A ROBUST NON-SYMMETRIC MIXTURE MODELS FOR IMAGE SEGMENTATION

Thanh Minh Nguyen, Q. M. Jonathan Wu, Senior Member, IEEE

Department of Electrical and Computer Engineering, University of Windsor
401 Sunset Avenue, Windsor, ON, N9B3P4, Canada
{nguyen1j, jwu}@uwindsor.ca

ABSTRACT

Finite mixture model with symmetric distribution has been widely used for many computer vision and pattern recognition problems. However, in many applications, the distribution of the data has a non-Gaussian and non-symmetric form. This study presents a new non-symmetric mixture model for image segmentation. The advantage of our method is that it is simple, easy to implement and intuitively appealing. In this paper, each label is modeled with multiple D-dimensional Student's-t distribution, which is heavily tailed and more robust than Gaussian distribution. Expectation maximization (EM) algorithm is adopted to estimate model parameters and to maximize the lower bound on the data log-likelihood from observations. Numerical experiments on various data types are conducted. The performance of the proposed model is compared to other mixture models, demonstrating the robustness, accuracy and effectiveness of our method.

Index Terms— Non-symmetric mixture model, non-Gaussian distribution, EM algorithm, and unsupervised image segmentation.

1. INTRODUCTION

Segmentation is one of the most important problems in computer vision. In literature, different techniques have been proposed for image segmentation. During the last decades, much attention has been given to clustering method based on the modeling of the probability density function of the data via finite mixture models [1–5].

Among these techniques, Gaussian mixture model (GMM) is a well-known method used in most applications. The major advantage of the GMM approach is that it provides a natural way to cluster data based on the components of the mixture that generated it. Another advantage is that the parameters can be efficiently estimated by adopting the expectation maximization (EM) algorithm [6, 7]. However, the GMM is sensitive to outliers and may lead to excessive sensitivity to small numbers of data points [8]. Also, for many applied problems, the tail of the Gaussian distribution is often shorter than required.

In order to improve the robustness of the algorithm, mixture models with the Student's-t distribution has been proposed in [8–10]. The main advantage of the Student's-t distribution is that it is heavily tailed than Gaussian distribution. Compared to the GMM, each component of the finite Student's-t mixture model (SMM) has one more parameter called the degrees of freedom (v). When v tends to infinity, the Student's-t distribution approaches the Gaussian distribution. Hence, SMM provides a more powerful and flexible approach for probabilistic data clustering compared to the GMM. However, in many real applications, the intensity distributions of each label type of the dataset do not exhibit exactly a Gaussian shape and are not symmetric [11]. For this reason, the results of the mixture models which are based on the symmetric distribution such as GMM, SMM are very poor in these non-symmetric situations.

Based on these considerations, in this paper, we propose a new finite mixture with non-symmetric distribution for image segmentation. Our approach differs from those discussed above by the following statements. Firstly, the Student's-t distribution, which is heavily tailed and more robust than Gaussian, is used in this paper. Secondly, each component density in our model is an asymmetric distribution that is modeled by multiple D-dimensional distribution. The advantage of the proposed distribution is that it has the flexibility to fit different shapes of observed data. Our model can be used for analyzing both univariate and multivariate data. Finally, EM algorithm is adopted to maximize the data log-likelihood and to optimize the parameters. We demonstrate through extensive simulations that the proposed model is superior to other clustering methods based on the modeling of the probability density function of the data via finite mixture model.

The rest of this paper is organized as follows. In Section 2, we present a brief introduction of the finite mixture model, commonly used in the literature for image segmentation. In Section 3, the proposed method and parameter estimation will be described in detail. The experimental results are demonstrated in Section 4 followed by conclusions in Section 5.

This research has been supported in part by the Canada Research Chair Program, AUTO21 NCE, and the NSERC Discovery grant.

2. STANDARD FINITE MIXTURE MODEL AND IMAGE SEGMENTATION

Notations used throughout the paper are as follows. The main objective is to segment an image consisting of N pixels into K labels. Let x_i, with dimension D, i=(1,2,...,N), denote an observation at the i-th pixel of an image. Labels are denoted by (Ω_1,Ω_2,...,Ω_K). The parameter π_j is the prior distribution of the pixel x_i belonging to the label Ω_j.

Consider the problem of estimating the posterior probability of x_i belonging to label Ω_j. The finite mixture models [1, 3, 5] assume that each pixel x_i is independent of the label Ω_j. The density function $f(\mathrm{x}_i|\Theta)$ at a pixel x_i is given by:

$$f(\mathrm{x}_i|\Theta) = \sum_{j=1}^{K} \pi_j p(\mathrm{x}_i|\Omega_j) \tag{1}$$

where, the prior probability that pixel x_i is in label Ω_j, which satisfies the constraints:

$$\pi_j \geq 0 \text{ and } \sum_{j=1}^{K} \pi_j = 1 \tag{2}$$

Each distribution $p(\mathrm{x}_i|\Omega_j)$ is called a component of the mixture. Note that, $p(\mathrm{x}_i|\Omega_j)$ can be any kind of distribution. In GMM, $p(\mathrm{x}_i|\Omega_j)$ is the Gaussian distribution $\Phi(\mathrm{x}_i|\mu_j, \Sigma_j)$. Each Gaussian distribution is written in the form:

$$\begin{aligned}\Phi(\mathrm{x}_i|\mu_j, \Sigma_j) &= \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma_j|^{1/2}} \\ &\times \exp\left\{-\frac{1}{2}(\mathrm{x}_i - \mu_j)^T \Sigma_j^{-1} (\mathrm{x}_i - \mu_j)\right\}\end{aligned} \tag{3}$$

where, the D-dimensional vector μ_j is the mean. The DxD matrix Σ_j is the covariance, and $|\Sigma_j|$ denotes the determinant of Σ_j.

In order to improve the robustness of the model, Student's-t distribution is used in SMM [8, 9]. In this model, $p(\mathrm{x}_i|\Omega_j)$ in Eq.(1) is the Student's-t distribution $S(\mathrm{x}_i|\mu_j, \Sigma_j, v_j)$ with longer tails and one more parameter compared to the Gaussian distribution $\Phi(\mathrm{x}_i|\mu_j, \Sigma_j)$. Each Student's-t distribution $S(\mathrm{x}_i|\mu_j, \Sigma_j, v_j)$ has its own mean μ_j, covariance Σ_j, and degree of freedom v_j. The Student's-t distribution $S(\mathrm{x}_i|\mu_j, \Sigma_j, v_j)$ is given by:

$$\begin{aligned}S(\mathrm{x}_i|\mu_j, \Sigma_j, v_j) &= \frac{\Gamma(v_j/2 + D/2)|\Sigma_j|^{-1/2}}{(v_j\pi)^{D/2}\Gamma(v_j/2)} \\ &\times \frac{1}{\left[1 + v_j^{-1}(\mathrm{x}_i - \mu_j)^{\mathrm{T}} \Sigma_j^{-1} (\mathrm{x}_i - \mu_j)\right]^{(v_j + D)/2}}\end{aligned} \tag{4}$$

The Student's-t distribution provides a longer tailed alternative to the Gaussian distribution. Each component in SMM originates from a wider class of elliptically symmetric distributions. Hence, SMM provides a more a powerful and flexible approach for probabilistic data clustering compared to the GMM.

3. PROPOSED METHOD

The key concept in the standard finite mixture model is the component $p(\mathrm{x}_i|\Omega_j)$. As shown in Section 2, all these statistical methods have relied on $p(\mathrm{x}_i|\Omega_j)$ for modeling the underlying distributions. However, all the component of GMM, SMM can only approximate a symmetric shape. In many real applications, the intensity distribution of each label type of the dataset does not exhibit exactly a Gaussian shape and are not symmetric. In order to overcome this problem, we propose a new finite Student's-t mixture with non-symmetric distribution, which is useful for modeling non-Gaussian image data.

First, we define a new non-Gaussian and non-symmetric distribution $p(\mathrm{x}_i|\Omega_j)$ that is used for the component of our mixture model. Differing from the above-mentioned mixture models, each component density in our model is modeled with multiple D-dimensional Student's-t distribution. The multi-dimensional distribution $p(\mathrm{x}_i|\Omega_j)$ in our model is defined as:

$$p(\mathrm{x}_i|\Omega_j) = \sum_{k=1}^{K_j} \eta_{jk} S(\mathrm{x}_i|\mu_{jk}, \Sigma_{jk}, v_{jk}) \tag{5}$$

where, K_j is the number of the Student's-t distribution that is used to model the label Ω_j. And η_{jk} is called the weighting factor that satisfies the following constraints:

$$\eta_{jk} \geq 0 \text{ and } \sum_{k=1}^{K_j} \eta_{jk} = 1 \tag{6}$$

In Eq.(5), $S(\mathrm{x}_i|\mu_{jk}, \Sigma_{jk}, v_{jk})$ is the Student's-t distribution

$$\begin{aligned}S(\mathrm{x}_i|\mu_{jk}, \Sigma_{jk}, v_{jk}) &= \frac{\Gamma(v_{jk}/2 + D/2)|\Sigma_{jk}|^{-1/2}}{(v_{jk}\pi)^{D/2}\Gamma(v_{jk}/2)} \\ &\times \frac{1}{\left[1 + v_{jk}^{-1}(\mathrm{x}_i - \mu_{jk})^{\mathrm{T}} \Sigma_{jk}^{-1} (\mathrm{x}_i - \mu_{jk})\right]^{(v_{jk} + D)/2}}\end{aligned} \tag{7}$$

where, the D-dimensional vector μ_{jk} is the mean. The DxD matrix Σ_{jk} is the covariance, Σ_{jk} denotes the determinant of Σ_{jk} and v_{jk} is the degree of freedom. The idea to define the distribution in Eq.(5) is based on a fact that non-symmetric distribution can be approximated by multiple Student's-t distributions.

Given the prior probability distribution π_j in Eq.(2) and the distribution $p(\mathrm{x}_i|\Omega_j)$ in Eq.(5), the log-likelihood function is written in the form.

$$L(\Theta) = \sum_{i=1}^{N} \log\left\{\sum_{j=1}^{K} \pi_j p(\mathrm{x}_i|\Omega_j)\right\} \tag{8}$$

The next objective is to optimize the parameter set $\Theta = \{\pi_j, \eta_{jk}, \mu_{jk}, \Sigma_{jk}, v_{jk}\}$ in order to maximize the log-likelihood function in Eq.(8). Application of the complete data condition in [1, 7], maximizing the log-likelihood function $L(\Theta)$ will lead to an increase in the value of the objective function $J(\Theta)$.

$$\begin{aligned} J(\Theta) &= \sum_{i=1}^{N}\sum_{j=1}^{K} z_{ij}^{(t)}\{\log\pi_j + \mathbb{E}_\Theta[\log p(\mathrm{x}_i|\Omega_j)]\} \\ &= \sum_{i=1}^{N}\sum_{j=1}^{K} z_{ij}^{(t)}\{\log\pi_j + \sum_{k=1}^{K_j} y_{ijk}^{(t)}\{\log\eta_{jk} \\ &\qquad + \mathbb{E}_\Theta[\log S(\mathrm{x}_i|\mu_{jk},\Sigma_{jk},v_{jk})]\}\} \end{aligned} \tag{9}$$

where, the posterior probability $z_{ij}^{(t)}$ in Eq.(10) at the iteration of the current step is

$$z_{ij}^{(t)} = \frac{\pi_j p(\mathrm{x}_i|\Omega_j)}{\sum\limits_{m=1}^{K}\pi_m p(\mathrm{x}_i|\Omega_m)} \tag{10}$$

and, the value of $y_{ijk}^{(t)}$ in Eq.(10) is given by

$$y_{ijk}^{(t)} = \frac{\eta_{jk}S(\mathrm{x}_i|\mu_{jk},\Sigma_{jk},v_{jk})}{\sum\limits_{m}^{K_j}\eta_{jm}S(\mathrm{x}_i|\mu_{jm},\Sigma_{jm},v_{jm})} \tag{11}$$

Note that, there is no closed form solution for maximizing the log-likelihood under a Student's-t distribution. To overcome this problem, the Student's-t distribution in previous models [8, 9] is represented as a Gaussian distribution with scaled precision u. The Student's-t distribution in our method is rewritten by:

$$\begin{aligned} &S(\mathrm{x}_i|\mu_{jk},\Sigma_{jk},v_{jk}) \sim \\ &\qquad \Phi(\mathrm{x}_i|\mu_{jk},\Sigma_{jk}/u_{ijk})\mathcal{G}(u_{ijk}|v_{jk}/2, v_{jk}/2) \end{aligned} \tag{12}$$

Where $\mathcal{G}(\cdot)$ is the Gamma distribution. Given the Student's-t distribution in Eq.(12), the objective function $J(\Theta)$ in Eq.(9) is written in the form.

$$\begin{aligned} J(\Theta) &= \sum_{i=1}^{N}\sum_{j=1}^{K} z_{ij}^{(t)}\{\log\pi_j + \sum_{k=1}^{K_j} y_{ijk}^{(t)}\{\log\eta_{jk} \\ &-\frac{D}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma_{jk}| + \frac{D}{2}\mathbb{E}_\Theta(\log u_{ijk}) - \log\Gamma(\frac{v_{jk}}{2}) \\ &+\frac{v_{jk}}{2}\log\frac{v_{jk}}{2} + (\frac{v_{jk}}{2}-1)\mathbb{E}_\Theta(\log u_{ijk}) - \frac{v_{jk}}{2}\mathbb{E}_\Theta(u_{ijk})\}\} \end{aligned} \tag{13}$$

Following [8], we have:

$$\mathbb{E}_\Theta(u_{ijk}) = u_{ijk}^{(t)} = \frac{v_{jk}^{(t)} + D}{v_{jk}^{(t)} + (\mathrm{x}_i - \mu_{jk}^{(t)})^T\Sigma_{jk}^{-1(t)}(\mathrm{x}_i - \mu_{jk}^{(t)})} \tag{14}$$

and,

$$\mathbb{E}_\Theta(\log u_{ijk}) = \log u_{ijk}^{(t)} - \log\left(\frac{v_{jk}^{(t)} + D}{2}\right) + \psi\left(\frac{v_{jk}^{(t)} + D}{2}\right) \tag{15}$$

In the M-step, to maximize the function in Eq.(13), the solution of $\partial J(\Theta)/\partial\Theta = 0$, after some manipulation, yields the estimates of μ_{jk} at the (t+1) step:

$$\mu_{jk} = \frac{\sum\limits_{i=1}^{N} z_{ij}^{(t)}y_{ijk}^{(t)}u_{ijk}^{(t)}\mathrm{x}_i}{\sum\limits_{i=1}^{N} z_{ij}^{(t)}y_{ijk}^{(t)}u_{ijk}^{(t)}} \tag{16}$$

The estimates of Σ_{jk} at the (t+1) step:

$$\Sigma_{jk} = \frac{\sum\limits_{i=1}^{N} z_{ij}^{(t)}y_{ijk}^{(t)}u_{ijk}^{(t)}(\mathrm{x}_i - \mu_{jk})(\mathrm{x}_i - \mu_{jk})^T}{\sum\limits_{i=1}^{N} z_{ij}^{(t)}y_{ijk}^{(t)}} \tag{17}$$

The estimates of the degrees of freedom v_{jk} are given by the solution of the equation.

$$\begin{aligned} &-\Psi\left(\frac{v_{jk}}{2}\right) + \log\left(\frac{v_{jk}}{2}\right) + 1 + \frac{\sum\limits_{i=1}^{N} z_{ij}^{(t)}y_{ijk}^{(t)}(\log u_{ijk}^{(t)} - u_{ijk}^{(t)})}{\sum\limits_{i=1}^{N} z_{ij}^{(t)}y_{ijk}^{(t)}} \\ &\qquad + \psi\left(\frac{v_{jk}^{(t)} + D}{2}\right) - \log\left(\frac{v_{jk}^{(t)} + D}{2}\right) = 0 \end{aligned} \tag{18}$$

Where $\Psi(\cdot)$ is the digamma function. Considering the constraints in Eq.(2) and Eq.(6), the prior distribution π_j, and the weighting factor η_{jk} at the (t+1) step are calculated by:

$$\pi_j = \frac{1}{N}\sum_{i=1}^{N} z_{ij}^{(t)}; \quad \eta_{jk} = \frac{\sum\limits_{i=1}^{N} z_{ij}^{(t)}y_{ijk}^{(t)}}{\sum\limits_{i=1}^{N} z_{ij}^{(t)}\sum\limits_{m=1}^{K_j} y_{ijm}^{(t)}} \tag{19}$$

In the next section, we will demonstrate the robustness, accuracy and effectiveness of the proposed model, as compared with other approaches.

4. EXPERIMENTS

To test the effectiveness of our method, the performance of our method with GMM [1], SMM [8]. In this paper, the proposed method is implemented with two D-dimensional Student's-t distribution for each label ($K_j = 2$). In the first experiment, one sample with 7936 simulated points from four labels is shown in Fig. 1(a). Each label has 1984 points. The ground truth distributions (two-dimensional view) of four labels are shown in Fig. 1(b). As shown in this figure, the intensity distribution of each label type of the data does not exhibit exactly a Gaussian shape and are not symmetric. In Fig. 1(c)-(e), the results of GMM, SMM, and our method are shown respectively. In this experiment, the error of the estimated distributions of GMM and SMM compared with the ground truth distributions in Fig. 1(b) remains quite high. The proposed method, as shown in Fig. 1(e), is more accurate as compared to other methods with the lowest misclassification ratio (MCR) [4].

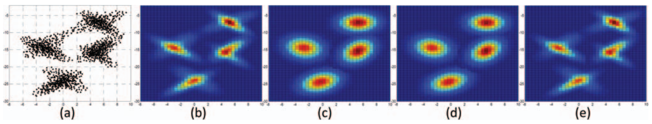

Fig. 1. The simulated point set experiment, (a): Original data, (b): Ground truth distribution, (c): GMM (MCR=1.10%), (d): SMM (MCR=1.08%), (e): Proposed Method (MCR=0.04%).

In Fig. 2, we show the segmentation results of a real-world color image from the Berkeley's image segmentation dataset. The original image, as shown in Fig. 2(a), was segmented into three labels: *water*, *sky*, and *other*. Extraction accuracies of the GMM, SMM, and our methods are shown in Fig. 2(b) to Fig. 2(d). As shown in Fig. 2(b) and Fig. 2(c), the segmentation accuracies for GMM and SMM methods are quite poor. The edge between the *water* and the *other* is lost. A closer inspection of the *water* area (look into the marked box) indicates that a small portion of pixels have been misclassified. The proposed method in Fig. 2(d), can better classify with more detail as compared with other methods.

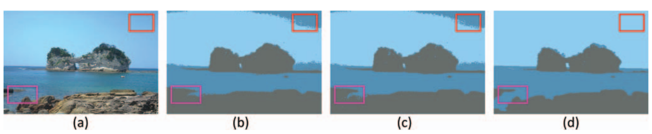

Fig. 2. Color natural image segmentation, (a): original image, (b): GMM, (c): SMM, (d): Proposed Method.

In the final experiment, one real brain image (IBSR01, slice=50), from the Center for Morphometric Analysis at Massachusetts General Hospital (http://www.cma.mgh.harvard.edu/ibsr/data.html), is used. The ground truth image is shown in Fig. 3(b). As seen from these results, the effect of noise on the final segmented image of GMM (MCR=29.06%) and SMM (MCR=28.53%) is high. We can see that many details are lost. Again, compared with other methods, the proposed method has the lowest MCR.

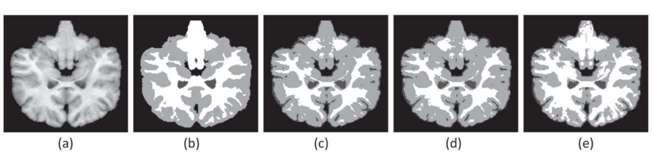

Fig. 3. The real brain image experiment, (a): Original image (IBSR01, slice=50), (b): Ground truth, (c): GMM (MCR= 29.06%), (d): SMM (MCR= 28.53%), (e): Proposed method (MCR= 19.73%).

5. CONCLUSIONS

We have presented a new non-symmetric mixture model for image segmentation in this paper. The distribution of our method has a non-Gaussian and non-symmetric form. Each label is modeled with multiple D-dimensional Student's-t distribution, which is heavily tailed and more robust than Gaussian distribution. The advantage of the proposed distribution is that it has the flexibility to fit different shapes of observed data. Our model can be used for analyzing both univariate and multivariate data. EM algorithm is adopted to maximize the data log-likelihood and to optimize the parameters. The proposed method has been tested on various data types, thereby demonstrating the excellent performance of the proposed model in segmenting the images.

6. REFERENCES

[1] McLachlan G., and Peel D., "*Finite Mixture Models*", New York, Wiley, 2000.

[2] Bishop C. M., "*Neural Networks for Pattern Recognition*", Oxford University Press, Walton Street, Oxford, 1995.

[3] Bishop C. M., "*Pattern Recognition and Machine Learning*", Springer, 2006.

[4] Thanh M. N., Wu Q. M. J., and Ahuja S., "An Extension of the Standard Mixture Model for Image Segmentation," *IEEE Trans. Neural Netw.*, vol. 21, no. 8, pp. 1326–1338, 2010.

[5] Titterington D. M., Smith A. F. M., and Makov U. E., "*Statistical Analysis of Finite Mixture Distributions*", Hoboken, Wiley, 1985.

[6] Dempster P., Laird N. M., and Rubin D. B., "Maximum likelihood from incomplete data via EM algorithm," *J. R. Stat. Soc.*, vol. 39, no. 1, pp. 1–38, 1977.

[7] Liu C., and Rubin D., "ML estimation of the t distribution using EM and its extensions, ECM and ECME," *Stat. Sinica*, vol. 5, no. 1, pp. 19–39, 1995.

[8] Peel D., and McLachlan G., "Robust Mixture Modeling Using the t Distribution," *Stat. Comput.*, vol. 10, pp. 339–348, 2000.

[9] Shoham S., "Robust clustering by deterministic agglomeration EM of mixtures of multivariate t-distributions," *Pattern Recognit.*, vol. 35, no. 55, pp. 1127–1142, 2000.

[10] 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. Imag.*, vol. 31, no. 1, pp. 103–106, 2012.

[11] Ashburner J., and Friston K. J., "Unified segmentation," *NeuroImage*, vol. 26, no. 3, pp. 839–851, 2005.