

# The Moments of the Mixel Distribution and Its Application to Statistical Image Classification

Asanobu Kitamoto<sup>1</sup>

National Institute of Informatics (NII)  
Room 1516, 2-1-2, Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan  
kitamoto@nii.ac.jp

**Abstract.** The mixel is a heterogeneous pixel that contains multiple constituents within a single pixel, and the statistical properties of a population of mixels can be characterized by the mixel distribution. Practically this model has a drawback that it cannot be represented in closed form, and prohibitive numerical computation is required for mixture density estimation problem. Our discovery however shows that the “moments” of the mixel distribution can be derived in closed form, and this solution brings about significant reduction of computation cost for mixture density estimation after slightly modifying a typical algorithm. We then show the experimental result on satellite imagery, and find out that the modified algorithm runs more than 20 times faster than our previous method, but suffers little deterioration in classification performance.

## 1 Introduction

The mixel, or the mixed pixel, is the “amalgam” of multiple constituents contained within a single pixel. Because of the finite resolution of sensors we use to observe the real world, a heterogeneous region, as well as a homogeneous region, is scanned as a single pixel. Hence we should regard the digital imagery as the spatially quantized (sampled) representation of the real world, and take for granted the presence of mixels. When the resolution of sensors is relatively coarse to the scale of objects in the real world, for example in remote sensing or medical images, the presence of mixels is usually inevitable.

To characterize the statistical properties of the mixels based on the probability theory, we can think of the probability distribution function (PDF) of mixels, or the mixel distribution. It is basically a phantom distribution not corresponding to any objects in the real world, because the mixture of multiple constituents takes place through the process of observation. The author insists that this implicit distribution should be taken into account explicitly for the statistical analysis of images. Our findings on the closed form moments of the standard mixel distribution (SMD) then opens the way to the efficient computation of mixture density estimation including the SMD.

The organization of the paper is as follows. Firstly, Sect. 2 introduces related works on mixels and also states several definitions. Secondly, Sect. 3 describes main results of the paper, namely the closed form moments of the SMD. Then

Sect. 4 explains the mixture density estimation modified for the efficient computation of the SMD. Next Sect. 5 illustrates experimental results on satellite imagery, and finally Sect. 6 concludes the paper.

## 2 Mixel Models

**Related Works.** Past literature on mixels mainly focused on the estimation of area proportions for each mixel from observed pixel values. Methods can be grouped into several approaches; for instance, image geometry-based methods [1], probability model-based methods [2, 3], fuzzy model-based methods [4, 5], linear projection-based methods [6], and non-linear regression-based methods [7]. The approach used in this paper shares some motivation with the probability model-based methods in that area proportions are estimated from parametric PDF models of constituent classes. However, the most important distinction of our method from those methods lies in the introduction of the mixel distribution. This model has not been paid much attention, probably because this distribution is hidden, and affects only when the percentage of mixels is relatively high. However, the mixel distribution shows interesting properties different from conventional PDF models [8, 9, 10], and hence requires special treatment for its efficient computation. Next we summarize definitions required afterwards.

**Definition 1 (Linear Model of the Mixel).** A “ $K$ -class mixel” is a pixel that contains  $K$  number of constituent classes. Let the area proportion of constituent class  $C_i$  be  $a_i \in (0, 1)$ . From the definition of the area proportion, the sum of area proportions for all constituent classes amounts to unity; namely  $\sum_{i=1}^K a_i = 1$ . Moreover let the radiation from class  $C_i$  be a  $D$ -dimensional vector  $\mathbf{x}_i$ . Then we hypothesize that the observed radiance of the mixel  $\mathbf{x}$  is represented by the linear combination of the radiance from the constituents. Neglecting the common noise term for simplifying the argument, we have

$$\mathbf{x} = \sum_{i=1}^K a_i \mathbf{x}_i . \quad (1)$$

**Definition 2 (Definition of the Mixel Distribution).** From (1), it is clear that the mixel distribution consists of two types of PDF models; namely a random vector  $\mathbf{x}_i$  drawn from the PDF of constituent class  $C_i$  represented by  $p(\mathbf{x}_i|C_i; \boldsymbol{\psi}_i)$  with parameters  $\boldsymbol{\psi}_i \in \Psi$ , and another random vector  $\mathbf{a} = (a_1, \dots, a_{K-1}) \in A$  drawn from the PDF of area proportions represented by  $f(\mathbf{a}; \phi)$ , where  $A = \{(a_1, \dots, a_{K-1}) | \sum_{i=1}^{K-1} a_i < 1, a_i > 0\}$  is the space of area proportions, and  $a_K = 1 - \sum_{i=1}^{K-1} a_i$ . Then the “ $D$ -band  $K$ -class mixel distribution” can be derived from  $K$  number of class-conditional  $D$ -dimensional multivariate PDF and one  $(K-1)$ -dimensional multivariate PDF of area proportions [8].

**Definition 3 (Definition of the SMD).** The term “standard mixel distribution” (SMD) is coined for representing a special mixel distribution derived

only from normal distributions (the PDF of constituent class) and the Beta distribution (the PDF of area proportions). Formally, the PDF of class  $C_i$  is a multivariate normal distribution:

$$p(\mathbf{x}|C_i; \boldsymbol{\psi}_i) = N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = \frac{1}{(2\pi)^{d/2} |\boldsymbol{\Sigma}_i|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) \right\}, \quad (2)$$

where  $\mathbf{x}$  is a  $D$ -dimensional column vector,  $\boldsymbol{\mu}_i$  is the  $D$ -dimensional mean vector,  $\boldsymbol{\Sigma}_i$  is the  $D$ -by- $D$  covariance matrix,  $(\mathbf{x} - \boldsymbol{\mu}_i)^T$  is the transpose of  $(\mathbf{x} - \boldsymbol{\mu}_i)$ ,  $\boldsymbol{\Sigma}_i^{-1}$  is the inverse of  $\boldsymbol{\Sigma}_i$ , and  $|\boldsymbol{\Sigma}_i|$  is the determinant of  $\boldsymbol{\Sigma}_i$ . On the other hand, the PDF of area proportions is the  $(K - 1)$ -dimensional Beta distribution:

$$f(\mathbf{a}; \boldsymbol{\phi}) = \frac{\Gamma(\phi_1 + \phi_2 + \cdots + \phi_K)}{\prod_{i=1}^K \Gamma(\phi_i)} \prod_{i=1}^K a_i^{\phi_i - 1} = \frac{\Gamma(\varphi)}{\prod_{i=1}^K \Gamma(\phi_i)} \prod_{i=1}^K a_i^{\phi_i - 1}, \quad (3)$$

where  $\boldsymbol{\phi}$  is a  $D$ -dimensional parameter vector with  $i$ -th component  $\phi_i > 0$  for all  $i$ ,  $\varphi = \sum_{i=1}^K \phi_i$ , and  $\Gamma(\cdot)$  is the Gamma function.

### 3 Central Moments of the SMD

**Derivation of the SMD.** First we choose a certain fixed vector for area proportions. Then (1) represents the linear combination of random vectors drawn from normal distributions. Because the normal distribution is a stable distribution, the mixel distribution also becomes the normal distribution with the following parameters:

$$\boldsymbol{\mu}_a = \sum_{i=1}^K a_i \boldsymbol{\mu}_i, \quad \boldsymbol{\Sigma}_a = \sum_{i=1}^K a_i^2 \boldsymbol{\Sigma}_i, \quad (4)$$

If we further assume mutual independence between random vectors  $\mathbf{x}_i$  and the random vector  $\mathbf{a}$ , the randomization of  $N(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a)$  with the a priori distribution of area proportions yields the SMD  $M(\mathbf{x})$  as follows [8]:

$$M(\mathbf{x}) = \int \int_A f(\mathbf{a}; \boldsymbol{\phi}) N(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a) d\mathbf{a}. \quad (5)$$

**Central Moments of the SMD.** Now we start to derive the moments of the SMD (5). Unfortunately the SMD itself cannot be represented in closed form because the integral involved in (5) cannot be solved. However, the central moments, namely the mean vector and the covariance matrix can be calculated in closed form, which is the main results of this paper. For the concise presentation of the paper, we will describe the detailed proof in the Appendix.

**Proposition 1.** *The mean vector  $\boldsymbol{\mu}_M$  of the SMD is represented by:*

$$\boldsymbol{\mu}_M = \frac{\sum_{i=1}^K \phi_i \boldsymbol{\mu}_i}{\sum_{i=1}^K \phi_i}. \quad (6)$$

**Proposition 2.** *The covariance matrix  $\Sigma_M$  of the SMD is represented by:*

$$\Sigma_M = \frac{\sum_{i=1}^K \phi_i(\phi_i + 1) \Sigma_i + \left\{ \sum_{i=1}^K \frac{\phi_i(\varphi - \phi_i)}{\varphi} \boldsymbol{\mu}_i \boldsymbol{\mu}_i^T - \sum_{i=1}^K \sum_{\substack{j=1 \\ j \neq i}}^K \frac{\phi_i \phi_j}{\varphi} \boldsymbol{\mu}_i \boldsymbol{\mu}_j^T \right\}}{\varphi(\varphi + 1)} . \quad (7)$$

In practice, above equations are applied in simpler forms with smaller number of dimensions and constituent classes. Hence we shall investigate several representative cases for the one-band two-class SMD. In this case, the area proportion distribution is equivalent to one-dimensional Beta distribution:

$$f(a; \phi) = \frac{\Gamma(\phi_1 + \phi_2)}{\Gamma(\phi_1)\Gamma(\phi_2)} a^{\phi_1-1} (1-a)^{\phi_2-1} . \quad (8)$$

where  $a_1 = a$  and  $a_2 = 1 - a$ . Then the moments of the mixel distribution can be derived from (6) and (7) with substituting  $\boldsymbol{\mu}_i \rightarrow \mu_i$ , and  $\Sigma_i \rightarrow \sigma_i^2$ .

$$\mu_M = \frac{1}{\phi_1 + \phi_2} (\phi_1 \mu_1 + \phi_2 \mu_2) , \quad (9)$$

$$\sigma_M^2 = \frac{\phi_1(\phi_1 + 1)\sigma_1^2 + \phi_2(\phi_2 + 1)\sigma_2^2 + \frac{\phi_1\phi_2}{\phi_1 + \phi_2} (\mu_2 - \mu_1)^2}{(\phi_1 + \phi_2)(\phi_1 + \phi_2 + 1)} . \quad (10)$$

*Example 1.* The simplest SMD corresponds to the case when the PDF of constituent class  $C_i$  is the delta function  $p(x|C_i) = \delta(x - \mu_i)$ , and the PDF of area proportions is the uniform distribution  $\phi_1 = \phi_2 = 1$ . This is a special case since we can analytically solve (5) to obtain the closed form mixel distribution:

$$M(x) = \frac{1}{\mu_2 - \mu_1} , \quad (11)$$

where  $\mu_2 > \mu_1$  is assumed without losing generality. The mean and the variance of this uniform distribution can be trivially calculated:

$$\mu_M = \frac{\mu_1 + \mu_2}{2}, \quad \sigma_M^2 = \frac{1}{12}(\mu_2 - \mu_1)^2 . \quad (12)$$

These results can also be obtained by substituting  $\sigma_1^2 = \sigma_2^2 = 0$  and  $\phi_1 = \phi_2 = 1$  in (9) and (10), which shows this is a special form of the SMD.

*Example 2.* Let us consider more ‘‘practical’’ case when the PDF of constituent class  $C_i$  is a univariate normal distribution  $N(\mu_i, \sigma_i^2)$ , and the PDF of area proportions is the uniform distribution. Substituting (9) and (10) with corresponding parameters in this case, we obtain the moments as follows:

$$\mu_M = \frac{\mu_1 + \mu_2}{2}, \quad \sigma_M^2 = \frac{1}{3}(\sigma_1^2 + \sigma_2^2) + \frac{1}{12}(\mu_2 - \mu_1)^2 . \quad (13)$$

Here the variance  $\sigma_M^2$  consists of two terms; the first term represents the effect of variance from constituent classes, while the second term is affected by the distance between means. The latter is in particular unique in the mixel distribution, because it is not caused by the random noise, but caused by the mixture of different constituents. This term also corresponds to (12).

Fig. 1 illustrates the shape of one-band two-class SMD for representative five cases. In spite of different skewness and kurtosis of the SMD, all the SMDs similarly extend between peaks of the constituent classes' PDFs. On the other hand, the right table summarizes the comparison between theoretically derived moments (6), (7) and empirical moments calculated from the simulated data. Consistency in this table intuitively verifies theoretical moments derived above.

## 4 Mixture Density Estimation by EM algorithm

**Finite Mixture Density and EM algorithm.** Image classification method used in this paper is based on a typical algorithm; that is, **1)** Estimate finite mixture density from the statistics of the observed image, **2)** Classify each pixel based on Bayes decision rule. Here the mixture density model is as follows:

$$p(\mathbf{x}|\boldsymbol{\psi}) = \sum_{i=1}^M p(\mathbf{x}|C_i, \boldsymbol{\psi}_i)P(C_i) , \quad (14)$$

where  $M$  is the number of distributions, and  $P(C_i)$  is the mixing parameter. It should now be emphasized that in our model  $M$  includes both the class-conditional PDF (normal distributions) and the SMD (not normal distributions). Then the problem is to search for parameters in (14) that best describes the statistics of the observed image.

One of the standard iterative algorithm for solving this problem is the EM algorithm [11]. Conceptually this algorithm consists of two steps, namely E-step and M-step. The update rules of parameters can be written as follows:

$$P^{t+1}(C_i) = \frac{1}{N} \sum_{k=1}^N \frac{P^t(C_i)p(\mathbf{x}_k|C_i, \boldsymbol{\psi}_i^t)}{p(\mathbf{x}_k|\boldsymbol{\psi}^t)} , \quad (15)$$

$$\boldsymbol{\psi}^{t+1} = \arg \max_{\boldsymbol{\psi} \in \Psi} \sum_{k=1}^N \log p(\mathbf{x}_k|C_i, \boldsymbol{\psi}_i) \frac{P^t(C_i)p(\mathbf{x}_k|C_i, \boldsymbol{\psi}_i^t)}{p(\mathbf{x}_k|\boldsymbol{\psi}^t)} , \quad (16)$$

where the superscript  $t$  denotes  $t$ -th iteration. In addition, more efficient update rules are available for parameters of the normal distribution;

$$\boldsymbol{\mu}_i^{t+1} = \frac{\left\{ \sum_{k=1}^N \mathbf{x}_k \frac{P^t(C_i)p(\mathbf{x}_k|C_i, \boldsymbol{\psi}_i^t)}{p(\mathbf{x}_k|\boldsymbol{\psi}^t)} \right\}}{\left\{ \sum_{k=1}^N \frac{P^t(C_i)p(\mathbf{x}_k|C_i, \boldsymbol{\psi}_i^t)}{p(\mathbf{x}_k|\boldsymbol{\psi}^t)} \right\}} , \quad (17)$$

$$\boldsymbol{\Sigma}_i^{t+1} = \frac{\left\{ \sum_{k=1}^N (\mathbf{x}_k - \boldsymbol{\mu}_i^{t+1})(\mathbf{x}_k - \boldsymbol{\mu}_i^{t+1})^T \frac{P^t(C_i)p(\mathbf{x}_k|C_i, \boldsymbol{\psi}_i^t)}{p(\mathbf{x}_k|\boldsymbol{\psi}^t)} \right\}}{\left\{ \sum_{k=1}^N \frac{P^t(C_i)p(\mathbf{x}_k|C_i, \boldsymbol{\psi}_i^t)}{p(\mathbf{x}_k|\boldsymbol{\psi}^t)} \right\}} . \quad (18)$$

**Approximation of the SMD.** However, efficient update rules (17), (18) or similar rules cannot be applied to the SMD because it is not a normal distribution nor cannot be represented in closed form. Numerical optimization (16) and

integral (5) requires heavy computation, so the incorporation of the SMD into mixture density model (14) has not been a practical choice.

However, in this paper, the results on the moments of the SMD play an important role in realizing efficient computation of mixture density estimation. The idea we try is very simple. That is, to exploit efficient update rules (17) and (18), we replace the SMD with the normal distribution  $N(\boldsymbol{\mu}_M, \boldsymbol{\Sigma}_M)$ ; in other words, we approximate the SMD with the normal distribution having identical mean and covariance.

A modified algorithm can be described as follows. **1)** All mixing parameters are updated by (15). **2)** Parameters of class-conditional distributions are updated by (17) and (18). **3)** Parameters of the mixel distributions are updated according to the theoretical moments (6) and (7) with updated parameters obtained above. Note that this algorithm is not just for increasing the number of normal distributions. Parameters  $\boldsymbol{\mu}_M, \boldsymbol{\Sigma}_M$  are not freely chosen but calculated so that it best describes the corresponding mixel distribution.

Fig. 2 compares the original and the approximated version of the SMD. This approximation is reasonable because it is theoretically guaranteed that the approximated version has the identical mean and covariance to the original version. Although some of the information is lost from the original SMD, especially around regions of skewed peaks, it is shown from experiments (which are not shown here) that unless Beta distribution parameters are small, the deviation does not affect fatally in terms of classification performance.

## 5 Experimental Results

We demonstrate the result of image classification on satellite images. The original image Fig. 3(c) contains sea and cloud regions, whose meteorological condition strongly suggests that many mixels should be present on the image. Fig. 3 (a) and (b) shows the result of mixture density estimation using the EM algorithm. In terms of goodness-of-fit to the image histogram, (b) is superior to (a). However please note that this better result is obtained from 20 to 100 times longer computation time than (a) using the approximated SMD. Here “20 times faster” may be a little exaggerated comparison because the previous method is not efficient in terms of speed, but the proposed algorithm is in fact sufficiently fast for practical use like this problem. More importantly, Bayesian borders are almost same in both cases, which suggests that classification performance is nearly equal in both cases.

## 6 Conclusion

Hence the proposed algorithm is much faster and almost equally effective compared to our previous method. Thus we conclude that the approximation of the SMD using the moments is a promising way to develop our classification algorithm to a more sophisticated algorithm that can be applied to a wide range of applications. With the proposed classification method, the statistics of the

image can be analyzed more appropriately and accurately under the presence of mixels. Moreover, the effect of speed-up we showed in this paper becomes more important when we increase the dimension  $D$  to apply our methods to multi-spectral images. The method itself can be easily extended to multi-spectral case, so the extension in this direction is practically the most important future work.

## References

- [1] Tabatabai, A.J. and Mitchell, R. Edge Location to Subpixel Values in Digital Imagery. *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol. 6, No. 2, pp. 188–201, 1984.
- [2] Santago, P. and Gage, H.D. Quantification of MR Brain Images by Mixture Density and Partial Volume Modeling. *IEEE Trans. Med. Img.*, Vol. 12, No. 3, pp. 566–574, 1993.
- [3] Choi, H.S., Haynor, D.R., and Kim, Y. Partial Volume Tissue Classification of Multichannel Magnetic Resonance Images — A Mixel Model. *IEEE Trans. Med. Img.*, Vol. 10, No. 3, pp. 395–408, 1991.
- [4] Wang, F. Fuzzy Supervised Classification of Remote Sensing Images. *IEEE Trans. Geo. Remote Sens.*, Vol. 28, No. 2, pp. 194–201, 1990.
- [5] Kent, J.T. and Mardia, K.V. Spatial Classification Using Fuzzy Membership Models. *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol. 10, No. 5, pp. 659–671, 1988.
- [6] 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.
- [7] Foody, G.M. Relating the Land-Cover Composition of Mixed Pixels to Artificial Neural Network Classification Output. *Photo. Eng. Rem. Sens.*, Vol. 62, No. 5, pp. 491–499, 1996.
- [8] Kitamoto, A. and Takagi, M. Image Classification Using Probabilistic Models that Reflect the Internal Structure of Mixels. *Patt. Anal. Appl.*, Vol. 2, No. 2, pp. 31–43, 1999.
- [9] Kitamoto, A. and Takagi, M. Image Classification Using a Stochastic Model that Reflects the Internal Structure of Mixels. In Amin, A., Dori, D., Pudil, P., and Freeman, H., editors, *Advances in Pattern Recognition*, Vol. 1451 of *Lecture Notes in Computer Science*, pp. 630–639. Springer, 1998.
- [10] Kitamoto, A. and Takagi, M. Area Proportion Distribution — Relationship with the Internal Structure of Mixels and its Application to Image Classification. *Syst. Comp. Japan*, Vol. 31, No. 5, pp. 57–76, 2000.
- [11] 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.

## A Proofs

**Corollary 1.**

$$\iint_A \frac{\Gamma(\varphi)}{\prod_{j=1}^K \Gamma(\phi_j)} \left( \prod_{j=1, j \neq i}^K a_j^{\phi_j - 1} \right) a_i^{\phi_i - 1 + 1} d\mathbf{a} = \frac{\phi_i}{\varphi} \quad (19)$$

$$\iint_A \frac{\Gamma(\varphi)}{\prod_{j=1}^K \Gamma(\phi_j)} \left( \prod_{j=1, j \neq i}^K a_j^{\phi_j - 1} \right) a_i^{\phi_i - 1 + 2} d\mathbf{a} = \frac{\phi_i(\phi_i + 1)}{\varphi(\varphi + 1)} \quad (20)$$

*Proof.* Because  $f(\mathbf{a}; \boldsymbol{\phi})$  is a probability density the following relationship holds:

$$\iint_A \prod_{i=1}^K a_i^{\phi_i-1} d\mathbf{a} = \frac{\prod_{i=1}^K \Gamma(\phi_i)}{\Gamma\left(\sum_{i=1}^K \phi_i\right)} = \frac{\prod_{i=1}^K \Gamma(\phi_i)}{\Gamma(\varphi)} \quad (21)$$

Then applying the following property of the Gamma function

$$\Gamma(n+1) = n\Gamma(n) \quad (22)$$

(19) can be calculated as follows:

$$\iint_A \left( \prod_{j=1, j \neq i}^K a_j^{\phi_j-1} \right) a_i^{\phi_i-1+1} d\mathbf{a} = \frac{\left( \prod_{j \neq i}^K \Gamma(\phi_j) \right) \Gamma(\phi_i+1)}{\Gamma(\varphi+1)} = \frac{\phi_i \prod_{j=1}^K \Gamma(\phi_j)}{\varphi \Gamma(\varphi)} \quad (23)$$

Then, by substituting (19) with (23), it is easy to show (19) holds. We can also derive (20) in the same manner.  $\square$

*Proof (of Proposition 1).* The mean vector of the standard mixel distribution is the first order central moment of  $M(\mathbf{x})$ , and it is calculated from the following formula:

$$\boldsymbol{\mu}_M = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{x} M(\mathbf{x}) d\mathbf{x} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{x} \left\{ \iint_A f(\mathbf{a}; \boldsymbol{\phi}) N(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a) d\mathbf{a} \right\} d\mathbf{x}$$

It is not possible to directly solve this integral with respect to  $\mathbf{a}$ . However, if we change the order of integral between  $\mathbf{a}$  and  $\mathbf{x}$  and first integrate (24) with respect to  $\mathbf{x}$ , this integral can be simplified as follows:

$$\begin{aligned} \boldsymbol{\mu}_M &= \iint_A f(\mathbf{a}; \boldsymbol{\phi}) d\mathbf{a} \int_{-\infty}^{\infty} \mathbf{x} N(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a) d\mathbf{x} \\ &= \iint_A f(\mathbf{a}; \boldsymbol{\phi}) d\mathbf{a} \int_{-\infty}^{\infty} (\mathbf{x} - \boldsymbol{\mu}_a + \boldsymbol{\mu}_a) N(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a) d\mathbf{x} \end{aligned} \quad (25)$$

Since the normal distribution is even and  $(\mathbf{x} - \boldsymbol{\mu}_a)$  is odd around  $\boldsymbol{\mu}_a$ ,

$$\int_{-\infty}^{\infty} (\mathbf{x} - \boldsymbol{\mu}_a) N(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a) d\mathbf{x} = 0 \quad (26)$$

Then (25) can be further simplified as follows:

$$\begin{aligned} \boldsymbol{\mu}_M &= \iint_A f(\mathbf{a}; \boldsymbol{\phi}) d\mathbf{a} \int_{-\infty}^{\infty} \boldsymbol{\mu}_a N(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a) d\mathbf{x} = \iint_A f(\mathbf{a}; \boldsymbol{\phi}) \boldsymbol{\mu}_a d\mathbf{a} \\ &= \iint_A \left( \frac{\Gamma(\varphi)}{\prod_{i=1}^K \Gamma(\phi_i)} \prod_{i=1}^K a_i^{\phi_i-1} \right) \left( \sum_{i=1}^K a_i \boldsymbol{\mu}_i \right) d\mathbf{a} \\ &= \sum_{i=1}^K \boldsymbol{\mu}_i \iint_A \frac{\Gamma(\varphi)}{\prod_{j=1}^K \Gamma(\phi_j)} \left( \prod_{j=1, j \neq i}^K a_j^{\phi_j-1} \right) a_i^{\phi_i-1+1} d\mathbf{a} \end{aligned} \quad (27)$$



Finally, (27) is integrated with respect to  $\mathbf{a}$  using (22),

$$\boldsymbol{\mu}_M = \sum_{i=1}^K \frac{\phi_i \boldsymbol{\mu}_i}{\varphi} = \frac{\sum_{i=1}^K \phi_i \boldsymbol{\mu}_i}{\sum_{i=1}^K \phi_i} \quad (28)$$

*Proof (of Proposition 2).* To obtain the covariance matrix, the second order moment around the mean vector, basically the same technique is applied; first integrating the equation with respect to  $\mathbf{x}$ , then with respect to  $\mathbf{a}$ . We solve the following equation:

$$\boldsymbol{\Sigma}_M = \iint_A f(\mathbf{a}; \boldsymbol{\phi}) d\mathbf{a} \iint_{-\infty}^{\infty} (\mathbf{x} - \boldsymbol{\mu}_M)(\mathbf{x} - \boldsymbol{\mu}_M)^T N(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a) d\mathbf{x} \quad (29)$$

Applying (26) into (29), we obtain a simplified form:

$$\begin{aligned} \boldsymbol{\Sigma}_M &= \iint_A f(\mathbf{a}; \boldsymbol{\phi}) d\mathbf{a} \iint_{-\infty}^{\infty} (\mathbf{x} - \boldsymbol{\mu}_a + \boldsymbol{\mu}_a - \boldsymbol{\mu}_M)(\mathbf{x} - \boldsymbol{\mu}_a + \boldsymbol{\mu}_a - \boldsymbol{\mu}_M)^T N(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a) d\mathbf{x} \\ &= \iint_A f(\mathbf{a}; \boldsymbol{\phi}) \boldsymbol{\Sigma}_a d\mathbf{a} + \iint_A f(\mathbf{a}; \boldsymbol{\phi}) (\boldsymbol{\mu}_a - \boldsymbol{\mu}_M)(\boldsymbol{\mu}_a - \boldsymbol{\mu}_M)^T d\mathbf{a} \end{aligned} \quad (30)$$

Splitting (30) into two terms, the first term can be calculated as follows:

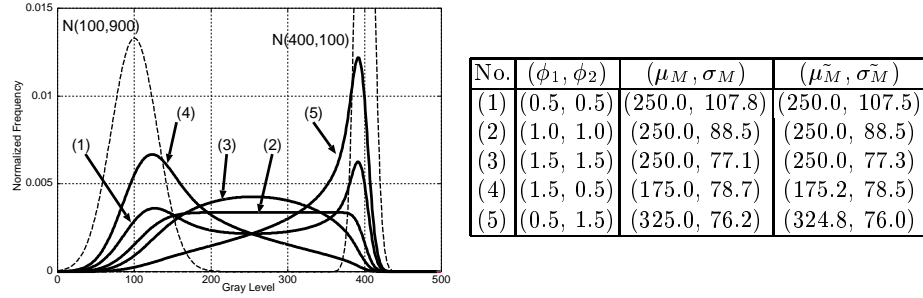
$$\begin{aligned} \iint_A f(\mathbf{a}; \boldsymbol{\phi}) \boldsymbol{\Sigma}_a d\mathbf{a} &= \iint_A \left( \frac{\Gamma(\varphi)}{\prod_{i=1}^K \Gamma(\phi_i)} \prod_{i=1}^K a_i^{\phi_i-1} \right) \left( \sum_{i=1}^K a_i^2 \boldsymbol{\Sigma}_i \right) d\mathbf{a} \\ &= \sum_{i=1}^K \boldsymbol{\Sigma}_i \iint_A \frac{\Gamma(\varphi)}{\prod_{i=1}^K \Gamma(\phi_i)} \prod_{j=1, j \neq i}^K a_j^{\phi_j-1} a_i^{\phi_i-1+2} d\mathbf{a} \\ &= \frac{1}{\varphi(\varphi+1)} \sum_{i=1}^K \phi_i(\phi_i+1) \boldsymbol{\Sigma}_i \end{aligned} \quad (31)$$

For the second term, we calculate the following relationship first:

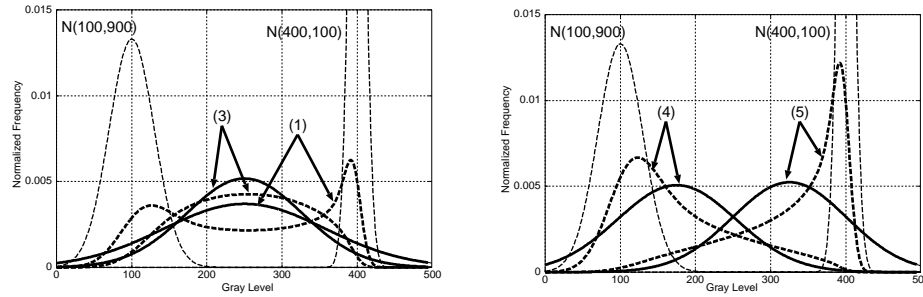
$$\boldsymbol{\mu}_a - \boldsymbol{\mu}_M = \sum_{i=1}^K \boldsymbol{\mu}_i a_i - \frac{\sum_{i=1}^K \phi_i \boldsymbol{\mu}_i}{\sum_{i=1}^K \phi_i} = \sum_{i=1}^K \left( a_i - \frac{\phi_i}{\sum_{i=1}^K \phi_i} \right) \boldsymbol{\mu}_i = \sum_{i=1}^K \left( a_i - \frac{\phi_i}{\varphi} \right) \boldsymbol{\mu}_i$$

Then, using (20), the second term can be simplified as follows:

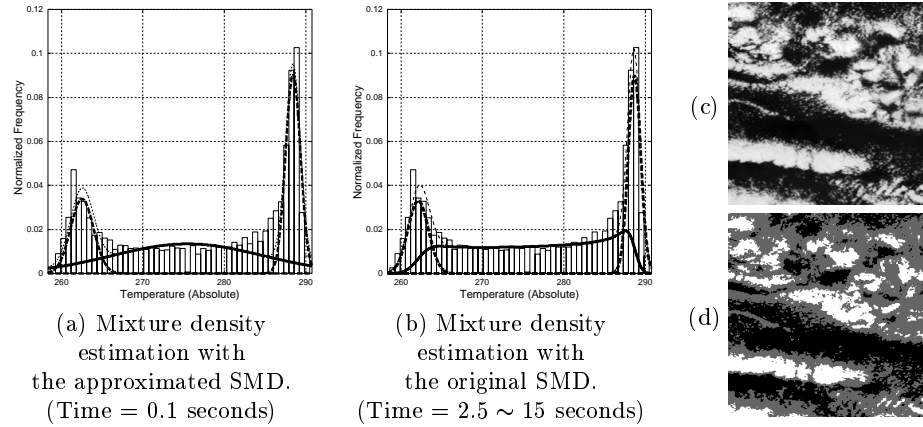
$$\begin{aligned} &\iint_A f(\mathbf{a}; \boldsymbol{\phi}) (\boldsymbol{\mu}_a - \boldsymbol{\mu}_M)(\boldsymbol{\mu}_a - \boldsymbol{\mu}_M)^T d\mathbf{a} \\ &= \iint_A \frac{\Gamma(\varphi)}{\prod_{i=1}^K \Gamma(\phi_i)} \prod_{i=1}^K a_i^{\phi_i-1} d\mathbf{a} \left\{ \sum_{i=1}^K \left( a_i - \frac{\phi_i}{\varphi} \right) \boldsymbol{\mu}_i \right\} \left\{ \sum_{i=1}^K \left( a_i - \frac{\phi_i}{\varphi} \right) \boldsymbol{\mu}_i^T \right\} \\ &= \sum_{i=1}^K \sum_{j=1}^K \boldsymbol{\mu}_i \boldsymbol{\mu}_j^T \iint_A \frac{\Gamma(\varphi)}{\prod_{i=1}^K \Gamma(\phi_i)} \prod_{i=1}^K a_i^{\phi_i-1} \left( a_i a_j - a_j \frac{\phi_i}{\varphi} - a_i \frac{\phi_j}{\varphi} + \frac{\phi_i \phi_j}{\varphi \varphi} \right) d\mathbf{a} \\ &= \sum_{i=1}^K \boldsymbol{\mu}_i \boldsymbol{\mu}_i^T \left\{ \frac{\phi_i(\phi_i+1)}{\varphi(\varphi+1)} - \frac{\phi_i \phi_i}{\varphi \varphi} \right\} + \sum_{i=1}^K \sum_{j=1, j \neq i}^K \boldsymbol{\mu}_i \boldsymbol{\mu}_j^T \left\{ \frac{\phi_i \phi_j}{\varphi(\varphi+1)} - \frac{\phi_j \phi_i}{\varphi \varphi} \right\} \\ &= \frac{1}{\varphi(\varphi+1)} \left\{ \sum_{i=1}^K \boldsymbol{\mu}_i \boldsymbol{\mu}_i^T \frac{\phi_i(\varphi - \phi_i)}{\varphi} - \sum_{i=1}^K \sum_{j=1, j \neq i}^K \boldsymbol{\mu}_i \boldsymbol{\mu}_j^T \frac{\phi_i \phi_j}{\varphi} \right\} \end{aligned} \quad (32)$$



**Fig. 1.** The standard mixel distributions for five cases. The left figure illustrates the shape of the SMD generated from same normal distributions  $N(100, 900)$  and  $N(400, 100)$ , and five different set of Beta distribution parameters shown in the right. Compare the moments obtained theoretically  $(\mu_M, \sigma_M)$  and empirically  $(\tilde{\mu}_M, \tilde{\sigma}_M)$ .



**Fig. 2.** Approximation of the SMD. Compare the original SMD (bold dotted line) to the approximated normal distribution (bold solid line). Beta distribution parameters for each case are identical to those used in Fig. 1.



**Fig. 3.** The result of image classification of satellite imagery. (a) and (b) shows the results of mixture density estimation, (c) is the original image, and (d) is the result of classification using (a), where white regions are cloud, black regions are sea and gray regions are mixels of both constituents. In (a),  $\phi_1 = \phi_2 = 0.5$  is fixed.