IEICE TRANS. INF. & SYST., VOL.E98–D, NO.4 APRIL 2015

PAPER

A Spatially Correlated Mixture Model for Image Segmentation

Kosei KURISU[†], ***Nonmember***, **Nobuo SUEMATSU**[†a)], **Kazunori IWATA**[†], ***and*** **Akira HAYASHI**[†], ***Members***

SUMMARY In image segmentation, finite mixture modeling has been widely used. In its simplest form, the spatial correlation among neighboring pixels is not taken into account, and its segmentation results can be largely deteriorated by noise in images. We propose a spatially correlated mixture model in which the mixing proportions of finite mixture models are governed by a set of underlying functions defined on the image space. The spatial correlation among pixels is introduced by putting a Gaussian process prior on the underlying functions. We can set the spatial correlation rather directly and flexibly by choosing the covariance function of the Gaussian process prior. The effectiveness of our model is demonstrated by experiments with synthetic and real images.

key words: *image segmentation, Gaussian processes, mixture models*

1. Introduction

There are many approaches to image segmentation. Among them, we address the approaches based on finite mixture modeling which models the probability density function of pixel luminance value with finite mixture models. In this framework, every pixel value is assumed to be drawn from one of the components of a mixture. The parameters of the finite mixture model are usually estimated using the Expectation-Maximization (EM) algorithm based on the maximum likelihood principle [1], [2]. However, since the simplest form of the finite mixture modeling does not take the spatial information of pixels into account, its segmentation results for noisy images are often over-segmented and include many segments that are too tiny.

To improve segmentation accuracy for noisy images, it is essential to consider the spatial information of pixels. In the previous work [3], the spatially variant finite mixture model (SVFMM) has been proposed. In this model, a set of finite mixture models corresponding pixels whose mixing proportions can vary individually is introduced. The spatial correlation is brought into the model by posing a Markov random field (MRF) prior on the mixing proportions. The EM-algorithm for SVFMM was improved afterwards in [4]. However, SVFMM still has the tendency of over segmentation for highly corrupted images.

In this paper, we propose a new model called spatially correlated mixture model (SCMM) and examine its effectiveness and properties on synthetic and real images. An important difference from SVFMM lies in the probabilistic model which expresses prior distribution of the mixing proportions. SCMM assumes that every class has an underlying function defined on image space, and the set of the functions for all classes govern mixing proportions. By putting a Gaussian process prior on the underlying functions, the spatial correlation among the mixing proportions is introduced. We develop a quasi EM-algorithm to obtain maximum a posteriori (MAP) estimation of the underlying functions.

A Gaussian process is an infinite set of random variables, where any finite number of elements have a joint Gaussian distribution [5]. We can use a Gaussian process to represent a probability distribution over function space. A Gaussian process prior is quite useful because the posterior given noise-free or noisy observations is also a Gaussian process whose mean function and covariance function can be obtained analytically. The covariance function of a Gaussian process expresses correlation among any finite values of the random function. Thus we can set the spatial correlation among pixel labels flexibly via the covariance function of the Gaussian process prior, and we can expect SCMM to alleviate the tendency of over segmentation SVFMM has by choosing the covariance function properly.

This paper is organized as follows. In Sects. 2 and 3, we review the simplest finite mixture modeling and the SVFMM. In Sect. 4, we describe our proposed model, SCMM. Then, in Sect. 5, we provide an empirical study with synthetic images and real images to examine the effectiveness and properties of our method, and we conclude in Sect. 6.

2. Simplest Finite Mixture Modeling

We briefly review the image segmentation technique with the simplest form of the finite mixture modeling.

Given a grayscale image, we want to label each pixel of the image as belonging to one of C classes. Let y_v denote the observation of at v-th pixel. In the simplest finite mixture model (SFMM), it is assumed that each pixel luminance value y_v is independently drawn from the finite mixture model

$$p(y_v|\Phi) = \sum_{c=1}^{C} \pi_c p(y_v|\theta_c),$$

where $\Phi = \{\pi_c, \theta_c | c = 1, \ldots, C\}$ and π_c are mixing proportions which sum to 1. $p(y_v|\theta_c)$ is the component of class c

Manuscript received September 10, 2014.
Manuscript revised November 19, 2014.
Manuscript publicized January 6, 2015.

[†]The authors are with the Graduate School of Information Sciences, Hiroshima City University, Hiroshima-shi, 731–3194 Japan.
a) E-mail: suematsu@hiroshima-cu.ac.jp

DOI: 10.1587/transinf.2014EDP7307

and $\theta_c = \{\mu_c, \sigma_c\}$ is the set of its mean and standard deviation. Throughout this paper, $p(y_v|\theta_c)$ denotes a univariate Gaussian distribution.

We can derive an EM algorithm [1] for maximum likelihood estimation of the parameters [2]. By introducing hidden variables

$$z_{cv} = \begin{cases} 1 & \text{if } y_v \text{ belongs to class } c \\ 0 & \text{otherwise,} \end{cases}$$

the expected complete-data log-likelihood given $\{y_v|v = 1, \ldots, V\}$ and Φ' can be written as

$$\begin{aligned} Q(\Phi|\Phi') &= \mathbb{E}\left[\log p(\{y_v\}, \{z_{cv}\}|\Phi)|\{y_v\}, \Phi'\right] \\ &= \sum_{v=1}^{V}\sum_{c=1}^{C} \zeta_{cv} \log\left[\pi_c p(y_v|\theta_c)\right], \end{aligned}$$

where V is the number of pixels in the image and

$$\zeta_{cv} = \mathbb{E}[z_{cv}|y_v, \Phi'] = \frac{\pi'_c p(y_v|\theta'_c)}{\sum_{c'=1}^{C} \pi'_{c'} p(y_i|\theta'_{c'})}. \tag{1}$$

Then, $Q(\Phi|\Phi')$ is maximized with respect to Φ when

$$\pi_c = \frac{1}{V}\sum_{v=1}^{V} \zeta_{cv}, \tag{2}$$

$$\mu_c = \frac{\sum_{v=1}^{V} \zeta_{cv} y_v}{\sum_{v=1}^{V} \zeta_{cv}}, \tag{3}$$

$$\sigma_c = \left(\frac{\sum_{v=1}^{V} \zeta_{cv}(y_v - \mu_c)^2}{\sum_{v=1}^{V} \zeta_{cf}}\right)^{1/2}. \tag{4}$$

The EM algorithm iterates between an E-step in which ζ_{cv} are evaluated according to Eq. (1) and an M-step in which the elements of Φ are updated according to Eqs. (2)-(4).

Once Φ is estimated, each pixel is labeled with the most probable class:

$$\arg\max_{c=1}^{C} p(z_{cv} = 1|y_v, \Phi).$$

Under the SFMM, pixel values $\{y_v\}$ are considered i.i.d samples from a finite mixture model and the spatial information of the pixels is not considered at all.

3. Spatially Variant Finite Mixture Model

The spatially variant finite mixture model (SVFMM) has been proposed to incorporate the spatial correlation of the pixels effectively [3] and its learning algorithm has been improved in [4].

The SVFMM defines the density of the pixels $\{y_v\}$ as

$$p(\{y_v\}|\Theta) = \prod_{v=1}^{V}\sum_{c=1}^{C} \pi_{cv} p(y_v|\theta_c),$$

where $\Theta = \{\pi_{cv}, \theta_c|c = 1, \ldots, C, v = 1, \ldots, V\}$. As we can see from the above density function, each pixel y_v has its own finite mixture model whose set of mixing proportions is $\{\pi_{v1}, \ldots, \pi_{vC}\}$ and the pixels are still independent given the SVFMM. The spatial correlation among pixel is brought through the prior distribution of the mixing proportions $\{\pi_{cv}|c = 1, \ldots, C, v = 1, \ldots, V\}$.

The prior distribution is modeled by an MRF and its density function is described by the Gibbs distribution:

$$p(\{\pi_{cv}\}) = \frac{1}{Z}\exp\left\{-\beta\sum_{s\in S} V_s(\{\pi_{cv}\})\right\},$$

where Z is the partition function, β is a parameter that controls the concentration of the distribution, S is the set of cliques, and $V_s(\cdot)$ is the potential function of clique $s \in S$.

As the prior distribution of $\{\pi_{cv}\}$ is introduced, the SVFMM has to be learned based on the maximum a posteriori (MAP) criterion. This is achieved by defining

$$\begin{aligned} Q_{\mathrm{MAP}}(\Theta|\Theta') = &\sum_{v=1}^{V}\sum_{c=1}^{C} \zeta_{cv} \log\left[\pi_{cv} p(y_v|\theta_c)\right] \\ &+ \log p(\{\pi_{cv}\}) \end{aligned} \tag{5}$$

and developing an EM algorithm to maximize Q_{MAP}. In Eq. (5), ζ_{cv} is slightly different from Eq. (1) and defined as

$$\zeta_{cv} = \mathbb{E}[z_{cv}|y_v, \Theta'] = \frac{\pi'_{cv} p(y_v|\theta'_c)}{\sum_{c'=1}^{C} \pi'_{c'v} p(y_i|\theta'_{c'})}. \tag{6}$$

The EM algorithm for SVFMM is similar to that for the SFMM. In its E-step, Eq. (6) is evaluated instead of Eq. (1), and in its M-step, μ_c and σ_c are updated using the same Eqs. (3) and (4). The main difference lies in the updating of $\{\pi_{cv}\}$ in its M-step. In [3], this updating is done using the gradient projection method [6]. On the other hand, in [4], it is done by finding the solutions of the equations $\frac{\partial Q_{\mathrm{MAP}}}{\partial \pi_{cv}} = 0$ and then solving linearly constrained quadratic programming problems to ensure that $\{\pi_{cv}\}$ satisfy the constraints

$$0 \le \pi_{cv} \le 1, \qquad \sum_{c=1}^{C} \pi_{cv} = 1.$$

4. Spatially Correlated Mixture Model

In this section, we describe the Spatially Correlated Mixture Model (SCMM) that we propose. The model involves underlying functions $f_1(\cdot), \ldots, f_C(\cdot)$ that are real valued functions defined over the image space and govern the mixing proportions of the SCMM. We first briefly review the Gaussian process regression and covariance functions of Gaussian processes, then describe details of the model and the learning algorithm for the model.

4.1 Gaussian Process Regression [5]

We can use a Gaussian process to describe a distribution of a random function, or a stochastic process. Formally, the random function $f(\cdot)$ defined in R^d is a Gaussian process

with mean function $m(\cdot)$ and covariance function $k(\cdot,\cdot)$, if for any natural number L and a set of L points in R^d $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_L\}$

$$\mathbf{f}(X) \sim \mathcal{N}(\mathbf{m}(X), K(X,X)),$$

where $\mathcal{N}(\mu, \Sigma)$ is the normal distribution with mean vector μ and covariance matrix Σ, $\mathbf{f}(X)$ and $\mathbf{m}(X)$ are vectors whose j-th elements are $f(\mathbf{x}_j)$ and $m(\mathbf{x}_j)$, respectively, and $K(X,X)$ is matrix whose (i,j)-th element is $k(\mathbf{x}_i, \mathbf{x}_j)$. When $f(\cdot)$ is a Gaussian process with mean function $m(\cdot)$ and covariance function $k(\cdot,\cdot)$, it is denoted as

$$f(\cdot) \sim \mathcal{GP}(m(\cdot), k(\cdot,\cdot)).$$

When we have some observations of an unknown function whose prior is a Gaussian process, we can obtain its posterior distribution using Gaussian process regression. Assume that we have a set of L observations of $f(\cdot)$, $\mathbf{y} = [y_1, \ldots, y_L]^{\mathrm{T}}$ at X, where

$$y_j = f(\mathbf{x}_j) + \varepsilon_j$$

and ε_j are independent Gaussian noises with mean 0 and variance $\sigma^2_{\varepsilon_j}$ (If some y_j is an exact observation, corresponding $\sigma^2_{\varepsilon_j} = 0$). The covariance between y_i and y_j is written as

$$\begin{aligned}\mathrm{Cov}(y_i, y_j) &= \mathbb{E}[(y_i - \mathbb{E}[y_i])(y_j - \mathbb{E}[y_j])] \\ &= \mathbb{E}[(f(\mathbf{x}_i) - m(\mathbf{x}_i))(f(\mathbf{x}_j) - m(\mathbf{x}_j))] + \mathbb{E}[\varepsilon_i \varepsilon_j] \\ &= k(\mathbf{x}_i, \mathbf{x}_j) + \delta_{ij}\sigma^2_{\varepsilon_i},\end{aligned}$$

where δ_{ij} is Kronecker's delta. Thus, given a set of test points $X^* = \{\mathbf{x}^*_1, \ldots, \mathbf{x}^*_M\}$, we can write the joint distribution of the vector $\mathbf{f}(X^*)$ and $\mathbf{y}$ as

$$\begin{bmatrix}\mathbf{f}(X^*) \\ \mathbf{y}\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\mathbf{m}(X^*) \\ \mathbf{m}(X)\end{bmatrix}, \begin{bmatrix}K(X^*, X^*) & K(X^*, X) \\ K(X, X^*) & K(X,X) + D\end{bmatrix}\right),$$

where D is the diagonal matrix whose diagonal elements are $\sigma^2_{\varepsilon_1}, \ldots, \sigma^2_{\varepsilon_L}$. Then, as shown in Appendix A.2 of [5], it is known that the conditional distribution of $\mathbf{f}(X^*)$ given $\mathbf{y}$ is also a multivariate Gaussian distribution and its mean vector and covariance matrix are

$$\begin{aligned}\mathbf{m}^* &= \mathbf{m}(X^*) \\ &\quad + K(X^*, X)[K(X,X) + D]^{-1}(\mathbf{y} - \mathbf{m}(X)), \qquad (7) \\ K^* &= K(X^*, X^*) - K(X^*, X)[K(X,X) + D]^{-1}K(X, X^*).\end{aligned}$$

4.2 Covariance Functions

The spatial correlation of any two pair of values of a underlying function is determined by the covariance function $k(\cdot,\cdot)$ of the Gaussian process prior. In this paper, we use the following two covariance functions.

$$k_{\mathrm{SE}}(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2l^2}\right), \qquad (8)$$

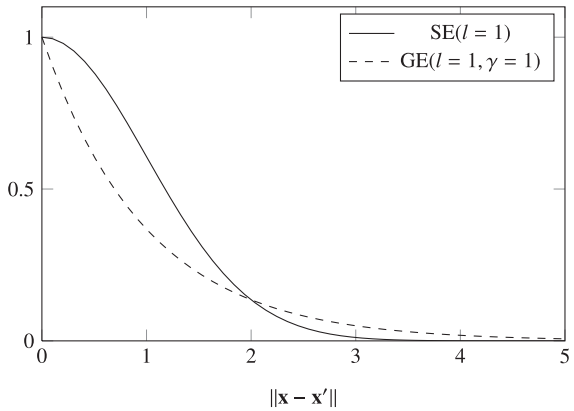


Fig. 1 Plots of two covariance functions, SE with $l = 1$, (solid) and GE with $l = 1, \gamma = 1$ (dashed).

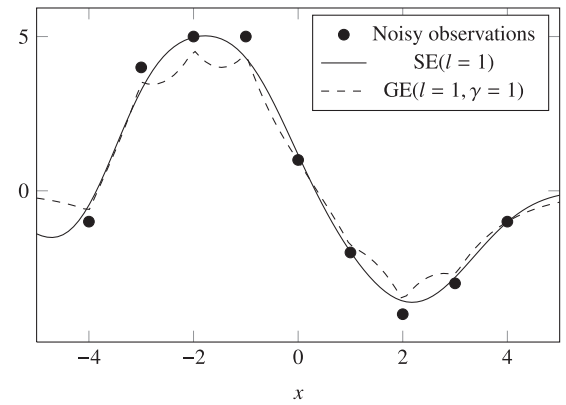


Fig. 2 Mean functions of the posterior Gaussian processes given noisy observations shown as filled circles. The covariance functions of the priors are SE with $l = 1$ (solid) and GE with $l = 1, \gamma = 1$ (dashed).

$$k_{\mathrm{GE}}(\mathbf{x}, \mathbf{x}') = \exp\left\{-\left(\frac{\|\mathbf{x} - \mathbf{x}'\|}{l}\right)^{\gamma}\right\} \quad \text{for} \quad 0 < \gamma \leq 2, \qquad (9)$$

which are referred to as Squared exponential (SE) and γ-exponential (GE), respectively.

Figure 1 shows plots of the two covariance functions of Eqs. (8) and (9). Figure 2 shows the posterior mean functions for Gaussian process priors with the covariance functions SE($l = 1$) and GE($l = 1, \gamma = 0.1$), which exhibit rather different shapes.

4.3 Spatially Correlated Mixture Model

In the SCMM, the density function of the pixels $\{y_v\}$ is written as

$$p(\{y_v\}|\Psi) = \prod_{v=1}^{V} \sum_{c=1}^{C} \rho_c(\mathbf{x}_v|\{f_c(\cdot)\}) p(y_v|\theta_c).$$

Here, $\{f_c(\cdot)\}$ is the set of underlying functions, $\Psi = \{f_c(\cdot), \theta_c | c = 1, \ldots, C\}$, and

$$\rho_c(\mathbf{x}_v|\{f_c(\cdot)\}) = \frac{\exp f_c(\mathbf{x}_v)}{\sum_{c'=1}^{C} \exp f_{c'}(\mathbf{x}_v)}. \qquad (10)$$

Note that $\{f_c(\cdot)\}$ is ambiguous up to a function $\phi(\mathbf{x})$ because if we define $f'_c(\mathbf{x}) = f_c(\mathbf{x}) + \phi(\mathbf{x})$ for $c = 1, \ldots, C$ then $\rho_c(\mathbf{x}|\{f_c(\cdot)\}) = \rho_c(\mathbf{x}|\{f'_c(\cdot)\})$ for any $\mathbf{x}$. To eliminate the ambiguity, we impose the following constraint:

$$\sum_{c=1}^{C} f_c(\mathbf{x}) = 0 \qquad \text{for any } \mathbf{x}.$$

Under this constraint, we can write the inverse of Eq. (10) as

$$f_c(\mathbf{x}_v) = \log \rho_c(\mathbf{x}_v|\{f_c(\cdot)\}) - \frac{1}{C}\sum_{c'=1}^{C} \log \rho_{c'}(\mathbf{x}_v|\{f_c(\cdot)\}). \quad (11)$$

In our model, the spatial smoothness is introduced by assuming Gaussian process prior on $\{f_c(\cdot)\}$. The assumption is described as follows:

$$\tilde{f}_c(\cdot) \overset{\text{i.i.d}}{\sim} \mathcal{GP}(m(\cdot), k(\cdot,\cdot)) \qquad \text{for } c = 1, \ldots, C, \quad (12)$$

$$f_c(\mathbf{x}) = \tilde{f}_c(\mathbf{x}) - \frac{1}{C}\sum_{c'=1}^{C} \tilde{f}_{c'}(\mathbf{x}). \quad (13)$$

We can rewrite this description into a more suitable form for later discussion. Let A_C be the centering matrix of order C, that is,

$$A_C = I_C - \frac{1}{C}U_C,$$

where I_C is the identity matrix of order C and U_C is the square matrix of order C whose elements are all 1. In addition, let W_C be the $C \times (C-1)$ matrix whose column vectors are the normalized eigen vectors of A_C corresponding to its eigen value 1. Then, Eqs. (12) and (13) are equivalent to

$$\tilde{f}_c(\cdot) \overset{\text{i.i.d}}{\sim} \mathcal{GP}(m_0(\cdot), k(\cdot,\cdot)) \quad \text{for } c = 1, \ldots, C-1, \quad (14)$$

$$\begin{pmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_C(\mathbf{x}) \end{pmatrix} = W_C \begin{pmatrix} \tilde{f}_1(\mathbf{x}) \\ \vdots \\ \tilde{f}_{C-1}(\mathbf{x}) \end{pmatrix} \quad (15)$$

where $m_0(\mathbf{x}) \equiv 0$.

4.4 Learning Algorithm

We devise a quasi EM algorithm for our model. The algorithm tries to maximize the expected log posterior given Ψ',

$$Q_{\text{MAP}}(\Psi|\Psi') = \sum_{v=1}^{V}\sum_{c=1}^{C} \zeta_{cv} \log\left[\rho_c(\mathbf{x}_v|\{f_c(\cdot)\})p(y_v|\theta_c)\right] + \log p(\{f_c(\mathbf{x}_v)\}).$$

Here,

$$\zeta_{cv} = \mathbb{E}[z_{cv}|y_v, \Psi'] = \frac{\rho_c(\mathbf{x}_v|\{f'_c(\cdot)\})p(y_v|\theta'_c)}{\sum_{c'=1}^{C}\rho_{c'}(\mathbf{x}_v|\{f'_c(\cdot)\})p(y_v|\theta'_{c'})}. \quad (16)$$

Since Eq. (16) corresponds to Eq. (6) in SVFMM, Eq. (16) is evaluated in the E-step of our algorithm.

Let F be the $C \times V$ matrix whose (c, v)-th element $f_{cv} = f_c(\mathbf{x}_v)$. In the M-step of our algorithm, F is updated by finding the maximum likelihood estimate and compromising it with the prior distribution described in Eqs. (14) and (15) using Gaussian process regression. This is performed as follows. As we can easily show that

$$\sum_{v=1}^{V}\sum_{c=1}^{C} \zeta_{cv} \log[\pi_{cv} p(y_v|\theta_c)]$$

is maximized with respect to $\{\pi_{cv}\}$ when $\pi_{cv} = \zeta_{cv}$, the ML estimate of F is obtained by using Eq. (11) as

$$F_{\text{ML}} = \left(\log \zeta_{cv} - \tfrac{1}{C}\textstyle\sum_{c'=1}^{C} \log \zeta_{c'v}\right)_{C\times V}.$$

Since $f_1(\cdot), \ldots, f_C(\cdot)$ are not independent, before applying Gaussian process regression, F_{ML} is transformed to $\tilde{F}_{\text{ML}}$ as

$$\tilde{F}_{\text{ML}} = W_C^{\text{T}} F_{\text{ML}}$$

using the relation $W_C^{\text{T}} W_C = I_{C-1}$. Then, Eq. (7) is applied to each row of $\tilde{F}_{\text{ML}}$ by regarding it as a sequence of noisy observations. When applying Eq. (7), D is set to be $\tau^2 I$ where I is the identity matrix and τ^2 is a design parameter of our algorithm. When we denote the result of Gaussian process regression by $\tilde{F}$, finally, the updated F is given by

$$F = W_C \tilde{F}.$$

The updating of θ_c is done by using Eqs. (3) and (4) again.

Since the updated F is obtained by compromising the maximum likelihood estimate with the prior distribution, we can expect that $Q_{\text{MAP}}(\Psi|\Psi')$ is high to a certain degree at the obtained F. However, since $Q_{\text{MAP}}(\Psi|\Psi')$ is not directly maximized or increased in our M-step, we call the procedure as quasi EM algorithm.

Once F is estimated, pixel v is labeled with the index corresponding to the largest element in v-th column of F, that is,

$$\arg\max_{c=1}^{C} f_{cv}.$$

5. Empirical Study

In this section, we examine the effectiveness and the validity of our proposing method using synthetic and real images. In the proposed method, long distance correlations are ignored so that we don't have to calculate very large covariance matrices of order of the number of pixels in images. In the following numerical experiments, we determine the initial values of the parameters based on the results of k-means method for all algorithms. In addition, we set $\tau^2 = 100$ for our algorithm, which defines D of Eq. (7) as $D = \tau^2 I$.

5.1 Experiment with Synthetic Images

In the first part of the empirical study, we use synthetic images to compare segmentation accuracy of SFMM,

Table 1 Comparison of error rates with the synthetic images for SFMM, SVFMM, and the proposed method.

C	σ	Error rates[%]		
		SFMM	SVFMM	SCMM
3	0.10	9.93	1.87	0.22
	0.20	27.80	14.94	1.29
	0.30	39.59	34.33	2.80
5	0.05	9.59	1.01	1.31
	0.10	27.04	10.76	2.94
	0.15	40.93	33.88	9.10

SVFMM, and SCMM. We corrupt synthetic images with Gaussian noise, and measure segmentation accuracy by how accurate is restoring noised images to original images. The segmentation accuracy is measured by a misclassification rate.

5.1.1 Experimental Setup

A C-class label image is generated by assigning label

$$c_v = \arg\max_{c=1}^{C} g_c(\mathbf{x}_v)$$

to v-th pixel located at $\mathbf{x}_v$ for $v = 1, \ldots, V$, where each function $g_c(\cdot)$ is obtained by interpolating, using cubic spline, independent standard Gaussian random values assigned to sparse grid points over the image space. Then, the noise free image corresponding to the label image is created by setting pixel values of the pixels whose label is c_v to

$$\frac{2c_v - 1}{2C},$$

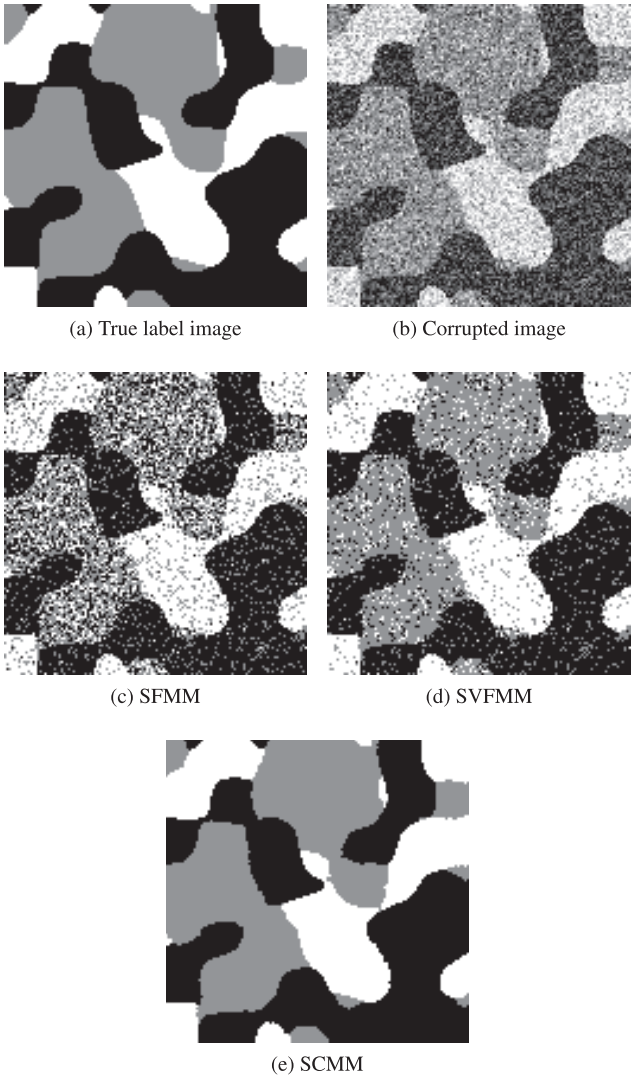
and a synthetic image is generated by corrupting the noise free image by zero mean Gaussian with standard deviation of σ. For each pair of C and σ, 100 synthetic images of 128×128 pixels are generated and used to measure the average error rate for each of SFMM, SVFMM, and SCMM.

The parameter β of SVFMM is chosen so that it performs best for each pair of C and σ. In the proposed method, a Gaussian process prior with GE covariance function of $l = 4, \gamma = 0.8$ is used throughout for synthetic images.

5.1.2 Results

Table 1 shows the measured error rates in this experiment. We can see that the proposed SCMM has lower error rates than SVFMM when image has relatively heavy noise. Since SFMM does not take spatial informations into account, SFMM provides poor results.

Figure 3 shows some examples of segmentation results for $(C, \sigma) = (3, 0.2)$. Figure 3 (a) is the true label image and Fig. 3 (b) is corrupted image which is used as the input image for each algorithms. Figure 3 (c) to 3 (e) are segmentation results for SFMM, SVFMM, and SCMM, respectively. Although SVFMM reduces errors compared to SFMM, we can still see the tendency of over segmentation and it includes many too tiny segments. On the other hand, the difference between the true label image and the result of SCMM is rather subtle.



(a) True label image (b) Corrupted image

(c) SFMM (d) SVFMM

(e) SCMM

Fig. 3 Segmentation results for an example of three class image. (a) True label image and (b) corrupted image by zero mean Gaussian noise with standard deviation of 0.2. Segmentation by SFMM (c), SVFMM (d), and the proposed SCMM (e).

5.2 Experiment with Real Images

From the Berkeley segmentation dataset (BSDS) [7], we picked 11 real images which look like whose luminance values give good clues to achieve decent segmentations. Note that all images were converted to grayscale ranging from 0 to 1 in advance. Figure 4 shows the selected images. We also use the images corrupted by Gaussian noise with standard deviation of 0.05 and 0.1.

Some ground truth segmentations produced by humans are provided for each image in BSDS. However, given a segmentation, we cannot define its misclassification rate because the correspondence between the segment labels in the

(a) #55067 (b) #86016 (c) #118035 (d) #135069

(e) #143090 (f) #167062 (g) #176035 (h) #241004

(i) #253036 (j) #60079 (k) #157036

Fig. 4 The real images chosen from BSDS.

segmentation and those in a ground truth cannot be determined. Thus, we compare SVFMM and SCMM using the following three criteria which come from rather different perspectives:

Probabilistic Rand (PR) Index [8] PR index is defined between a segmentation and the set of ground truth segmentations. PR index takes values in $[0, 1]$ and higher values indicate better results.

Variation of Information (VI) [9] Since VI is a measure defined between two segmentations, the mean of VI values obtained between a segmentation and all of the human segmentations is used as the VI value for the segmentation. VI takes nonnegative values and lower values indicate better results.

F-measure of Boundary Detection [10] This measure is defined based on the precision-recall framework for the task to detect segment boundary pixels, in which true positives are determined by the human segmentations. Precision and recall are calculated based on the counts accumulated over the 11 real images, and F-measure is obtained from precision and recall. F-measure takes values in $[0, 1]$ and higher values indicate better results.

5.2.1 Experimental Setup

In this experiment, two covariance functions, SE and GE described in Sect. 4.2 are used for our proposing SCMM.

Table 2 The parameters of SVFMM and the covariance functions SE and GE of SCMM.

SVFMM	SCMM	
	SE	GE
$\beta = 0.5$	$l = 2$	$l = 0.5, \gamma = 0.1$

In this section SCMM with SE and GE are referred to as SCMM-SE and SCMM-GE, respectively. The parameters of the covariance functions and the parameter β of SVFMM are chosen so that they perform best in terms of mean PR index. The tuned parameters are shown in Table 2.

In human segmentations, each connected segment has its own label regardless of the luminance values of the pixels in the segment. On the other hand, in the segmentations produced by SVFMM or SCMM, some disconnected segments can have the same label when they correspond to the same component of the finite mixture. Thus, in this experiment, the segmentations produced by SVFMM and SCMM are relabeled so that any connected segment has a unique label.

Throughout the experiment with real images, number of classes, C is set to 4 because SVFMM and SCMM both produce decent segmentations for the selected real images when $C = 4$.

5.2.2 Results

First, we compare the segmentation performance of SVFMM and the proposed methods in terms of PR index. Table 3 shows the means of the measured PR index values. While SCMM-SE is comparable to SVFMM when the images not corrupted, the proposed methods achieve higher values than SVFMM in other cases. A remarkable fact is that although the performance of SVFMM is degraded for the corrupted images as normally expected, the performance of the proposed methods is comparable or even better.

Next, we examine the segmentation performance in terms of VI. Table 4 shows the means of the measured VI values. Although when the images are not corrupted, SCMM-SE was comparable to SVFMM in terms of PR index, it outperforms in terms of VI. We again see the phenomenon in which the proposed methods perform better when the images are corrupted by noise.

Table 3 Means of PR index values for SVFMM and the proposed methods, SCMM-SE and SCMM-GE.

σ	PR Index (standard deviation)		
	SVFMM	SCMM-SE	SCMM-GE
0.00	0.704 (0.130)	0.704 (0.133)	0.723 (0.149)
0.05	0.703 (0.139)	0.705 (0.132)	0.721 (0.148)
0.10	0.681 (0.128)	0.712 (0.108)	0.736 (0.133)

Table 4 Means of VI values for SVFMM and the porposed methods, SCMM-SE and SCMM-GE.

σ	VI (standard deviation)		
	SVFMM	SCMM-SE	SCMM-GE
0.00	2.18 (0.82)	2.07 (0.74)	1.89 (0.72)
0.05	2.24 (0.88)	2.08 (0.75)	1.90 (0.73)
0.10	2.51 (0.83)	2.00 (0.78)	1.81 (0.84)

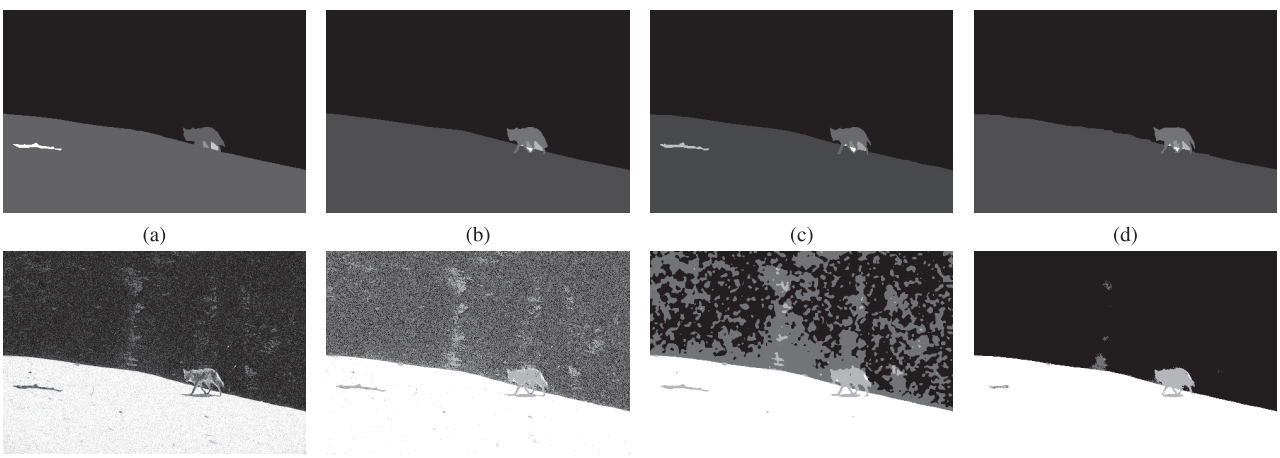
Finally, we evaluate the quality of the segmentation results with the evaluating methodology developed in [10]. Table 5 shows F-measures along with precisions and recalls, which indicates that the proposed methods outperform SVFMM. Note that the recalls of SVFMM is very high while its precisions are very low. We can understand this fact from Fig. 5. As shown in Fig. 5 (f), the result of SVFMM is highly over-segmented, which would cause many false positives and few undetected boundary pixels. The result of SCMM-GE shown in Fig. 5 (h) is similar to the human segmentations shown in Fig. 5 (a)- 5 (d) and is not so over-segmented and thus, its precision (recall) is much higher (lower) than that of SVFMM.

This time, as normally expected, the segmentation performance of our methods degenerate as the added noise increases, although it dose so much more moderately than that of SVFMM does. As seen from Table 5, adding more noise leads to more over segmented results for any methods. Taking into account this fact, from the mysterious improvement of the segmentation evaluations of the proposed methods based on PR index and VI by adding noise, we can conjecture that PR index and VI has the tendency of overestimation for over segmented results.

We also measured CPU time of SVFMM and our method for $\sigma = 0$. On average, SVFMM and SCMM took roughly 50 and 10 minutes per image, respectively. Although CPU time would depend on the parameters and the details of implementation of the algorithms, it can be said

Table 5 F-measures with precisions and recalls for SVFMM and the proposed methods, SCMM-SE and SCMM-GE.

σ	F-measure (Precision, Recall)		
	SVFMM	SCMM-SE	SCMM-GE
0.00	0.44 (0.30, 0.84)	0.57 (0.44, 0.78)	0.62 (0.63, 0.62)
0.05	0.29 (0.17, 0.88)	0.50 (0.37, 0.80)	0.61 (0.59, 0.63)
0.10	0.23 (0.13, 0.97)	0.44 (0.31, 0.79)	0.52 (0.43, 0.67)



(a) (b) (c) (d)

(e) #167062 with noise (f) SVFMM, PRI=0.828 (g) SCMM-SE, PRI=0.799 (h) SCMM-GE, PRI=0.989

Fig. 5 Example of segmentation results for the image #167062. (a)-(d) are examples of the ground truth segmentations. (e) is the input real image corrupted by Gaussian noise of $\sigma = 0.1$. (f)-(h) are segmentation results of SVFMM, SCMM-SE, and SCMM-GE, respectively.

that our method is not so computationally intensive in comparison with SVFMM.

6. Conclusions and Future Work

We have proposed a new image segmentation algorithm based on finite mixture modeling. In this algorithm, image segmentation is achieved by finding MAP estimates of the underlying functions that govern the mixing proportions of a spatially correlated mixture model.

The effectiveness of the proposed approach has been demonstrated by the experiments with synthetic and real images. The proposed approach achieved significantly higher accuracy than SVFMM for noisy synthetic images. Moreover, in the experiment with real images, the proposed method produced segmentation results which more closely resemble human segmentations than those of SVFMM do.

The covariance function of the Gaussian process prior of a SCMM plays a major role in the proposed method. We can control behavior of the underlying functions, and hence segmentation results by choosing a covariance function. Although there is a wide variety of options for the covariance function, only a small portion of it has been investigated. There would be room for improvement in its segmentation accuracy by choosing the covariance function considering the statistics of the images to be segmented.

A possible extension to our technique is to incorporate some information from other techniques such as edge detection. Since Gaussian process regression is based on Bayesian framework, it would be relatively straightforward. For example, if an edge segment which lies on the boundary between consecutive image segments is detected for certain, it might be beneficial to add a set of imaginary observations of zeros located on the edge segment to the original observations $\tilde{F}_{\mathrm{ML}}$ so that interchange of the largest one among underlying functions is facilitated on the edge.

References

[1] A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," J. Royal Statistical Society. Series B (Methodological), vol.39, pp.1–38, 1977.
[2] G. McLachlan and D. Peel, Finite Mixture Models, Wiley-Interscience, 2000.
[3] S. Sanjay-Gopal and T.J. Hebert, "Bayesian pixel classification using spatially variant finite mixtures and the generalized em algorithm," IEEE Trans. Image Process., vol.7, no.7, pp.1014–1028, 1998.
[4] K. Blekas, A. Likas, N.P. Galatsanos, and I.E. Lagaris, "A spatially constrained mixture model for image segmentation," IEEE Trans. Neural Netw., vol.16, no.2, pp.494–498, 2005.
[5] C.E. Rasmussen and C.K.I. Williams, Gaussian Process for Machine Learning, MIT Press, 2006.
[6] D.G. Luenberger and Y. Ye, Linear and Nonlinear Programming, third ed., Springer, 2008.
[7] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," Proc. 8th International Conference on Computer Vision, pp.416–423, 2001.
[8] R. Unnikrishnan and M. Hebert, "Measures of similarity," Proc. 7th IEEE Wrokshop on Application of Computer Vision, pp.394–400, 2005.
[9] M. Meilă, "Comparing clusterings by the variation of information," Learning Theory and Kernel Machines, vol.2777, pp.173–187, 2003.
[10] D.R. Martin, C.C. Fowlkes, and J. Malik, "Learning to detect natural image boundaries using local brightness, color, and texture cues," IEEE Trans. Pattern Anal. Mach. Intell., vol.26, no.5, pp.530–549, 2004.

Kosei Kurisu received the B.S. degree in Information Science from Hiroshima City University in 2013. He is currently studying information sciences at the postgraduate program and working toward the M.S. degree at Hiroshima City University. His main interests and activities are in pattern recognition.

Nobuo Suematsu received the B.S. degree and the M.S. degree in physics from Kyushu University in 1988 and 1990, respectively. He received his Ph.D in engineering from Osaka University in 2000. From 1990-1994, he joined Fujitsu Laboratories LTD. In 1994, he joined the faculty member of Hiroshima City University, and currently he is an associate professor of the Graduate School of Information Sciences, Hiroshima City University. His current research interests include machine learning, pattern recognition.

Kazunori Iwata received a BE and an ME degree from Nagoya Institute of Technology, Aichi, Japan in 2000 and 2002, respectively, and a PhD degree in Informatics from Kyoto University, Kyoto, Japan, in 2005. He was a JSPS research fellow from April 2002 to March 2005. He has been with Hiroshima City University, Hiroshima, Japan, since April 2005. His research interests include machine learning, statistical inference, and information theory. He received the IEEE Kansai-Section Student Paper Award in February 2005. He is an Associate Editor of the IEICE Transactions on Information and Systems.

Akira Hayashi received a B.S. degree in Mathematics from Kyoto University, Kyoto, Japan, in 1974, an M.S. degree in Computer Science from Brown University, Providence, RI, in 1988, and a Ph.D. degree in Computer Science from the University of Texas at Austin, in 1991. He joined IBM Japan in April 1974. He was a visiting Associate Professor at Kyushu Institute of Technology, Fukuoka, Japan, until March 1994. Currently, he is a Professor at the Graduate School of Information Sciences, Hiroshima City University, Hiroshima, Japan. His research interests include machine learning and pattern recognition.