Automatic Classification and Compression of GMS Cloud Imagery in Heavy Rainfall Monitoring Application

Li Jun (李 俊), *Zhou Fengxian* (周凤仙) and *Wang Luyi* (王路易)
Institute of Atmospheric Physics, Academia Sinica, Beijing 100029
Received January 14, 1992; revised April 27, 1992

ABSTRACT

The Bayes Decision method (BD) has been implemented on GMS Stretched VISSR (S.VISSR) images. Several objects are identified. These are earth's surface, stratiform, convective and cirrus clouds. Overlaid 6-hour surface rainfall observations show that the rain area on the ground matches with precipitable cloud area identified by BD method. Finally the data of classified image are compressed to about one-eighth amount by use of Line-Run-Coding technique. It is helpful for data archive and remote transmission.

I. INTRODUCTION

The geostationary satellite has the unique ability to frequently observe the atmosphere and its cloud cover from the synoptic scale down to the cloud scale. Hourly visible and infrared cloud imagery of geostationary satellite is not only used to analyze the distribution of clouds, the evolution of synoptic system, but also used to study the structure and activity of mesoscale cloud system. Therefore the cloud imagery is an important data source for analyzing the weather systems such as heavy rainfall and severe thunderstorms. In recent years with widespread installment of GMS receiving systems and availability of personal computer in satellite data application the computer processing of satellite cloud imagery has been extensively developed. But the methods mainly focus on image enhancement and image composition, and the cost of color composition of different images is expensive. In view of characteristics of plentiful information and large data amount, the effective image processing and recognition methods should be studied to extract the accurate and useful information from satellite data for meteorological or oceanographical purposes. The technique to reduce the data amount is meaningful in saving storage space. Research works on those aspects have been conducted lately at the Institute of Atmospheric Physics, Chinese Academy of Sciences (Li et al., 1990, 1991).

This paper uses Bayes Decision method to classify the simultaneous GMS visible and infrared images into several objects. The classified image identifies the earth's surface, stratiform, convective and cirrus clouds. After each classification procedure, the mean vector and prior probability, the covariance matrix of each class are recalculated to act as the criteria of next time classification procedure. Overlaid 6-hour surface rainfall observations show that the rain area on the ground matches with precipitable cloud area: stratiform or convective cloud area. This method is useful in objective heavy rainfall monitoring. Furthermore, the Line-Run-Coding technique is adopted to code the classified image and reduces the data amount greatly, which is helpful in data archive and remote transmission.

Finally, the software based on the procedure mentioned above was implemented on a real-time GMS Stretched VISSR receiving system which is owned by the Hydrological Forecasting and Water Control Center, Ministry of Water Resources in summer 1991. Experiments carried out by applying this method on several cases which caused flood disaster in Anhui Province during June 1991 show that the results of image classification are promising and the precipitable cloud area fits the observational rainfall region well. Further application of this method is necessary for verification and improvement.

II. CLASSIFICATION OF CLOUD IMAGE DATA

The purpose of classification is to sort GMS visible and infrared image data into several classes. For the sake of simplicity, we rewrite the $M \times N$ pixel VIS and IR image data in matrix form:

$$\begin{bmatrix} V_{11} & V_{12} & \cdots & V_{1N} & V_{21} & V_{22} & \cdots & V_{2N} & \cdots & V_{M1} & V_{M2} & \cdots & V_{MN} \\ I_{11} & I_{12} & \cdots & I_{1N} & I_{21} & I_{22} & \cdots & I_{2N} & \cdots & I_{M1} & I_{M2} & \cdots & I_{MN} \end{bmatrix}$$

where the first line of the matrix represents VIS image and the second one is for IR image, every (V_{ij}, I_{ij}) is the sample of classification, we denote it $\mathbf{X}=(V,I)$ for convenience. The Bayes Decision rule is a criterion for partitioning the feature vector $\mathbf{X}$ to class $\mathbf{W}_i$ from classes $i=1,2,...,m$. Let $P(X|W_i)$ be the conditional probability that the feature vector has value $\mathbf{X}$ and and is in the class W_i, this is the probability density function of feature vectors from training areas of class W_i. Let $P(W_i)$ be the prior probability of class W_i in given image. This probability of observing a feature vector from class W_i, is independent of any other information.

Before Bayes Decision is used the training feature, a representative area for each desired class, must be selected from the image to determine the class boundaries so that the classification of unknown vectors results in minimum error rate. It is important that the training area should be a homogeneous sample of the respective class, but at the same time, the range of variability of the training data set should not be too small for keeping the covariance matrix nonsingular. In this study, the training areas are selected for the results of clustering based on both spectral and textural characteristics of the imagery (Li et al., 1990).

The Bayes Decision rule states that:

$\mathbf{X} \in W_i$, if and only if

$$P(\mathbf{X}|W_i)P(W_i) \geq P(\mathbf{X}|W_j)P(W_j) \tag{1}$$

for all $j=1,2,...,$m.

We denote $P(\mathbf{X}|W_i)P(W_i)$ the discrimination function. In case the feature vectors of the sample matrix are multivariate normal distribution, the discrimination function becomes

$$G_i(\mathbf{X}) = P(\mathbf{X}|W_i)P(W_i) = P(W_i)(2\pi)^{-1}\left|\sum_i\right|^{-\frac{1}{2}}\exp[-\frac{1}{2}(\mathbf{X}-\mathbf{M}_i)\sum_i^{-1}(\mathbf{X}-\mathbf{M}_i)] \,, \tag{2}$$

where $\mathbf{M}_i$ is the mean vector of class W_i from training area and $\sum_i$ is the covariance matrix of the same class. Let $g_i(\mathbf{X})$ be the logarithm of $G_i(\mathbf{X})$, then

$$g_i(\mathbf{X}) = \ln P(W_i) - \frac{1}{2}\ln\left|\sum_i\right| - \frac{1}{2}(\mathbf{X}-\mathbf{M}_i)^T\sum_i^{-1}(\mathbf{X}-\mathbf{M}_i) \,. \tag{3}$$

Because both $G_i(\mathbf{X})$ and $g_i(\mathbf{X})$ are monotonic decreasing functions, taking the largest value of $G_i(\mathbf{X})$ or $g_i(\mathbf{X})$ is the same. Therefore Bayes Decision rule can be calculated as follows: the random vector $\mathbf{X}$ should be classified to class W_i if $g_i(\mathbf{X})$ is the largest for $Ki = 1,2,\ldots,m$.

In our application, we denote $P(W_i)$ as follow

$$P(W_i) = \frac{1}{\bar{D}_i} / \sum_{j=1}^{m} \frac{1}{\bar{D}_j} \quad , \tag{4}$$

where $\bar{D}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} (\mathbf{X}_j - \mathbf{M}_i)(\mathbf{X}_j - \mathbf{M}_i)^T$

n_i is the vector number of class W_i. When applying this classification procedure to GMS-4 image data, we take $m=4$, and $P(W_i)$, $\mathbf{M}_i$ etc. are recalculated after each classification to act as the criteria of next time classification, which makes the classification procedure have the function of self-study. Thus, it improves the effectiveness of Bayes Decision method, and also, the results of classification will be less affected by some specific conditions such as geographical latitude, season, synoptic situation etc.

III. COMPRESSION OF CLASSIFIED IMAGE

We use Line-Run-Coding method to compress the classified image. This technique is described in another paper (Li et al., 1991). For example, the line run code of the 6 × 4 2-value image *I*(6,4) (see Fig.1) is (0,3)(1,2)(0,1)(0,1)(1,3)(0,2)(1,2)(0,1)(1,3)(1,3)(0,1)(1,2). In our application to GMS-4 classified cloud image, each pair of numbers is stored in a byte, and the image data are compressed to one-eighth amount by use of this technique. In general, the compression rate is directly related to the form of image.

0	0	0	1	1	0
0	1	1	1	0	0
1	1	0	1	1	1
1	1	1	0	1	1

Fig.1. 6 × 4 2-value image.

IV. APPLICATION

The classification method mentioned above was applied to the heavy rainfall cases occurred in Huaihe River during mid-June, which brought about flood disaster in Anhui Province and caused the decision to evacuate the local residents and open the embankment. Several pairs of simultaneous 350 × 640 pixel GMS-4 IR and VIS images are collected during this period for experiment. The VIS image of S.VISSR has 64-gray level which represents albedo, and IR image has 256-gray level which represents brightness temperature of earth's surface or cloud top. For the sake of simplicity, IR range keeps its original 256 levels and VIS range has been expanded to 256 levels from its original 64 values.

The classification method was applied to GMS-4 images at time 15:00 (local time) of June 12, 1991. Four classes have been identified, whose parameter values of gravity centers are given in Table 1.

Table 1. Classification Results at 15:00 LT of June 12

CLASS	VIS	IR	ALBEDO(%)	BRIGHTNESS TEMP.(K)	OBJECT
1	59.51	114.07	5.67	288.70	earth surface
2	98.01	177.21	15.13	254.32	cirrus
3	125.66	141.30	25.00	275.38	mid-low clouds
4	154.20	207.16	37.85	231.54	convec. clouds

The identifications given to the classes are:

Class 4: the coldest and brightest clouds corresponding to deep convective clouds which usually bring heavy rainfall

Class 3: not too cold but quite bright corresponding to mid-low stratiform clouds which may also bring about precipitation

Class 2:quite cold but not bright enough reflecting characteristics of semi-transparent cirrus which has no rain related with

Class 1: the warmest and darkest part corresponding to earth's surface i.e., land or water

The classification results at time 14:00 LT of June 13 (see Table 2) are similar to time 15:00 LT of June 12, but the parameters of classes between two times have some changes especially the parameters of class 3 and class 4 which are related to rainfall. The reason is that the weather system was on developing stage during June 12 and the IR count 207 of class 4 showed that the cloud top was cold and high. On the next day the weather system began to decay and the cloud top height decreased (201). This was testified by 6-hour surface rainfall observations, on June 12 6-hour (8-14 LT) maximum rainfall was 136 mm and the next 6-hour (14-20 LT) rainfall increased to 151 mm, whereas on June 13 2-8 LT maximum rainfall was 163 mm and 8-14 LT value decreased to 84 mm, the tendency from developing to degrading is evident.

Table 2. Classification Results at 14:00 LT of June 13

CLASS	VIS	IR	ALBEDO(%)	BRIGHTNESS TEMP.(K)	OBJECT
1	60.65	118.15	5.77	286.82	earth's surface
2	99.41	179.33	15.51	252.99	cirrus
3	132.57	145.73	27.73	272.71	mid-low clouds
4	161.71	201.27	40.98	236.69	convec. clouds

Figure 2 is the classified cloud image at 14:00 LT of 13 June with 8-14 LT 6-hour surface rainfall observations overlaid. Since in operational environment the classified objects are displayed by different colors on TV screen they are easy to be distinguished by users. On a Black / White image four objects can still be recognized: black part corresponds to deep convective clouds, white part shows mid-low clouds, light gray denotes cirrus and dark gray represents earth's surface. The fibrous cirrus in southeast and north of the cloud band is distinguished clearly. The 6-hour ground rainfall observations fall into deep convective cloud and mid-low cloud areas completely (It is hardly to be seen on black and white picture). Figure 3 is the plot of 6-hour (8-14 LT) rainfall observed by hydrological stations on June 13. The main rain area was located within the middle-low reaches of the Yangtze River and Yellow River. So it is evident that the automatic classification is an objective recognition

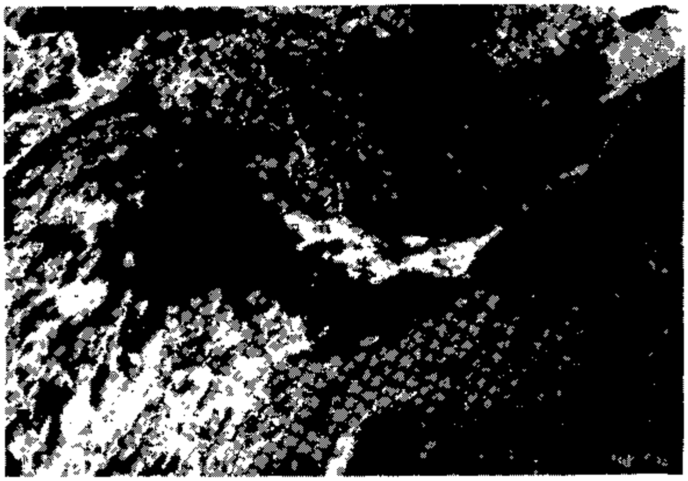

Fig.2. Classified cloud image at 14:00 LT of 13 June.

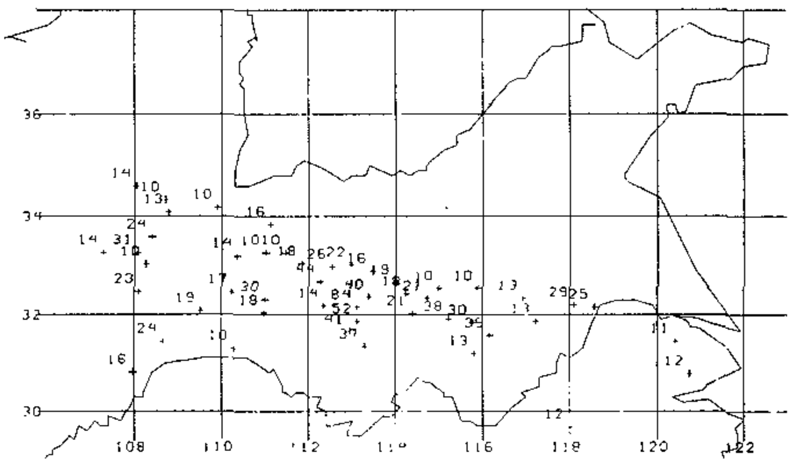

Fig.3. 6–hour (8–14 LT) rainfall (>10mm) on June 13, 1991.

technique. Because both VIS and IR information are used for classification, it avoids the disadvantage of mixing up deep convective clouds with cirrus on IR image due to similarity of cloud top temperature, and with mid–low clouds on VIS image due to similarity of albedo. It becomes especially serious in today's enhanced false color image products available at most of the meteorological and hydrological units because there is only one image (mostly the IR image) being used for color enhancement. Therefore the automatic classification using both VIS and IR images has the advantage. It also makes the cloud pattern more simple to be recognized.

As touching the probability of classification error, considering the classes we wish to discriminate may overlap in feature space, the probability of error is given by an error matrix (E_{ij}), where E_{ij} is the probability of classifying the feature vector as class W_j but it really is from class W_i. The error matrix can be obtained by two steps: first, we assume the results of present classification are standard and the Bayes procedure is re–taken based on the results; second, compare the two procedures and count the pixels erroneously classified to form the matrix. Table 3. is a classification error matrix at 14:00 LT of June 13, from which it can be seen that the classification is quite accurate except some pixels of class 2 are erroneously classified to class 3, this will not affect the heavy rainfall monitoring because the class of convective clouds is very stable in the two classification procedures.

Table 3. Classification Error Matrix of 14:00 LT, June 13

E_{ij} Class \ Class	1	2	3	4
1	0.993	0.007	0.000	0.000
2	0.033	0.785	0.136	0.046
3	0.066	0.000	0.910	0.024
4	0.000	0.001	0.001	0.998

V. ANALYSIS AND DISCUSSION

By use of Bayes Decision method, the GMS S.VISSR images are classified into several objects in order to identify the heavy precipitable region, general precipitable region and non-precipitable region. Experiments show that deep convective clouds are well separated from cirrus and mid-low clouds, and the results of classification keep more accurate due to updating the prior probability and density function after each classification procedure. Overlaid 6-hour ground rainfall observations on the classified cloud image show that the rainfall area matches with deep convective cloud or mid-low cloud area. By use of Line-Run-Coding technique the satellite data are reduced to about one-eighth amount, which saves the storage disk space and transmitting time to remote users. In this study, we consider especially the usefulness of the software and time efficiency of computation. The programme is run on PC-386 with EGA or VGA color monitor. Firstly, the simultaneous 350×640 pixel VIS and IR GMS cloud images are classified by using Bayes Decision method and time cost is 5 minutes. Secondly, the classified image was compressed in order to transfer the data to remote users rapidly. It spends one minute to compress a 350×640 pixel image (220 KB) into Line-Run form data (30 KB). Less than 3 minutes are spent to transfer a classified image with 2400 baud rate and redisplay the compressed data into image. Thus the whole procedure spends no more than 10 minutes, which satisfies the requirement of real-time GMS data processing in operational environment.

During nighttime, the daytime classification results are used as thresholds to classify the IR image since the VIS image is not available. Of course the result is not as good as that in daytime. Further improvement will be based on the spectral and texture information from IR image for the nighttime classification. With correctness of classification, high speed of computation and effective data compression our study has the capability to provide forecasters with quantitative satellite products which are very important to monitor the severe weather. It is also significant to synoptic study. So the prospects of the software are obvious.

Further study should be focused on rainfall rate and area determination quantitatively based on image classification. We wish to transfer the research result to operational use as soon as possible.

The authors wish to express their sincere thanks to Ms. Yang Yang for cooperation in this study. They are also grateful to Hydrological Forecasting Water Control Center for providing the GMS-4 satellite cloud images and rainfall observation data for the skill development and experiment.

REFERENCES

Adler, R. F. and A. J. Negri (1988), A satellite infrared tchnique to estimate tropical convective and stratiform rainfall, *J. Applied Meteorology*, **27**: 31-51.

Coakley, J. A. and D. G. Baldwin (1984), Towards the objective analysis of clouds from satellite imagery data, *J. Climate Appli. Meteor.*, **23**: 1065–1099.

Li, J. and F. X. Zhou (1990), Computer identification of multispectral satellite cloud imagery, *Advances in Atmospheric Sci.*, **7**: 366–375.

Li, J. and F. X. Zhou (1991), Satellite data reduction using entropy–preserved image compression technique, *Advances in Atmospheric Sci.*, **8**: 237–242.

Li, J. and F. X. Zhou (1992), Cloud motion estimation from visual and IR data of geosynchronous satellite using fourier technique, *Advances in Space Research*, **12**: 123–126.

Rossow, W. B., E. Kinsella, A. Wolf and L. Garder (1985), *World Climate Research Program*, International Satellite Cloud Project (ISCCP) Description of Reduced Resolution Radiance Data, July 1985.