Man–Computer Interactive Method on Cloud Classification Based on Bispectral Satellite Imagery

Yu Fan (郁 凡), *Liu Changsheng* (刘长盛)
Department of Atmospheric Sciences, Nanjing University, Nanjing 210093
and *Chen Weimin* (陈渭民)
Department of Atmospheric Physics, Nanjing Institute of Meteorology, Nanjing 210044
Received November 11, 1996; revised December 24, 1996

ABSTRACT

A bispectral cloud classification method based on man–computer interactive way, i.e. a unit feature space classification method (UFSCM), has been presented in this paper. Apart from land and water, six types of clouds including cumulonimbus, multilayer cloud system, thin / think cirrus, middle and low level clouds are recognized. The method has been tested by using more than two hundred samples, with total accuracy reaching 87.1%.

Key words: Bispectral satellite imagery, Cloud classification, Unit feature space classification, Man–computer interactive method

I. INTRODUCTION

Satellite images reveal the global temporal and spatial cloud information with the most succinct way. The types, shapes and evolution of clouds exhibited by satellite imagery comprehensively reflect the inherent dynamic and thermodynamic processes in the atmosphere. The reasonable recognition of cloud types and the accurate estimation of cloud amount by computer, are the basis to introduce the useful information of satellite imagery into numerical analysis. Lin et al. (1987) conducted the computer classification of satellite images by use of dynamic clustering approach. Li et al. (1990, 1992) implemented respectively the cloud classification of polar orbit and geostationary satellite cloud imagery and the monitoring of intensive rainfall. We (1994) also did the work on the cloud classification and on the discrimination of Meiyu frontal rainfall grade by use of box classification method. This paper presents a bispectral cloud classification method based on man–computer interactive way, i.e. a unit feature space classification method (UFSCM), in order to improve the accuracy of the cloud classification on infrared and visible two dimension spectral feature space (2DSFS).

Of meteorological satellite data, the infrared (IR) and visible (VS) spectral images are most extensively used. The IR–VS 2DSFS is a measurement space which consists of the above two orthogonal components, and the abscissa and ordinate respectively represent the gray values of IR and VS images. Any pixel of the same name (pixel with the same spatial coordinates (m,n)) in the bispectral images constitutes a 2–D spectral feature vector which is composed of IR gray value $G_I(m,n)$ and VS gray value $G_V(m,n)$ components and represents a point in the 2DSFS. Each type of cloud in satellite images has specific IR and VS spectral feature, and all the pixels of a type of cloud in satellite images, form a cluster in the 2DSFS which corresponds to the type of cloud. Pixels of different types of clouds with different IR and / or VS spectral features form clusters of different locations and different patterns on 2DSFS. However, due to the influence of a variety of factors in remote sensing process, clusters indicating different types of clouds are often overlapped. The purpose of cloud classification is to reasonably recognize the genus of each spectral vector on the bispectral feature space, the focus is to identify the clusters indicating different types of clouds and the underlying surface as correctly as possible.

The cloud classification methods based on 2DSFS mainly include threshold method and statistical method. Box classification is one of threshold methods. In the box classification method, the optimal IR and VS thresholds for each cluster are determined by analyzing a large number of known samples, and then the area of each cluster is enveloped with a rectangular box. Each of rectangular areas on 2DSFS respectively represents a cretain type of cloud or the underlying surface, and the genus of each spectral feature vector is determined in cloud classification process by the area where its corresponding point falls into. The processing speed of box classification is fast, and the result is displayed simultaneously. However, the limitation of rectangular box has obviously influence on the accuracy of genus discrimination, and it is especially inadequate to the cluster whole vectors have a interrelation, such as thin cirrus. Maximum likelihood method is representative of statistical methods. It based on the hypothesis that the IR and VS gray values of any type of cloud follow a normal distribution, and therefore the cluster distribution on 2DSFS can be described by the mean value and covariance matrix. If training samples are numerous enough, the statistical probability of each known feature vector to each cluster can then be calculated and the elliptic equiprobability line which will bound a type of cloud can be drawn on 2DSFS. During the course of the cloud classification, the probability that an unknown spectral feature vector belongs to every genus is calculated by use of those probability functions and then maximum likelihood genus of the vector is determined. Maximum likelihood method is theoretically more reasonable and accurate than box classification method. However, it is required that the selected training samples of each genus must be representative and the primary hypothesis of normal distribution must be satisfied. Those requirements produce some difficulties in practical operation, because the training samples which are selected from one satellite image or even a part of a image in common practice are hard to satisfy the above requirements. However, the theory shows that the correct mean value vector and covariance matrix of each type of cloud can be obtained only if the prerequisite is met, otherwise erroneous classification may happen. In maximum likelihood method, the probability to every genus of each spectral feature vector must be calculated, therefore the computation is great. Box classification method and maximum likelihood method are both based on training enormous samples, and they delineate the cluster distribution areas, representing various types of clouds and the underlying surface, on 2DSFS with the straight lines parallel to coordinate axes or with quadratic curves respectively.

How can we define the distribution area of the cluster of each type of cloud reasonably? Platt (1983) made a meaningful discussion on water cloud and ice cloud with the aid of radiation transfer theory. He analyzed the influence of the variations of the cloud amount and optical depth of water cloud and ice cloud in a field of view on the distribution of feature vector in spectral feature space. He found that for the thin cloud filling a field of view, regardless of water cloud or ice cloud, as optical depth increases, the hodograph of spectral feature vectors which increase monotonously with increasing IR and VS gray values, and the differences between water cloud and ice cloud mainly exist in the cloud top temperature and the upwards concave extent of the curve. For thick cloud, the hodograph of spectral feature vectors displays an oblique line which increases with increasing cloud amount in a filled of view. The oblique lines and concave curves of water cloud and ice cloud enclose respectively an area which defines the cluster of the type of cloud (Fig. 1). For two overlapped layers of clouds, the circumstances are more complicated, and the spectral feature vectors often form irregular distribution areas due to the different spectral and microphysical characteristics of up–layer cloud and low–layer cloud. The above conclusion has been verified in our analysis.

II. UNIT FEATURE SPACE CLASSIFICATION METHOD

In a specific region and a certain season, the distribution areas of the cluster of various types of clouds on 2DSFS are generally unchanged, this also is the premise for the supervised classification methods including box classification method and maximum likelihood method. This paper presents a new method for cloud classification i.e. unit feature spectral classification method (UFSCM) and attempts to improve the cloud classification by fitting the distribution area of the cluster of various types of clouds on 2DSFS. In this method, after VS gray value is expanded from 6 bit to 8 bit like IR gray value, the IR–VS 2DSFS is divided into 4 × 4 units of 64 × 64, and this unit is the unit feature space. As long as the distribution area of the cluster of a certain type of cloud on the spectra feature space is correctly determined by enough representative training samples, it can be marked by the group of corresponding unit feature spaces, then being used to identify the type of cloud.

For the purpose of easy and correctly determining the genus of all unit feature spaces in IR–VS 2DSFS, we designed a man–computer interactive analysis algorithm. Essentially, the

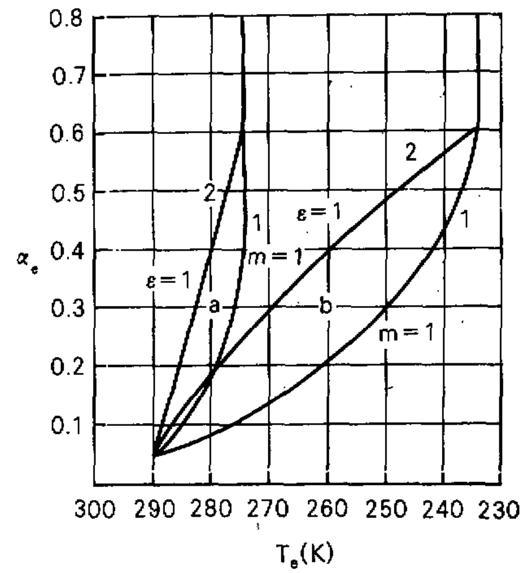

Fig. 1. Bispectral curves of albedo α_e versus brightness temperature T_e for (a) low cloud and (b) a high cloud. Curve 1 unbroken layer with variable optical depth: Curve 2 broken layer of cloud of constant optical depth($\varepsilon \approx 1$).

cluster consisting of spectral feature vectors on 2DSFS is the mapping of all pixels of corresponding type of cloud in IR–VS bispectral images. A pixel of same name on bispectral images uniquely maps into a corresponding unit feature space, however the unit feature space can correspond to many pixels of different spatial coordinates on bispectral images, if only those pixels have similar IR and VS spectral characteristics. In addition, there are some unit feature spaces which are not relevant with any pixels on IR and VS images, the man–computer interactive analysis algorithm is just established on the above mapping concepts.

Fig. 2 shows a block diagram of the man–computer interactive analysis program. "The recognition of cluster centers" is realized by analyzing the 2–D histogram of the bispectral images of selected areas and then determines the location of high frequency distribution area (HFDA) of each type cloud. In "unit feature space classification", the genus of the unit feature spaces representing various types of clouds is determined by further analyzing the 2–D histogram mapping of local images of single type of cloud. Finally, the unit feature space classification table is synthesized for further application to directly output cloud classification images, and to calculate the cloud amount for each type of cloud.

The 2–D histogram is a concrete expression form of the IR–VS 2DSFS divided into 4×4 units of 64×64, and its distribution function is defined by the following expression:

$$H(i,j) = \sum_{m=1}^{M} \sum_{n=1}^{N} \lambda_{ij}(G_I(m,n), G_V(m,n)) \qquad i = 1,2,\cdots,64; \; j = 1,2,\cdots,64$$

where N is the number of pixels of each row in the image of selected area, M the total row number in the image, and i and j are sequential numbers of row and column respectively of a unit feature space in the whole 2DSFS. $\lambda_{ij}(G_I, G_V)$ is defined as (including the left end):

$$\lambda_{ij}(G_I, G_V) = \begin{cases} 1 & (G_{Ii-1} \leqslant G_I < G_{Ii}) \bigcap (G_{Vj-1} \leqslant G_V < G_{Vj}) \\ 0 & \text{otherwise} \end{cases}$$

In the man–computer interactive analysis program, the 2–D histogram mapping of regional image (image⇒2–D histogram) can be displayed on screen respectively with 16–color frequency distribution 2–D histogram or the frequency number data of all spectral feature

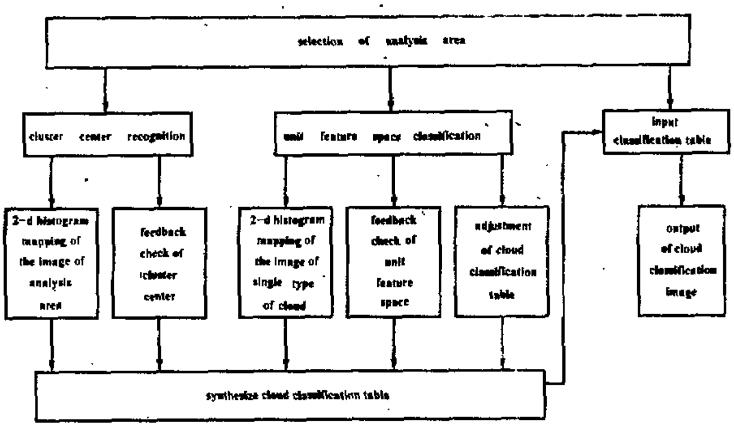

Fig. 2. Block diagram for man–computer interactive program.

vectors on each unit feature space. The high frequency distribution areas on 2–D histogram indicate the large probability areas and cluster centers of the spectral feature vectors corresponding to each type cloud or underlying surface.

Table 1 shows a real example of 2–D histogram analysis (Changjiang–Huaihe valley and a part of the southeast coastal area of China; 14:00 LT 23 July 1993). Eight main high frequency areas labeled with A,B,C,D,E,F,G and H, respectively reveal the center locations and distribution patterns of eight clusters representing intense cumulonimbus, multilayer cloud system, middle cloud, thick cirrus, thin cirrus, low cloud, water surface and land. Region A on the top right of the histogram is an HFDA with high values of IR and VS gray level, and it reveals the cluster center of the intense cumulonimbus of mature stage with its top reaching the tropopause. Region G and Region H, lying over the bottom left, have small values of IR and VS gray leve, and represent water surface and land respectively. Region E and its neighborhood high value points form the HFDA of thin cirrus, and it is the transmission character of thin cirrus than results in a distribution pattern of arc shape on 2–D histogram. This

Table 1. 2–D Histogram of Case Analysis

[illegible]

prolonged pattern from the bottom left to the top right reflects thin cirrus′ character that its IR and VS gray values increase with increasing depth of cirrus. After the depth of cirrus exceeds a certain value, the increase in the IR gray value, which represents the sum of cirrus′ emittance and the transmissive radiation of the lower layer of cloud, becomes slow, and the effect of the increase in cirrus′ depth mainly reflects in the rapid increase of VS gray value. Region D is such an HFDA of vertical stripe shape of thick cirrus. Region F on the left of the histogram is the HFDA of low cloud. Region C on the top middle is the HFDA of middle could. Because most of middle clouds in the image of the selected area are under the cirrus layer, only a few can be seen from the edge and gaps of high cloud layers, therefore, the area of the HFDA of middle cloud on the histogram is small, and it leans to the high cloud area. Region B in the top right is a convergent area of differnt types of cloud systems, and mainly represents the multilayer cloud system consisting of low cloud, middle cloud, cirrus and the weak cumulonimbus in initial and growing stages.

It can be seen from the above analysis of 2–D histogram that although the clusters join together, the high frequency distribution areas can still reveal the corresponding center location and distribution pattern of various types of clouds well. "Feedback check of cluster center" is used to test whether the HFDA analyzed is unique for corresponding type of cloud. This is an inverse process of the 2–D histogram mapping of images, its function is to perform the feedback from histogram to image (2–D histogram ⇒image). After an HFDA is selected with a movable window, all pixels corresponding to this selected HFDA will display a specific color on an IR image. By comparing the IR image with initial IR image, it can be determined whether the HFDA represents a certain type of cloud only.

The correct positioning of the centers of the clusters representing various types of clouds and the surface is the premise of correctly classifying. Only after the genus of every unit feature space is determined, cloud classification can be finally realized. The direct mapping from

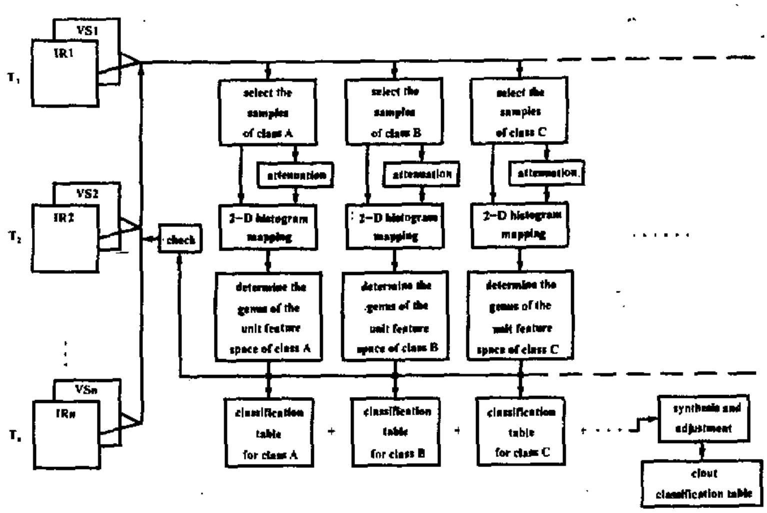

Fig. 3. Flow chart of unit feature space classification.

the known samples of images is used to determine the genus of all unit feature spaces. Like all supervised classification methods, the crux of the matter is to select a representative training set which includes enough known samples. The man–computer interactive way opens up a shortcut for selecting the training set. Fig. 3 is the flow chart of unit feature space classification. The representative known samples are read directly from microcomputer screen by use of a movable window with changeable size, and the number of samples may be expanded from any historical or more recent cloud image data. Once the known samples of a type of cloud are determined, their spectral feature vectors are simultaneously mapping on to the 2–D histogram. Only the HFDA of those samples is selected by the aid of multistage attenuation, then the genus of the relevant unit feature space is set up. The reliability of analysis is checked by the inverse process from the unit feature space on 2–D histogram to the images, and the genus of relevant unit feature space is finally determined. With the same procedures gradually extending the range of each cluster, after all the unit feature space are covered, the classification is completed. The unit feature space classification table for each cluster is recorded in file form, and after the adequate adjustment for overlapped and blank unit feature spaces, the cloud classification table is finally synthesized. Because the whole training process of known sample is visible and able to be checked, this has laid the foundation for a stable and reliable cloud classification table.

III. RESULTS

In this paper, 10 pairs of GMS–4 geostationary satellite infrared–visible bispectral images in June to August of 1992 to 1994 (Table 2) are analyzed, and 7 pairs of them are used to create the cloud classification table by use of UFSCM. The results are shown in Fig. 4. The bottom boundary of cluster distribution of low cloud, middle cloud and thin cirrus is a concave curve, which is an agreement with the theoretical analysis by Platt (1983). 30 samples for each type of cloud and the underlying surface are picked out from the rest three pairs of images (marked with *) with a 4×4 or 8×8 selected window for the purpose to test. The result of the accuracy with a criterion of 85% is listed in Table 3.

Table 2. Time of Cloud Images for Analyzing and Testing(*)

11:00LT 21 June 1992
14:00LT 21 June 1992
13:00LT 13 July 1992
14:00LT 13 Aug. 1992 *
14:00LT 21 June 1993
12:00LT 29 June 1993 *
14:00LT 23 July 1993
12:00LT 7 June 1994
14:00LT 7 June 1994 *
12:00LT 26 Aug. 1994

Table 3 shows that 209 samples (the sum of the numbers on the main diagonal) among 240 samples analyzed are correctly classified, with the whole accuracy reaching 87.1%. The advantage of bispectral cloud classification is able to distinguish cumulonimbus with thick cirrus, especially the cirrus nothus left by dissipated cumulonimbus. Comparing the cluster of thick cirrus and cirrus nothus with cumulonimbus, their IR gray values are almost the same,

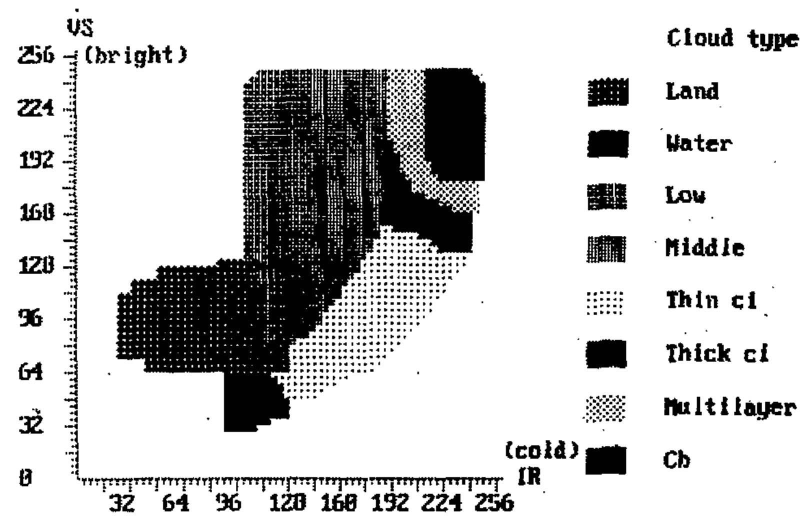

Fig. 4. 2–D histogram of cloud classification table.

Table 3. Testing Results for Each Class

	C_b	multi -layer	thick cirrus	middle cloud	thin cirrus	low cloud	land	water	real success ratio
C_b	26	4	0	0	0	0	0	0	86.7%
multilayer	0	27	3	0	0	0	0	0	90.0%
thick cirrus	0	2	25	3	0	0	0	0	83.3%
middle cloud	0	0	1	26	2	1	0	0	86.7%
thin cirrus	0	0	2	3	23	1	0	1	76.7%
low cloud	0	0	0	1	0	27	2	0	90.0%
land	0	0	0	0	0	2	26	2	86.7%
water	0	0	0	0	0	0	1	29	90.7%
analysis success ratio	100%	81.8%	80.6%	78.8%	92.0%	87.5%	89.7%	90.6%	

and it is difficult to distinguish them only by use of IR image. For a dissipating cumulonimbus, the top part forms the cirrus nothus, in the meantime its IR spectral characteristics are not changed to any great extent, but its lower part has dissipated or has tended to be loose and thus the transmission of visible light greatly reduces the VS gray value of the cirrus nothus. Therefore, the cluster of cirrus nothus on 2–D histogram obviously lies below the clusters of cumulonimbus, and the situation of thick cirrus is the same. With regard to 4 samples of cumulonimbus which fall into the cluster of multilayer cloud systems, they are generally weak and can still be recognized from their shapes on the cloud classification output image. Errors of the recognition of multilayer cloud system are mainly due to its top layer cirrus being thin, so that the multilayer cloud system merges into the cluster of thick cirrus. The transmission property of the thin cirrus renders its spectral characteristics strongly dependent on the lower layer. The analysis success rate of thin cirrus over the sea surface is high due to the uniform underlying surface, however, the thin cirrus over land may be erroneously recognized as thick cirrus, middle cloud or even low cloud because of the varied albedo of land. In the analyses, we found that a stretch of cumulus groups may frequently fall into the clusters of middle cloud and low cloud.

Fig. 5 gives an example of cloud classification analysis. The local area of 192 × 192 pixels taken from the IR and VS images at 12:00LT 13 August 1992 are analyzed, and the cloud classification images, cloud amount and the histogram of classification are shown in Fig. 5. It can be seen that the classification results are satisfactory.

IV. CONCLUSIONS

(1) Because man−computer interactive way is adopted in UFSCM, sufficient samples can be taken from historical and more recent cloud image data in training samples to realize the fitting of unit feature spaces to various types of clusters. This ensures that the classification has a higher accuracy and is consistent with application.

(2) The man−computer interactive analysis program is coded with C++ language, and it can be operated in a 486 microcomputer with a display of 1 M display memory and 1024 × 768 resolution, and timely processing the GMS infrared and visible cloud images. The processing area of cloud classification can be selected from the whole image with a movable rectangular window of changeable size (maximum size: 192 × 192 pixels). The output of cloud classification image and the calculation of cloud amount are completed through looking up the table, the results are simultaneously displayed on screen or printed out. After a little modification, this software can be used to analyze the polar orbit satellite multispectral images.

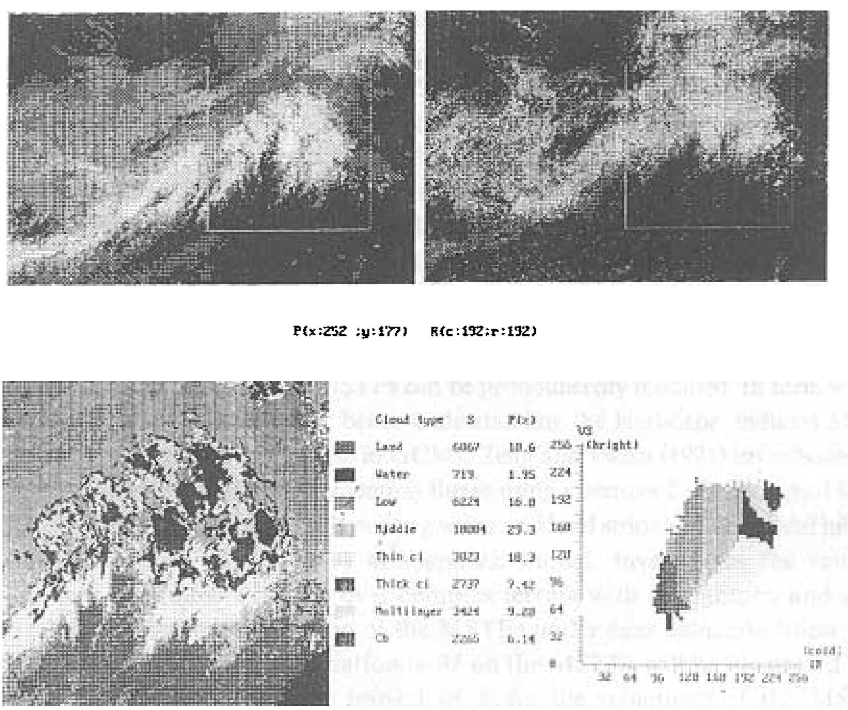

Fig. 5. An example for cloud classification.

(3) Owing to variations in season, area and solar zenith angle, the shapes and positions of various clusters may change a little, therefore, after finishing the correction of solar zenith angle for visible cloud images, the cloud classification tables adequate to various areas and seasons can be established and then saved in a file. In the latter application, the corresponding cloud classification table can be called according to real circumstances, and cloud classification images and cloud amount statistics can be rapidly and accurately manifested.

(4) The UFSCM based on man–computer interactive way is practicable and flexible. For example, one uses various grades of rainfall to train the cloud classification table, then it can output the cloud distribution image and cloud amount for various grades of rainfall.

Finally, due to the limitation of data, the correction of solar zenith angle to visible images is not performed in this paper, therefore only the data near the midday are used. If the correction is done, it may further improve the analysis accuracy and expand the applicable range.

The authors wish to express their sincere thanks to Mr. Xiao Wenan and Zhang Yungang for supplying the image data used in this paper.

REFERENCES

Li Jun and Zhou Fengxian (1990), Computer identification of multispectral satellite cloud imagery, *Advances in Atmospheric Sciences*, **7**: 366–375.

Li Jun, Zhou Fengxian and Wang Luyi (1992), Automatic classification and compression of GMS cloud imagery in heavy rainfall monitoring application, *Advances in Atmospheric Sciences*, **9**: 459–464.

Lin Xijian and Liu Changsheng (1987), Extraction of cloud parameters from bispectral measurements, *Atmospheric Radiation Progress and Prospects, Edited by Kuo–Nan Liou and Zhou XiuJ*, Science Press, Beijing, China, 452–458.

Platt C.M.R. (1983), On the bispectral method for cloud parameter determination from satellite VISSR data: separating broken cloud and semitransparent cloud, *J. Climate Appl. Meteor.*, **22**: 429–439.

Yu Fan, Chen Weimin (1994), Research on the cloud classification for the bispectral cloud picture, *J. of Nanjing Institute of Meteorology*, **17**: 117–124.

Yu Fan, Chen Weimin (1994), Category identification of rainfall from Meiyu front clouds by use of bispectral cloud maps, *J. of Nanjing Institute of Meteorology*, **17**: 171–176.