

# Image Classification Using a Stochastic Model that Reflects the Internal Structure of Mixels

Asanobu KITAMOTO<sup>1</sup> and Mikio TAKAGI<sup>2</sup>

<sup>1</sup> R&D Department, National Center for Science Information Systems (NACSIS),  
3-29-1, Otsuka, Bunkyo-ku, 112-8640 Tokyo, Japan

kitamoto@rd.nacsis.ac.jp

<sup>2</sup> Department of Applied Electronics, Science University of Tokyo,  
2641, Yamazaki, Noda-shi, 278-8510 Chiba, Japan

takagi@te.noda.sut.ac.jp

**Abstract.** This paper proposes new ideas for the classification of images with the presence of *mixels*, or mixed pixels. Based on the internal structure of mixels, we first propose a stochastic model called *area proportion density*, and we demonstrate that Beta distribution is an appropriate model for this density. Next, based on the linear model of a mixel, we derive another stochastic model called *mixel density*. This model is then incorporated into the mixture density model of the image histogram, and we show the peculiar flat shape of this model works particularly effective for image histograms with long tail. Finally we present experiments on satellite imagery, and the goodness-of-fit of the proposed model is evaluated from the viewpoint of information criterion.

## 1 Introduction

The basic assumption of most conventional image processing algorithms is that an image pixel is an *atom*, which is the smallest element of the image not dividable any further. However, in reality, an image pixel corresponds to a certain region in the real world, and the region is not always homogeneous. If the resolution of the sensor is coarser than the scale of spatial variation in the real world, even a single pixel consists of multiple constituents. Such heterogeneous pixel with underlying internal structure is called a *mixed pixel* or a *mixel*, whereas a homogeneous pixel containing only one constituent is called a *pure pixel*.

Fundamental differences of statistical properties between mixels and pure pixels must be recognized for the analysis of images with the presence of mixels. Therefore the purpose of this paper is to propose new stochastic models, such as *area proportion density* or *mixel density*, for the proper handling of mixels from the viewpoint of probability theory. These models are rarely noticed in previous studies on mixels, but in this paper, we discuss the importance, characteristics and derivation of the models. Since one of the main application of this paper is the analysis of cloud mixels on satellite images, which are taken with coarse resolution sensors, we apply the proposed method to the classification of satellite images, and the effectiveness of the method is quantitatively evaluated from the viewpoint of information criterion.

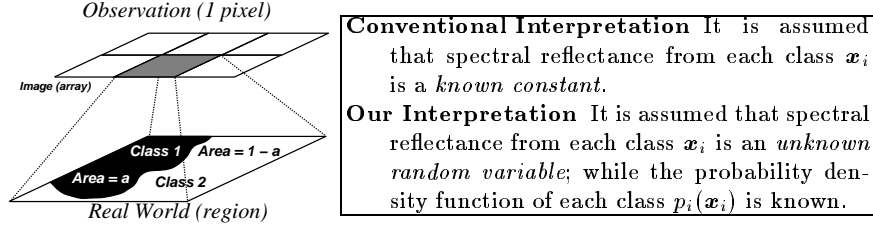


Fig. 1. The mixel model and the interpretation of the mixel model

## 2 The Model of a Mixel

Suppose observing the real world by a sensor with  $p$  different bands. The pixel value of each pixel is then represented by a  $p$  dimensional observation vector  $\mathbf{r} = (r_1, \dots, r_p)^t$ . In this paper, a mixel containing  $k$  classification classes is called a  $k$ -class mixel. For example, Fig. 1 illustrates a 2-class mixel. Moreover, spectral reflectance originating in each classification class within a  $k$ -class mixel is denoted as  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^t$ , and the area proportion of each classification class is denoted as  $a_i$ , where  $0 < a_i < 1$  and  $\sum_{i=1}^k a_i = 1$  is satisfied.

In general, it is natural to assume that the pixel value of a mixel is the weighted sum of spectral reflectance originating in each class  $\mathbf{x}_i$ , weighted by the function of area proportions. Based on this assumption, this paper adopts the simplest model, namely the *linear model* as follows:

$$\mathbf{r} = \sum_{i=1}^k a_i \mathbf{x}_i + \varepsilon, \quad (1)$$

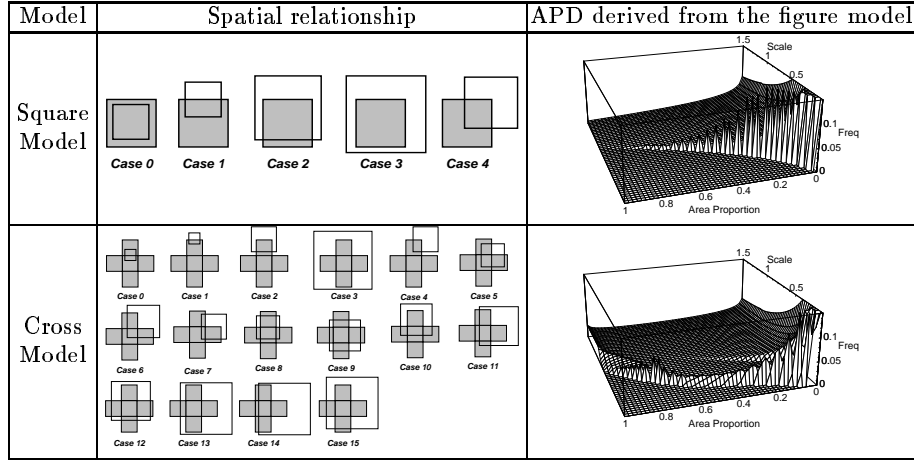
where the first term is the linear combination of spectral reflectance weighted by area proportions, and the second term represents noise independent of classes.

However, because of the limited space of the paper, we focus on the most representative case, 2-class mixel case observed with 1-band. The linear model in this case is represented as follows:

$$r = ax_1 + (1 - a)x_2 + \varepsilon, \quad (2)$$

where the area proportion of class 1 is denoted as  $a$ .

Here, the most important distinction between our interpretation and conventional interpretation (for example [1], [2], [3]) on Eq. (1) is summarized in Fig. 1. The assumption that  $\mathbf{x}_i$  are known for all mixels, as conventional interpretation assumes, is based on the assumption that  $\mathbf{x}_i$  are constants. In other words, the sources of variation with regard to the pixel value of mixels are assumed to be confined in class-independent noise term  $\varepsilon$ . In reality, however, radiation from each class shows some variation due to the causes of its own, and the sum of such variations results in the variation of pixel value of mixels. Therefore, the basic assumption of this paper is that " $\mathbf{x}_i$  are random variables, and the variation of radiation from each class contributes to the variation of pixel value of mixels."



**Fig. 2.** Spatial relationship between the figure model and the observing pixel, and APD derived from the figure model

We now think of a probability density function (PDF) of each classification class  $p_i(\mathbf{x}_i)$ , which we call a pure pixel density (PPD), and develops a new theory based on the probability theory. From this standpoint,  $x_i$  in Eq. (2) becomes a random variable with PDF  $p_i(x_i)$ , and four variables in Eq. (2) are unknown; the only known variable is the pixel value of a mixel  $r$ . This indicates that the solution of  $a$  can be obtained from Eq. (2) through the analysis of PDF whose corresponding random variables appear on the right-hand side of Eq. (2). We start with investigating the stochastic model of  $a$ , which we call *area proportion density* (Sect. 3). We next examine the stochastic model appearing on the left-hand side, which we call *mixel density* (Sect. 4). After proposing two types of distributions, we discuss a method for estimating the area proportion  $a$  from the observation value  $r$  based on expectation (Sect. 5).

### 3 Area Proportion Density Derived From the Internal Structure of Mixels

Area proportion density (APD) of mixels is a stochastic model closely associated with the internal structure of pixels. This density was sometimes investigated by studies on meteorology or remote sensing; however, only empirical distribution was proposed in those works. We, on the other hand, derive APD analytically using such simple figure models as the square model and the cross model representing a convex figure and a concave figure respectively. Possible spatial relationship between the figure model and the observation pixel is illustrated in Fig. 2. Supposing that the observation pixel takes any relative position to the figure model, we can think of a PDF  $f(a)$  of area proportion  $a$  ( $0 < a < 1$ ), where area proportion is the percentage of area of the figure model captured inside the observation pixel.

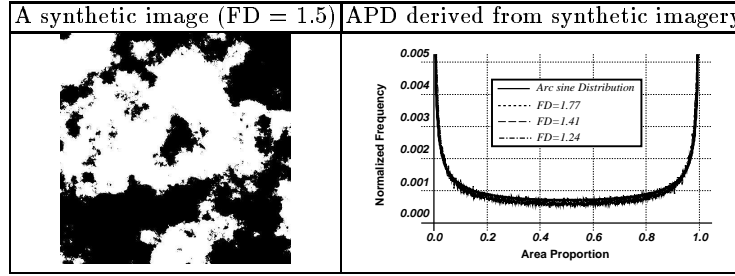


Fig. 3. APD derived from fractal synthetic imagery (FD : Fractal Dimension)

For example, in case 4 of the square model, we can derive cumulative area proportion density  $F(a)$  in an exact analytical form as follows:

$$F(a) = a - a \log a. \quad (3)$$

Due to the limited space of this paper, we omit the detail of analytical derivation process and illustrate only the final result of APD in Fig. 2. In this figure, the APD of particular observation scale is the intersection along the "area proportion" axis. Although this figure looks complex, APD in a particular scale shows the variety of shape such as L-shape, J-shape and uniform-shape.

Second we derive APD from simulation using synthetic images. Since one of the main application of this paper is the analysis of cloud mixels on satellite images, and cloud shape can be described using fractal, we adopt fractal synthetic imagery as a simulation model. First successive random addition algorithm [4] is used to create synthetic images, and created images are then "virtually" observed in block basis to simulate low resolution sensors. APD obtained through this protocol is illustrated in Fig. 3. This U-shaped density is very well described by *Arc sine distribution* [5], [6], known in the field of random walk, described by:

$$f(a) = \frac{1}{\pi \sqrt{a(1-a)}}. \quad (4)$$

We finally propose Beta distribution [7] as the model of APD because of its capability for describing all types of distributions derived thus far. Beta distribution is represented as follows:

$$f(a) = \frac{1}{B(\alpha, \beta)} a^{\alpha-1} (1-a)^{\beta-1}, \quad (5)$$

where  $0 \leq a \leq 1$ ,  $\alpha > 0$ ,  $\beta > 0$ , and  $B(\alpha, \beta) = \int_0^1 a^{\alpha-1} (1-a)^{\beta-1} da$  is beta function. This distribution represents a wide range of distributions by changing two parameters, such as uniform distribution at  $(\alpha, \beta) = (1, 1)$  and Arc sine distribution at  $(\alpha, \beta) = (0.5, 0.5)$ . Furthermore, the mean  $\mu_B$  and the variance  $\sigma_B^2$  of Beta distribution is calculated as  $\mu_B = \frac{\alpha}{\alpha+\beta}$ , and  $\sigma_B^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ . This means that if  $\alpha = \beta$ ,  $\mu = 0.5$  and the distribution becomes symmetric. Otherwise, two parameters  $\alpha$ ,  $\beta$  are related to the convexity and concavity of boundary shape, and can be used as shape or textural attributes of an image.

## 4 Mixel Density

Incorporating APD proposed in Eq. (5), and assuming the PDF of  $x_i$  in Eq. (2) as normal distribution, we can now calculate the stochastic model of  $r$ , which appears on the left-hand side of Eq. (2). First, we discuss the PDF of  $r$  by fixing the value of  $a$ . If we assume that PPD  $p_i(x_i)$  of all classes are known, then we can calculate a new PDF  $p(r; a)$  by referring to a formula for calculating the PDF of the sum of random variables [6]. If we omit the noise term in Eq. (2), we get the solution as follows:

$$p(r; a) = \int_{-\infty}^{\infty} q_1(r-t)q_2(t)dt = q_1(r) * q_2(r), \quad (6)$$

where  $q_1(r) = p_1(r/a)$  and  $q_2(r) = p_2(r/(1-a))$ , and  $*$  denotes convolution. The noise-included version of  $p(r; a)$  can be easily obtained by just convolving the PDF of the noise term  $p_\epsilon(r)$  as follows  $p(r; a) = q_1(r) * q_2(r) * p_\epsilon(r)$ .

For instance, if PPD is represented by normal distribution  $N(\mu_i, \sigma_i^2)$ , and the PDF of the noise term is also assumed to be normal distribution  $N(0, \sigma_\epsilon^2)$ , then  $p(r; a)$  also becomes normal distribution  $N(\mu_a, \sigma_a^2)$  with the following parameters:

$$\begin{aligned} \mu_a &= a\mu_1 + (1-a)\mu_2, \\ \sigma_a^2 &= \{a\sigma_1\}^2 + \{(1-a)\sigma_2\}^2 + \sigma_\epsilon^2, \end{aligned} \quad (7)$$

The fundamental reason that  $p(r; a)$  can be so easily derived in this case originates in the fact that normal distribution is a member of *stable distributions*. Moreover, even in the case that PPD is not represented by normal distribution, we can derive  $p(r; a)$  by means of *characteristic functions* on continuous case, or FFT (Fast Fourier Transform) on discrete case.

We now consider the set of mixels present in the image as a population. In this context, area proportion  $a$  should be regarded as an event of a random variable defined in the parameter space  $\Psi$ . Then *mixel density* (MD)  $d(r)$ , which gives the probability of the pixel value of a mixel taking  $r$  independent of the value of area proportion  $a$ , is derived by the following equation:

$$d(r) = \int_{\Psi} f(a)p(r; a)da, \quad (8)$$

To facilitate readers' understanding on Eq. (8), we introduce the simplest case in which radiation from each class is assumed to be exactly the same for all mixels; that is the case that the variance within each class is 0 and PPD is represented by delta function  $p_i(r) = \delta(r - \mu_i)$ . In this case,  $p(r; a)$  is calculated as  $p(r; a) = \delta(r - \{a\mu_1 + (1-a)\mu_2\})$ . To further simplify the derivation process, we assume APD to be uniform distribution  $f(a) = 1$ . Then the exact form of MD is obtained as the following:

$$d(r) = \int_0^1 f(a)p(r; a)da = \begin{cases} \frac{1}{\mu_2 - \mu_1} & \mu_1 < r < \mu_2 \\ 0 & r \leq \mu_1, r \geq \mu_2 \end{cases}. \quad (9)$$

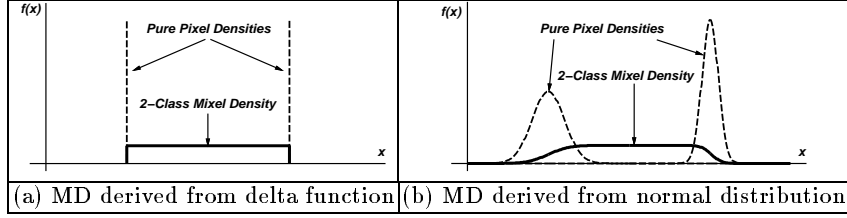


Fig. 4. Examples of 2-class mixel density with 1-band

This exact solution represents uniform distribution, as shown in Fig. 4 (a), with the interval  $(\mu_1, \mu_2)$ . We show another example in Fig. 4 (b), when PPD is represented by normal distribution.

As demonstrated above, MD is not a density derived from the simple summation of PPD; rather derived from integrating convolved PPD using Eq. (6). In other words, MD is the consequence of *spatial quantization effect* of digitized imagery. Therefore, MD should be regarded as a new stochastic model that represents a fundamentally different stochastic process compared to PPD. The significant characteristics of this model may be summarized as follows.

1. MD extends between peaks of a set of related PPD, and its flat shape is very peculiar and hardly used as a PDF.
2. Since the shape of MD is determined by both the shape parameters of a set of related PPD and APD, MD itself does not have its own shape parameters.

In particular, the former characteristic is important, since there are few statistical image classification techniques that utilizes PDF of this shape. For image histograms with long-tail flat regions between keen peaks, the idea of MD suggests a new interpretation that peak regions and flat regions correspond to PPD and MD respectively, as shown in Fig. 4.

## 5 Expected Area Proportion

Finally we proceed to the estimation of area proportion appeared in Eq. (2). We propose a new method of estimating area proportion as the expectation of  $a$  in terms of  $p(r; a)$ , which we call *expected area proportion*, as follows:

$$a_e = \int_{\Psi} af(a)p(a|r)da, \quad (10)$$

where  $p(a|r)$  is a conditional PDF obtained by  $p(a|r) = \frac{p(r;a)}{p(r)} = \frac{p(r;a)}{\int_{\Psi} p(r;a)da}$ .

Again we introduce the simplest case to demonstrate the concrete calculation of expected area proportion. Assuming that PPD is represented by delta function, we can calculate expected area proportion  $a_e$  based on Eq. (10) using the form of  $p(r; a)$  shown previously:

$$a_e = \int_0^1 af(a)p(a|r)da = \frac{\mu_2 - r}{\mu_2 - \mu_1}, \quad (11)$$

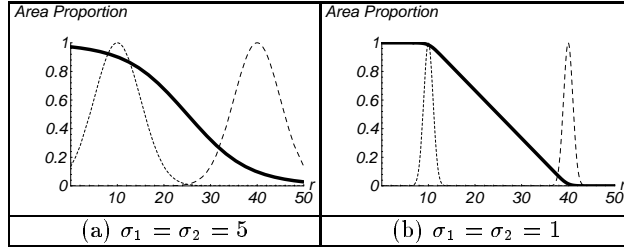


Fig. 5. The expected area proportion of 2-class mixels with 1-band

where we assume  $\mu_1 < r < \mu_2$  without losing generality. This exact solution shows that area proportion is linearly proportional to the pixel value of a mixel.

Next we assume PPD to be normal distribution. The result illustrated in Fig. 5 shows that as the variance of PPD becomes smaller, the graph of expected area proportion changes from a curve to a linear function. This result is the natural extension of Eq. (11) because normal distribution with zero variance is identical to delta function.

## 6 Mixture Density Estimation and Image Classification

To apply our proposed method, it is required to estimate all the PDF from image data. Hence we regard the image histogram as a set of  $M$  number of PDF, namely *finite mixture density*, as follows:

$$p(r|\phi, \alpha) = \sum_{i=1}^M \alpha_i p_i(r|\phi_i), \quad (12)$$

where  $p_i(r|\phi_i)$  is the PPD with parameters  $\phi_i$ , and  $\alpha_i$  is a mixture where  $\alpha_i > 0$  and  $\sum_{i=1}^M \alpha_i = 1$ . Here important distinction between conventional models and ours must be noticed; the number of classification classes is not identical to the number of PDF appeared in Eq. (12). That is because we have to consider MD *in addition to* PPD to incorporate the effect of mixels. After estimating parameters of PPD by applying EM algorithm for mixture density estimation [8], [9], we apply Bayes decision rule for pixel-based image classification. Because of the limited space, we address the detail of the method in another paper [10].

The images used in the experiments are infra-red images observed by AVHRR sensor (Channel 4) on meteorological satellite NOAA. These images are mainly filled with streak clouds around Japan in winter, which meteorological condition likely produces many mixels on the image. First we show the result of mixture density estimation in Fig. 6. In all cases, the mixture density model shows good fitness to a long-tail image histogram because of the adoption of MD, and we can observe especially good fit when Beta distribution is used as APD. This goodness-of-fit is then quantitatively evaluated using Akaike Information criterion (AIC) [11]. The figures at the bottom of Fig. 6 illustrate relationship

between Beta distribution parameters and AIC. These figures show that optimal  $\alpha$  and  $\beta$  are dependent on images; however our mixture density estimation method found nearly optimal parameters as shown in the middle of Fig. 6.

Finally we describe the result of image classification in Fig. 7. In all cases, satellite images consist of large number of mixels, which spread around pure pixels in a planar form. Our method works effective both when the number of sea class pixels dominates as (a),(b) and vice versa as (c). This result shows the wide applicability and robustness of our proposed method.

## 7 Conclusion

This paper proposed an image classification method considering the presence of mixels on an image. A new stochastic model *area proportion density* were proposed so that they reflect the underlying internal structure of mixels. Moreover, *mixel density* is derived from the linear model of mixels with the concrete derivation process, and its characteristics, peculiarity of shape is summarized. Furthermore, our proposed method of estimating area proportion of mixels called *expected area proportion*, together with other proposed models, was applied to the classification of satellite images. We believe this method paves the way to the direct modelling of complex image histogram based on the probability theory.

## References

1. Settle, J.J. and Drake, N.A.: Linear Mixing and the Estimation of Ground Cover Proportions. *Int. J. Remote Sensing*, Vol. 14, No. 6, pp. 1159–1177, 1993.
2. Foody, G.M. and Cox, D.P. Sub-Pixel Land Cover Composition Estimation Using a Linear Mixture Model and Fuzzy Membership Functions. *Int. J. Remote Sensing*, Vol. 15, No. 3, pp. 619–631, 1994.
3. Bosdogianni, P., Petrou, M., and Kittler, J.: Mixed Pixel Classification in Remote Sensing. In *SPIE Proceedings of the EUROPTO series, Image and Signal Processing for Remote Sensing*, Vol. 2315, pp. 494–505, 1994.
4. Voss, R.F.: Random Fractal Forgeries. In Earnshaw, R.A., editor, *Fundamental Algorithms for Computer Graphics*, pp. 805–835. Springer-Verlag, 1985.
5. Feller, W.: An Introduction to Probability Theory and Its Applications, Vol. 1. John Wiley & Sons, Inc., 1957.
6. Feller, W.: An Introduction to Probability Theory and Its Applications, Vol. 2. John Wiley & Sons, Inc., 1966.
7. Falls, L.W.: The Beta Distribution: A Statistical Model for World Cloud Cover. *Journal of Geophysical Research*, Vol. 79, No. 9, pp. 1261–1264, 1974.
8. Redner, R.A. and Walker, H.F.: Mixture Densities, Maximum Likelihood and the EM algorithm. *SIAM Review*, Vol. 26, No. 2, pp. 195–239, 1984.
9. Schlesinger, M.I.: Interdependence of Learning and Self-Learning in Pattern Recognition. *Cybernetics*, No. 2, pp. 81–87, 1968.
10. Kitamoto, A. and Takagi, M.: A Stochastic Model of Mixels and Image Classification. In *Proc. of 13th ICPR*, pp. 745–749, Wien, 1996.
11. Akaike, H.: Information Theory and an Extension of the Maximum Likelihood Principle. In Petrov, B.N. and Csaki, F., editors, *2nd Inter. Symp. on Information Theory*, pp. 267–281, 1973.



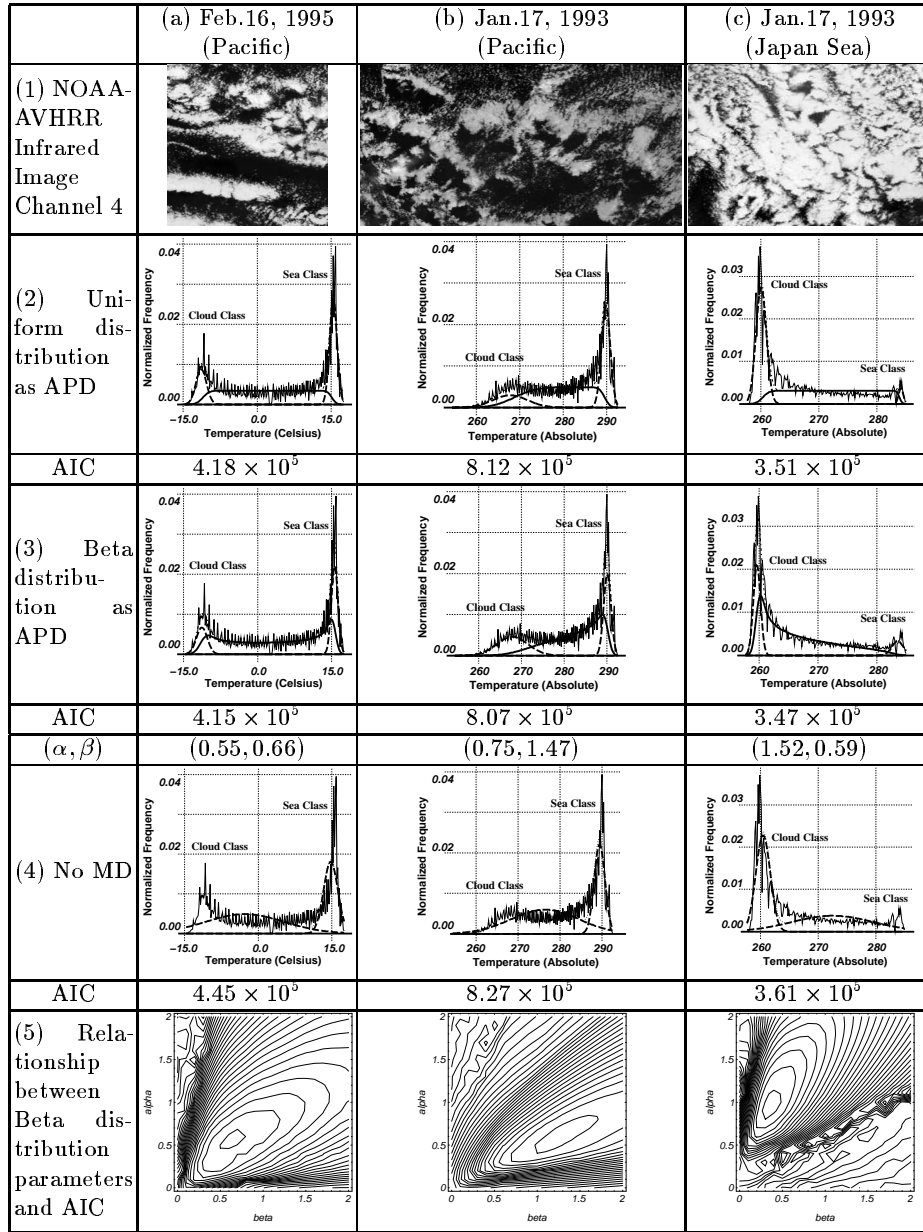


Fig. 6. The result of mixture density estimation applied to the image histogram (thin solid line: histogram, bold solid line: MD, bold dotted line: PPD), and the bottom line is the relationship between Beta distribution parameters and AIC. The ellipses in those figure correspond to the optimal region in terms of AIC

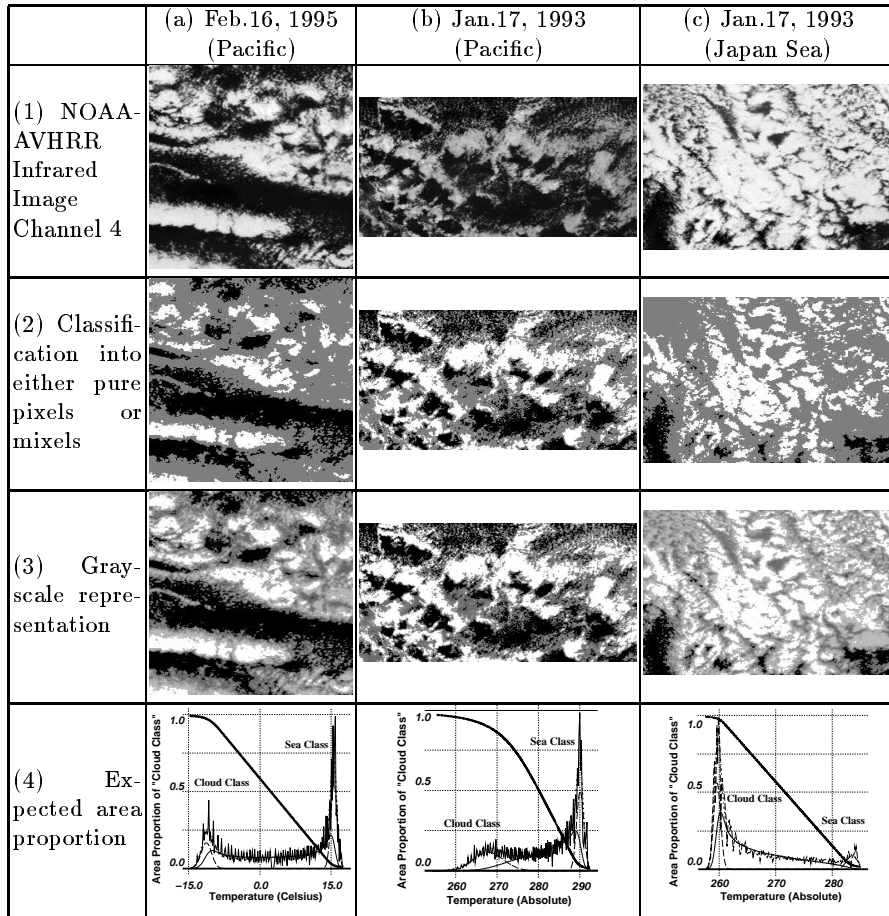


Fig. 7. The result of image classification. In (2), white means cloud, black means sea and gray means a mixel of both; whereas in (3), gray means cloud, black means sea, and mixels are shown with gray level as the whiter, the bigger the area proportion of cloud. In (4), bold solid line represents the area proportion of "cloud class"