www.ietdl.org

Published in IET Image Processing
Received on 26th March 2013
Revised on 29th May 2013
Accepted on 29th August 2013
doi: 10.1049/iet-ipr.2013.0232

ISSN 1751-9659

Image segmentation by Dirichlet process mixture model with generalised mean

Hui Zhang[1,2,3], Qing Ming Jonathan Wu[2], Thanh Minh Nguyen[2]
[1]*School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, People's Republic of China*
[2]*Department of Electrical and Computer Engineering, University of Windsor, Windsor, ON, Canada*
[3]*Jiangsu Engineering Center of Network Monitoring, Nanjing University of Information Science and Technology, Nanjing, People's Republic of China*
E-mail: nrzhanghui@gmail.com

Abstract: The Dirichlet process mixture model (DPMM) with spatial constraints – e.g. hidden Markov random field (HMRF) model – has been considered as an effective algorithm for image processing application. However, the HMRF model is complex and time-consuming for implementation. A new DPMM has been introduced, where a generalised mean (GDM) is selected as the spatial constraints function. The GDM is applied not only on prior probability (and posterior probability) to incorporate local spatial information and component information, but also on conditional probability to incorporate local spatial information and observation information. The purpose of the HMRF model and GDM are the same for incorporating some spatial constraints into the system. However, compared to HMRF, GDM is easier, faster and simpler to implement. Finally, a variational Bayesian approach has been adopted for parameters estimation and model selection. Experimental results on image segmentation application demonstrate the improved performance of the proposed approach.

1 Introduction

As a Bayesian non-parametric model, the Dirichlet process mixture model (DPMM) was introduced by Ferguson [1] and has been very popular in statistics over the last few years, for providing a Bayesian framework for clustering problems with an unknown number of groups [2, 3]. The theory behind the DPMM is based on the observation that a countable infinite number of component distributions in an ordinary finite mixture model tends on the limit to a Dirichlet process (DP) [1] prior. The Markov chain Monte Carlo (MCMC) method [4] and variational Bayesian (VB) inference [5] are two common useful inference techniques for parameter learning for DPMM. In [6], the DPMM is applied in brain MRI tissue classification with more than encouraging results obtained. In [7–9], spatial constraints are incorporated into DPMM for image segmentation. DPMM based not only on Gaussian distribution but also on different distributions is introduced in [10, 11]. In [12–16], the DPMM is adopted to the hidden Markov random field (HMRF) model and mixture of generalised Dirichlet distributions.

A remaining challenge of clustering approaches for image segmentation is related to their lack of spatial structure in an image. To overcome this shortcoming, a wide variety of approaches have been proposed to incorporate spatial information into the image [17–23]. A common approach is the use of a Markov random field (MRF) [24, 25]. Such a method aims to impose spatial smoothness constraints on the image pixel labels. Recently, a special case of the MRF model – the HMRF model – has been proposed [26, 27]. The state sequence of HMRF cannot be observed directly, but can be indirectly observed through a field of observations. In the HMRF model, the spatial information in an image is encoded through the contextual constraints of neighbouring pixels, which are characterised by conditional MRF distributions. Parameter estimation in HMRF models usually relies on maximum likelihood (ML) or Bayesian methods [28, 29]. Besag [30] introduces the idea of pseudo likelihood approximation when ML estimation is intractable. Based on this well-known approximation, various HMRF model estimation approaches have been proposed [31–37].

The recent work for the combination of DPMM and MRF/HMRF is illustrated in [6–11]. In this paper, we combine DPMM with a generalised mean (GDM) instead of MRF/HMRF for image segmentation. One drawback of MRF/HMRF models is that they are computationally expensive to implement, and require the additional parameter β to control the degree of image smoothness. This additional parameter β is usually determined by researcher's experience. The chosen parameter β has to be both large enough to tolerate the noise, and small enough to preserve image sharpness and details. With the help of GDM, our model is fully free of the empirically adjusted parameter β. Although image segmentation is the motivation and specific application in this paper, the idea of combining GDM and DPMM can also be applied to any other clustering analysis applications.

The remainder of this paper is organised as follows: In Section 2, we briefly introduce the mathematical background of GDM and DPMM. In Section 3, we introduce the proposed new DPMM and then we illustrate how to incorporate the local spatial information into DPMM with GDM. We also show the relationship between our algorithm and other spatial constraints works based on the MRF/HMRF model. The parameter learning estimated by the VB inference algorithm is given in Section 4. The experimental results of the proposed approach are given in Section 5. Finally, some concluding remarks are provided.

2 Mathematical background

2.1 Generalised mean

In mathematics, a GDM, also known as a power mean is an abstraction of the Pythagorean means, including arithmetic, geometric and harmonic means. The GDM of $a_1, a_2, \ldots, a_n$ is defined as

$$M_p(a_1, a_2, \ldots, a_n) = \left(\frac{1}{n}\sum_{i=1}^{n} a_i^p\right)^{1/p} \tag{1}$$

where $a_i \geq 0$, $p \in [-\infty, +\infty]$ and $\sum_{i=1}^{n} a_i = 1$.

For $p \to 0$, (1) approaches the geometric mean (GM)

$$M_G(a_1, a_2, \ldots, a_n) = \left(\prod_{i=1}^{n} a_i\right)^{1/n} \tag{2}$$

For $p = 1$, (1) results in the arithmetic mean (AM)

$$M_A(a_1, a_2, \ldots, a_n) = \frac{1}{n}\sum_{i=1}^{n} a_i \tag{3}$$

There are some other special cases of GDM based on different p values. For example, if $p = -1$, M is a harmonic mean; for $p = 2$, M is a quadratic mean; under the condition $p \to -\infty$, $M_{-\infty} = \min(a_1, a_2, \ldots, a_n)$ and the condition $p \to \infty$, $M_{\infty} = \max(a_1, a_2, \ldots, a_n)$.

2.2 Dirichlet process mixture model

The DP prior introduced by Ferguson [1] is a commonly used prior on the parameters of a mixture model with an unknown number of mixture components. Based on the DP prior for the random variables $\{\theta_n^*\}_{n=1}^{N}$, DPMM assumes that the sample distribution (random measure) G is drawn from a DP(G_0, α), with a base distribution (measure) G_0 and a precision parameter α. The formal notation of DP is given as follows

$$\begin{gathered} G|\{G_0, \alpha\} \sim DP(G_0, \alpha) \\ \theta_n^*|G \sim G \end{gathered} \tag{4}$$

Based on the relationship between the DP and generalised Pólya urn schemes, we introduce the Pólya urn representation [38] of DPMM, where the DP is viewed as the limit of Pólya urn schemes. Let $\{\theta_c\}_{c=1}^{K}$ be the set of distinct values taken by the variables $\{\theta_n^*\}_{n=1}^{N-1}$. Denoting as f_c^{N-1} the number of values in $\{\theta_n^*\}_{n=1}^{N-1}$ that equal to θ_c, the conditional distribution of θ_N^* given $\{\theta_n^*\}_{n=1}^{N-1}$ has the form

$$p\left(\theta_N^*|\{\theta_n^*\}_{n=1}^{N-1}, G_0, \alpha\right) = \frac{\alpha}{\alpha + N - 1}G_0 + \sum_{c=1}^{K}\frac{f_c^{N-1}}{\alpha + N - 1}\delta_{\theta_c} \tag{5}$$

where δ_{θ_c} denotes the distribution concentrated at single point θ_c. Equation (5) shows that when considering θ_N^* given all other observations $\{\theta_n^*\}_{n=1}^{N-1}$, this new sample is either drawn from base distribution G_0 with probability $\alpha/(\alpha + N - 1)$, or is selected from the existing draws θ_c according to a multinomial allocation, with probabilities proportional to existing groups size f_c^{N-1}.

The parameter α plays a balancing role between sampling a new parameter from the base distribution G_0, or sharing a previously sampled parameter. A larger α indicates more components, and in the limit $\alpha \to \infty$, $G \to G_0$. On the contrary, as $\alpha \to 0$, all $\{\theta_n^*\}_{n=1}^{N-1}$ tend to cluster to a single component and take on the same value.

A draw from DP may also be represented in terms of a stick-breaking construction [39] which provides the explicit characterisation of G. Consider two infinite collections of independent random variables, $v = \{v_c\}_{c=1}^{\infty}$ and $\{\theta_c\}_{c=1}^{\infty}$, where v_c is drawn from the beta distribution Beta(1, α), and θ_c is drawn independently from the base distribution G_0. The stick-breaking representation of G is then defined as

$$G = \sum_{c=1}^{\infty} \pi_c(v)\delta_{\theta_c} \tag{6}$$

where

$$\pi_c(v) = v_c\prod_{j=1}^{c-1}\left(1 - v_j\right) \tag{7}$$

with

$$\pi_c(v) \in [0, 1] \text{ and } \sum_{c=1}^{\infty}\pi_c(v) = 1$$

Under the stick-breaking representation (6) of the DP, the atoms θ_c can be seen as the parameters of the component distributions of a mixture model comprising an unbounded number of component densities. With the finite component numbers, the modified model is known as the generalised Dirichlet mixture model with finite Dirichlet distributions.

Let $y = \{y_n\}_{n=1}^{N}$ be a set of observations models of DPMM. Then, each observation y_n is assumed to be drawn from its own conditional probability density function $p(y_n|\theta_n^*)$ parameterised by the parameter set θ_n^*. Introducing the indicator variables $x = (x_n)_{n=1}^{N}$, with $x_n = c$ denoting that θ_n^* takes on the value of θ_c, the DPMM with DP priors can be expressed as

$$\begin{gathered} y_n|x_n = c; \quad \theta_c \sim p(y_n|\theta_c) \\ x_n|\pi(v) \sim \text{Mult}(\pi(v)) \\ v_c|\alpha \sim \text{Beta}(1, \alpha) \\ \theta_c|G_0 \sim G_0 \end{gathered} \tag{8}$$

IET Image Process., 2014, Vol. 8, Iss. 2, pp. 103–111
doi: 10.1049/iet-ipr.2013.0232

where $\pi(v) = \left(\pi_c(v)\right)_{c=1}^{\infty}$ is given by (7), and Mult($\pi(v)$) is a multinomial distribution with parameter $\pi(v)$.

3 Proposed method

3.1 New DPMM with GDM (DPMM_GDM)

It is noted that finite mixture model can be considered as a linear combination of prior probability and conditional probability from the expression of its mathematical formula. Traditional spatial constraints methods such as MRF/HMRF only pay special attention on the former [40]. For incorporating more local spatial information, in this paper, we adopt the GDM on these two items (prior probability and conditional probability) to make the traditional DPMM more robust to noise for image segmentation application.

For image-processing application, let y_n, with dimension d, $n = (1, 2, \ldots, N)$, denotes the intensity value at the ith pixel of an image and let x_n, $n = (1, 2, \ldots, N)$, denotes the labels of the components that y_n belongs to. Then, $x_n = c$ $(c = 1, 2, \ldots, K)$ denotes the corresponding c class label of the nth pixel. Different from traditional DPMM, in our model, the new conditional probability is defined as GDM[$p(y_n|\theta_c)$], where the original conditional probability $p(y_n|\theta_c)$ is defined in (8). Let us assume it satisfies the Gaussian distribution and such a model corresponds to the Gaussian DPMM. For incorporating local spatial information in image processing, we modify the GDM in (1) to make it not focus on pixel n, but on the neighbourhood widow of the nth pixel. With the help of the geometric mean, we have

$$p(y_n|\theta_c) = \prod_{m\in\mathcal{N}_n} \left[p(y_m|\theta_c)\right]^{1/\mathcal{N}_n} \tag{9}$$

where $\mathcal{N}_n$ is the neighbourhood of the nth pixel, including the nth pixel itself. In traditional DPMM, the observation y_n satisfies the conditional probability $p(y_n/\theta_c)$. In our model, this item is influenced by the conditional probabilities of nth neighbourhood pixels for incorporating the local spatial information.

It is noted that the convolution of two Gaussian functions is still a Gaussian function. Thus, (9) can be modified as

$$\prod_{m\in\mathcal{N}_n} \left[p(y_m|\theta_c)\right]^{1/\mathcal{N}_n} = p(y_{n'}|\theta_c) \tag{10}$$

$y_{n'}$ denotes the modified observation generated by the nth neighbourhood pixels y_m. One possible approximation may be $\bar{y}_n = \sum_{m\in\mathcal{N}_n} y_m/\mathcal{N}_n$ (application of AM on y_n). In this case, the proposed model degrades to the standard DPMM after a pre-processing by applying an arithmetic mean on image intensity value y_n. However, this approximation processing may lead to computation error. Thus, we adopt (9) instead of (10) in this paper. It is significant to point out that the kernel of Gaussian function $p(y_n|\theta_c)$ can be considered as a distance measure from observation y_n to parameter θ_c in standard DPMM. In our model, this distance is measured by the modified observation $y_{n'}$ to parameter θ_c for incorporating spatial information and observation information to make the model more robust to noise.

The new DPMM with the GDM can be expressed as

$$y_n|x_n = c; \quad \theta_c \sim \prod_{m\in\mathcal{N}_n} \left[p(y_m|\theta_c)\right]^{1/\mathcal{N}_n}$$

$$x_n|\pi(v) \sim \mathrm{Mult}(\pi(v))$$

$$v_{nc}|\alpha_n \sim \mathrm{Beta}(1, \alpha_n)$$

$$\theta_c|G_0 \sim G_0 \tag{11}$$

where $\pi(v) = \left(\pi_c(v)\right)_{c=1}^{\infty}$ is given by (7), and Mult($\pi(v)$) is a multinomial distribution with parameter $\pi(v)$. Here, each observation y_n is assumed to be drawn from the modified conditional probability density function $\prod_{m\in\mathcal{N}_n} \left[p(y_m|\theta_c)\right]^{1/\mathcal{N}_n}$ parameterised by the parameter set θ_c.

Another limitation of traditional DPMM is its lack of an explicit consideration of the spatial dynamics (interdependencies) between the neighbouring sites on the input lattice. One possible solution is to impose an additional MRF distribution over the component densities [7, 8]. However, such an HMRF model is complex and time-consuming for implementation. In this paper, we introduce an alternative method by applying a GDM on the component prior probability for incorporating the spatial information and component information. Thus, the new component prior probability is generalised by

$$p(x_n = c|\pi(v)) = \sum_{m\in\mathcal{N}_n} \frac{1}{\mathcal{N}_n} p(x_m = c|\pi(v)) \tag{12}$$

It is noted that the probability $p(x_n = c/\pi(v))$ denotes the 'possibility' that observation y_n belongs to class x_n. This probability may generate a wrong value under the noise effect. In our model, this probability is influenced by the probabilities in its neighbourhoods. As long as the signal strength is greater than the noise strength in the neighbourhood domain, the correct probability can always be estimated.

The graphic model of DPMM and DPMM_GDM are shown in Figs. 1*a* and *b*, respectively.

3.2 Connection to existing methods

Non-parametric Bayesian approaches based on DPMM have been used in much research for image processing applications. Among these recent works, few consider incorporating the spatial constraints into the DPMM. The work in [7] is based on the introduction of a spatially constrained variant of the DP where spatial smoothness constraints on the class assignments are enforced by an MRF. Based on the Pólya urn representation, we have

$$p\left(\theta_N^* | \{\theta_n^*\}_{n=1}^{N-1}, G_0, \alpha\right) \propto \frac{\alpha}{Z} G_0 + \sum_{c=1}^{K} M\left(\theta_N^* | \{\theta_n^*\}_{n=1}^{N-1}\right) f_c^{N-1} \delta_{\theta_c} \tag{13}$$

where f_c^{N-1} is the number of distinct values in $\{\theta_n^*\}_{n=1}^{N-1}$ that equal to θ_c, Z is a normalising constant, and $M\left(\theta_N^* | \{\theta_n^*\}_{n=1}^{N-1}\right)$ is an MRF prior integrated in the mechanics of the derived DP variant that enforces the smoothness constraint (for the variables θ_n^*).

The work in [8, 9] introduces a model of the spatial dynamics (interdependencies) between the neighbouring sites on the input lattice by imposing an additional MRF (Gibbsian) distribution over the DPMM component densities emitting the observable data. The component prior

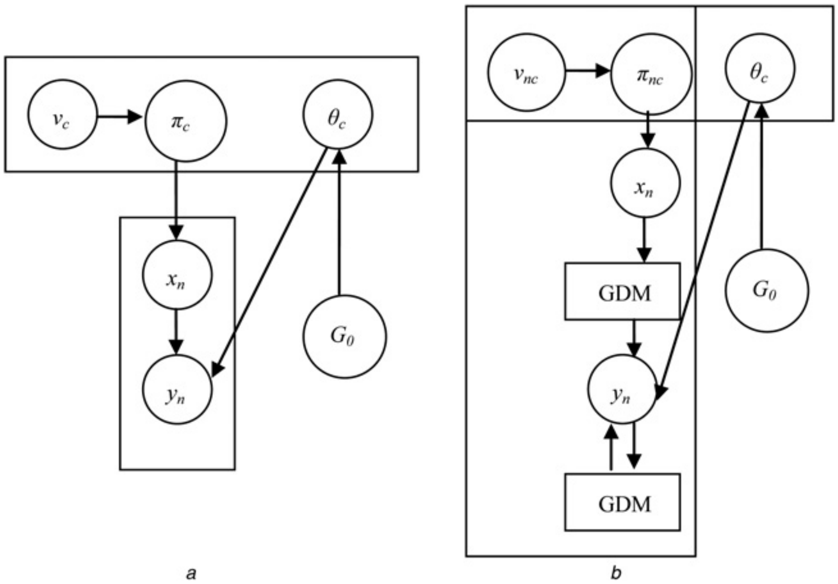

Fig. 1 *Graphic model of DPMM and DPMM_GDM*
a Graph of DPMM
b Graph of proposed DPMM_GDM

probability is given as

$$p\left(x_n = c|\pi(v),\ \hat{x}_{\partial_n}\right) = p\left(x_n = c|\hat{x}_{\partial_n};\ \beta\right)p(x_n = c|\pi(v)) \tag{14}$$

where $p\left(x_n = c|\hat{x}_{\partial_n};\ \beta\right)$ are the pointwise MRF prior probabilities of the model states obtained by application of the mean-field like approximation [37].

From (13), we can see that [7] introduces a spatially constrained variant of the original DP where the MRF is imposed internally in the DP process mechanics. Different from [7], we can observe from (14) that [8, 9] imposes an (approximate) Markov–Gibbsian field directly on the component labels x, but not on the parameter θ_n^*. Comparing (12) with (13) and (14), in our model, we not only apply the GDM on the component labels x but also on the conditional probability $p(y_n|\theta_c)$. It is noted that [7] adopts Pólya urn representation and the MCMC algorithm, and [8, 9] exploits the stick-breaking representation and VB inference algorithm. However, our method does not concern the DP representation and parameter learning algorithm, but focuses on the spatial constraints to incorporate local spatial information.

4 Variational approximation

To obtain the estimation of parameters, we maximise the marginal likelihood $p(y)$ by integrating out the variables as follows

$$p(y) = \int p\left(y,\ \Psi\right)\mathrm{d}\Psi \tag{15}$$

where $\Psi = \{v, \alpha, x, \theta\}$ denotes the set of variables. Here, both the latent variables and model parameters are treated as stochastic model variables. The distribution of variables v and x is given in (11). We then choose the Gamma and the normal-Wishart distribution as the priors for variables α and θ as follows

$$p\left(\alpha_n\right) = \mathcal{G}\left(\alpha_n|\eta_{n1},\ \eta_{n2}\right) \tag{16}$$

$$p\left(\theta_c\right) = p\left(\mu_c,\ R_c\right) = \mathrm{NW}\left(\mu_c,\ R_c|\lambda_c,\ m_c,\ \omega_c,\ \varphi_c\right) \tag{17}$$

It is noted that separate normal and Wishart priors can also be imposed over the means and precisions of the Gaussian distribution, respectively [41, 42]. However, considering the correlations of these variables, we select the normal-Wishart distribution priors to make the model less sensitive to outliers [43].

The integral in (15) denotes the joint integration over continuous variables $\{v,\ \alpha,\ \mu,\ R\}$ and the summation over discrete variables x. Since the integration in (15) is intractable, an alternative way to solve this problem is by using the VB method, which aims to maximise a lower bound L of the logarithmic marginal likelihood $p(y)$

$$L(q) = \int q(\Psi)\log\frac{p(y,\ \Psi)}{q(\Psi)}\mathrm{d}\Psi \le \log p(y) \tag{18}$$

where $q(\Psi)$ is an arbitrary distribution that provides an approximation to the true posterior distribution $p(\Psi/y)$. We see that the function $L(q)$ forms a rigorous lower bound on the true log marginal likelihood. Although the computation of the original log likelihood function log $p(y)$ is not tractable, the lower bound $L(q)$ may be tractable enough to compute by choosing a suitable form for the q distribution. The difference between the lower bound $L(q)$ and the true log likelihood log $p(y)$ is the Kullback–Leibler (KL) divergence. The KL divergence is non-negative and is zero when the variational posterior is equal to the true posterior $q(\Psi) = p(\Psi/y)$. The goal in a variational approach is to choose a suitable form for q, such that the lower bound becomes maximised – that is, the KL divergence becomes minimised. For this purpose, we approximate p while optimising q by minimising the KL divergence. To make progress, we assume a factorised variational distribution of

IET Image Process., 2014, Vol. 8, Iss. 2, pp. 103–111
doi: 10.1049/iet-ipr.2013.0232

the form

$$q(\Psi) = q(x)q(\alpha)q(v)q(\theta) \tag{19}$$

Factorisation of $q(\Psi)$ of the form (19) is a common approach and has been successfully used in VB inference [42–44]. Minimising the KL divergence with respect to all possible function forms of q, the standard variational approach provides the following general form of the solutions

$$q(\Psi_i) = \frac{\exp\langle \log p(\boldsymbol{y}, \Psi)\rangle_{k \neq i}}{\int \exp\langle \log p(\boldsymbol{y}, \Psi)\rangle_{k \neq i} d\Psi_i} \tag{20}$$

where $\langle\cdot\rangle_{k \neq i}$ denotes an expectation with respect to the distributions $q_k(\Psi_k)$ for all $k \neq i$. In (20), the marginal distributions under which the expectations are taken are the Markov blanket of the marginal in question. It is noticed that the optimal variational posteriors are expected to take the same functional form as the corresponding conjugate priors [45]. Thus, the factors of the variational posterior are given by calculation of (20) as follows

$$q(v_{nc}) = \mathrm{Beta}(\beta_{nc,1}, \beta_{nc,2}) \tag{21}$$

where

$$\beta_{nc,1} = 1 + q(x_n = c)$$
$$\beta_{nc,2} = \langle \alpha_n \rangle + \sum_{c'=c+1}^{K} q(x_n = c') \tag{22}$$

and

$$q(\alpha_n) = \mathcal{G}(\alpha_n | \tilde{\eta}_{n1}, \tilde{\eta}_{n2}) \tag{23}$$

where

$$\tilde{\eta}_{n1} = \eta_{n1} + K + 1$$
$$\tilde{\eta}_{n2} = \eta_{n2} - \sum_{c=1}^{K-1} \left[\psi(\beta_{nc,2}) - \psi(\beta_{nc,1} + \beta_{nc,2})\right] \tag{24}$$

and $\langle \alpha_n \rangle = \tilde{\eta}_{n1}/\tilde{\eta}_{n2}$, $\psi(\cdot)$ denotes the digamma function.

Similarly, let us consider the posteriors over the variable θ

$$q(\theta_c) = q(\mu_c, R_c) = \mathrm{NW}(\mu_c, R_c | \tilde{\lambda}_c, \tilde{m}_c, \tilde{\omega}_c, \tilde{\varphi}_c) \tag{25}$$

where we first introduce the notation

$$\tilde{\gamma}_c = \sum_{n=1}^{N} q(x_n = c)$$

$$\bar{y}_c = \frac{\sum_{n=1}^{N} \sum_{m \in \mathcal{N}_n} \frac{1}{\mathcal{N}_n} q(x_n = c) y_m}{\tilde{\gamma}_c}$$

$$\Delta_c = \sum_{n=1}^{N} \sum_{m \in \mathcal{N}_n} \frac{1}{\mathcal{N}_n} q(x_n = c)(y_m - \bar{y}_c)(y_m - \bar{y}_c)^{\mathrm{T}} \tag{26}$$

Then, we have

Fig. 2 *Some segmentation results obtained from the evaluated algorithms*

First column: Original images from Berkeley image data set. Second column: Results of the MDP/MRF in [7]. Third column: Results of the IHMRF in [8]. Fourth column: Results of the DPMM_GDM

$$\tilde{\lambda}_c = \lambda_c + \tilde{\gamma}_c$$

$$\tilde{m}_c = \frac{\lambda_c m_c + \tilde{\gamma}_c \bar{y}_c}{\tilde{\lambda}_c}$$

$$\tilde{\omega}_c = \omega_c + \tilde{\gamma}_c$$

$$\tilde{\varphi}_c = \varphi_c + \Delta_c + \frac{\lambda_c \tilde{\gamma}_c}{\lambda_c + \tilde{\gamma}_c}\left(m_c - \bar{y}_c\right)\left(m_c - \bar{y}_c\right)^{\mathrm{T}} \quad (27)$$

Finally, the component posterior probability is generalised by

$$q(x_n = c) = \sum_{m \in \mathcal{N}_n} \frac{1}{\mathcal{N}_n}\left[\tilde{\pi}_{mc}(v)\tilde{p}\left(y'_m|\theta_c\right)\right] \quad (28)$$

$$\begin{aligned}\tilde{\pi}_{mc}(v) &= \exp\left(\langle \log \pi_{mc}(v)\rangle\right) \\ &= \exp\left[\sum_{j=1}^{c-1}\left\langle \log\left(1 - v_{mj}\right)\right\rangle + \langle \log v_{mc}\rangle\right]\end{aligned} \quad (29)$$

with

$$\begin{aligned}\langle \log v_{mc}\rangle &= \psi\left(\beta_{mc,1}\right) - \psi\left(\beta_{mc,1} + \beta_{mc,2}\right) \\ \langle \log(1 - v_{mc})\rangle &= \psi\left(\beta_{mc,2}\right) - \psi\left(\beta_{mc,1} + \beta_{mc,2}\right)\end{aligned} \quad (30)$$

and

$$\begin{aligned}\tilde{p}\left(y'_m|\theta_c\right) &= \exp\left(\langle \log p\left(y'_m|\theta_c\right)\rangle\right) \\ &= \exp\left[\sum_{m' \in \mathcal{N}_m} \frac{1}{\mathcal{N}_m}\left\langle \log p\left(y_{m'}|\theta_c\right)\right\rangle\right]\end{aligned} \quad (31)$$

where

$$\begin{aligned}\langle \log p(y_{m'}|\theta_c)\rangle = &-\frac{d}{2}\log 2\pi + \frac{1}{2}\left\langle \log\left|R_c\right|\right\rangle \\ &-\frac{1}{2}\left\langle \left(y_{m'} - \mu_c\right)^{\mathrm{T}} R_c\left(y_{m'} - \mu_c\right)\right\rangle\end{aligned}$$

Table 1 Comparison of different methods for Berkeley image data set, probabilistic rand (PR) index (%)

Image #	MDP/MRF	IHMRF	DPMM_GDM
126007	80.32	81.76	90.87
220075	71.05	75.15	76.11
38092	80.91	81.83	81.26
130026	50.02	49.22	50.67
385039	74.54	82.24	85.87
170057	76.97	77.05	81.29
223061 + Gaussian noise (mean = 0, variance = 0.02)	70.01	68.50	71.17
101085 + Gaussian noise (mean = 0, variance = 0.02)	76.59	72.05	73.33
mean	72.55	73.48	76.32
computation time	488.99 s	1065.1 s	390.98 s

where

$$\begin{aligned}\langle (y_{m'} - \mu_c)^{\mathrm{T}} R_c (y_{m'} - \mu_c)\rangle = &\frac{d}{\tilde{\lambda}_c} + \tilde{\omega}_c\left(y_{m'} - \tilde{m}_c\right)^{\mathrm{T}} \\ &\times \tilde{\varphi}_c^{-1}\left(y_{m'} - \tilde{m}_c\right)\end{aligned}$$

$$\left\langle \log\left|R_c\right|\right\rangle = -\log\left|\frac{\tilde{\varphi}_c}{2}\right| + \sum_{k=1}^{d}\psi\left(\frac{\tilde{\omega}_c + 1 - k}{2}\right)$$

5 Experimental results and discussion

In this section, we experimentally evaluate our algorithm in a set of real images and a multidimensional noised image. We also evaluate MDP/MRF [7] and IHMRF [8, 9] for fair comparison. The source codes for the MDP/MRF and IHMRF algorithms can be downloaded from the authors' websites. Our experiments have been developed in MATLAB R2009b, and are executed on an Intel Pentium Dual-Core 2.2 GHz CPU, 2GB RAM.

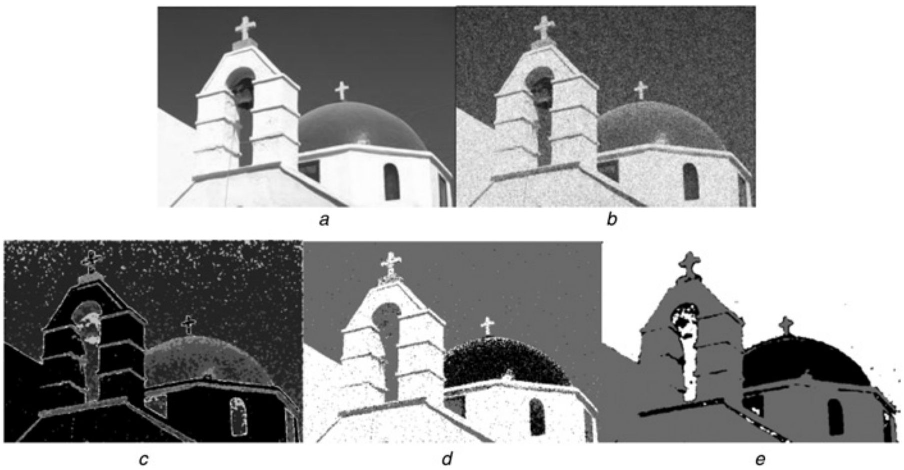

Fig. 3 *RGB image segmentation with image noise*

a Original image
b Noised image
c MDP/MRF, PR = 0.8065, t = 400 s
d IHMRF, PR = 0.7952, t = 305.91 s
e Proposed, PR = 0.8577, t = 132.11 s

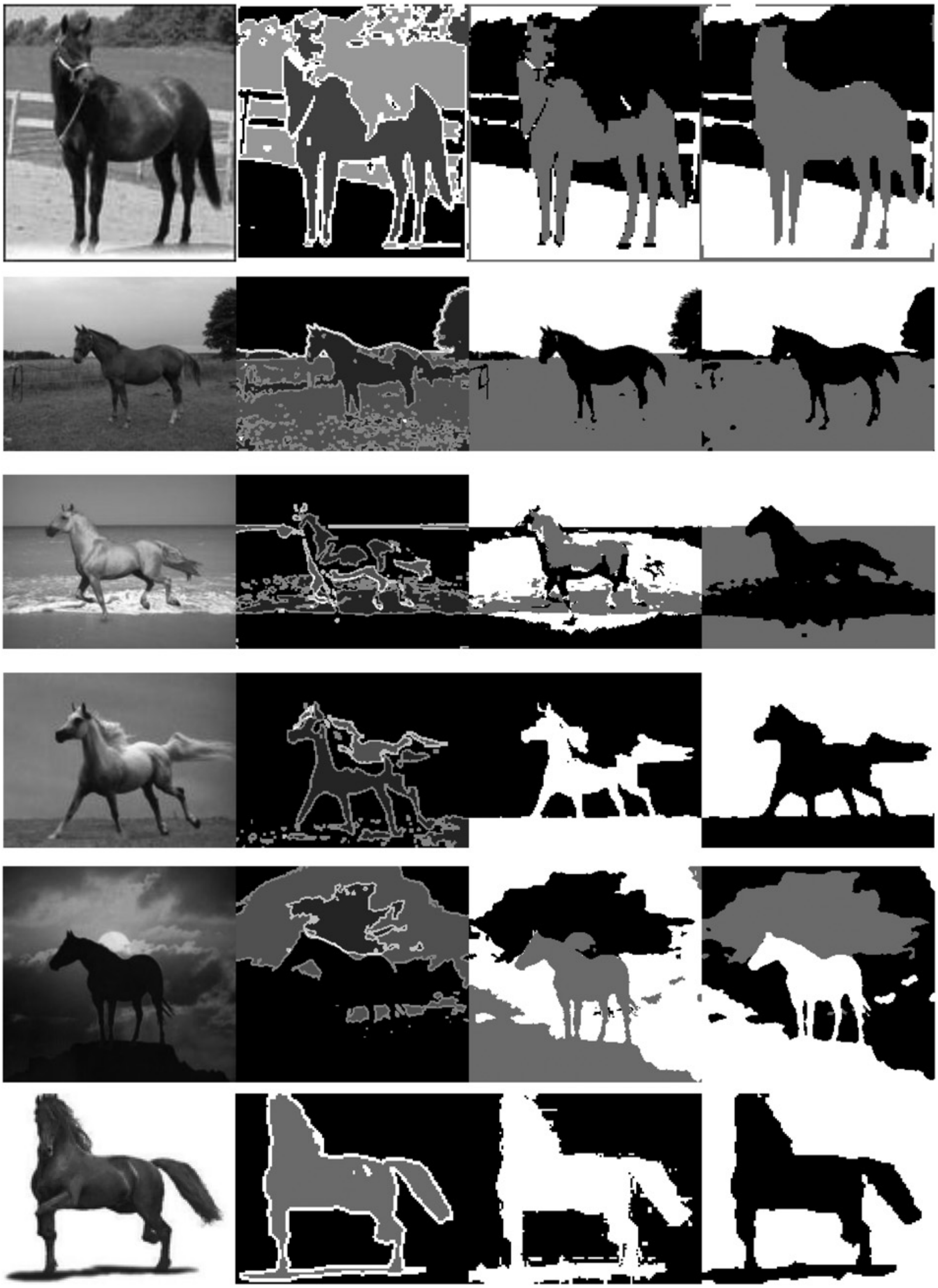

Fig. 4 *Some segmentation results of horses*

First column: Original images from Weizmann horse data set. Second column: Results of the MDP/MRF in [7]. Third column: Results of the IHMRF in [8]. Fourth column: Results of the proposed DPMM_GDM

In the first experiment, we evaluate the performance of different methods based on a subset of the Berkeley image data set [46], which comprises a set of real-world colour images along with segmentation maps provided by different individuals. We employ the probabilistic rand (PR) index [47] to evaluate the performance of the proposed method, with the multiple ground truths available for each image within the data set. It has been shown that the PR index possesses the desirable property of being robust to segmentation maps that result from splitting or merging segments of the ground truth [48]. The PR index takes values between 0 and 1, with values closer to 0 (indicating an inferior segmentation result) and values closer to 1 (indicating a better result).

Fig. 2 shows the original images and the segmentation results obtained by the evaluated methods. Table 1 presents the average PR values for all methods. Compared to other methods, the proposed method yields the best segmentation results with the highest PR values. From the images on the second row of Fig. 2, it can be seen that MDP/MRF and IHMRF misclassify some portions of pixels at the edge region between the sky and the mountain at the middle-left of the image. In contrast, the proposed DPMM_GDM segments the image well and shows perfect results. From the images on the fourth row of Fig. 2, we can observe that our DPMM_GDM can distinguish the sky and the top parts of the building well even if the image is corrupted by the Gaussian noise. This is expected due to the fact that the existing methods only consider some constrains on component labels. Different from previous methods, we not only consider component labels but also consider some constraints on the conditional probabilities (distributions).

We also evaluate the computation time for all methods in the previous experiment. The average computation time t (in seconds) of the different methods is presented on the last line of Table 1. It is noted that the computation of our methods is much faster than that of other methods. This may be the contribution of the usage of simple GDM instead of complex MRF in our model. Compared to other methods, our algorithm can be calculated most quickly and achieve the best segmentation results.

In the second experiment, we try to segment the multidimensional RGB colour image into three classes: the blue sky, the red roof and the white wall. The original image (481 × 321) shown in Fig. 3*a* is corrupted by heavy Gaussian noise, with mean = 0 and variance = 0.05. The noised image is shown in Fig. 3*b*, and the segmentation results of MDP/MRF, IHMRF and our DPMM_GDM are shown in Figs. 3*c*–*e*, respectively. The accuracy of segmentation for MDP/MRF is quite poor. Although IHMRF obtains better results, it is still sensitive to heavy noise. The accuracy of the segmentation result from the proposed method, as shown in Fig. 3*e*, is better than that of other methods, obtaining the highest PR values.

In the last experiment, we use the Weizmann data set [49, 50], which contains 328 images of horses with different poses, sizes, face directions, backgrounds and illumination conditions. There is only one horse in each image, and there is a single object category in the data set: horse. The segmentation results of different methods are shown in Fig. 4, from which we can see that the proposed method obtains the best performance compared with its competitors. From Fig. 4, we can observe that MDP/MRF always misclassifies some pixels in the background – for example, lawn in the second and fourth row image, water in the third row image. One major problem with IHMRF is that it is more likely to separate parts from the same object into different segments. For example, it misclassifies some back parts of the horse in the first and fourth row image. IHMRF also ‘ignores’ some pixels of the horse’s legs and tail in the second and third row images, respectively. Moreover, in the fifth row image, IHMRF segments the object (horse) and background (sun) into the same group. However, the segmentation results of DPMM_GDM do not show these phenomena. Therefore, we argue that enforcing spatial coherence between adjacent regions via DPMM_GDM avoids separating parts of the same object into different groups.

6 Conclusions

In this paper, we first introduce a new DPMM and then use GDM for incorporating the local spatial information, observation information and component information. Compared to traditional spatial constraints methods – for example, MRF/HMRF algorithm – our algorithm is simple, easy and effective for implementation. Moreover, MRF/HMRF needs the additional parameter β to keep a balance between robustness to noise and image sharpness and details. Different from the MRF/HMRF model, our model is fully free of the empirically adjusted parameter β. Empirical studies on two image data sets demonstrate the improvement of our model in image segmentation. Finally, in this paper, we focus on Gaussian DPMM. In fact, our algorithm is general enough and can be applied to other finite mixture models [45, 51].

7 Acknowledgments

This work was supported in part by the Canada Chair Research Program and the Natural Sciences and Engineering Research Council of Canada. This work was supported in part by the National Natural Science Foundation of China under grants 61232016, 61105007 and 61103141. This work was supported in part by the National Basic Research Programme 973 (2011CB311808), 201301030 and 2013DFG12860. This work was supported in part by PAPD (a project funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions).

8 References

1 Ferguson, T.: ‘A Bayesian analysis of some nonparametric problems’, *Ann. Stat.*, 1973, **1**, pp. 209–230
2 Antoniak, C.: ‘Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems’, *Ann. Stat.*, 1974, **2**, (6), pp. 1152–1174
3 Blei, D.M., Jordan, M.I.: ‘Variational inference for Dirichlet process mixtures’, *Bayesian Anal.*, 2006, **1**, (1), pp. 121–144
4 Neal, R.: ‘Markov chain sampling methods for Dirichlet process mixture models’, *J. Comput. Graph. Stat.*, 2000, **9**, pp. 249–265
5 Blei, D., Jordan, M.: ‘Variational methods for the Dirichlet process’. Proc. 21st Int. Conf. Machine Learning, New York, 2004, pp. 12–19
6 da Silva, A.R.F.: ‘A Dirichlet process mixture model for brain MRI tissue classification’, *Med. Image Anal.*, 2007, **2**, pp. 169–182
7 Orbanz, P., Buhmann, J.: ‘Nonparametric Bayes image segmentation’, *Int. J. Comput. Vis.*, 2008, **77**, pp. 25–45
8 Chatzis, S.P., Tsechpenakis, G.: ‘The infinite hidden Markov random field model’. Proc. 12th Int. IEEE Conf. Computer Vision, Kyoto, Japan, 2009, pp. 654–661
9 Chatzis, S.P., Tsechpenakis, G.: ‘The infinite hidden Markov random field model’, *IEEE Trans. Neural Netw.*, 2010, **21**, (6), pp. 1004–1014
10 Chatzis, S.P., Korkinof, D., Demiris, Y.: ‘A spatially-constrained normalized gamma process prior’, *Expert Syst. Appl.*, 2012, **39**, pp. 13019–13025
11 Wei, X., Li, C.: ‘The infinite student’s *t*-mixture for robust modeling’, *Signal Process.*, 2012, **92**, (1), pp. 224–234
12 Qi, Y., Paisley, J.W., Carin, L.: ‘Music analysis using hidden Markov mixture models’, *IEEE Trans. Signal Process.*, 2007, **55**, (11), pp. 5209–5224
13 Paisley, J., Carin, L.: ‘Hidden Markov models with stick-breaking priors’, *IEEE Trans. Signal Process.*, 2009, **57**, (10), pp. 3905–3917

doi: 10.1049/iet-ipr.2013.0232

14 Du, L., Chen, M., Lucas, J., Carin, L.: 'Sticky hidden Markov modeling of comparative genomic hybridization, signal processing', *IEEE Trans. Signal Process.*, 2010, **58**, (10), pp. 5353–5368
15 Bouguila, N., Ziou, D.: 'A Dirichlet process mixture of generalized Dirichlet distributions for proportional data modeling', *IEEE Trans. Neural Netw.*, 2010, **21**, (1), pp. 107–122
16 Bouguila, N.: 'Count data modeling and classification using finite mixtures of distributions', *IEEE Trans. Neural Netw.*, 2011, **22**, (2), pp. 189–198
17 Destrempes, F., Angers, J.F., Mignotte, M.: 'Fusion of hidden Markov random field models and its Bayesian estimation', *IEEE Trans. Image Process.*, 2006, **15**, (10), pp. 2920–2935
18 Nikou, C., Likas, A., Galatsanos, N.P.: 'A Bayesian framework for image segmentation with spatially varying mixtures', *IEEE Trans. Image Process.*, 2010, **19**, (9), pp. 2278–2289
19 Chen, S.F., Cao, L.L., Wang, Y.M., Liu, J.Z., Tang, X.O.: 'Image segmentation by MAP-ML estimations', *IEEE Trans. Image Process.*, 2010, **19**, (9), pp. 2254–2264
20 Hedjam, R., Mignotte, M.: 'A hierarchical graph-based Markovian clustering approach for the unsupervised segmentation of textured color images'. ICIP 2009, pp. 1365–1368
21 Nguyen, T.M., Jonathan Wu, Q.M.: 'Dirichlet Gaussian mixture model: application to image segmentation', *Image Vis. Comput.*, 2011, **29**, (12), pp. 818–828
22 Nguyen, T.M., Jonathan Wu, Q.M.: 'Fast and robust spatially constrained Gaussian mixture model for image segmentation', *IEEE Trans. Circuits Syst. Video Technol.*, 2013, **23**, (4), pp. 621–635
23 Peng, B., Zhang, L., Zhang, D.: 'A survey of graph theoretical approaches to image segmentation', *Patt. Recognit.*, 2013, **46**, (3), pp. 1020–1038
24 Geman, S., Geman, D.: 'Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1984, **6**, (6), pp. 721–741
25 Clifford, P.: 'Markov random fields in statistics', in: Grimmett, G., Welsh, D. (eds): 'Disorder in physical systems. A volume in honour of John M. Hammersley on the occasion of his 70th birthday' (Clarendon Press, Oxford Science Publication, Oxford, UK, 1990)
26 Rabiner, L.R.: 'A tutorial on hidden Markov models and selected applications in speech recognition', *Proc. IEEE*, 1989, **77**, (2), pp. 257–286
27 Zhang, Y., Brady, M., Smith, S.: 'Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm', *IEEE Trans. Med. Imag.*, 2001, **20**, (1), pp. 45–57
28 Dempster, P., Laird, N.M., Rubin, D.B.: 'Maximum likelihood from incomplete data via EM algorithm', *J. R. Stat. Soc. B*, 1977, **39**, (1), pp. 1–38
29 Archer, G.E.B., Titterington, D.M.: 'Parameter estimation for hidden Markov chains', *J. Stat. Plan. Inference*, 2002, **108**, pp. 364–390
30 Besag, J.: 'Statistical analysis of non-lattice data', *The Statistician*, 1975, **24**, pp. 179–195
31 Zhang, J., Modestino, J.W., Langan, D.: 'Maximum-likelihood parameter estimation for unsupervised stochastic model-based image segmentation', *IEEE Trans. Image Process.*, 1994, **3**, pp. 404–420
32 McLachlan, G.J., Krishnan, T.: 'The EM algorithm and extensions'. Series in Probability and Statistics, New York, Wiley, 1997
33 Zhou, Z., Leahy, R., Qi, J.: 'Approximate maximum likelihood hyperparameter estimation for Gibbs priors', *IEEE Trans. Image Process.*, 1997, **6**, (6), pp. 844–861
34 Besag, J.: 'On the statistical analysis of dirty pictures', *J. R. Stat. Soc. B*, 1986, **48**, pp. 259–302
35 Qian, W., Titterington, D.: 'Estimation of parameters in hidden Markov models', *Philos. Trans. R. Soc. Lond. A*, 1991, **337**, pp. 407–428
36 Zhang, J.: 'The mean field theory in EM procedures for blind Markov random field image restoration', *IEEE Trans. Image Process.*, 1993, **2**, (1), pp. 27–40
37 Forbes, F., Peyrard, N.: 'Hidden Markov random field model selection criteria based on mean field-like approximations', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2003, **25**, (9), pp. 1089–1101
38 Blackwell, D., MacQueen, J.: 'Ferguson distributions via Pólya urn schemes', *Ann. Statist.*, 1973, **1**, (2), pp. 353–355
39 Sethuraman, J.: 'A constructive definition of the Dirichlet prior', *Stat. Sin.*, 1994, **2**, pp. 639–650
40 Zhang, H., Wu, Q.M.J., Thanh, M.N.: 'Incorporating mean template into finite mixture model for image segmentation', *IEEE Trans. Neural Netw. Learn. Syst.*, 2013, **24**, (2), pp. 328–335
41 van Lieshout, M.N.M.: 'Markovianity in space and time'. Dynamics and Stochastics: Festschrift in Honor of Michael Keane, 2006, pp. 154–167
42 Corduneanu, A., Bishop, C.M.: 'Variational Bayesian model selection for mixture distributions'. Artificial Intelligence and Statistics 2001, San Mateo, CA: Morgan Kaufmann, 2001, pp. 27–34
43 Svensen, M., Bishop, C.M.: 'Robust Bayesian mixture modelling', *Neurocomputing*, 2005, **64**, pp. 235–252
44 Archambeau, C., Verleysen, M.: 'Robust Bayesian clustering', *Neural Netw.*, 2007, **20**, (1), pp. 129–138
45 Bishop, C.M.: 'Pattern recognition and machine learning' (Springer, 2006)
46 Martin, D., Fowlkes, C., Tal, D., Malik, J.: 'A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics'. Proc. 8th IEEE Int. Conf. Computer Vision, Vancouver, BC, Canada, 2001, vol. 2, pp. 416–423
47 Unnikrishnan, R., Pantofaru, C., Hebert, M.: 'A measure for objective evaluation of image segmentation algorithms'. IEEE Conf. Computer Vision and Pattern Recognition, 2005, vol. 3, pp. 34–41
48 Unnikrishnan, R., Hebert, M.: 'Measures of similarity'. Proc. IEEE Workshop Computer Vision Application, 2005, pp. 394–400
49 Shotton, J., Blake, A., Cipolla, R.: 'Contour-based learning for object detection'. Proc. Ninth Int. IEEE Conf. Computer Vision, Washington, DC, USA, 2005, pp. 503–510
50 Zhao, B., Li, F.-F., Xing, E.P.: 'Image segmentation with topic random field'. Proc. 11th European Conf. Computer Vision, Crete, Greece, 2010
51 McLachlan, G., Peel, D.: 'Finite mixture models' (Wiley, New York, 2000)