Study on clustering of satellit images by K-means algorithm

° 설 상 동 * 김 정 선 **
Sang Dong, SUL Jung Seon, KIM

*, ** Dept. of Avionics Eng., Hankuk Aviation College

ABSTRACT: K-means algorithm was used to classify 3 cloud-type, that is low, mix and cumulonimbus. Initial cluster centers and K parameter is given in this paper by coarse computing and Fisher's algorithm. Results indicate that performance index is minimized and mix cloud is well classified.

1. INTRODUCTION

Cloud-type classification is an important component of meteorological and hydrological programs which require estimates of parameters such solar radiation, rainfall, moisture and sea-surface temperature. One advantage of the satellite images is that it covers abroad range of scales in time and space. [4]

The contribution of the research described in this paper is that a method is developed for incorporating segment features into automatic cloud-classification procedures.

The basic scheme is outlined in Fig. 1. The algorithm consists of two parts: the seggmentation and the classification.

The segmentation is described in Section 2 [1][3] and the classification is described in Section 3. [5][6] The results of testing the algorithm on a sample set of 44 infrared GMS (geostationary meteorological satellite)-3 cloud windows are presented in Section 4. As conclusion in Section 5, [2], directions for future research are considered.

2. SEGMENTATION

It was indicated a need for segmentation prior to the feature extraction process by Parikh and Rosenfeld. [1]

This algorithm clusters ordered sets of data points as gray-level values by frequence of occurrence and minimizes performance index which is defined as the sum of the squared distance from all points within the cluster to the cluster center.

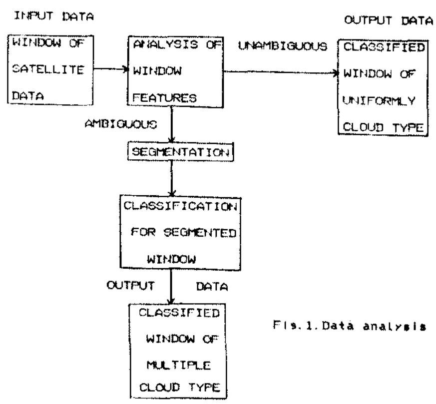

Fig. 1. Data analysis scheme

The procedure consists of the following steps.

step1. Choose K initial cluster centers

$M(1,1), M(1,2), , , , M(1,K)$.

step2. At the kth iterative step, distribute the samples (x) among the K cluster domains, using the following relation

$x \in S(it, j)$

$$\text{if } \|x-M(it,j)\| < \|x-M(it,i)\| \text{ ---------- (1)}$$

for all $i, j=1,2,,,K, i \neq j$

where $S(it, j)$ denotes the set of samples whose cluster center is $M(it, j)$.

step3. From the results of step2, compare the new cluster centers $M(it+1, j), j=1,2,,K$, such that the sum of the squared distances from all points in $S(it, j)$ to the new cluster center is minimized. that is, the new cluster center $M(it+1, j)$ is computed so that the performance index

$$VAR(j) = \sum_{x \in S(it,j)} \|x-M(it+1,j)\|^2 \quad j=1,2,,,K \quad -(2)$$

is minimized.

The $M(it+1, j)$ which minimizes this performance index is simply the sample mean of $S(it, j)$. Therefore, the new cluster center is given by

$$M(it+1, j) = 1/N(j) \sum_{x \in S(it,j)} x$$

where $N(j)$ is the number of samples in $S(it, j)$

step4. If $m(it+1, j) = M(it, j)$ for $j=1,2,,,k$ the procedure is terminated. otherwise goto step2.

The behavior of the K-means algorithm is influenced by the value of parameter K and the choice of initial cluster centers.

We determined the segmentation parameter as K=10 by Fisher's algorithm. [5]

The most significant difference between Fisher algorithm and other clustering algorithms is that globally optimal partitions of the data are obtained by dynamic programming procedure, rather than locally optimal partition by iterative optimization. The value of initial cluster centers are computed by dividing total sample sets by K.

Because this algorithm is derived from the manner in which cluster centers are sequentially updated, this is locally optimal partition. Gray-level values by frequency of occurrence is presented in Fig.3. The result for segmentation of GMS-3 cloud C4 window is presented in Fig.4 and 5.

3. CLASSIFICATION

The algorithm for segmentation and classification of infrared GMS-3 cloud window is presented in Fig.2.

At the first stage of the algorithm, all samples of satellite images in which the maximum infrared reading is less than or equal to 100 are classified as low clouds.

The segmentation procedure for a value of K=10 is then applied to all other samples.

At the second stage of the algorithm, threshold values are used to determine possible cloud-type within each of the remaining samples.

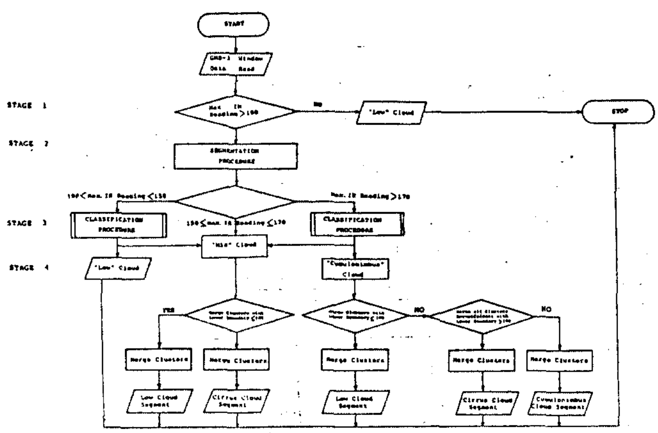

Fig.2. Automatic clustering algorithm

The third stage of the algorithm is the classification procedure. Samples for which the maximum infrared reading lay between 100 and 150 were tested to determine if they contained cumulonimbus clouds.

Samples for which the maximum infrared readings lay between 150 and 170 were classified as mix clouds which contained both low cloud types and cirrus cloud types.

The classification procedure classified each window into one of three categories:

(1) LOW-window containing only cumulus, stratus and stratocumulus clouds.

(2) MIX-windows containing cirrus and lower clouds.

(3) CUMULONIMBUS-windows containing cumulonimbus and lower clouds.

The texture feature which was used to determine the cloud-type category of the coldest could segment in a window was edge strength per unit area. Edge strength per unit area E was defined as the average value of the Roberts gradient over all points within a given area. [6]

For the classification procedure, the value of the Roberts Gradient at a point A in a 2 x 2 array

A	B
C	D

was approximated by the quantity

$$\max(|A - D|, |B - C|) \quad \text{------------(4)}$$

At the fourth stage of the decision procedure, clusters were merged together using a temperature threshold value to determine low cloud segments and cirrus cloud segments. The cumulonimbus cloud segment of samples which were classified as cumulonimbus by the classification procedure consisted of the coldest cluster.

The low cloud segment was obstained by merging together all clusters for which the minimum infrared observation within the cluster was less than or equal to 100.

Classification result of C4 window by K-means and Fisher's algorithm is presented in Fig.6 and.7, respectively.

4. COMPUTER SIMULATION

The automatic clustering algorithm was tested on 44 sample sets of 88 infrared windows of GMS-3 satellite images.

The temporal resolution of the satellite images is 30 minutes and spatial resolution is 4x4 km.

The image fixed geographical locations ranging from 15N to 55N latitude and from 100E to 140E longitude is consisted of 1434 x1078 registered resolution.

The image dimension of 1434x1078 array is computed by partition into 128x128 array windows. Partition infrared images into 88 windows is shown in Fig.8.

Classification results for each windows are shown in table 1.

TABLE 1. CLASSIFICATION RESULT

K-MEAN(FISHER)

WD	C M L	M I X	L O W	WD	C M L	M I X	L O W
A1	14(15)	31(30)	54(55)	A2	15	27	57
B1	11(5)	29(63)	60(32)	B2	11	10	79
C1	15(23)	44(53)	41(24)	C2	1	18	81
D1	0(16)	49(36)	51(48)	D2	9	15	76
E1	1(2)	11(2)	89(96)	E2	9	42	48
F1	11(2)	17(23)	72(75)	F2	30	36	34
G1	8(9)	12(14)	80(77)	G2	15	23	61
H1	2(3)	11(11)	86(86)	H2	1	16	83
I1	0(1)	34(32)	66(67)	I2	0	10	90
J1	4(6)	14(6)	81(88)	J2	2	10	92
K1	2(3)	22(46)	76(52)	K2	2	33	65
A4	8(24)	49(0)	43(24)	A3	1	11	87
B4	55(0)	34(81)	11(0)	B3	2	7	91
C4	27(27)	41(41)	32(32)	C3	3	13	84
D4	0(2)	13(0)	87(2)	D3	6	27	67
E4	0(0)	66(100)	34(0)	E3	0	66	34
F4	41(0)	57(100)	2(0)	F3	22	65	13
G4	32(0)	62(97)	6(0)	G3	0	94	6
H4	14(3)	52(71)	34(3)	H3	3	31	66
I4	2(3)	34(16)	64(3)	I3	0	33	67
J4	3(5)	31(0)	66(5)	J3	6	0	94
K4	NO CLOUD WINDOW			K3	1	64	35

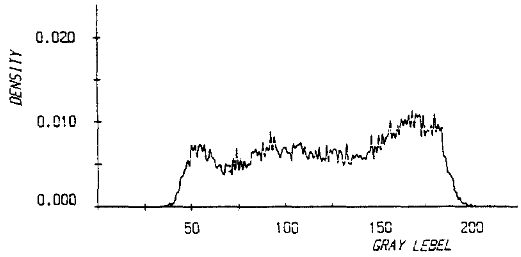

Fig.3.Histogram of C4 window

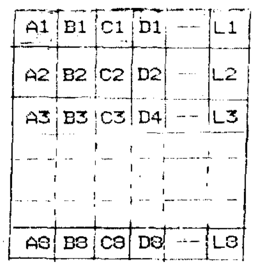

Fig.8. Windows of GMS-3 image

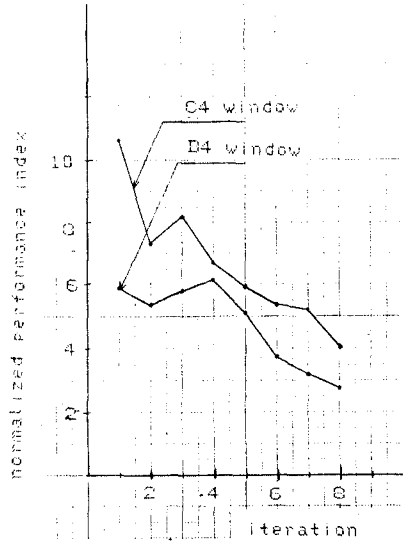

Fig.4.Relation of performance index and iteration

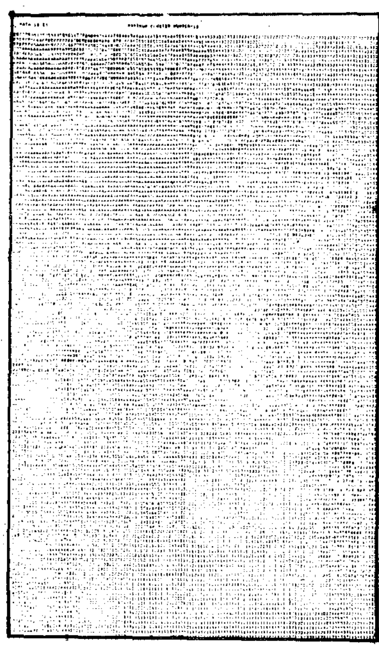

Fig.5. Segmentation of C4 window by K-means

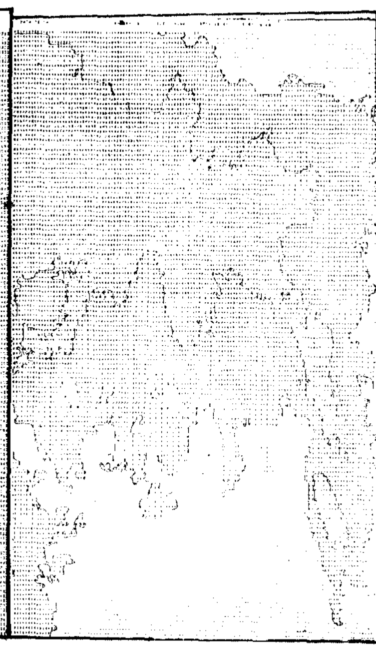

Fig.6.Classification of C4 window by K means

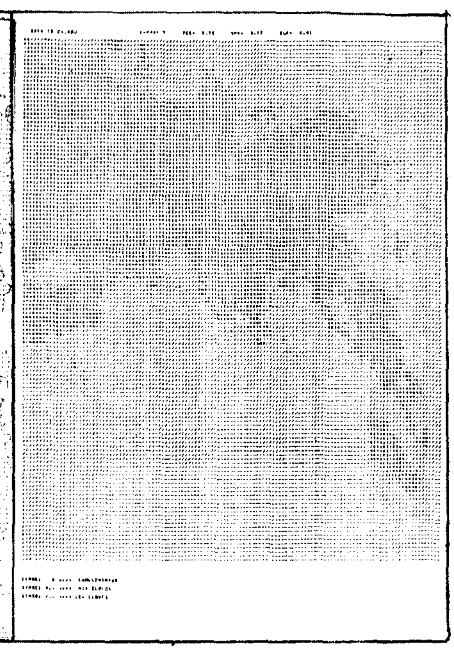

Fig.7.Classification of C4 window by Fisher's algorithm

5. CONCLUSION

The algorithm for cloud-type segmentation and classification based on an examination of clusters in the infrared histogram and features extracted both from the window and from the clusters.

A texture feature was used to determine whether or not a cluster represented a distinct cloud object or should be merged with other clusters to form a cloud-type object. Statistical pattern recognition techniques were conbined with the segmentation technique.

We, by K-means algorithm, could well distinguish mix clouds, which were difficult to be distinguished from another clouds by Fisher's algorithm.

To be locally optimal partitions by iterative optimization procedures of K-means algorithm is verified by minimizing of performance index. In this study we examined infrared images only. However there are two wavelength bands in GMS-3 images: visible(0.55~0.7 μm) and infrared(10.5~12.6 μm). The variation of brightness in visible images reveals the reflective character of objects and provides information on cloud thickness. The brightness of the infrared images is proportional to the temperature of the objects in the field of view and can be translated cloud-top height.[2]

If future studies treat visible and infrared satellite images, it is possible to determine rain rate by means of stastical pattern recognition.

REFERENCE

[1] J.A.Parikh, A.Rosefeld, "Automatic segmentation and classification of infrared meteorogical satellite data," IEEE Trans. on S.M.C., vol.SMC-8, no.10, pp.736~743, OCT. 1978.

[2] B.G.Lee, R.Chin & D.W.Martin, "Automated rain-rate class. of satellite images using statistical P.R.," IEEE Trans. on geo. and remote sensing, vol.GE-23, no.3, pp. 315~323, MAY 1985.

[3] J.Tou, R.Gonzalez, Pattern recognition principles, Addison-Wesley, 1974.

[4] E.C.Barrett & D.W.Martin, The use of satellite data in rain fall monitoring, New York: Academic. 1981.

[5] 김은간, "기상위성 영상의 자동분할 및 분류에 관한 연구" 항공 대학원 석사학위논문. 1986.

[6] R.O.Duda and P.E.Hart, Pattern class. and scene analysis, New York: Wiley, 1973.

[7] W.K.Pratt, Digital image processing, John Wiley and son, 1987.