www.ietdl.org

Published in IET Image Processing
Received on 21st June 2012
Revised on 15th October 2012
Accepted on 8th November 2012
doi: 10.1049/iet-ipr.2012.0340

ISSN 1751-9659

Image segmentation by a new weighted Student's *t*-mixture model

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, People's Republic of China*
[2]*Department of Electrical and Computer Engineering, University of Windsor, Canada N9B 3P4*
[3]*Jiangsu Engineering Center of Network Monitoring, Nanjing University of Information Science and Technology, People's Republic of China*
E-mail: jwu@uwindsor.ca

Abstract: In this study, the authors introduce a new weighted Student's *t*-mixture model (WSMM) for image segmentation. Gaussian distribution and Student's *t*-distribution are the two commonly used probabilities in the finite mixture model (FMM). The Student's *t*-mixture model has come to be regarded as an alternative to Gaussian mixture models, as it is heavily tailed and more robust for outliers. Moreover, the pixels are considered independent of each other in the FMM. Although some existing methods incorporate the spatial relationship between neighbouring pixels, they do not consider the relationship between spatial information and clustering information, thus those reported methods remain sensitive to noise. The advantages of the authors method are as follows: first, the authors introduce WSMM to incorporate the local spatial information, pixel intensity value and clustering information in an image. Second, the authors model is simple, easy to implement and has a good balance between noise insensitiveness and image detail preservation. Third, they adopt the gradient method and expectation maximisation algorithm, which allow for simultaneous estimation of optimal parameters. Finally, the most useful statistical tool for image segmentation, the well-known hidden Markov random field model, is a special case of their model. Thus, their method is general enough for model-based techniques construction. Experimental results on synthetic and real images demonstrate the improved robustness and effectiveness of their approach.

1 Introduction

Image segmentation is one of the most important and difficult problems in image processing. Although different methodologies [1–6] have been proposed for image segmentation, it remains a challenge because of overlapping intensities, low contrast of images and noise perturbation. Generally, image segmentation can be divided into four categories: thresholding, clustering, edge detection and region extraction. Clustering is essentially a process of pixel classification to segment the image pixels into subsets. Given a known or estimated number of classes, all image pixels can be clustered into the above classes and a unique numeric label will be assigned to each class.

One of the most widely used clustering models for image segmentation is the well-known finite mixture model (FMM). The Gaussian mixture model (GMM) is most commonly selected as a particular case of the FMM by assuming the conditional probability as a Gaussian distribution [7–9]. The GMM is a flexible and powerful statistical modelling tool for multivariate data. The main advantage of the standard GMM is that it is easy to implement and the small number of parameters can be efficiently estimated by adopting the expectation maximisation (EM) algorithm. However, the GMM is weak in parameter estimation in the case of heavy outliers. In many practical applications the robustness for outlying data are crucial, because the mixture of outliers and observable data may severely affect the estimation of the model parameters as well as the model's complexity. Thus, additional components are needed to capture the tails of the distributions. When outliers are present in the available fitting datasets (as often happens in real-word applications), GMMs tend to require excessively high number of mixture components to capture the long tails of the approximated distributions so as to retain their pattern recognition effectiveness [10].

Student's *t*-mixture model (SMM), recently proposed by Peel and McLachlan [11] as an alternative to GMM, provides high robustness against noise. The SMM is more robust than the GMM for heavier tails, and the distribution for each component in the SMM is embedded in a wider class of elliptically symmetric distributions with an additional parameter called degrees of freedom. Hence, the FMM of the longer tailed multivariate Student's *t*-distribution provides a much more robust approach than the GMM. However, as a histogram-based model, FMM assumes each pixel in an image is independent of its neighbours. Thus, neither GMM nor SMM take into account the spatial dependencies in the image. Moreover, they do not use the prior knowledge that adjacent pixels most likely belong to the same cluster. Thus, the

performances of GMM and SMM remain sensitive to noise and image contrast levels.

To overcome this shortcoming, a wide variety of approaches have been proposed to incorporate spatial information in the image A common approach is the use of a Markov random field (MRF) [12]. Such methods aim to impose spatial smoothness constraints on the image pixel labels. However, most reported methods are limited to using MRF as a general prior in an FMM-based approach. Recently, a special case of the MRF model – the hidden MRF (HMRF) model – has been proposed [13, 14]. The state sequence of HMRF is directly unobservable but can be indirectly observed through a field of observations. In the HMRF model, the spatial information in an image is encoded through contextual constraints of neighbouring pixels, which are characterised with conditional MRF distributions. Parameter estimation in HMRF models usually relies on maximum likelihood (ML) or Bayesian methods [15, 16]. Besag [17] introduces the idea of the pseudo likelihood approximation when ML estimation is intractable. Based on this well-known approximation, various HMRF model estimation approaches have been proposed. To deal with different HMRF models, we can estimate characteristic parameters with iterative methods such as the EM algorithm or iterative conditional estimation [18–23]. In the E-step iteration, iterated conditional modes (ICM), the maximum a posterior (MAP) estimate and the maximum posterior marginal are alternative ways to replace image pixel labels by current estimation [24–26]. To overcome the biased parameter estimates, Qian and Titterington [27] propose an improved approximation to the EM solution called the point-pseudo-likelihood EM (PPL-EM) algorithm. This is closely related to a more recently proposed mean-field approximation of the MRF [28–31]. Dunmur and Titterington [32] reveal the relationship between these two methods.

Another type of mixture model, the spatially variant FMM (SVFMM) has also been successfully applied to image segmentation [33, 34]. Instead of imposing the MRF-based smoothness constraint on the pixel labels, this method considers imposing the smoothness on the contextual mixing proportions. In [33], Sanjay and Hebert adopt an MAP estimation with a Gibbs MRF-based prior, which enhances spatial smoothness and generates spatially continuous clusters. However, in this situation, the MAP algorithm cannot find the pixel labels in closed form, and the authors resort to the gradient project algorithm. In [35], quadratic programming is used to improve the problem of pixel label computation for the SVFMM. In [36], a family of smoothness priors for the contextual mixing proportions based on the Gauss–MRFs is proposed to guarantee different smoothness strength for each cluster. This model is also refined to capture information in different spatial directions. The inference is obtained by EM algorithm with closed form update equations. As a hierarchical Bayesian model with spatial constraints, Nikou *et al.* [37] considers the constraint of prior probabilities: they are positive and must sum to one. This model assumes that the contextual mixing proportions follow a Dirichlet compound multinomial (DCM) distribution. The parameters of the multinomial distribution are integrated out in a fully Bayesian framework, and the iterative updates for the parameter estimation of the Dirichlet are computed in closed form through the EM algorithm.

Although some studies argue that the spatial information enhances insensitivity to noise, these studies still lack sufficient robustness for noise and outliers. In this paper, we propose a new weighted SMM (WSMM) for image segmentation. Our approach differs from methods mentioned above in a number of ways: first, the prior probability in our method is different for each pixel and depends on the neighbours of the pixel; moreover, our approach considers the spatial relationship between neighbouring pixels and the relationship between the spatial information and clustering information. Second, for spatial information, we add weighting for distant pixels in order to establish a good balance between noise and image sharpness and details, as the weighted parameters decrease with the increasing distance. Owing to the weighted factor used to incorporate the local spatial information, we call our method the WSMM. Third, our approach is based on the FMM – it is simple and can be easily implemented. We adopt an EM algorithm and gradient method for parameter estimation. Furthermore, we construct a flexible framework, analyse the standard FMM in a new way, and reveal the intrinsic relationship between our approach and the HMRF model. In fact, the HMRF can be considered a special case of our model. The proposed approach is applied to segment synthetic and real images. The performance of our proposed approach compared with the state-of-art technology demonstrates its improved robustness and effectiveness.

The remainder of this paper is organised as follows: in Section 2, we briefly introduce the mathematical background of several methods commonly used in the literature for image segmentation. We also analyse the main problems caused by FMM in image segmentation. In Section 3, the proposed approach is described in detail. Parameters estimation is presented in Section 4. The experimental results for evaluating the performance of the proposed approach are given in Section 5. Finally, some concluding remarks are provided.

2 Preliminary theories

2.1 Student's t-distribution

The Student's t-distribution provides a heavy-tailed alternative to the Gaussian distribution for potential outliers. The probability density function (PDF) of a multivariate Student's t-distribution, with mean μ, covariance matrix $\mathbf{\Sigma}$ and degrees of freedom v is [1]

$$t(y|\mu, \mathbf{\Sigma}, v) = \frac{\Gamma(v/2 + p/2)}{\Gamma(v/2)} \frac{|\Sigma|^{-1/2}}{(\pi v)^{p/2}} \times \left[1 + \frac{(y-\mu)^T \Sigma^{-1} (y-\mu)}{v}\right]^{-(v+p)/2} \quad (1)$$

where p is the dimensionality of the variable y and $\Gamma(s)$ is the Gamma function

$$\Gamma(s) = \int_0^\infty t^{s-1} \mathrm{e}^{-t} \mathrm{d}t \quad (2)$$

From [11, 38], it can be shown that a Student's t-distribution $t(\mu, \mathbf{\Sigma}, v)$ is equivalent to a Gaussian distribution $\mathcal{N}(\mu, \mathbf{\Sigma}/u)$, with the precision scaling factor u, which is a Gamma-distributed latent variable, depending on the degree

of freedom v, defined as

$$t(y|\mu, \Sigma, v)\tilde{\mathcal{N}}(y|\mu, \Sigma/u)\mathcal{G}(u|v/2, v/2) \quad (3)$$

The PDF of the Gamma distribution $\mathcal{G}$ is given by

$$\mathcal{G}(u|\alpha, \beta) = \frac{1}{\Gamma(\alpha)}\beta^{\alpha}u^{\alpha-1}\mathrm{e}^{-\beta u} \quad (4)$$

Fig. 1 shows a graphical illustration of the Student's t-distribution, with μ, Σ fixed and for various values of the degrees of freedom v. It can be shown that for $v \to \infty$, the Student's t-distribution tends to a Gaussian distribution with the same mean μ and covariance Σ. On the contrary, as v tends to zero, the tails of the distribution become longer, providing a heavy-tailed alternative for potential outliers without affecting the mean or the covariance of the distribution. Thus, SMM can provide better computational efficiency, better algorithm stability and improved model parameter estimates compared to GMM [10, 39].

2.2 Finite mixture model

Let us first consider two sets $Q = \{1, 2,\ldots, K\}$ and $L = \{1, 2,\ldots, D\}$. Let S be a finite index set, $S = \{1, 2,\ldots, N\}$. We shall refer to set S as the set of sites or locations. Let X and Y be two random fields, their state space $\mathcal{X}$ and $\mathcal{Y}$ are indexed by the supposed set of sites S (every site $i \in S$), given by

$$\mathcal{X}_i = \{x_i : x_i \in Q\} \text{ and } \mathcal{Y}_i = \{y_i : y_i \in L\}$$

Their product space

$$\mathcal{X} = \prod_{i=1}^{N} x_i \text{ and } \mathcal{Y} = \prod_{i=1}^{N} y_i$$

shall be denoted as the space of the configurations of the state values of the considered site set, $\boldsymbol{x} = (x_i)$ and $\boldsymbol{y} = (y_i)$.

For image segmentation application, an image consisting of N pixels is segmented into K classes. y_i denotes the observation (intensity value) at the ith pixel of an image and x_i denotes the corresponding class label of the ith pixel.

For every $j \in Q$ and $i \in S$, the probability

$$p(x_i = j) = \pi_j \quad (5)$$

is the prior distribution of the pixel y_i, belonging to the class x_i, which satisfies the constraints

$$0 \le \pi_j \le 1 \text{ and } \sum_{j=1}^{K} \pi_j = 1 \quad (6)$$

$x_i = j$, y_i follows a conditional probability distribution $p(y_i|\theta_j)$, in which θ_j is the set of parameters. Specific to the SMM, the conditional probability $p(y_i|\theta_j)$ is selected as a Student's t-distribution. Under the independent assumption, we have

$$p(\boldsymbol{y}, \boldsymbol{x}) = \prod_{i=1}^{N} p(y_i, x_i) \quad (7)$$

and with the help of (5)–(7), the FMM can be calculated as

$$p(y_i|\boldsymbol{\pi}, \boldsymbol{\theta}) = \sum_{j=1}^{K} \pi_j p\left(y_i|\theta_{x_i}\right) \quad (8)$$

Although FMM is widely used for its simplicity and effectiveness as a model [1], it only describes the data statistically; no spatial information about the data is utilised. In other words, it does not take into account the spatial correlation between the neighbouring pixels in the decision process. However, images with the same intensity distribution may have totally different structural properties. To overcome the problems with the FMM, and to reduce the sensitivity of the segmentation result with respect to noise, several researchers have suggested modifications to incorporate the local spatial information between the neighbouring pixels. Two typical models are introduced in the subsection below.

2.3 Hidden MRF model

MRF accounts for contextual constraints by using spatial information with conditional MRF distributions. In an MRF, the sites in S are related to one another via a neighbourhood system, which is defined as $\mathcal{N} = \{\mathcal{N}_i, i \in S\}$, where $\mathcal{N}_i$ is the set of sites neighbouring $I - i \notin \mathcal{N}_i$ and $i \in \mathcal{N}_j$ if and only if $j \in \mathcal{N}_i \forall i, j \in S$. A random field X is said to be an MRF on S with respect to a neighbourhood system $\mathcal{N}$ if and only if [25]

$$p(\boldsymbol{x}) > 0, \quad \forall \boldsymbol{x} \in \mathcal{X}$$

$$p\left(x_i|\boldsymbol{x}_{S-\{i\}}\right) = p\left(x_i|\boldsymbol{x}_{\mathcal{N}_i}\right)$$

According to the Hammersley–Clifford theorem [40], a given random field is an MRF if and only if its probability distribution is a Gibbs distribution, thus

$$p(\boldsymbol{x}) = Z^{-1}\exp\left(-U(\boldsymbol{x})\right) \quad (9)$$

where Z is a normalising constant called the partition function

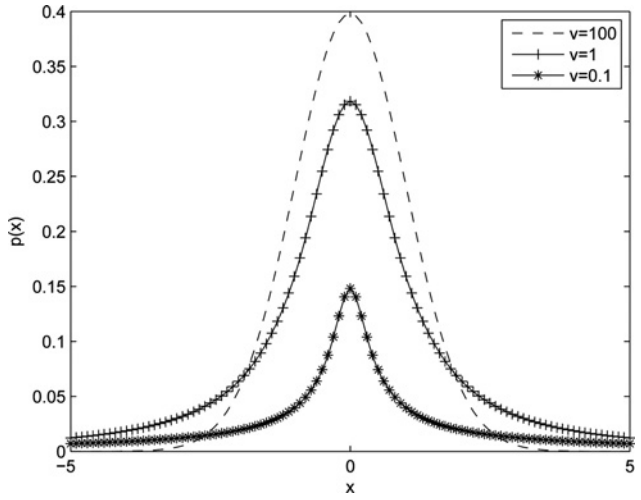

Fig. 1 *Univariate Student's t-distribution with μ and Σ fixed for various degrees of freedom*

As $v \to \infty$ the distribution tends to a Gaussian
For small values of v, the distribution has heavier tails than a Gaussian

doi: 10.1049/iet-ipr.2012.0340

given by

$$Z = \sum_{\boldsymbol{x}\in\mathcal{X}} \exp\left(-U(\boldsymbol{x})\right) \tag{10}$$

and $U(\boldsymbol{x})$ is an energy function of the form

$$U(\boldsymbol{x}) = \sum_{c\in C} V_c(\boldsymbol{x}) \tag{11}$$

where C is the all possible cliques and $V_c(\boldsymbol{x})$ stands for the clique potential associated with the clique c. A clique c is defined as a subset of sites in S in which every pair of distinct sites are neighbours, except for single-site cliques.

HMRF is a special case of HMMs, which are defined as stochastic processes with an unobservable state field X and an observable (or emitted) random field Y. The probability distribution $p(\boldsymbol{x})$ satisfies (9). Moreover, for any $\boldsymbol{x} \in \mathcal{X}$, the random variables y_i satisfy conditional independent assumption

$$p(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^{N} p(y_i|x_i) \tag{12}$$

which provides a convenient approximation of the posterior field $p(\boldsymbol{x}|\boldsymbol{y})$, still guaranteeing its Markovianity.

The computation of the term Z defined in (10) involves all possible realisations of $\boldsymbol{x}$ of the HMRF, which is hardly ever feasible in terms of computational requirements. Besag [17] introduces the idea of pseudo likelihood approximation of the Markov field prior to solve this problem. Under this approximation, the prior of the Markov field is

$$p(\boldsymbol{x}) = \prod_{i=1}^{N} p\left(x_i|\boldsymbol{x}_{\mathcal{N}_i}\right) \tag{13}$$

Relevant studies [1, 2, 14, 27] show that optimisation of the Markov field prior under the pseudo likelihood approximation (13) offers good estimates of the HMRF model parameters.

2.4 MRF on contextual mixing proportions

Another model, the SVFMM, imposes the MRF-based smoothness constraint on the contextual mixing proportions [33–37]. The main difference between FMM and SVFMM is the definition of prior probability, which equals π_j in FMM and π_{ij} (also called contextual mixing proportion) in SVFMM. MRF is used on contextual mixing proportion π_{ij} in SVFMM to incorporate the spatial information between the neighbouring pixels, and to consider the relationship between the pixel y_i and the class x_i. Here, the contextual mixing proportion π_{ij} still needs to satisfy the constraints in (6). Let $\boldsymbol{\pi}$ denote the configuration of the states π_{ij}, that is, $2\boldsymbol{\pi} = (\pi_{ij})_{i\in S,\ j\in Q}$. Similar to HMRF, the SVFMM method introduces the Gibbs function for prior probability

$$p(\boldsymbol{\pi}) = Z^{-1} \exp\left(-U(\boldsymbol{\pi})\right) \tag{14}$$

where

$$U(\boldsymbol{\pi}) = \sum_{c\in C} V_c(\boldsymbol{\pi}) \tag{15}$$

Many realisations of (15) are proposed in [33–37]. A possible solution is the Gauss–MRF given by

$$U(\boldsymbol{\pi}) = \beta \sum_{i=1}^{N} \sum_{m\in\mathcal{N}_i} \sum_{j=1}^{K} \left(\pi_{ij} - \pi_{mj}\right)^2$$

The traditional iterative EM algorithm is used for parameter learning in HMRF. However, because of the complexity of the log-likelihood function of SVFMM, the M step of the EM algorithm cannot be directly applied to the contextual mixing proportion π_{ij}. In [33], the authors apply a gradient projection algorithm to solve this problem. Quadratic programming is used to compute the pixel label priors in [35].

3 Weighted SMM

Let y_i, with dimension p, $i = (1, 2, \ldots, N)$, denote the intensity value at the ith pixel of an image and j ($j = 1, 2, \ldots, K$) denote the corresponding class label of the ith pixel. The WSMM is defined as

$$p(y_i|\boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{\theta}) = \sum_{j=1}^{K} \pi_{ij}\left[\gamma_{ij} t\left(y_i|\theta_j\right)^{\alpha_{ij}}\right] \tag{16}$$

where the Student's distribution $t(y_i/\theta_j)$ is defined in (1) with parameters $\theta_j = \{\mu_j, \Sigma_j, \nu_j\}$ as

$$t\left(y_i|\theta_j\right) = t\left(y_i|\mu_j, \Sigma_j, \nu_j\right) = \frac{\Gamma\left(\nu_j/2 + p/2\right)}{\Gamma\left(\nu_j/2\right)} \frac{\left|\Sigma_j\right|^{-1/2}}{\left(\pi\nu_j\right)^{p/2}} \times \left[1 + \frac{\left(y_i - \mu_j\right)^T \Sigma_j^{-1}\left(y_i - \mu_j\right)}{\nu_j}\right]^{-\left(\nu_j+p\right)/2} \tag{17}$$

where γ_{ij} is defined as

$$\gamma_{ij} = \left[\frac{\Gamma\left(\nu_j/2\right)}{\Gamma\left(\nu_j/2 + p/2\right)}\right]^{\alpha_{ij}} \times \frac{\Gamma\left(\left(\alpha_{ij}\nu_j + \alpha_{ij}p\right)/2\right)(\pi)^{p\left(\alpha_{ij}-1\right)/2} \nu_j^{\left(\alpha_{ij}p-1\right)/2}}{\Gamma\left(\left(\alpha_{ij}\nu_j + \alpha_{ij}p - p\right)/2\right)\left(\alpha_{ij}\nu_j + \alpha_{ij}p - p\right)^{(p-1)/2} \Sigma_j^{-\left(\alpha_{ij}-1\right)/2}} \tag{18}$$

to make our model satisfy the constraint which is proved in the appendix. It is worth to point out that this idea is inspired by the previous work [9]. The main difference between our model and [9] is: we use α_{ij} instead of constant parameter λ in [9]. It is due to the fact that this parameter may change values for different image pixel i and component j. In other words, this parameter may have different values in different i, j conditions. So it is suitable to select a variable than a constant value for this parameter.

The prior probability π_{ij} represents the prior distribution of the pixel y_i belonging to class j, which satisfies the constraint

in (6), that is

$$0 \leq \pi_{ij} \leq 1 \text{ and } \sum_{j=1}^{K} \pi_{ij} = 1 \tag{19}$$

Compared with standard SMM, which assumes the component of the mixture (conditional probability) satisfies the Student's t-distribution, our model assumes it satisfies the Student's t-distribution to power alpha with corresponding pixel label i and class label j. For the same pixel intensity value and jth class, our model gives a different component probability, as discussed in the previous section. The spatial information is considered by the calculation of prior probability π_{ij}. Thus, our model incorporates local spatial information, pixel intensity value and class information in an image.

With the intrinsic relationship between the prior probability π_{ij} $(p(x_i = j))$ and the posterior probability, the prior probability can be expressed as a linear combination of the posterior probabilities as

$$\pi_{ij} = \sum_{m=1}^{N} w_{im} z_{mj} \tag{20}$$

where z_{ij} is the posterior probability $p(x_i = j/y_i)$ and w_{im} is the weighted parameters to control the influence of other (mth) pixels to the ith pixel. We introduce the weighted parameters w_{im} to incorporate the spatial information in order to preserve robustness and noise insensitiveness, and to control the balance between the image noise and the image details. As such, the strength of weighted parameters should be decreased as the distance between the pixel m and i increases. We define the weighted parameters w_{im} as the function of d_{im}, which is the spatial Euclidean distance between pixels m and i. The possible choices for these functions are $1/d_{im}$, $1/x(1 + d_{im}^{\beta})$, $1/\exp(d_{im}^{\beta})$, $(\beta = 1, 2, \ldots)$ etc. Here, we adopt

$$w_{im} = \frac{1}{1 + d_{im}^2} \tag{21}$$

Considering constraint in (19), the normalisation of the prior probability should be applied. Here, it is possible to use the DCM process [37] for (20) as follows

$$\pi_{ij} = \frac{\sum_{m=1}^{N} w_{im} z_{mj}}{\sum_{j=1}^{K} \sum_{m=1}^{N} w_{im} z_{mj}} \tag{22}$$

We normalise the prior probability with the exponential function as

$$\pi_{ij} = \frac{\exp\left(\beta \sum_{m=1}^{N} w_{im} z_{mj}\right)}{\sum_{j=1}^{K} \exp\left(\beta \sum_{m=1}^{N} w_{im} z_{mj}\right)} \tag{23}$$

With the help of (20) and (23), we can find the relationship between our method and the HMRF model [1]. Our model degrades to the HMRF model if we set $\alpha_{ij} = 1$ and select the proper weighted parameters w_{im} as

$$w_{im} = \begin{cases} 1, & d_{im} \leq \sqrt{2} \\ 0, & \text{otherwise} \end{cases} \tag{24}$$

All pixels in (24) build up to a 3×3 square window, where the centre of the window is the ith pixel and the neighbourhood of the ith pixel is the mth pixel. It is worth pointing out that the first-order neighbourhood (four-neighbour, including horizontal and vertical pixels) can be obtained under the condition $d_{im} \leq 1$, and the second-order neighbourhood (eight-neighbour, plus diagonal pixels) can be obtained by setting $d_{im} \leq \sqrt{2}$. The effect of direction information in the refined model is discussed in [36]. The neighbourhood size is the main difference between our model and the HMRF, which has to select only the 3×3 square window because of its Markov property. For a one-dimensional (1D) chain, the Markov property means that the probabilistic behaviour of the chain at some time i, given knowledge of its complete past, depends only on its state in the immediate past $i - 1$ [41]. For a 2D image, the definition of neighbours ($i - 1$ and $i + 1$) extends to horizontal, vertical and diagonal pixels, which becomes a 3×3 square window. In our model, 3×3, 5×5, 7×7 etc. square windows can be used by the proper selection of weighted parameters, w_{im}. In general, the large window aims to tolerate the noise and the small window tends to preserve image sharpness and details. For a good balance, in this paper, we use the 5×5 square window

$$w_{im} = \begin{cases} \dfrac{1}{1 + d_{im}^2}, & d_{im} \leq 2\sqrt{2} \\ 0, & \text{otherwise} \end{cases} \tag{25}$$

If we set prior probability π_{ij} independent of i, our model degrades to SMM. If we use the 3×3 square window under the condition $d_{im} \leq \sqrt{2}$, the HMRF model can be considered a special case of our model. If we set α_{ij} as equal to constant λ, our model degrades to HMRF with fuzzy c-means algorithm (HMRF-FCM) [9]. Moreover, if we set $d_{im} < 1$ and $w_{im} = 1$, (20) is simplified to $\pi_{ij} = z_{ij}$, where the prior probability sets the same value as the posterior probability. In this case, our model can be seen as a special case of SVFMM (MAP is equivalent to ML in [33]). Thus, the proposed model can be considered an extension of the classical SMM, SVFMM or HMRF models and is general enough to describe mixture model-based methods.

A contribution of our idea is the establishment of (16), which incorporates the spatial information, the pixel intensity value and the clustering information in an image. Another improvement is that we reveal the relationship between the prior probability and the posterior probability in (20). Moreover, normalisation with the exponential function is used in (23). In the end, weighted parameters w_{im} used to control the influence of the neighbourhood pixels, depending on their distance from the central pixel, are shown in (25).

4 Parameter learning

With the help of the scaling factor, Student's t-distribution can be represented as a mixture of Gaussians with the same mean and scaled covariance. Equation (16) can be rewritten as

$$p(y_i|\boldsymbol{\Theta}) = \sum_{j=1}^{K} \pi_{ij} \gamma_{ij} \mathcal{N}\left(y_i|\mu_j, \Sigma_j/u_{ij}\right)^{\alpha_{ij}} \tag{26}$$

where $\boldsymbol{\Theta} = \left\{\pi_{ij}, \alpha_{ij}, \mu_j, \Sigma_j, v_j, u_{ij}\right\}_{i,j=1}^{N,K}$. The scaling factors

doi: 10.1049/iet-ipr.2012.0340

u_{ij} follow the Gamma distribution, which depends only on the degree of freedom v_j for the pixel y_i corresponding to the jth component

$$u_{ij} \sim \mathcal{G}\left(\frac{v_j}{2}, \frac{v_j}{2}\right) \tag{27}$$

Let h_{ij} be a set of latent component labels, where $h_{ij} = 1$ if y_j is generated from the ith component and $h_{ij} = 0$ otherwise. According to [11], the complete-data likelihood can be factored into the product of the marginal densities of the $\boldsymbol{h}$, the conditional densities of the $\boldsymbol{u}$ given the h_{ij}, and the conditional densities of the $\boldsymbol{y}$ given the u_{ij} and the h_{ij}

$$L = \prod_{i=1}^{N}\prod_{j=1}^{K}\left[\pi_{ij}\gamma_{ij}\mathcal{N}\left(y_i|\mu_j, \Sigma_j/u_{ij}\right)^{\alpha_{ij}} p\left(u_{ij}|v_j\right)^{\alpha_{ij}}\right]^{h_{ij}} \tag{28}$$

where

$$\mathcal{N}\left(y_i|\mu_j, \Sigma_j/u_{ij}\right) = \frac{\left|\Sigma_j\right|^{-1/2} u_{ij}^{p/2}}{(2\pi)^{p/2}} \times \exp\left\{-\frac{1}{2}u_{ij}\left(y_i - \mu_j\right)^T \Sigma_j^{-1}\left(y_i - \mu_j\right)\right\} \tag{29}$$

$$p\left(u_{ij}|v_j\right) = \frac{1}{\Gamma\left(v_j/2\right)}\left(v_j/2\right)^{v_j/2}\left(u_{ij}\right)^{v_j/2-1}\mathrm{e}^{-u_{ij}v_j/2} \tag{30}$$

From (29) and (30), the complete-data log-likelihood (ignoring constant terms) can be written as (see (31))

To maximise the log-likelihood function in (31), we apply the EM algorithm, which requires calculation of the quantity Q, the expectation of the complete data log-likelihood function.

The E-step

$$\begin{aligned} Q = E^{(k)}(\log L) = \sum_{i=1}^{N}\sum_{j=1}^{K} E\left(h_{ij}\right)\Bigg\{ & \log\pi_{ij}^{(k)} + \log\Gamma\left(\frac{\alpha_{ij}^{(k)}v_j^{(k)} + \alpha_{ij}^{(k)}p}{2}\right) \\ & + \left(\frac{\alpha_{ij}^{(k)}p - p}{2}\right)\log\pi + \left(\frac{\alpha_{ij}^{(k)}p - 1}{2}\right)\log v_j^{(k)} \\ & - \log\Gamma\left(\frac{\alpha_{ij}^{(k)}v_j^{(k)} + \alpha_{ij}^{(k)}p - p}{2}\right) \\ & - \left(\frac{p-1}{2}\right)\log\left(\alpha_{ij}^{(k)}v_j^{(k)} + \alpha_{ij}^{(k)}p - p\right) - \frac{1}{2}\log\left|\Sigma_j^{(k)}\right| \\ & + \alpha_{ij}^{(k)}\left[-\frac{1}{2}E\left(u_{ij}\right)\left(y_i - \mu_j^{(k)}\right)^{\mathrm{T}}\Sigma_j^{-1(k)}\left(y_i - \mu_j^{(k)}\right)\right] \\ & + \alpha_{ij}^{(k)}\left[-\log\Gamma\left(\frac{v_j}{2}\right) + \frac{v_j}{2}\left(\log\left(\frac{v_j}{2}\right) + E^{(k)}\left(\log u_{ij}\right) - E^{(k)}\left(u_{ij}\right)\right)\right]\Bigg\} \end{aligned} \tag{32}$$

where k denotes the kth iteration estimation of the corresponding quantities. The posterior probabilities of latent component h_{ij} can be calculated as

$$z_{ij} = E\left(h_{ij}\right) = p\left(h_{ij} = 1|y_i\right) = \frac{p\left(h_{ij} = 1, y_i\right)}{p\left(y_i\right)} \tag{33}$$

Using (17), we have

$$z_{ij}^{(k)} = \frac{\pi_{ij}^{(k)}\gamma_{ij}^{(k)}t\left(y_i|\mu_j^{(k)}, \Sigma_j^{(k)}, v_j^{(k)}\right)^{\alpha_{ij}^{(k)}}}{\sum_{h=1}^{K}\pi_{ih}^{(k)}\gamma_{ih}^{(k)}t\left(y_i|\mu_h^{(k)}, \Sigma_h^{(k)}, v_h^{(k)}\right)^{\alpha_{ih}^{(k)}}} \tag{34}$$

From [11], the posterior probabilities of the scaled factors and its log function can be calculated as

$$\begin{aligned} E^{(k)}\left(u_{ij}\right) &= u_{ij}^{(k)} \\ &= \frac{v_j^{(k)} + p}{v_j^{(k)} + \left(y_i - \mu_j^{(k)}\right)^T\Sigma_j^{-1(k)}\left(y_i - \mu_j^{(k)}\right)} \end{aligned} \tag{35}$$

$$\begin{aligned} E^{(k)}\left(\log u_{ij}\right) = \log u_{ij}^{(k)} - \log\left(\frac{v_j^{(k)} + p}{2}\right) \\ + \psi\left(\frac{v_j^{(k)} + p}{2}\right) \end{aligned} \tag{36}$$

The M-step is calculated by maximising the posterior expectation Q of the complete data log-likelihood. The expression of means μ_j can be obtained by maximising the posterior expectation Q over μ_j, which yields

$$\frac{\partial Q}{\partial \mu_j^{(k+1)}} = \frac{1}{2}\sum_{i=1}^{N}\alpha_{ij}^{(k)}z_{ij}^{(k)}u_{ij}^{(k)}\left(2\Sigma_j^{-1(k)}\mu_j^{(k)} - 2\Sigma_j^{-1(k)}y_i\right) \tag{37}$$

The solution of $\partial Q/\partial\mu_j^{(k+1)} = 0$ yields the estimation of μ_j

$$\mu_j^{(k+1)} = \frac{\sum_{i=1}^{N}\alpha_{ij}^{(k)}z_{ij}^{(k)}u_{ij}^{(k)}y_i}{\sum_{i=1}^{N}\alpha_{ij}^{(k)}z_{ij}^{(k)}u_{ij}^{(k)}} \tag{38}$$

Next, when we consider the partial derivative of Q with

$$\begin{aligned} \log L = \sum_{i=1}^{N}\sum_{j=1}^{K} h_{ij}\Bigg\{ & \log\pi_{ij} + \log\gamma_{ij} + \alpha_{ij}\left[-\frac{1}{2}\log\left|\Sigma_j\right| - \frac{1}{2}u_{ij}\left(y_i - \mu_j\right)^{\mathrm{T}}\Sigma_j^{-1}\left(y_i - \mu_j\right)\right] \\ & + \alpha_{ij}\left[-\log\Gamma\left(\frac{v_j}{2}\right) + \frac{v_j}{2}\left(\log\left(\frac{v_j}{2}\right) + \log u_{ij} - u_{ij}\right)\right]\Bigg\} \end{aligned} \tag{31}$$

respect to Σ_j^{-1} at the $(k+1)$ iteration step, we have

$$\frac{\partial Q}{\partial \Sigma_j^{-1(k+1)}} = \frac{1}{2}\sum_{i=1}^{N} z_{ij}^{(k)}\left(\Sigma_j^{(k)} - \alpha_{ij}^{(k)} u_{ij}^{(k)}\left(y_i - \mu_j^{(k+1)}\right) \times \left(y_i - \mu_j^{(k+1)}\right)^{\mathrm{T}}\right) \tag{39}$$

Let $\partial Q/\partial \Sigma_j^{-1(k+1)} = 0$ yields

$$\Sigma_j^{(k+1)} = \frac{\sum_{i=1}^{N} \alpha_{ij}^{(k)} z_{ij}^{(k)} u_{ij}^{(k)}\left(y_i - \mu_j^{(k+1)}\right)\left(y_i - \mu_j^{(k+1)}\right)^{\mathrm{T}}}{\sum_{i=1}^{N} z_{ij}^{(k)}} \tag{40}$$

Setting the partial derivative of Q with respect to v_j, we have

$$\begin{aligned}\frac{\partial Q}{\partial v_j^{(k+1)}} = \frac{1}{2}\sum_{i=1}^{N} \alpha_{ij}^{(k)} z_{ij}^{(k)}\Bigg[&\psi\left(\frac{\alpha_{ij}^{(k)} v_j^{(k)} + \alpha_{ij}^{(k)} p}{2}\right) - \psi\left(\frac{\alpha_{ij}^{(k)} v_j^{(k)} + \alpha_{ij}^{(k)} p - p}{2}\right) - \frac{\alpha_{ij}^{(k)}(p-1)}{\alpha_{ij}^{(k)} v_j^{(k)} + \alpha_{ij}^{(k)} p - p} \\ &+ \frac{\alpha_{ij}^{(k)} p - 1}{v_j^{(k)}} + \log\left(\frac{v_j}{2}\right) - \psi\left(\frac{v_j}{2}\right) + 1 + \log u_{ij}^{(k)} - u_{ij}^{(k)} \\ &+ \psi\left(\frac{v_j^{(k)}}{2}\right) - \log\left(\frac{v_j^{(k)} + p}{2}\right)\Bigg]\end{aligned} \tag{41}$$

Setting (41) equals to zero, it follows that v_j needs to be computed iteratively to solve the equations

$$\begin{aligned}&\psi\left(\frac{\alpha_{ij}^{(k)} v_j^{(k)} + \alpha_{ij}^{(k)} p}{2}\right) - \psi\left(\frac{\alpha_{ij}^{(k)} v_j^{(k)} + \alpha_{ij}^{(k)} p - p}{2}\right) - \frac{\alpha_{ij}^{(k)}(p-1)}{\alpha_{ij}^{(k)} v_j^{(k)} + \alpha_{ij}^{(k)} p - p} + \frac{\alpha_{ij}^{(k)} p - 1}{v_j^{(k)}} \\ &+ \log\left(\frac{v_j}{2}\right) - \psi\left(\frac{v_j}{2}\right) + 1 + \frac{\sum_{i=1}^{N} \alpha_{ij}^{(k)} z_{ij}^{(k)}\left[\log u_{ij}^{(k)} - u_{ij}^{(k)}\right]}{\sum_{i=1}^{N} \alpha_{ij}^{(k)} z_{ij}^{(k)}} \\ &+ \psi\left(\frac{v_j^{(k)}}{2}\right) - \log\left(\frac{v_j^{(k)} + p}{2}\right) = 0\end{aligned} \tag{42}$$

The final step is to update the estimation of weighted parameters alpha using the gradient method as follows

$$\alpha_{ij}^{(k+1)} = \alpha_{ij}^{(k)} - \eta \frac{\partial Q}{\partial \alpha_{ij}} \tag{43}$$

where η is the learning rate and its value is sufficiently small. In this paper, we have selected $\eta = 10^{-6}$ and

$$\begin{aligned}\frac{\partial Q}{\partial \alpha_{ij}} = z_{ij}^{(k+1)}\Bigg[&\frac{\left(v_j^{(k+1)} + p\right)}{2}\left(\psi\left(\frac{\alpha_{ij}^{(k)} v_j^{(k+1)} + \alpha_{ij}^{(k)} p}{2}\right) - \psi\left(\frac{\alpha_{ij}^{(k)} v_j^{(k+1)} + \alpha_{ij}^{(k)} p - p}{2}\right)\right) + \frac{p}{2}\left(\log \pi + \log v_j^{(k+1)}\right) \\ &+ \frac{(p-1)\left(v_j^{(k+1)} + p\right)}{2\left(\alpha_{ij}^{(k)} v_j^{(k+1)} + \alpha_{ij}^{(k)} p - p\right)} - \frac{1}{2} u_{ij}^{(k)}\left(y_i - \mu_j^{(k+1)}\right)^{T} \\ &\times \Sigma_j^{-1(k+1)}\left(y_i - \mu_j^{(k+1)}\right) - \log\Gamma\left(\frac{v_j^{(k+1)} + p}{2}\right) \\ &+ \frac{v_j^{(k+1)}}{2}\left(\log\left(\frac{v_j^{(k+1)}}{2}\right) + \log u_{ij}^{(k)} - \log\left(\frac{v_j^{(k+1)} + p}{2}\right)\right. \\ &\left.+ \psi\left(\frac{v_j^{(k+1)} + p}{2}\right) - u_{ij}^{(k)}\right)\Bigg]\end{aligned} \tag{44}$$

With consideration of the weighted parameters alpha, our model can perform more robustly, as shown in the experimental results. We conclude our model as the following (see Fig. 2).

5 Experimental results

In this section, we experimentally evaluate our WSMM in a set of synthetic images and real images. We also evaluate SMM [11], FLICM [42], SVFMM [35], HMRF-FCM [9] and SMM-SC [43] for comparison. The source codes for FLICM, SVFMM and HMRF-FCM algorithms can be downloaded from the authors' websites [44–46]. Our

Algorithm:

1. Initialise the algorithm with the k-means method to obtain initial values and set parameters α_{ij} equal to 1.2, k:=1.
2. Use Eqs. (25) and (23) to calculate the weighted parameters $w_{mi}^{(k)}$ and prior probability $\pi_{ij}^{(k)}$.
3. Compute the posterior probability $z_{ij}^{(k)}$ and latent variable $u_{ij}^{(k)}$ according to the E-step, using Eqs. (34)-(36), respectively.
4. Compute the quantities $\mu_j^{(k+1)}$, $\Sigma_j^{(k+1)}$ and $v_j^{(k+1)}$ according to the M-step, using Eqs. (38), (41), and (42) respectively. Update $\alpha_{ij}^{(k+1)}$ by Eq. (43).
5. Terminate the iterations if the EM algorithm converges; otherwise, increase the iteration (k=k+1) and repeat steps 2-4.

Fig. 2 *Algorithm: The EM Algorithm for the WSMM*

doi: 10.1049/iet-ipr.2012.0340

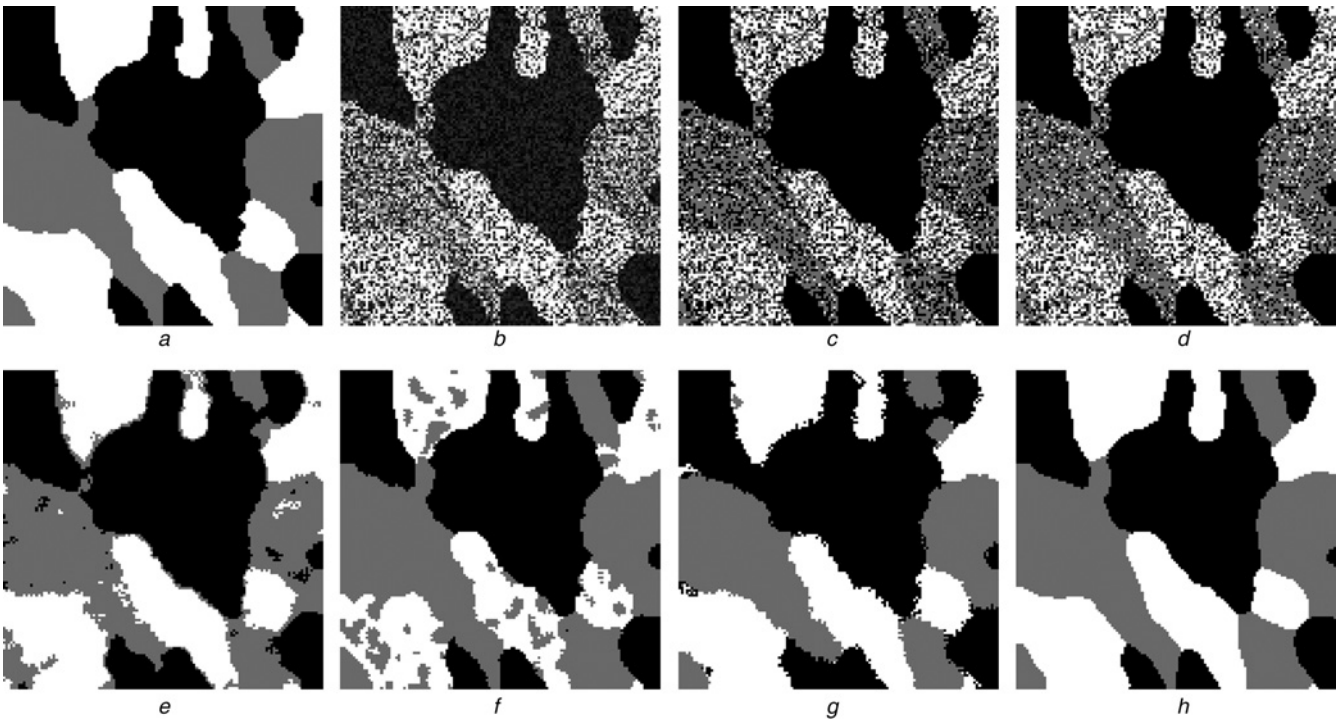

Fig. 3

a Original three-class image (128 × 128 size)
b Corrupted by multiplicative noise with noise density = 0.4
c SMM (MCR = 31.46%)
d SVFMM (MCR = 28.14%)
e FLICM (MCR = 9.33%)
f HMRF-FCM (MCR = 10.19%)
g SMM-SC (MCR = 5.82%)
h WSMM (MCR = 3.19%)

experiments have been developed in Matlab R2009b, and were executed on an Intel Pentium Dual-Core 2.2 GHZ CPU, 2 G RAM.

5.1 Synthetic images

In the first experiment, a four-class synthetic image (128 × 128, shown in Fig. 3*a*) is used to compare the performance of the proposed method with others. Fig. 3*b* shows the same image corrupted by multiplicative noise with noise density = 0.4. To evaluate the segmentation results, we employ the misclassification ratio (MCR) [14] in our experiments, which is defined as

$$\text{MCR} = \frac{\text{number of misclassified pixels}}{\text{total number of pixels}} \times 100\% \qquad (45)$$

The value of MCR is in the [0–100] range, where lower values indicate better segmentation performance. The segmentation results of the noise image (Fig. 3*b*) by SMM, SVFMM, FLICM, HMRF-FCM, SMM-SC and WSMM are shown in Fig. 3*c*–*h*. The class number is set to 3, based on previous experience. It can be seen that SMM and SVFMM cannot segment the image well. FLICM erroneously classifies the pixels in the grey parts of the image and HMRF-FCM erroneously classifies the pixels in the white parts. SMM-SC creates three new small regions at the top right, the top left and the middle bottom of the image. However, the proposed WSMM shows remarkable robustness under heavy noise conditions. Table 1 gives the segmentation results for all methods with different noise intensities. We observe that the proposed method outperforms the other methods, with the lowest MCR.

5.2 Real images

In this experiment, we evaluate the performance of the proposed WSMM based on a subset of the Berkeley image dataset [47], which is comprised of a set of real-world colour images along with their segmentation maps provided by different individuals. We employ the probabilistic rand (PR) index [48] to evaluate the performance of the proposed method with the multiple ground truths available for each image within the dataset. It has been shown that

Table 1 MCR, % and computation time, s of synthetic image with additive various multiplicative noise for different methods

Methods	0.1	0.2	0.3	0.4	Computation time (average)
SMM	20.72	23.51	28.38	31.46	1.76
SVFMM	15.21	19.64	24.65	28.14	72.49
FLICM	2.89	5.31	8.51	9.33	11.54
HMRF-FCM	1.84	4.59	6.56	10.19	64.54 s
SMM-SC	1.78	1.92	3.53	5.82	5.01 s
WSMM	1.45	1.89	2.65	3.19	17.38

Fig. 4 *Original image from the Berkeley image segmentation dataset*

a 22090
b 24063
c 78019
d 105053
e 108073
f 124084
g 135069
h 253036
i 46076
j 302003
k 61086
l 106025

the PR index possesses the desirable property of being robust to segmentation maps that result by splitting or merging segments of the ground truth [49]. The PR index takes values between 0 and 1, with values closer to 0 (indicating a worse segmentation result) and values closer to 1 (indicating a better result).

Fig. 4 shows the original Berkeley images used for the image segmentation experiment. The corresponding segmentation results by WSMM are illustrated in Fig. 5. For fair comparison, we also evaluate the performance of SMM, SVFMM, FLICM, HMRF-FCM and SMM-SC in addition to our method. The class number K is set by human vision. Table 2 contains the PR values for all methods corresponding to each one of the test images in Fig. 4. Compared with other methods, it can be seen that the proposed WSMM yields the best segmentation results with the highest PR values.

6 Conclusions

In this paper, we see the FMM in a new way and reveal intrinsic relationships between our model and HMRF model. The HMRF model can be considered as a special case in our general model. In our model, the prior probability depends on weighted parameters and the component of the mixture is assumed by Student's t-distribution to parameter power alpha. Thus, we call our approach the WSMM. Weighted parameters are used to control the influence of the neighbouring pixels, depending

doi: 10.1049/iet-ipr.2012.0340

Fig. 5 *Image segmentation results by proposed method*

a 22090
b 24063
c 78019
d 105053
e 108073
f 124084
g 135069
h 253036
i 46076
j 302003
k 61086
l 106025

Table 2 Comparison of different methods for Berkeley image dataset, PR index

Image no.	Class	SMM	SVFMM	FLICM	HMRF-FCM	SMM-SC	WSMM
108 073	4	0.5137	0.5763	0.5824	0.5855	0.6197	0.6214
124 084	4	0.6058	0.606	0.6	0.7203	0.7231	0.7426
135 069	2	0.987	0.9834	0.9834	0.9873	0.9882	0.9913
302 003	3	0.6985	0.717	0.7172	0.7169	0.7115	0.7223
105 053	3	0.5391	0.5495	0.51	0.5546	0.6358	0.6387
22 090	4	0.7644	0.7757	0.7675	0.7777	0.793	0.7946
46 076	6	0.8322	0.8320	0.8381	0.8602	0.8972	0.9250
61 086	5	0.6704	0.7043	0.7213	0.7220	0.7002	0.7313
106 025	4	0.8153	0.8082	0.7833	0.8241	0.8246	0.85
253 036	4	0.6818	0.711	0.704	0.7181	0.7821	0.7834
78 019	7	0.8234	0.8148	0.8162	0.8231	0.8322	0.859
24 063	4	0.8015	0.8486	0.8139	0.8559	0.8629	0.8737

on their distance from the central pixel. This influence decreases as the distance increases. We not only consider the spatial relationship between neighbouring pixels, but also the relationship between the spatial information and component information. The local spatial information, pixel intensity value and component information are incorporated for segmentation in an image. Moreover, we simultaneously use the EM algorithm for mixture model parameters estimation and the gradient method for the rest parameters estimation. The proposed approach, with good balance between the noise and image details, is compared to other state-of-the-art technologies for image segmentation. Experimental results demonstrate the superior performance of our approach.

7 Acknowledgments

The authors would like to thank the comments and the suggestions of the associate editor and the reviewers. 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 grant numbers 61105007 and 61103141. 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 McLachlan, G., Peel, D.: 'Finite mixture models' (John Willey and Sons, New York, 2000)
2 Bishop, C.M.: 'Pattern recognition and machine learning' (Springer, 2006)
3 Fjørtøft, R., Delignon, Y., Pieczynski, W., Sigelle, M., Tupin, F.: 'Unsupervised classification of radar images using hidden Markov chains and hidden random fields', *IEEE Trans. Geosci. Remote Sens.*, 2003, **41**, pp. 675–686
4 Carson, C., Belongie, S., Greenspan, H., Malik, J.: 'Blobworld: image segmentation using expectation-maximization and its application to image querying', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002, **24**, pp. 1026–1038
5 Permuter, H., Francos, J., Jermyn, I.: 'A study of Gaussian mixture models of color and texture features for image classification and segmentation', *Pattern Recognit.*, 2006, **39**, pp. 695–706
6 Ma, Z., Tavares, J.M.R.S., Jorge, R.N., Mascarenhas, T.: 'A review of algorithms for medical image segmentation and their applications to the female pelvic cavity', *Comput. Methods Biomech. Biomed. Eng.*, 2010, **13**, pp. 235–246
7 Thanh, M.N., Wu, Q.M.J., Ahuja, S.: 'An extension of the standard mixture model for image segmentation', *IEEE Trans. Neural Netw.*, 2010, **21**, pp. 1326–1338
8 Thanh, M.N., Wu, Q.M.J.: 'Gaussian-mixture-model-based spatial neighborhood relationships for pixel labeling problem', *IEEE Trans. Syst. Man Cybern. B*, 2012, **42**, pp. 193–202
9 Chatzis, S.P., Varvarigou, T.A.: 'A fuzzy clustering approach toward hidden Markov random field models for enhanced spatially constrained image segmentation', *IEEE Trans. Fuzzy Syst.*, 2008, **16**, pp. 1351–1361
10 Chatzis, S.P., Kosmopoulos, D.I., Varvarigou, T.A.: 'Robust sequential data modeling using an outlier tolerant hidden Markov model', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2009, **31**, pp. 1657–1669
11 Peel, D., McLachlan, G.: 'Robust mixture modeling using the t-distribution', *Stat. Comput.*, 2000, **10**, pp. 335–344
12 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)
13 Rabiner, L.R.: 'A tutorial on hidden Markov models and selected applications in speech recognition', *Proc. IEEE*, 1989, **77**, pp. 257–286
14 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. Imaging*, 2001, **20**, pp. 45–57
15 Dempster, P., Laird, N.M., Rubin, D.B.: 'Maximum likelihood from incomplete data via EM algorithm', *J. Roy. Stat. Soc. B*, 1977, **39**, pp. 1–38
16 Archer, G.E.B., Titterington, D.M.: 'Parameter estimation for hidden Markov chains', *J. Stat. Plan. Inference*, 2002, **108**, pp. 365–390
17 Besag, J.: 'Statistical analysis of non-lattice data', *Statistician*, 1975, **24**, pp. 179–195
18 Baum, L.E., Petrie, T., Goules, G., Weiss, N.: 'A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains', *Ann. Math. Stat.*, 1970, **41**, pp. 164–171
19 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
20 McLachlan, G.J., Krishnan, T.: 'The EM algorithm and extensions', '*Series in Probability and Statistics*' (Wiley, New York, 1997)
21 Zhou, Z., Leahy, R., Qi, J.: 'Approximate maximum likelihood hyperparameter estimation for Gibbs priors', *IEEE Trans. Image Process.*, 1997, **6**, pp. 844–861
22 Pieczynski, W.: 'Statistical image segmentation', *Mach. Graph. Vis.*, 1992, **1**, pp. 261–268
23 Delmas, J.P.: 'An equivalence of the EM and ICE algorithm for exponential family', *IEEE Trans. Signal Process.*, 1997, **45**, pp. 2613–2615
24 Besag, J.: 'On the statistical analysis of dirty pictures', *J. Roy. Stat. Soc. B*, 1986, **48**, pp. 259–302
25 Geman, S., Geman, D.: 'Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1984, **6**, pp. 721–741
26 Hjort, N., Mohn, E., Storvik, G.: 'A simulation study of some contextual classification methods for remotely sensed data', *IEEE Trans. Geosci. Electron.*, 1987, **25**, pp. 796–804
27 Qian, W., Titterington, D.: 'Estimation of parameters in hidden Markov models', *Philos. Trans. Roy. Soc. London A*, 1991, **337**, pp. 407–428
28 Zhang, J.: 'The mean field theory in EM procedures for Markov random fields', *IEEE Trans. Signal Process.*, 1992, **40**, pp. 2570–2583
29 Zhang, J.: 'The mean field theory in EM procedures for blind Markov random field image restoration', *IEEE Trans. Image Process.*, 1993, **2**, pp. 27–40
30 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**, pp. 1089–1101
31 Celeux, G., Forbes, F., Peyrard, N.: 'EM procedures using mean field-like approximations for Markov model-based image segmentation', *Pattern Recognit.*, 2003, **36**, pp. 131–144
32 Dunmur, A.P., Titterington, D.: 'Parameter estimation in latent profile models', *Comput. Stat. Data Anal.*, 1998, **27**, pp. 371–388
33 Sanjay, G.S., Hebert, T.J.: 'Bayesian pixel classification using spatially variant finite mixtures and the generalized EM algorithm', *IEEE Trans. Image Process.*, 1998, **7**, pp. 1014–1028
34 Diplaros, A., Vlassis, N., Gevers, T.: 'A spatially constrained generative model and an EM algorithm for image segmentation', *IEEE Trans. Neural Netw.*, 2007, **18**, pp. 798–808
35 Blekas, K., Likas, A., Galatsanos, N.P., Lagaris, I.E.: 'A spatially constrained mixture model for image segmentation', *IEEE Trans. Neural Netw.*, 2005, **16**, pp. 494–498
36 Nikou, C., Galatsanos, N.P., Likas, A.: 'A class-adaptive spatially variant mixture model for image segmentation', *IEEE Trans. Image Process.*, 2007, **16**, pp. 1121–1130
37 Nikou, C., Likas, A., Galatsanos, N.P.: 'A Bayesian framework for image segmentation with spatially varying mixtures', *IEEE Trans. Image Process.*, 2010, **19**, pp. 2278–2289
38 Liu, C., Rubin, D.: 'ML estimation of the t-distribution using EM and its extensions, ECM and ECME', *Statistica Sinica*, 1995, **5**, pp. 19–39
39 Gerogiannis, D., Nikou, C., Likas, A.: 'The mixtures of student's t-distribution as a robust framework for rigid registration', *Image Vis. Comput.*, 2009, **27**, pp. 1285–1294
40 Besag, J.: 'Spatial interaction and the statistical analysis of lattice systems', *J. R. Stat. Soc. B*, 1974, **36**, pp. 192–326
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 Krinidis, S., Chatzis, V.: 'A robust fuzzy local information C-means clustering algorithm', *IEEE Trans. Image Process.*, 2010, **5**, pp. 1328–1337
43 Thanh, M.N., Wu, Q.M.J.: 'Robust student's t-mixture model with spatial constraints and its application in medical image segmentation', *IEEE Trans. Med. Imaging*, 2011, **31**, pp. 103–116
44 Krinidis, S., Chatzis, V.: Available at http://www.infoman.teikav.edu.gr/~stkrini/pages/develop/FLICM/FLICM.html, last access, October 2012

doi: 10.1049/iet-ipr.2012.0340

45 Blekas, K.: Available at http://www.cs.uoi.gr/~kblekas/sw/MAPsegmentation.html, last access, January 2010
46 Chatzis, S.P.: Available at http://www.web.mac.com/soteri0s/Sotirios_Chatzis/Software.html, last access, May 2010
47 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. Eighth IEEE Int. Conf. on Comput. Vis., Vancouver, BC, Canada, 2001, vol. 2, pp. 416–423
48 Unnikrishnan, R., Pantofaru, C., Hebert, M.: 'A measure for objective evaluation of image segmentation algorithms'. IEEE Conf. on Comput. Vis. Pattern Recognit., 2005, vol. 3, pp. 34–41
49 Unnikrishnan, R., Hebert, M.: 'Measures of similarity'. Proc. IEEE Workshop Comput. Vis. Appl., 2005, pp. 394–400

9 Appendix

Let $p(y) = \gamma t(y/\theta)^{\alpha}$. Here, it is obvious that $p(y) > 0$. Now we aim to prove $\int p(y) = 1$

$$\int p(y) = \int \gamma t(y|\theta)^{\alpha}\, \mathrm{d}y = \int_{-\infty}^{+\infty} \left[\frac{\Gamma(v/2)}{\Gamma(v/2+p/2)}\right]^{\alpha} \times \frac{\Gamma\big((\alpha v+\alpha p)/2\big)(\pi)^{p(\alpha-1)/2} v^{(\alpha p-1)/2}}{\Gamma\big((\alpha v+\alpha p-p)/2\big)(\alpha v+\alpha p-p)^{(p-1)/2}\Sigma^{-(\alpha-1)/2}} \times \left[\frac{\Gamma(v/2+p/2)}{\Gamma(v/2)}\frac{\Sigma^{-1/2}}{(\pi v)^{p/2}}\right]^{\alpha} \times \left[1+\frac{(y-\mu)^T\Sigma^{-1}(y-\mu)}{v}\right]^{-\alpha(v+p)/2} \mathrm{d}y$$

$$= \int_{-\infty}^{+\infty} \frac{\Gamma\big((\alpha v+\alpha p)/2\big)\Sigma^{-1/2}}{\Gamma\big((\alpha v+\alpha p-p)/2\big)(\alpha v+\alpha p-p)^{(p-1)/2}(\pi)^{p/2}v^{1/2}} \times \left[1+\frac{(y-\mu)^T\Sigma^{-1}(y-\mu)}{v}\right]^{-\alpha(v+p)/2} \mathrm{d}y$$

$$= \int_{-\infty}^{+\infty} \frac{\Gamma\big(v'/2+p/2\big)}{\Gamma\big(v'/2\big)}\frac{\Sigma'^{-1/2}}{(\pi v')^{p/2}} \times \left[1+\frac{()y-\mu^T\Sigma'^{-1}(y-\mu)}{v'}\right]^{-(v'+p)/2} \mathrm{d}y = 1$$

The proof has been completed.