Mixture Density Estimation Under the Existence of Mixels

Asanobu Kitamoto Mikio Takagi
Institute of Industrial Science, University of Tokyo
7-22-1, Roppongi, Minato-ku, Tokyo, 106 Japan
Tel: +81-3-3402-6231 Ext. 2647 kitamoto@tkl.iis.u-tokyo.ac.jp

Abstract – This paper focuses on a mixed pixel or *mixel*. In the mixture density estimation problem, we include implicit distributions produced by mixels, in addition to the explicit distributions produced by pure pixels. NOAA-AVHRR ch.5 images are classified into several categories including mixel categories based on a Bayesian decision rule, and three optimization schemes are compared.

1 INTRODUCTION

If a single category entirely covers a pixel, this pixel is called a pure pixel; otherwise, it is called a mixed pixel or *mixel*. Since the ground resolution of NOAA-AVHRR data, which we will use in experiments, is approximately 1.1 km, it is natural to assume that one pixel is constituted by multiple ground-cover categories. In this case, pixel value is a function of the radiation from the mixture of component categories. Under the existence of mixels, as a result, a radiance distribution deviates from a distribution consisting of pure pixel distributions only. Mixels are thereby misclassified into one of the pure pixel categories, causing unsatisfactory results. To deal with this problem, many papers have been presented in remote sensing [1, 2], or medical imaging [3], for example.

Here we assume that the radiance distribution of an image is a parametric family, *finite mixture densities*

$$p(x|\Phi) = \sum_{i=1}^{K} \alpha_i p_i(x|\phi_i),\ x = (x_1, \cdots, x_n)^T \in R^n, \quad (1)$$

where each α_i is nonnegative and $\sum_{i=1}^{K} \alpha_i = 1$, and where each p_i is itself a density function parameterized by $\phi_i \in \Omega_i \subseteq R^{n_i}$, and K is the number of categories. We denote $\Phi = (\alpha_1, \cdots, \alpha_K, \phi_1, \cdots, \phi_K)$ and set

$$\Omega = \left\{ (\alpha_1, \cdots, \alpha_K, \phi_1, \cdots, \phi_K) : \sum_{i=1}^{K} \alpha_i = 1 \right. \text{ and } \left. \alpha_i \geq 0, \phi_i \in \Omega_i \text{ for } i = 1, \cdots, K \right\}. \quad (2)$$

Estimating Φ using a sample is referred to as the mixture density estimation problem [4].

In previous works, since only pure pixel distributions were considered in their model, mixels may be assigned to a proper category only through increasing the number of categories in a tentative way. We argue that mixture densities inherently contain *implicit* mixel distributions as well as *explicit* pure pixel ones. Since, in our model, mixel distributions are clearly separated from pure pixel ones, we can properly deal with the effects caused by mixels.

2 MIXEL MODEL

In this paper, only one channel is used, i.e. $x \in R^1$, and p is assumed to have a normal density $N(\phi_i)$:

$$p(x|\phi_i) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-(x-\mu_i)^2/2{\sigma_i}^2},\ \phi_i = (\mu_i, {\sigma_i}^2)^T \in R^2, \quad (3)$$

where p represents within-category variability caused by random changes in conditions. The use of more general distributions such as pearson system distributions [5] may be more accurate. However, a normal distribution is commonly used because of its simplicity and its useful mathematical properties such as *stability*, which we will refer to later.

Suppose that a pixel contains a fraction ρ of category C_i and the remainder $(1-\rho)$, is another category C_j. We call it a *two-category mixel* m_{ij}. Let p_i denote the probability density function (pdf) of category C_i. Let X_i denote a random variable whose pdf is p_i, which is assumed to be a normal distribution $N(\mu_i, {\sigma_i}^2)$ as stated earlier. Our mixel model further assumes that the computed radiation value r_{ij} of m_{ij} is a linear combination of two events:

$$r_{ij}(\rho) = \rho x_i + (1-\rho)x_j, \quad (4)$$

where ρ $(0 \leq \rho \leq 1)$ is the mixture proportion of category C_i in the mixel m_{ij}, and x_i is a possible value of X_i. Since x_i and x_j are both possible values of random variables X_i and X_j, r_{ij} is also a possible value of a random variable, which we denote M_{ij} and its pdf p_{ij}.

Usually the "shape" of p_{ij} is not known. However, if the pdfs of random variables X_i and X_j both belong to the same type of a *stable distribution*, then the pdf of a random variable $c_iX_i + c_jX_j$ $(c_i, c_j > 0)$ will also belong to the same type of distribution. In particular, if the pdfs of X_i and X_j are both normal distributions, the random variable M_{ij} also has a normal distribution as its pdf. p_{ij} is therefore a normal distribution $N(\mu_{ij}, {\sigma_{ij}}^2)$ with

$$\mu_{ij}(\rho) = \rho\mu_i + (1-\rho)\mu_j, \quad (5)$$

$${\sigma_{ij}}^2(\rho) = \rho^2{\sigma_i}^2 + (1-\rho)^2{\sigma_j}^2. \quad (6)$$

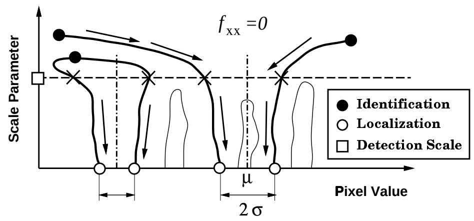


Figure 1: Scale-space representation of zero-crossings in the second derivative.

Thus the use of a stable distribution, in particular a normal distribution, greatly simplifies the problem.

Furthermore, we regard ρ as a uniform random variable in the interval [0,1]. Then p_{ij} with random ρ becomes:

$$p_{ij}(x) = \int_0^1 N(\mu_{ij}(\rho), {\sigma_{ij}}^2(\rho))d\rho = \int_0^1 \frac{1}{\sqrt{2\pi {\sigma_{ij}}^2(\rho)}} \exp\left[-\frac{(x-\mu_{ij}(\rho))^2}{2{\sigma_{ij}}^2(\rho)}\right] d\rho. \quad (7)$$

Note that Equation (7) does not have its own shape parameters; the shape of p_{ij} is determined only by α_{ij}.

We argue that mixture densities of K classification categories inherently contain ${}_K C_2 = K(K-1)/2$ implicit mixel distributions introduced above as well as K explicit pure distributions. Similarly, a *three-category mixel* distribution may be defined; however, to keep our model simple, we omitted mixel distributions whose components have more than three categories. In summary, finite mixture densities are modified from Equation (1) as follows:

$$p(x|\Phi) = \sum_{i=1}^{K} \alpha_i^P p_i^P(x|\phi_i^P) + \sum_{j=1}^{{}_K C_2} \alpha_j^M p_j^M(x|\phi_k^M, \phi_l^M), \quad (8)$$

where superscripts M and P denote mixel and pure pixel distributions respectively, and k, l specifies pure pixel distributions associated with a mixel distribution j.

3 ESTIMATION OF PARAMETERS USING SCALE-SPACE

The mixture density estimation problem requires a good initial estimate of parameters to attain a good local maximum. Many methods have been proposed, for example method of moments, but the method used in this paper is a scale-space based approach proposed by Carlotto[6]. The basic idea of scale-space is "coarse-to-fine" strategy, in which salient peaks or valleys are first identified in coarse scales and then localized in fine scales.

A useful property of scale-space representation with regard to a normal distribution is a "funnel-like" shape as Figure 1 illustrates. Since trajectories of zero-crossing (zc) in the second derivative ($f_{xx}=0$) shows such shape, distance between paired trajectories decreases as scale goes down to fine scales. In the end, the distance d reaches 2σ, where σ is the standard deviation of a normal distribution. In the case of mixture densities, however, this relationship is no more than an estimate because d may deviates from 2σ due to interactions with other densities. In the following a brief estimation procedure is presented:

1. Identify salient zcs and pair adjacent zcs.
2. Localize the position of paired zcs x_1 and x_2.
3. Measure distance d between paired zcs and calculate standard deviation by $\sigma = d/2$.
4. Calculate mean value μ by $\mu = (x_1 + x_2)/2$.
5. Calculate mixture proportion α by Area$(\mu - \sigma \le x \le \mu + \sigma) =$ Area$(x_1 \le x \le x_2) = 0.683\alpha$.

4 PARAMETER OPTIMIZATION

In this section, we introduce two approaches, maximum likelihood (ML) and distance minimization (DM), used to determine parameters. Assume that a particular $\Phi^* = (\alpha_1^*, \cdots, \alpha_n^*, \Phi_1^*, \cdots \Phi_n^*) \in \Omega$ is the "true" parameter value to be estimated. Suppose that $\{x_k\}_{k=1,\cdots,N}$ is an independent sample of N unlabeled observations on the mixture, i.e., a set of N observations on independent, identically distributed random variables with density $p(x|\Phi^*)$. Then the *log-likelihood function* is:

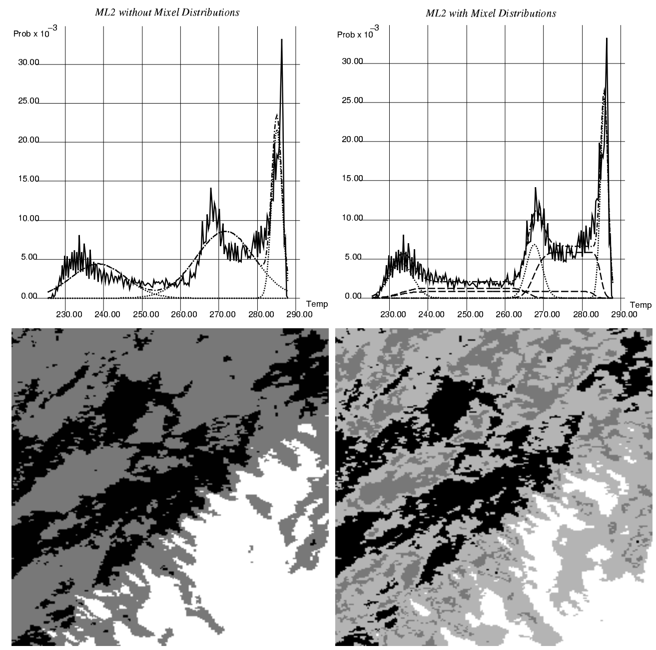
$$L(\Phi) = \sum_{i=1}^{N} \log p(x_k|\Phi), \quad (9)$$

where N is the size of an image in this paper. By a maximum likelihood estimate of Φ^*, we mean any choice of Φ in Ω at which the log-likelihood function $L(\Phi)$ attains its largest local maximum in Ω. However, as is often the case, N is too large to carry out naive computation of Equation (9) in reasonable time. In the following we will show two ideas for reducing the redundancy in Equation (9).

The first scheme (ML1) simply reduces the number of observations, N, by sampling pixels from an image with sampling rate r. Then the number of observation is reduced to $N_s = Nr$. The second scheme (ML2) makes use of the histogram of an image. Let histogram $h(z_i)$ has N_L level $\{z_1, \cdots, z_{N_L}\}$. In this case, the log-likelihood function is modified as follows:

$$L(\Phi) = \sum_{i=1}^{N_L} h(z_i) \log p(z_i|\Phi). \quad (10)$$

In spite of inherent quantization error, an advantage of this scheme may be the significant reduction of computation time. Estimated parameters are then iteratively optimized using EM algorithm [4].



(a) (b)

Figure 2: Mixture density estimation without (a) / with (b) mixel distributions. Dotted lines indicate estimated density functions, while a solid line describes the histogram. The white and black regions in image (a) correspond to the high cloud and sea categories, respectively.

Another scheme (DM) is based on a least-square fitting, considering the histogram to be a sampled version of a smooth pdf. Squared error E between $p(x|\Phi)$ and histogram is

$$E(\Phi) = \sum_{i=1}^{N_L} \left[p(z_i|\Phi) - h(z_i)\right]^2 , \quad (11)$$

where the definition of the histogram is identical to the previous one. Since minimization of E is a non-linear optimization problem, we used simulated annealing.

5 RESULTS

Based on Bayesian decision criterion, the pixel i will be classified into category C_j, if

$$j = \arg\left\{ \max_{1 \le k \le K} \alpha_k p(x_i|\phi_k) \right\} , \quad (12)$$

where K is the number of categories including mixels.

First we compare classification results to evaluate how effective mixel distributions were. The image used is an image of size 256×256 cropped from NOAA-AVHRR ch.5 data. Based on our observation that the image contains a sea category and several cloud categories, we set the number of categories beforehand to $K = 3$.

Table 1: The number of pixels classified into each category.

Category	ML1	ML2	DM
Pure 1	8579	8972	8070
Pure 2	13590	9626	7306
Pure 3	14793	13303	10398
Mixel 1 + 2	11756	11852	12628
Mixel 1 + 3	0	0	0
Mixel 2 + 3	16818	21783	27134

Figure 2 shows a comparison between with/without mixel distributions. It shows that the boundary of cloud regions were classified as mixels (black regions in the image (b)). This result is reasonable because mixels tend to occur on region boundary. To validate the use of mixel distribution, we calculated the value of the information criterion, AIC (Akaike's Information Criterion). The value was 4999 without mixel distributions, while 5076 with mixel distributions (ML1). The smaller the AIC value, the better the model is. This result hence supports the superiority of the model with mixel distributions.

Finally Table 1 shows the number of pixels classified into each category. Three schemes produced similar results except for a slight difference of DM in classification of mixels. Moreover, ML2 was the fastest of all.

6 CONCLUSION

In practical use, the number of categories should be produced automatically. Use of information criterion such as AIC may be a hint to this problem. Further, use of multi-spectral data will achieve more accurate classification.

References

[1] F. Wang. **Fuzzy Supervised Classification of Remote Sensing Images**. *IEEE Trans. Geo. Remote Sens.*, Vol. 28, No. 2, pp. 194–201, 1990.

[2] K. Kim and M.M. Crawford. **Adaptive Parametric Estimation and Classification of Remotely Sensed Imagery Using a Pyramid Structure**. *IEEE Trans. Geo. Remote Sens.*, Vol. 29, No. 4, pp. 481–493, 1991.

[3] P. Santago and H.D. Gage. **Quantification of MR Brain Images by Mixture Density and Partial Volume Modeling**. *IEEE Trans. Med. Img.*, Vol. 12, No. 3, pp. 566–574, 1993.

[4] R.A. Redner and H.F. Walker. **Mixture Densities, Maximum Likelihood and the EM algorithm**. *SIAM Review*, Vol. 26, No. 2, pp. 195–239, 1984.

[5] H.C. Quelle, Y. Delignon, and A. Marzouki. **Unsupervised Bayesian Segmentation of SAR Images Using the Pearson System Distributions**. In *IGARSS'93*, pp. 1538–1540, 1993.

[6] M.J. Carlotto. **Histogram Analysis Using a Scale-Space Approach**. In *CVPR'85*, pp. 334–340, 1985.