Satellite Data Reduction Using Entropy–preserved Image Compression Technique

Li Jun (李 俊), *Zhou Fengxian* (周凤仙)
Institute of Atmospheric Physics, Academia Sinica, Beijing 100029, China
and *Gao Qinghuai* (高清怀)
Information Science Center, Peking University, Beijing 100871, China
Received October 9, 1990; revised November 21, 1990

ABSTRACT

In this paper, three techniques, line run coding, quadtree DF (Depth–First) representation and H coding for compressing classified satellite cloud images with no distortion are presented. In these three codings, the first two were invented by other persons and the third one, by ourselves. As a result, the comparison among their compression rates is given at the end of this paper. Further application of these image compression technique to satellite data and other meteorological data looks promising.

I. INTRODUCTION

With the progress of the image processing and pattern recognition, the atmospheric information including satellite observation data stored as images became more and more, which stimulated the birth and development of the study on image database. When an image database is established, the first problem to be solved is how to compress the images efficiently, this is very important for transmitting information to remote users. In this paper, some image coding methods are put forward for compressing the classified satellite cloud images. The coding methods are generally classified into two classes: one makes the image distorted when the compressed image is decoded, and another makes no distortion. The first class of compression methods (Musmann, 1984; Kunt et al., 1985) is applied to the images on which the values of the pixels are intensity of light, radiation, etc. The second class, being also called entropy–preserved coding, is suitable for the classified images on which the values of pixels are class numbers. In this paper, we will discuss the second class of compression technique applied to classified satellite cloud images such as these from NOAA / AVHRR and GMS / VISSR.

There are many entropy–preserved coding methods, all of which are based on the following principles: the probability of equality of the neighbor pixels′ values is larger than that of the non–neighbor pixels′ values, for example, the line run coding considers the coincidence of the value of a pixel with those of its right neighbor and left neighbor; the quadtree coding uses the equality of the values of the pixels in a quadrant; the H coding we invented is based on both the equality of the values of the pixels in every quadrant and the one of the values of the 4–adjacent pixels or 4–adjacent homogeneous quadrants.

In the second session, we will give the coding methods of line run, quadtree DF representation and H coding, and in the third session, the comparison between their compression rates is presented.

	1	2	3	4
1	1	1	0	0
2	1	1	0	0
3	0	1	1	0
4	0	1	1	1

Fig.1

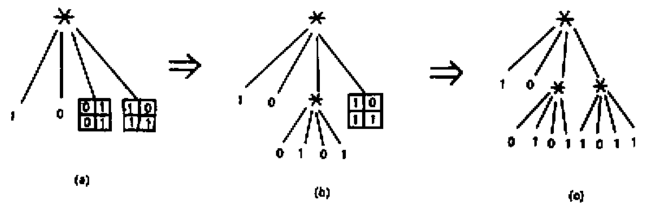

Fig.2 Generating process of quadtree

II. THE CODING TECHNIQUES

1. *Line Run Coding*

For given an image $I(x,y)$, where $x=1,2,\cdots,M$, $y=1,2,\cdots,N$, we can obtain its line run coding line by line. For xth line, if $y=1,2,\cdots,y_l$ and $I(x,y)=I(x,l)$, then we use the pair $(I(x,l),y_l)$ to represent the pixels $I(x,1),\cdots,I(x,y_l)$. Generally, we call $(C_{x_1},L_{x_1}),\cdots,(C_{x_k},L_{x_k})$ the run code or the xth line of image, if it satisfies both conditions:

$$C_{x_i}=I(x,\text{sum}_i) \quad \text{and} \quad I(x,k)=I(x,\text{sum}_i)$$

for $k=\text{sum}_{i-1}+1,\ \text{sum}_{i-1}+2,\cdots,\ \text{sum}_i$, where $\text{sum}_i=\sum_{j=1}^{i}L_j$, $\text{sum}_0=0$, $\text{sum}_{k_x}=N$, $i=1,2,\cdots,k_x$.

And we define the string $(C_{11},L_{11}),(C_{12},L_{12}),\cdots,(C_{1k_1},L_{1k_1}),(C_{21},L_{21}),(C_{22},L_{22}),\cdots,(C_{2k_2},L_{2k_2}),\cdots,(C_{M1},L_{M1}),(C_{M2},L_{M2}),\cdots,(C_{Mk_M},L_{Mk_M})$ the line run code of the image I. For example, the line run code of the 4×4 2–value image I(4,4) (see Fig.1) is (1,2), (0,2), (1,2), (0,2), (0,1), (1,2), (0,1), (0,1), (1,3).

2. *The Quadtree DF (Depth–First) Representation*

For given an image $I(2^n,2^n)$, where n is an integer larger than 0, we contineously divide it into equally sized quadrants, subquadrants, $\cdots$, until homogeneous blocks (possibly pixels) are obtained. For describing easily, we call the quadrant and subquadrant of I in the same term of quadrant in the following description. We record the dividing process using the generating process of a tree with four out-degree whose every leaf node represents a quadrant in which the pixels have a same value to label this node, as shown in Fig.2 (The example image is shown in Fig.1), Fig. 2(c) is the quadtree representation of the image.

There are many different quadtree representations, such as QuadCode (Li et al., 1984), DF representation (Kawaguchi et al., 1980) and CD (color and depth) representation (Shi, 1986), each of them has its own merit and shortcoming: the quadCode representation facilitates image processing but requires much more storage, while the DF representation has the highest compression rate, and the CD coding has much higher compression rate than QuadCode representation's and more facilitates image processing than DF coding does. Because we only consider data compression, the DF coding was chosen.

If we visit every node of a quadtree in the depth-first order and only record the value labelling the node (the inner node is labelled by " * "), we will obtain a string, being the quadtree's DF representation. For example, the DF representation of image show in Fig.1 is * 10 * 0101 * 1011 (Also refer to Fig.2(c) where the corresponding quadtree is illustrated).

3. *H Coding*

To present the H coding, we define an order to the pixels in the image. When we arrange all the pixels into this order, every four pixels following a pixel, a big pixel consisted of the four pixels order arranged, are 4-adjacent with the big pixel in the raster image, and the pixels in a quadrant Q correspond with a continuous string which is followed by the string corresponding to one of the 4-adjacent quadrants of Q.

We construct this order recursively.

For given an image I whose size is $2^n \times 2^n$,

1) suppose Q_0, Q_1, Q_2 and Q_3 are the four quadrants of I whose size is $2^{n-1} \times 2^{n-1}$, we define, $Q_0 < Q_1 < Q_3 < Q_2$, or (0,1,3,2) for short (see Fig.3).

2) suppose Q' is a quadrant in I whose size is $2^j \times 2^j$ where $j < n$, and Q'_0, Q'_1, Q'_2 and Q'_3 are four quadrants of Q' with size $2^{j-1} \times 2^{j-1}$. The order of Q'_0, Q'_1, Q'_2, Q'_3 is written as (a_0, a_1, a_2, a_3) which is a permutation of (0,1,2,3), then the order of the four quadrants of Q'_{a_0} is defined as (a_0, a_3, a_2, a_1), the order of the four quadrants of Q'_{a_1} is defined as (a_0, a_1, a_2, a_3), the order of the four quadrants of Q'_{a_2} is defined as (a_0, a_1, a_2, a_3), and the order of the four quadrants of Q'_{a_3} is defined as (a_2, a_1, a_0, a_3),

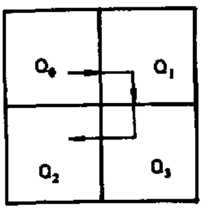

Fig.3

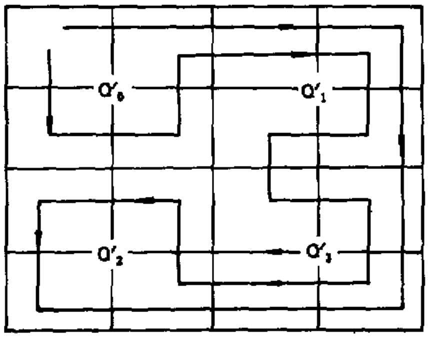

Fig.4 illustration of this rule.

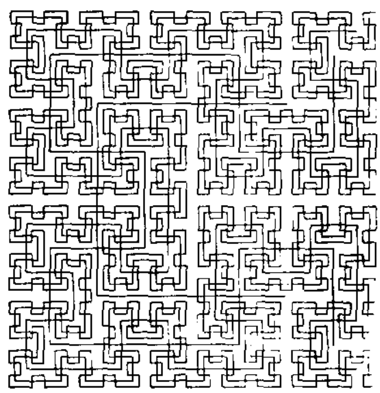

Fig.5. Hilbert curves $H_1 \cdots H_5$.

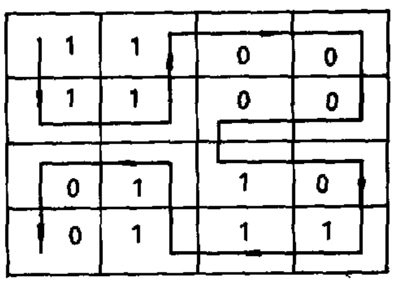

Fig.6

3) the subquadrants inherit the order of its father quadrants, that is, if Q'_1 is a quadrant of quadrant Q_1, Q'_2 is a quadrant of quadrant Q_2, and $Q_1 < Q_2$, then $Q'_1 < Q'_2$.

When we visit every pixel in this order, we draw an interesting curve (see Fig.5) which was invented by the great mathematician Hilbert, so we call this order Hilbert order or H–order for short.

If we arrange all the pixels into H–order, their coordinates make a line $(x_1,y_1)(x_2,y_2)\cdots(x_{2^{2n}},y_{2^{2n}})$. We call the string $I(x_1,y_1)I(x_2,y_2)\cdots I(x_{2^{2n}},y_{2^{2n}})$ the *H*–order representation of image *I*, and the run coding $(C_1,L_1)(C_2,L_2)\cdots(C_m,L_m)$ of $I(x_1,y_1)I(x_2,y_2)\cdots I(x_{2^{2n}},y_{2^{2n}})$ is called the *H* code of image *I*. For example, Fig.6 defines the *H*–order of the 4×4 image shown in Fig.1, and the *H* code of this image is (1,4)(0,4)(1,1)(0,0)(1,4)(0,2).

III. EXPERIMENTAL RESULTS AND ANALYSIS

Because the amount of satellite observed data is very large, it is necessary to extract useful information from satellite imagery to form productive images of all kinds, and methods have been put forward to obtain classified images and feature images (Liljas, 1982; 1986; Li et al., 1990). The classified images and feature images not only have less bits than the original satellite data have, but also are very convenient to be transmitted to remote users when they are compressed by using coding methods described above. To show the coding efficiency of the three methods to satellite imagery, we did some experiments on a group of classified satellite cloud images, these classified cloud images were obtained from GMS / VISSR by using fuzzy clustering method based on texture feature analysis (Zhou et al., 1990). The size of the images is 256×256, the values of the pixels are less than 16, the experiments were computed by MicroVAX and it will spend about one minute in coding and then decoding a classified image. The results are listed in Table 1, where the coding efficiency is measured by

$C = \text{codelength} / 2^{2n}$, if we suppose the image size is $2^n \times 2^n$.

Table 1. Results of Coding Techniques Applied to Satellite Classified Images

C \ Coding / Image	*H*–coding	DF Representation	Line Run Coding
Image 1	0.178	0.228	0.179
Image 2	0.182	0.230	0.183
Image 3	0.208	0.248	0.204
Image 4	0.200	0.246	0.183

From Table 1, it can be seen that the results of compression of using H–coding and line run coding methods are more efficiently than that of using quadtree DF representation when these methods are applied to satellite cloud images, but the DF representation has been widely used in GIS (Geography Information System), and shown many advantages in geoscience information processing. The application of DF representation to atmospheric information should be carefully studied. On the other hand, the advantages and disadvantages of each method applied to satellite images are still not clear, in general, they depend on the structure of an image to be coded. Theoretical analysis and more practical experiments should be carefully carried out in order to determine which method is more effective in specific satellite data reduction.

IV. CONCLUSIONS

The application of entropy–preserved image compression techniques to classified images or feature images can reduce large amount of satellite imagery data, thus it becomes a convenient and rapid approach to transmit the main information of satellite imagery to the remote stations for various purposes. Further studies should be focused on applying these compression techniques to other atmospheric observation data (include conventional meteorological observation data). Compression techniques applied to the original satellite imagery will also be carried out in the near future in the Institute of Atmospheric Physics, Chinese Academy of Sciences.

REFERENCES

Cornillon, P. et al. (1988), Processing, compression and transmission of satellite IR data for near–real–time use at sea, *J. Atmos.and Ocean. Tech.*, **5**: 320–327.

Kawaguchi, E. and T. Endo (1980), *On a method of binary–picture representation and its application to data compression*, *IEEE Trans. on PAMI*, **PAMI–2**: 77–86.

Kunt, M. et al. (1985), Second–generation image–coding techniques, *Proc. IEEE*, **73**: 547–574.

Li Jun and Zhou Fengxian (1990), Computer identification of multispectral satellite cloud imagery, *Adv. Atmos. Sci.* **7**: 366–375.

Li Jun and Zhou Fengxian (1990), Delineation of mesoscale features of ocean on satellite IR image, *Adv. Atmos. Sci.* **7**: 423–432.

Li Jun and Zhou Fengxian (1990), Cloud motion estimation from visual and IR data of geosynchronous satellite using fourier technique. XXVIII COSPAR.

Li, S. X. and M. H. Leow (1984), The quad code and its application, *Proc. 7th ICPR*, 227–229.

Liljas, E. (1982), *Automated techniques for the analysis of satellite cloud imagery*, *Nowcasting*, K. A. Browning, Academic Press, 177–190.

Liljas, E. (1986), Use of the AVHRR 3.7 micrometer channel in multispectral cloud classification, SMHI, 1–23.

Musmann, H. G. (1984), Predictive image coding, in: Image transmission techniques, Edit. by W. K. Pratt, Academic Press.

Shi, Q. Y. (1986), Image processing using CD representation, *Proc. 8th ICPR*, 320–322.

Wirth, N. (1982), "Algorithm+Data Structure=Program".

Zhou Fengxian and Li Jun (1990), Identification of deep convective clouds from GMS / IR image using texture feature analysis method. (Submitted to AAS).