Image Segmentation by Student's-t Mixture Models Based on Markov Random Field and Weighted Mean Template

Xu Pan, Hongqing Zhu^{*} and Qunyi Xie

School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China hqzhu@ecust.edu.cn

Abstract

Finite mixture model (FMM) with Gaussian distribution has been widely used in many image processing and pattern recognition tasks. This paper presents a new Student's-t mixture model (SMM) based on Markov random field (MRF) and weighted mean template. In this model, the Student's-t distribution is considered as an alternative to the Gaussian distribution due to the former is heavily tailed than Gaussian distribution, thus providing robustness to outliers. With the help of the weighted mean template, the spatial information between neighboring pixels of an image is considered during the learning step. In addition, the proposed method is able to impose the smoothness constraint on the pixel label by using MRF. Furthermore, an efficient energy function and a novel factor are applied in current model to decrease the computational complexity. Numerical experiments are presented on simulated and real world images, and the results are compared with other FMM-based models.

Keywords: Student's-t mixture model, Markov random field, image segmentation, spatial information, mean template

1. Introduction

Segmentation is one of the most difficult problems in image processing [1] and pattern recognition [2]. The purpose of image segmentation is to cluster all image pixels into non-overlapped groups with respect to some real world objects. One of the most commonly used clustering methods is finite mixture model (FMM). Due to the ease of implementation, the standard Gaussian mixture model (GMM) has been selected most widely as a particular case of FMM. Applying GMM had good segmentation results on images without noise. However, its accuracy in noisy images is not enough mainly because the prior probability π_i is not related to pixel i so that the spatial relationship between neighboring pixels is not taken into account. For this reason, the segmentation result of GMM is extremely sensitive to noise. To reduce the sensitivity of the noise in segmented image, the finite Student's-t mixture model (SMM) has been recently introduced in [3] as an alternative to GMM. It is because that the Student's-t distribution has heavily tailed than Gaussian distribution. Compared to the GMM, each component of the SMM has one more parameter called the degrees of freedom v. However, both GMM and SMM don't consider the fact that spatially adjacent pixel points most likely should belong to the same cluster. Recently, Markov Random Field (MRF) has been applied to impose spatial smoothness constraints on the image segmentation. But one main difficulty concerning the use of MRF as smoothness constraints is their high computational complexity.

In this paper, we present a new finite Student's-t mixture model, based on MRF and weighted mean template. In this model, the M-step of the EM algorithm [4] can be

^{*} Corresponding Author

directly applied for the maximization of the log-likelihood with respect to the parameters. The proposed model is easy to implement and compared with the existing MRF model, require lesser number of parameters. In addition, the factor G_{ij} defined in [5] is used as multiplication of both posterior probabilities and prior distributions, which lead to the proposed method is easy to implementation. To accurately evaluate the influence of the neighboring pixels during the learning step, the proposed method incorporate the weighted mean template into the model. Hence, it improves segmentation results, particularly when an image is corrupted by high levels of noise. Experimental results obtained on synthetic, real world grayscale images, and magnetic resonance (MR) images demonstrate the robustness, accuracy and effectiveness of the proposed approach in image segmentation.

The rest of the paper is organized as follows: Section 2 briefly reviews the related works. In Section 3, the proposed method and the parameter estimation will be discussed in detail. Section 4 presents the parameter learning. The experimental results are shown in Section 5. Finally, in the last section we summarize our results and conclude this paper.

2. Standard Finite Mixture Model

Let x_i , i = (1, 2, ..., N), with dimension D, denote an observation at the *i*-th pixel of an image. To classify N pixels of an image into K labels, it is assumed that x_i is independent of the label Ω_j . The density function $f(x_i | \prod, \Theta)$ at each pixel x_i can be expressed by

$$f\left(x_{i}|\Pi,\Theta\right) = \sum_{j=1}^{K} \pi_{ij} \Phi\left(x_{i}|\Theta_{j}\right),\tag{1}$$

where $\pi i j$ is the prior probability of xi belong to label Ωj and satisfies the following constraints:

$$0 \le \pi_{ij} \le 1 \text{ and } \sum_{j=1}^{K} \pi_{ij} = 1.$$
 (2)

According to (1), we can derive the joint conditional density of the data set $X=(x_1, x_2, ..., x_N)$

$$p(X|\Pi,\Theta) = \prod_{i=1}^{N} f\left(x_i|\Pi,\Theta\right) = \prod_{i=1}^{N} \left[\sum_{j=1}^{K} \pi_{ij} \Phi\left(x_i|\Theta_j\right)\right],\tag{3}$$

where $\Phi(x_i|\Theta_j)$ is the Student's-*t* distribution.

$$\Phi\left(x_{i}|\Theta_{j}\right) = \frac{\Gamma\left(\frac{v_{j}+D}{2}\right)\left|\Sigma_{j}\right|^{-\frac{1}{2}}}{\left(\pi v_{j}\right)^{\frac{D}{2}}\Gamma\left(\frac{v_{j}}{2}\right)\left[1+v_{j}^{-1}\delta\left(x_{i},\mu_{j};\Sigma_{j}\right)\right]^{\frac{v_{j}+D}{2}},$$
(4)

where the D-dimensional vector μ j is mean and vj is the degree of freedom. The D×D matrix $\sum j$ are the covariance. $\delta(x_i, \mu_j; \Sigma_j) = (x_i - \mu_j)^T \sum_{j=1}^{-1} (x_i - \mu_j)$ is squared Mahalanob is distance, and Γ is the Gamma function. Figure 1 illustrates Student's-t distribution with different degrees of freedom v for the same mean μ and covariance Σ . Comparing to the Gaussian distribution, their overall shapes are very similar, however, the Student's-t distribution provides a longer tail. With the number of degrees of freedom v growing, the Student's-t distribution tends to the Gaussian distribution.



Figure 1 Plot of the 1-D distributions (zero mean and variance one). Left: Student's *t*-distribution; Right: Gaussian distribution.

3. Proposed Methods

For standard finite mixture model, the spatial relationship between the neighboring pixels is not integrated into the segmentation procedure. To impose spatial smoothness constraints among neighboring pixels, this paper proposes a new density function $f(x_i|\Pi,\Theta)$ at each observation point x_i given by

$$f\left(x_{i}|\Pi,\Theta\right) = \sum_{j=1}^{K} \pi_{ij} \left(\sum_{m \in N_{i}} \frac{\omega_{m}}{R_{i}} \Phi\left(x_{m}|\Theta_{j}\right)^{p}\right)^{1/p},$$
(5)

where N_i is the neighborhood of the *i*-th pixel (including itself), and *p* is a non-zero real number. The weighting factor ω_m is used to control the influence of neighbor's term depending on their distance from the central pixel. With the decreasing of the distance between neighborhood pixel and the central pixel, the value of ω_m should increase. Therefore, ω_m can be expressed as a function of spatial Euclidean distance d_{mi} .

$$\omega_m = \frac{1}{\left(2\pi\delta^2\right)^{1/2}} \exp\left(-\frac{d_{mi}^2}{2\delta^2}\right), \text{ with } \delta = \frac{q-1}{4}, \tag{6}$$

Here q stands for the neighborhood window size. R_i is a normalized parameter of the form

$$R_i = \sum_{m \in N_i} \omega_m \,. \tag{7}$$

Based on Bayes' rule, the posterior probability density function can be represented by

$$p(\Pi, \Theta | X) \propto p(X | \Pi, \Theta) p(\Pi).$$
(8)

According to (5), the new joint conditional density in (8) can be rewritten as

$$p(X|\Pi,\Theta) = \prod_{i=1}^{N} \left[\sum_{j=1}^{K} \pi_{ij} \left(\sum_{m \in N_i} \frac{\omega_m}{R_i} \Phi\left(x_m |\Theta_j\right)^p \right)^{1/p} \right].$$
(9)

Based on the theorem of Hammersley-Clifford [6], a given random field can be defined as an MRF if and only if its probability distribution is a Gibbs distribution. Thus,

$$p(\Pi) = Z^{-1} \exp\left\{-\frac{1}{T}U(\Pi)\right\},\tag{10}$$

where Z is a normalizing constant, T is a temperature constant. In current paper, a new smooth prior $U(\Pi)$ is chosen to incorporate the spatial correlation as

$$U(\Pi) = -\sum_{i=1}^{N} \sum_{j=1}^{K} G_{ij}^{(i)} \log \pi_{ij}^{(i+1)}, \qquad (11)$$

where t is the iteration step. For easy implementation purpose, a novel factor G_{ij} is defined by

$$G_{ij}^{(t)} = \exp\left[\frac{\beta}{2q} \sum_{m \in N_i} \left(z_{mj}^{(t)} + \pi_{mj}^{(t)}\right)\right].$$
(12)

Substituting (11) into (10) yields

$$p(\Pi) = Z^{-1} \exp\left\{\frac{1}{T} \sum_{i=1}^{N} \sum_{j=1}^{K} G_{ij}^{(t)} \log \pi_{ij}^{(t+1)}\right\}.$$
(13)

Thus, the log-likelihood function of (8) can be written as follows.

$$L(\Pi, \Theta | X) = \sum_{i=1}^{N} \log \left\{ \sum_{j=1}^{K} \pi_{ij}^{(t+1)} \left(\sum_{m \in N_i} \frac{\omega_m}{R_i} \Phi \left(x_m | \Theta_j^{(t+1)} \right)^p \right)^{1/p} \right\} - \log Z + \frac{1}{T} \sum_{i=1}^{N} \sum_{j=1}^{K} G_{ij}^{(t)} \log \pi_{ij}^{(t+1)}$$
(14)

After applying the complete-data condition, the optimization problem of $L(\Pi, \Theta | X)$ is equivalent to maximizing the following objective function.

$$S(\Pi, \Theta | X) = \sum_{i=1}^{N} \sum_{j=1}^{K} z_{ij}^{(t)} \left\{ \log \pi_{ij}^{(t+1)} + \log \left[\sum_{m \in N_i} \frac{\omega_m}{R_i} \Phi \left(x_m | \Theta_j^{(t+1)} \right)^p \right]^{1/p} \right\} - \log Z + \frac{1}{T} \sum_{i=1}^{N} \sum_{j=1}^{K} G_{ij}^{(t)} \log \pi_{ij}^{(t+1)}$$
(15)

4. Parameter Learning

Due to the high computational complexity of (15), it cannot be calculated directly. In order to overcome these disadvantages, the Jensen's inequality [7] is applied, which states that, given a set of numbers $\lambda_i \ge 0$ and $\Sigma \lambda_i = 1$, one has log $(\Sigma \lambda_i x_i) \ge \Sigma \lambda_i \log (x_i)$. For simplicity purpose, the parameter Z and T are always set as one. Thus, considering $\omega_m/R_i \ge 0$ and $\Sigma_{m \in \mathbb{N}_i} \omega_m/R_i = 1$, we can derive the new objective function as follows.

$$S(\Pi, \Theta | X) = \sum_{i=1}^{N} \sum_{j=1}^{K} z_{ij}^{(t)} \left\{ \log \pi_{ij}^{(t+1)} + \sum_{m \in N_i} \frac{\omega_m}{R_i} \log \Phi\left(x_m | \Theta_j^{(t+1)}\right) \right\} + \sum_{i=1}^{N} \sum_{j=1}^{K} G_{ij}^{(t)} \log \pi_{ij}^{(t+1)} .$$
(16)

In E-step, the posterior probability of the hidden variables can be calculated as

$$z_{ij}^{(t)} = \frac{\pi_{ij}^{(t)} \left(\sum_{m \in N_i} \frac{\omega_m}{R_i} \Phi\left(x_m | \Theta_j^{(t)}\right)^p \right)^{1/p}}{\sum_{h=1}^K \pi_{ih}^{(t)} \left(\sum_{m \in N_i} \frac{\omega_m}{R_i} \Phi\left(x_m | \Theta_h^{(t)}\right)^p \right)^{1/p}}$$
(17)

Due to there is no analysis solution for maximizing the log-likelihood under a Student's-*t* distribution, in this paper, Student's-*t* distribution can be regarded as a mixture of Gaussian distribution with the same mean u_{ij} and scaled covariance.

$$u_{ij}^{(t)} = \frac{v_j^{(t)} + D}{v_j^{(t)} + \delta\left(x_i, \mu_j^{(t)}; \Sigma_j^{(t)}\right)}$$
(18)

In order to calculate the means μ_j at the (*t*+1) iteration step, setting the partial derivative of (16) with respect to $\mu_j^{(t+1)}$ equals to zero $(\partial S / \partial \mu_j^{(t+1)} = 0)$, we have

$$\mu_{j}^{(t+1)} = \frac{\sum_{i=1}^{N} z_{ij}^{(t)} \sum_{m \in N_{i}} \frac{\omega_{m}}{R_{i}} u_{mj}^{(t)} x_{m}}{\sum_{i=1}^{N} z_{ij}^{(t)} \sum_{m \in N_{i}} \frac{\omega_{m}}{R_{i}} u_{mj}^{(t)}}$$
(19)

Similarly, let $\partial S / \partial \sum_{j}^{(t+1)} = 0$. One can obtain the covariance $\sum_{j}^{(t+1)}$ from

$$\Sigma_{j}^{(t+1)} = \frac{\sum_{i=1}^{N} z_{ij}^{(t)} \sum_{m \in N_{i}} \frac{\omega_{m}}{R_{i}} u_{mj}^{(t)} \left(x_{m} - \mu_{j}^{(t)}\right) \left(x_{m} - \mu_{j}^{(t)}\right)^{T}}{\sum_{i=1}^{N} z_{ij}^{(t)}}$$
(20)

If the derivation of (16) with respect to $v_j^{(t+1)}$ is equated to zero, we can obtain the degrees of freedom $v_j^{(t+1)}$ using

$$\log\left(\frac{v_{j}^{(t+1)}}{2}\right) - \psi\left(\frac{v_{j}^{(t+1)}}{2}\right) + 1 - \log\left(\frac{v_{j}^{(t)} + D}{2}\right) + \frac{\sum_{j=1}^{N} z_{ij}^{(t)} \sum_{m \in N_{i}} \frac{\mathcal{O}_{m}}{R_{i}} \left(\log u_{mj}^{(t)} - u_{mj}^{(t)}\right)}{\sum_{j=1}^{N} z_{ij}^{(t)}} + \psi\left(\frac{v_{j}^{(t)} + D}{2}\right) = 0 \quad (21)$$

where $\psi(x) = \{\partial \Gamma(x) / \partial(x)\} / \Gamma(x)$ is the digamma function. Taking the constraint of the prior probability (2) into account, we obtain the following expression by using the Lagrange's multiplier λ_i .

$$S_{\pi} = S - \sum_{i=1}^{N} \lambda_i \left(\sum_{j=1}^{K} \pi_{ij}^{(t+1)} - 1 \right).$$
(22)

To obtain the prior probability $\pi_{ij}^{(t+1)}$, the derivation of (16) with respect to $\pi_{ij}^{(t+1)}$ is set to zero. Thus,

$$\pi_{ij}^{(t+1)} = \frac{z_{ij}^{(t)} + G_{ij}^{(t)}}{1 + \sum_{h=1}^{K} G_{ih}^{(t)}}$$
(23)

So far, for maximizing the objective function, the steps of proposed method based on weighted mean template and MRF are finished. The various steps of the proposed algorithm can be summarized as follows.

Algorithm:

Step 1. Initialize the parameters: using fuzzy *c*-means method, we obtain the initial means μ_j , the covariance Σ_j , the freedom of degree v_j , and the prior probability π_{ij} , respectively. Setting β =12.

Step 2. In **E-step**, calculate the posterior probability z_{ij} using (17); evaluate the weight u_{ij} using (18); update the novel factor G_{ij} using (12).

Step 3. In **M-step**, calculate the means μ_j , the covariance Σ_j , and the freedom of degree v_i using (19), (20), and (21), respectively. Update the prior probability π_{ij} using (23).

Step 4.. Compute the log-likelihood function using (14). If the convergence of (14) is satisfied, terminate the iteration. Otherwise, t=t+1 and go to step 2.

5. Experimental Results

In this section, three experiments are conducted to evaluate the effectiveness of the proposed methods, and the results are compared with the GMM, SMM, and ACAP [8].

5.1. Data Clustering

We begin with experiments on a set of noisy data to investigate the robustness of the proposed approach in noisy environment. The results obtained by using the standard GMM, SMM, ACAP and proposed method (β =6) are demonstrated in Figure 2. In this experiment, the sample with 2500 simulated points from five bivariate Gaussian distribution is shown in Figure 2. These data points are corrupted by 150 noise points (outliers) drawn from a bivariate uniform distribution, with each of its components in the interval [-0.4, 1.4]. The means of Gaussian distributions are μ_1 =(5.1291,7.0924)^T, μ_2 =(4.6048,1.1597)^T, μ_3 =(3.5040,0.7808)^T, μ_4 =(0.9505,3.6925)^T, and μ_5 =(4.3367,0.3363)^T, respectively. The covariances are $\Sigma_1 ~ \Sigma_5 = (1.5, 0; 0, 1.5)$. To evaluate objectively and compare the performance of the proposed method, we illustrate the simulated data points along with the contours of the clusters obtained using the evaluated algorithms. As can be seen in Figure 2, compared to GMM, SMM, and ACAP, we find that our method is very robust to the effect of the outliers. It is because that the proposed method takes the influence of the neighborhood pixels



Figure 1. Noisy Data Clustering. (a) GMM; (b) SMM; (c) ACAP; (d) Proposed Method

5.2. Segmentation of Real World Images

In the second experiment, four real world images from the Berkeley image dataset are selected randomly to compare different methods. To evaluate the performance of the proposed method, the probabilistic rand (PR) index [9] is used. The value of PR index ranges from 0 to 1. The higher the value is, the better the segmentation results are. The segmentation results using different methods are illustrated in Table 1. The number of class K is set according to human vision system characteristics. As can be seen in Table 1, the segmentation accuracy of GMM and SMM, along the objective boundaries is modestly poor. Table 2 lists the PR index. From Table 2, we can find that the proposed method has the highest PR values. It indicates that the proposed method yields the best segmentation results.

5.3. Segmentation of MR Images

In the last experiment, four real MR images from the Internet Brain Segmentation Repository (ISBR07, 256×256) are used to test the effectiveness of our methods. We apply GMM, SMM, ACAP, and the proposed method to segment the whole brain. The segmentation results are shown in Table 3. As can be seen from this table, the effects of noise on the final segmented images of GMM and SMM are high. The proposed method, on the other hand, can better classify with more robust to this noise. The quantitative results are illustrated in Table 4. As can be seen, the proposed method has the highest PR values for the segmented images.



Table 1. Real World Images Segmentation

Images	K	GMM	SMM	ACAP	Proposed Models
135069	2	0.9733	0.9732	0.9760	0.9860
238011	3	0.8170	0.8199	0.8264	0.8361
106025	3	0.8114	0.8357	0.8379	0.8597
55069	4	0.8778	0.8820	0.8854	0.8943

|--|

6. Conclusions

In this paper, we proposed an effective fuzzy clustering method for grayscale image segmentation. The novel approach incorporates the weighted mean template and MRF into the standard SMM model. The advantage of such a model is that it considers the spatial relationship among neighboring pixels. Furthermore, an efficient energy function and a novel factor Gij are used so as to the EM algorithm can be directly applied to calculate the new objective function. Thus, the proposed approach is simple and easy to implement. The proposed method was tested with simulated, real world grayscale and real MR images, demonstrating excellent performance in noisy conditions, compared to other mixture model-based approaches.

Acknowledgments

This work has been supported by the National Natural Science Foundation of China under Grant No. 61371150.



Table 3. MR Images Segmentation

Images	Κ	GMM	SMM	ACAP	Proposed Models
Skull1	3	0.7862	0.8089	0.7856	0.8092
Skull2	3	0.7867	0.8101	0.7857	0.8097
Skull3	3	0.8037	0.8201	0.8049	0.8183
Skull4	3	0.8216	0.8367	0.8220	0.8347

Table 4. Comparison of Different Methods for MR Images (PR Index)

References

- [1] A. Ribbens, J. Hermans, F. Maes, D. Vandermeulen and P. Suetens, "Unsupervised segmentation, clustering, and groupwise registration of heterogeneous populations of brain MR images", IEEE Trans. Med. Imaging, vol. 33, (2014), pp. 201-224.
- [2] D. Gong, G. Medioni and X. Zhao, "Structured time series analysis for human action segmentation and recognition", IEEE Trans Pattern Anal. and Mach. Intell., vol. 36, (2014), pp. 1414-1427.
- [3] C. Nikou, G. Sfikas and N. Galatsanos, "Robust image segmentation with mixtures of Student's-*t* distributions", in International Conference on Image Processing. IEEE, (**2007**), pp. 273-276.
- [4] D. Peel and G. McLachlan, "Robust mixture modeling using the *t*-distribution", Statistics and Computing, vol. 10, (2000), pp. 335-344.
- [5] M. N. Thanh and Q. M. J. Wu, "Fast and robust spatially constrained Gaussian mixture model for image segmentation", IEEE Trans. Circuits and Systems for Video Technology, vol. 23, (2013), pp. 621-635.
- [6] J. Besag, "Spatial iteraction and the statistical analysis of lattice systems", J. R. Stat. Soc. B, vol. 36, (1974), pp. 192-326.
- [7] W. Rudin, "Real and Complex Analysis", New York: McGraw-Hill, (1987).
- [8] H. Zhang, Q. M. J. Wu, and M. N. Thanh, "Incorporating mean template into finite mixture model for image segmentation", IEEE Trans. Neural Networks and Learning Systems, vol. 24, (2013), pp. 328-335.
- [9] R. Unnikrishnan, C. Pantofaru and M. Hebert, "A measure for objective evaluation of image segmentation algorithms", Computer Vision and Pattern Recognition, vol. 24, (**2005**), pp. 1-34.