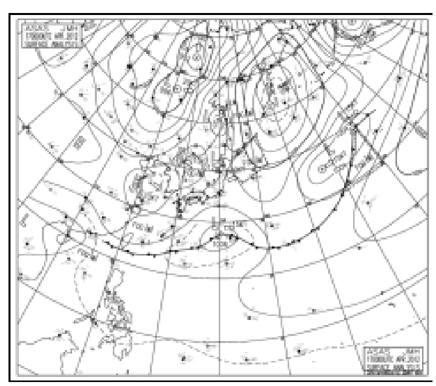

Fig.1 Meteorological facsimile map

* This work is partially supported by a grant from the National High Technology Research and Development Program of China (863 Program) (No. 2011AA110201)

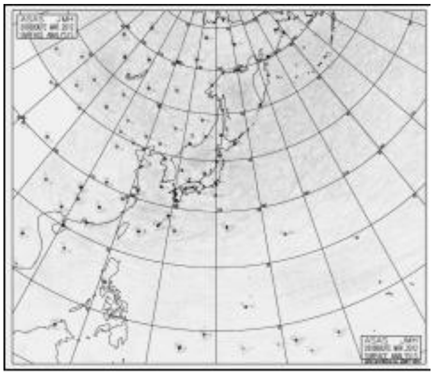

Fig.2 The base map of meteorological facsimile map

The binary image is shown in Fig.3 and the relevant base map is shown in Fig.4(since the complete map is too large, to show the details more clear, in the following parts of the paper we will only show a fraction of the map.).

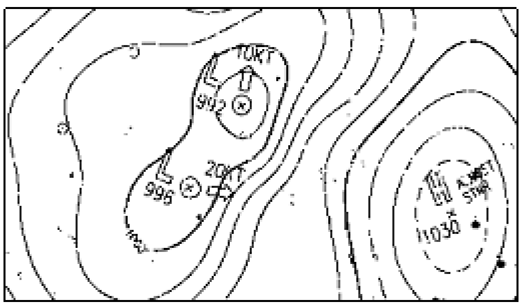

Fig.3 The original image after removing base map

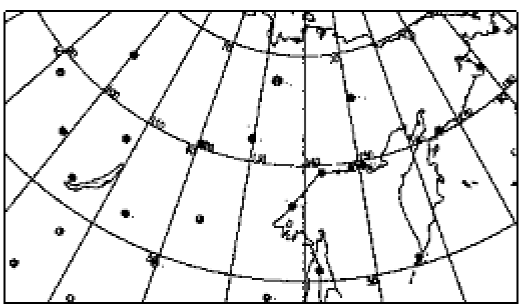

Fig.4 The base map of the original image

The latitude lines are concentric and the isobar annotations are along the tangent direction of latitude lines(i.e., they are not horizontal) and close to latitude lines. So the commonly used morphology method is not appropriate for extracting these annotations. The direction information is useful for annotations extraction. As shown in Fig.4, longitude lines can be extracted by Hough transform. The crossing point of the longitude lines is the center of concentric circles of latitude lines. Once the center is determined, the tangent direction along latitude line at any point can be obtained.

The size of each character for the annotations of isobar is fixed (6 * 11 pixels), the interval between two characters is 2 pixels. Thus, for example, a four-character annotation occupies 30 * 11 pixels. This indicates that in the area of the annotation exists more non-zero pixels than in other areas. According to this feature, the candidate annotation area can be detected in the following way. The preprocessed image is scanned by an appropriate size of a rectangular window. If the number of pixels in the current scan region is greater than the detection threshold, this region is detected as the candidate annotation region. Since annotations rotate with different positions, so the size of the rectangular area in this process is set to 24 * 24 pixels, the detection threshold we used in the experiment is 60. In the extraction results, parts of isobars that are close to annotation will be reserved. Since these lines are smooth and do not intersect with each other, so they can be removed using circle tracking method[7]. The principle of circle tracking method is shown in Fig.5. The intersection point of tracking circle and isobars is recorded while moving the circle every time, if the trace points exceeds the set threshold value, and the row vectors of the start and end point of the line or column vectors also exceed a certain threshold value, then the result can be identified as line and removed. So that the rough extraction of characters is accomplished. In the process of circle tracking, the radius of the circle is selected as 2 or 3 pixels, the length threshold is set to 4, the row vector difference threshold value is 10 and the column vector difference threshold value is 8.

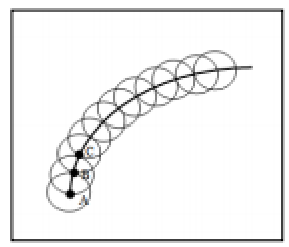

Fig.5 The principle of circle track method

The morphological operation, dilation, is carried out on the result image of rough extraction with a 3*3 cross structural element. Then a rectangular frame with a size of 30 * 11 with the center of every foreground-point is designed. The direction of rectangular frame is determined by the center of latitude circles and the current point. The width direction of the rectangle is parallel to the line which connecting the center of circles and the current point. Thus the rectangle is ensured to go along the tangential direction of the latitude line. As shown in Fig.6, the four vertices of the rectangle are gotten by the four lines which consist the rectangle. A new image is constituted with the number of pixels in every rectangle as the current point pixel values. The center of the annotation has the largest value within its surrounding area. Fig.7 shows the three - dimensional projection of an image with the size of 200 * 200 containing several annotations. After dilation, for annotations containing four characters(most isobar annotations are consist of 4 numbers), the rectangular area should be greater than 30 * 11, 330 pixels in total, while as for the three-character annotation, the rectangular area should be greater than 22 * 11, 242 pixels in total. Because after dilation the foreground points occupies most of the area of the rectangle, as shown in Fig.7, the peaks are more than 200, so processing the pixel above 200 sequentially as the center of annotation, the extraction results are shown in Fig.8.

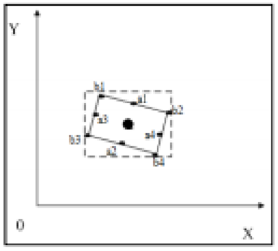

Fig.6 Determination of the center of annotation.

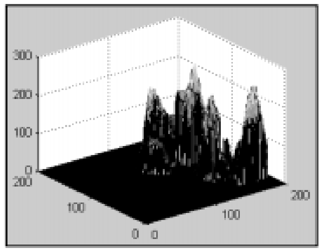

Fig.7 The three-dimensional projection

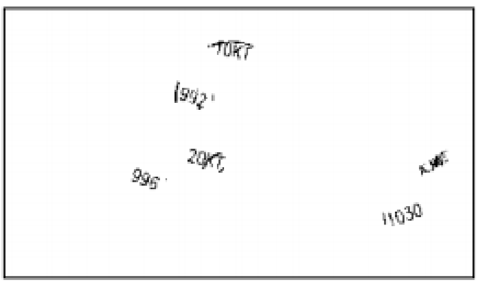

Fig.8 The extraction results of annotation

By character recognition technology, annotations can be recognized and coded for further application. The mark of high and low pressure also can be extracted in a similar way as annotations, so that interference can be reduced for isobar processing.

III . Vectorization of isobars

A . Extraction of isobars

Fig.9 shows the original image that has removed annotation information and the marks of high and low pressure.

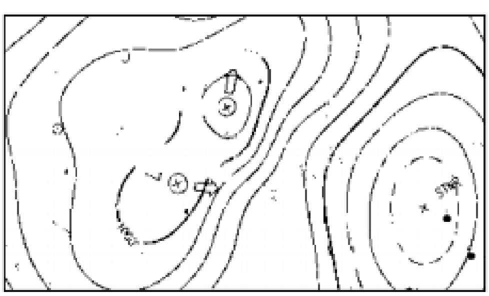

Fig.9 The meteorological facsimile map after the removal of annotation

The rest information of the image is basically isobars. The interference information (noise points and noise blocks) should be eliminated before further processing. Since these noise parts are in small size, the method of morphology can be applied. Specifically, close-operation can be done firstly, then corrosion. After scanning the foreground-points, the number of the background points in its area can be obtained. If its boundary is belongs to background and the scanned square is within a given threshold, then it is regarded as noise and eliminated.

In this paper, the operation of isobar vectorization is based on thinning. Therefore, before vectorization, isobars with width of one pixel should be extracted. There are many thinning algorithms, such as Pota algorithm, Rosenfeld algorithm, Pavlidis algorithm and the algorithm that based on mathematically morphology, etc[8-12]. Since there are no certain directions and definite bending degree of isobars, the thinning algorithm applied in this paper is different from the above ones. A filter is used to track the direction of isobars before thinning operation is carried out.

As shown in Fig.10, there are 8 directions departing from one point. Generally, the width directions of lines should be vertical to the directions of lines. In Fig.10, the black bold lines represent the width directions of every direction. In program, it is relatively difficult to describe the lean width directions points. Taking direction 2 as an example, as shown in Fig.11, direction 2 is 45 degrees, every small block represents a pixel, the width of the line is 3 pixels. The direction of the arrow is the line direction, and the pixels occupied by the arrow are the framework that needs to be kept. A significant feature can be observed from the width of isobars, that is the width directions can be simplified to vertical ones and horizontal ones. Actually, in Fig.11 the width direction can be assigned as horizontal direction. If it is processed line by line, it obeys the law that extracting the central line regardless of a few deviation of only 1 pixel. Direction 4, 6 and 8 are essentially processed in the same way as direction 2. Hence, there are only two width directions needed to be assigned, namely, the width direction 1 and 5 can be assigned as vertical direction, the rest can be assigned as horizontal direction. In the tracking process, the information needs to be recorded is the distance increment of the tracked point and former point and the distance increment of the former point and the point before that. If the two increments are both (0, 1) or both (0,-1), or one of them is (-1,1), the other is (1,1), or one is (1,-1), the other is (-1,-1), the width direction is vertical direction, or it is horizontal direction.

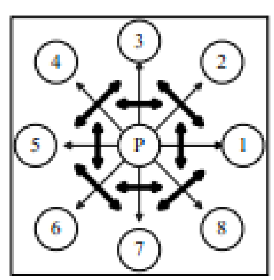

Fig.10 Line directions and width

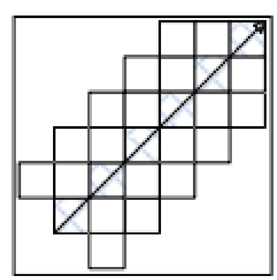

Fig.11 The relationship of the lean direction and the width

After thinning, isobars are already one pixel in width. However, there are some gaps in lines. For the sake of the following vectorization process, the extracted isobars should be continuous. In this paper, α-β filter is applied to extract isobars. In the process of filtering, the predicted value can be used to find the matched broken point, so that the lines can be connected.

The main equation of α-β filters [13] are as followed:

$$\hat{x}_n = 2(2n-1)x_n/[n(n+1)] + \{1-2(2n-1)/[n(n+1)]\}(\hat{x}_{n-1}+\Delta\hat{x}_{n-1}) \quad (1)$$

$$\Delta\hat{x}_n = 6[x_n-(\hat{x}_{n-1}+\Delta\hat{x}_{n-1})]/[n(n+1)]+\Delta\hat{x}_{n-1} \quad (2)$$

In the equation, x_n is the direction value of x-axis of the n point on the isobar, $\hat{x}_n$ is the direction filter value and $\Delta\hat{x}_n$ is the incremental position filter value of the n point on the isobar.

So that $\hat{x}_n$ and $\Delta\hat{x}_n$ can be recurred from $\hat{x}_{n-1}$ and $\Delta\hat{x}_{n-1}$, in the formula (1) and (2) set α_n and β_n

$$\alpha_n = 2(2n-1)/[n(n+1)] \quad (3)$$

$$\beta_n = 6/[n(n+1)] \quad (4)$$

α_n and β_n are filter coefficients which change with n. Thus, (1) and (2) can be expressed as

$$\hat{x}_n = \alpha_n x_n + (1-\alpha_n)\hat{x}_{n|n-1} \quad (5)$$

$$\Delta\hat{x}_n = \Delta\hat{x}_{n-1} + \beta_n(x_n - \hat{x}_{n|n-1}) \quad (6)$$

The equation $\hat{x}_{n|n-1} = \hat{x}_{n-1} + \Delta\hat{x}_{n-1}$ indicates the predicted position of the next point from the current point.

There are two forms of α-β filter, the variable gain form and the constant gain form. In terms of variable gain form, when tracking number n is over 20, the value of α and β will become extremely small, then it is necessary to use fixed value instead. In this paper, every isobar contains so many points, n is too big, for such a long isobar, the first 20 tracking points is not that important, so the value of α and β can be set at first. It not only can reduce the complexity of program, but also it does not affect the effect of tracking and extraction. So in this paper, constant gain α-β filter is implemented. After several experiments, the values of α and β are ultimately determined as 0.8 and 0.3. During the process of tracking, the measured value in the field of 5 * 5 of the prediction value is being sought. If there are more than one measured value, then the one which is nearest to the predicted value is recorded. Filter value can be figured out by measured value and predicted value. If there is no measured value, take the final filter point as a breakpoint, the predicted value as measured value, continuing to track. The length of broken tracking is set to be 15, namely, if forecast tracking more than 15 points and there is still no matching point to the broken point, then the broken point is considered as an end point and finish the tracking of this isobar. Vectorization of isobars is completed with polygonal lines, therefore, during the filter process, the interpolation function used to connect broken isobar is “liner”, so that the gap between the breakpoints can be filled.

Fig.12 shows the extracting result, burrs are removed and short disconnections are linked smoothly. The process of extracting annotation results in some lines disconnected longer, these discontinuousness are too long to be connected. However, without annotation interference, isobars can be extracted smoothly and continuously.

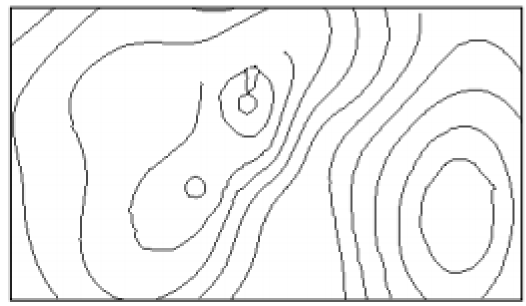

Fig.12 Isobars extracted by α-β filter

B . Vectorization of isobars

The vectorization of a curved line is essentially to compress the line data. From the perspective of information theory, curved line compression is to extracted a subset A' from the data set A which make up the line. Within a specified accuracy range, the subset is considered as a new source of information which should reflect set A approximately and reaches a compression ratio as much as possible. The principle of vectorization is shown in Fig.13, P_i is the starting point of a tracked line, P_n is the ending point. $P_k(k=i+1,i+2,\cdots n-1)$ is a point which is located between P_i and P_n. Searching point P_k from the starting point P_i, make sure that the distance d from every point between P_i and P_k to the segment $\overline{P_iP_k}$ is no more than a certain threshold value. So the segment $\overline{P_iP_k}$ can substitute for curve segment $\overline{P_iP_k}$, and the recorded points P_i and P_k are enough to redraw the original curve[14-15].In this paper, the distance threshold value is 2, the final vectorization image is shown in Fig.14.

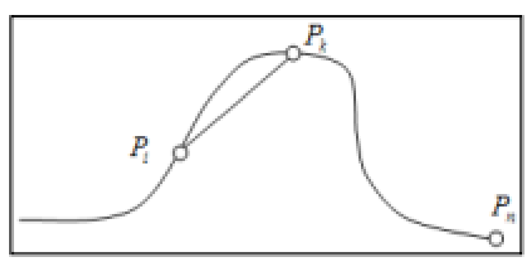

Fig.13 Vectorization principle

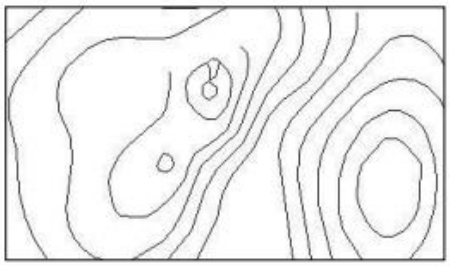

Fig.14 The vectorization image

C. Analysis of result

For the above image, isobars share 2837 pixels, accounting for 3.78% of the image. The number of points to describe Fig.14 is 217, so the achieved compression ratio is 13.0737.

In order to analyze the vectorization result, another image which is less interfered by annotations is experimented. The original image is shown in Fig.15, isobars occupy 2436 points, take the proportion of 4.54% of the image. The vectorization result of isobars is shown in Fig16.

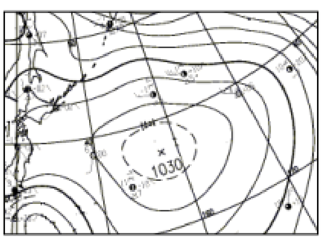

Fig.15 The original image for vectorization

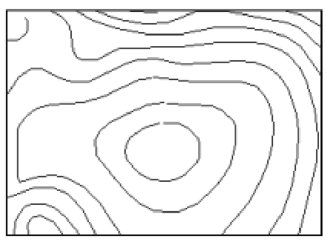

Fig.16 The result of vectorization

In this paper, compression ratio and similarity are chosen to evaluate vectorization results. During the process of vectorization, different threshold value d will result in different compression ratio and similarity, as shown in Tab.1.

Tab.1 The relationship between threshold value d and compression ratio and similarity

threshold value d	1	2	2.5	3	5
compression ratio of Fig.14	8.37	13.08	14.62	15.42	19.70
Similarity of Fig.14	0.98	0.98	0.97	0.97	0.95
compression ratio of Fig.16	9.79	13.61	15.72	17.28	20.64
Similarity of Fig.16	0.98	0.97	0.96	0.96	0.94

As Tab.1 shows, the larger the threshold value is the greater compression ratio. But when the threshold value is too large, the vectorization information will be distorted, losing some details of isobar. So considering the compression ratio and accuracy simultaneously, the distance threshold value d is finally set to be 2 in this paper.

IV . Conclusion

In this paper, an algorithm for meteorological facsimile map vectorization is proposed. By extracting annotations and vectoring isobars, the information related to isobars is successfully extracted. Noises in the map are removed and broken lines are connected by means of α-β filtering before extracting information from isobars in detail. Then the extracted isobars are approximated by polygonal lines and denoted by the coordinates position of the turning points. Thus the map is converted from its original bitmap format into vector format and can easily be stored and displayed together with other navigation information.

References

[1] Hu Xiaotong, Study on the Vectorization Technology of Dot Matrix Meteorological Facsimile Maps. Harbin, Harbin engineering university, 2007.1

[2] Huang Chengbo, The Research and Design Implementation of Meteorological Facsimile Maps Vectorization. Dalian, Dalian Maritime University.2011.5

[3] Liu Huimin, Research and Realization of Pivotal Techniques in the Image Processing and Information Recognizing for Meteorological Facsimile Chart. Harbin, Harbin engineering university, 2008.1

[4] Li Cundong, *et al*, Information extraction from meteorological facsimile maps. Journal of Image and Graphics. No.10, 2012,pp: 1268-1273

[5] Website of the Meteorological facsimile map: http://www.imocwx.com/in dex.htm

[6] Dong Li-ju, YU Ge, An Efficient Approach to Image Binarization. Journal of Northeastern University (Natural Science) No. 25(12), 2004, pp:1149-1152.

[7] Chen Zhengguang, Wu Yushu, Wang Yufang. A New Schema of Vectorization of Contour. Journal of Computer Engineering and Applications.No.3, 2004, pp: 84-86

[8] Li Ying, Duan Shan, Study of Binary Image Thinning Based on Mathematical Morphology. Journal of South-central College For Nationalities(Natural Sciences).No.4,2005, pp:96-99

[9] A. Rosenfeld and A. C. Kak, Digital Picture Processing. Orlando, FL: Academic, 1982, ch. 11, vol. 2.

[10] Pavlidist. Algorithms for graphics and image processing.Washington DC: Rockville Computer Science Press, 1982

[11] P. S. P. Wang and Y. Y. Zhang, A fast and flexible thinning algorithm, IEEE Trans. Computers, vol. 38, pp. 741-745, May 1989.

[12] Raghu Krishnapuram, Ling-Fan Chen, Implementation of Parallel Thinning Algorithms Using Recurrent Neural Networks. IEEE Transactions on Neural Networks, No.1 1993, pp: 142-147

[13] Qi Guoqing, Signal detection and estimation - Principles and Applications. Electronic Industry Press, December 2010.

[14] Zouxiu Ming, Zhang Yuexin, Research and Realization about Grid Map Vector. Journal of Computer Engineering and Applications, No.19, 2003, pp: 102-104

[15] Yu Shengchen, Liu Dayou, Separation of Contour Lines and Longitudinal and Latitudinal Lines and Contour's Vectorization. Journal of Image and Graphics. No.5(2),2000,pp:167-169

[16] Zhang Zheng, Wang Yanping, Xue Guixiang, Digital image processing and machine vision-Visual C++ and Matlab. People's Posts and Telecommunications Press, April 2010.