Image Segmentation Using a Trimmed Likelihood Estimator in the Asymmetric Mixture Model Based on Generalized Gamma and Gaussian Distributions

Yi Zhou and Hongqing Zhu

School of Information Science & Engineering, East China University of Science and Technology, Shanghai 200237, China

Correspondence should be addressed to Hongqing Zhu; hqzhu@ecust.edu.cn

Received 24 July 2017; Revised 14 December 2017; Accepted 26 December 2017; Published 24 January 2018

Academic Editor: Xiangyu Meng

Copyright © 2018 Yi Zhou and Hongqing Zhu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Finite mixture model (FMM) is being increasingly used for unsupervised image segmentation. In this paper, a new finite mixture model based on a combination of generalized Gamma and Gaussian distributions using a trimmed likelihood estimator (GGMM-TLE) is proposed. GGMM-TLE combines the effectiveness of Gaussian distribution with the asymmetric capability of generalized Gamma distribution to provide superior flexibility for describing different shapes of observation data. Another advantage is that we consider the spatial information among neighbouring pixels by introducing Markov random field (MRF); thus, the proposed mixture model remains sufficiently robust with respect to different types and levels of noise. Moreover, this paper presents a new component-based confidence level ordering trimmed likelihood estimator, with a simple form, allowing GGMM-TLE to estimate the parameters after discarding the outliers. Thus, the proposed algorithm can effectively eliminate the disturbance of outliers. Furthermore, the paper proves the identifiability of the proposed mixture model in theory to guarantee that the parameter estimation procedures are well defined. Finally, an expectation maximization (EM) algorithm is included to estimate the parameters of GGMM-TLE by maximizing the log-likelihood function. Experiments on multiple public datasets demonstrate that GGMM-TLE achieves a superior performance compared with several existing methods in image segmentation tasks.

1. Introduction

Segmenting an object from its background has an important role in machine learning and computer vision [1]. In recent years, several unsupervised image segmentation algorithms have been presented [2, 3]. Statistical approach, particularly finite mixture model (FMM), is one of the most widely known ones [4]. This can be used to model arbitrary univariate or multivariate observed data. In particular, modelling the probability density function of pixel attributes with Gaussian mixture model (GMM) has proved successful in segmentation tasks [5]. It is mainly because the parameters of GMM can be easily estimated by maximizing the maximum likelihood (ML) of the observed data using the expectation maximization (EM) algorithm. However, there remain limitations preventing GMM from achieving improved performance. The first challenge is sensitivity to noise, which is caused by the independence of the spatial relationship of pixels during parameter learning. The second challenge derives from the difficulty in fitting asymmetric observed data. In addition to these limitations, GMM is sensitive to outliers and can lead to excessive sensitivity to a small number of data points.

Recent studies have attempted to overcome the above disadvantages. The existing schemes can be categorised as follows.

(1) Schemes Based on Markov Random Field (MRF). A wide variety of approaches, especially Markov random field, have been introduced for resisting noise. These schemes utilize MRF smoothness prior to modelling the joint prior distribution of pixel labels; hence, the spatial information of the pixels is considered via the contextual constraints of the neighbouring pixels [6, 7]. Therefore, the MRF-based mixture model has stronger ability to resist noise. However, MRF suffers from the fact that parameter estimation is difficult and suffers high computational complexity. Recently,
mean template is employed along with a spatially varying mixture model to alleviate the influence of noise in image segmentation [8]. It is a natural approach to prevent noise because it automatically filters noise using a mean filter.

(2) Schemes Based on Asymmetric Probability Distribution. In general, GMM does not fit well if the shapes of the observed data are asymmetric [9]. Indeed, in many real applications, the intensity distribution of the observed data is not symmetric. Thus, it would seem that FMM with asymmetric distribution such as Gamma distribution [3], Weibull distribution [10], and Rayleigh distribution [11] could overcome this limitation. Another typical case is to obtain the asymmetric distribution via the linear weighted aggregative method using two or more symmetric probability distributions. One typical example is asymmetric Student’s $t$ mixture model (NSMM) [12], where each component density is modelled with multiple Student’s $t$ distributions. Another example of this case is the Bayesian-bounded asymmetric mixture model (BAMM) [13], which was developed by a subset of the authors [12] and other coauthors, for unsupervised image segmentation. Each component of their approach can model different shapes of observed data with different bounded support regions. A close relative of this framework involves a bounded asymmetrical Student’s $t$ mixture model [14]. Peculiarly, we note that the mixture of two or more different distributions has caused great concern and yet has developed rapidly in recent years. Typical algorithms include Zhou et al.’s statistical model [15], which is a mixture of $K$-distribution and lognormal distribution. This also includes De Angelis et al. [16] who offered a robust time interval measurement method based on a Gaussian–uniform mixture model. Browne et al. [17] incorporated a multivariate Gaussian and uniform distributions as the component density, which allowed for superior mixture possessing. These methods demonstrate a competitive performance in fitting different shapes of observed data.

(3) Schemes Based on Trimming Method. In general, for GMM-based algorithms, the parameters are estimated by the ML estimator through the EM algorithm. However, the ML estimator is overly sensitive to outliers and GMM cannot address outliers properly. Therefore, outliers seriously deteriorate the performance of Gaussian-based clustering algorithms. To overcome this shortcoming, a common approach is to consider a mixture model with Student’s $t$ distribution (SMM), which provides a longer-tailed alternative to the Gaussian distribution [18]. Therefore, SMM is more robust to outliers than the GMM for heavier tails. Another model-based method, which presents a theoretically well-based segmentation criterion in presence of outliers, is the trimming method [19]. The main principle in trimming is to locate and discard the outliers from the likelihood function. Segmentation results benefit from the trimming approach. Müller and Neykov proposed the fast trimmed likelihood estimator (FAST-TLE) [20]. Galimzianova et al. developed the confidence level ordering trimmed likelihood estimator (CLO-TLE) [21]. However, they do not function effectively in noisy samples, especially when each group has a different size of observations.

Motivated by the aforementioned considerations, in this paper, we present a two-step procedure (GGMM-TLE), beginning with a component-based confidence level ordering trimmed likelihood estimator. Because the majority of observed data contains outliers, it is necessary to discard these in a previous step before robustly estimating parameters. As a new algorithm, the proposed technique considers the components with lower mixture weights. This avoids eliminating the samples belonging to the components with a small number of observations as the outliers. Then, we propose a novel finite mixture model based on a mixture of generalized Gamma and Gaussian distributions (GGMM).

The proposed GGMM with Markov random fields has high flexibility and can be used to fit the asymmetric data owing to the introduction of the asymmetric generalized Gamma distribution. Moreover, we theoretically prove the property of identifiability of the GGMM through the strategy presented by Atienza et al. [22, 23], which indicates that the GGMM’s mixture representation is unique. This property is crucial to ensure that the parameter estimation problem is well posed. Therefore, the proposed algorithms can be effectively applied for segmenting images. Moreover, by imposing spatial smoothness constraints among neighboring pixels using MRF, the neighboring pixels should have the same label. Therefore, the proposed model reduces the segmentation sensitivity to noise in a still image. We demonstrate through simulation study that the proposed framework is superior to other related methods in terms of the misclassification ratio and Dice similarity coefficient.

The remainder of this paper is organized as follows. Section 2 introduces the proposed mixture model in detail. In Section 3, we prove the identifiability of the proposed mixture model. The process of parameter learning is described in Section 4. The ordering method for likelihood trimming is reported in Section 5. Section 6 provides the experimental results and analysis. Finally, we conclude with a discussion in Section 7.

2. Model Formulation

Assume a set of data $X = \{x_i \mid i = 1, 2, \ldots, N\}$, where each $x_i$ denotes an observation at the $i$th pixel of an image. $N$ is the total number of pixels in an image. The proposed mixture model assumes that the density function at pixel $x_i$ is given by

$$
 f(x_i \mid \Theta) = \sum_{j=1}^{K} \pi_j \left\{ \sum_{k=1}^{K} h_{jk} p_{jk}(x_i \mid \theta_{jk}) \right\}, \tag{1}
$$

where $\Theta = \{\pi_j, h_{jk}, \theta_{jk}\}$ is the complete parameter set of the proposed mixture model; $K$ denotes the number of mixture components. The prior $\pi_j$ represents the probability that the observation $x_i$ belongs to the $j$th label and $h_{jk}$ is called the weighting factor; they satisfy the following constraints:
The Markov random field based on Gibbs distribution can be formulated, we define the following maximum a posteriori probability density function of the proposed model as

\[
L(\Theta | X) = \log \left\{ \frac{1}{Z} \prod_{j=1}^{K} \pi_j p(X_j | \Theta) \right\},
\]

where \( Z \) is the normalizing constant. To train the proposed mixture model, based on the above formulation, we define the following log-likelihood function as

\[
L(\Theta | X) = \sum_{i=1}^{N} \log \left\{ \frac{\pi_i}{Z} \prod_{j=1}^{K} p(X_j | \Theta) \right\} - \log Z + \frac{1}{T} \sum_{i=1}^{N} \sum_{j=1}^{K} G_{ij} \log \pi_j.
\]

The above scheme contains two parts, where the first denotes the proposed mixture model and the last is the Markov model. In general, the EM algorithm is an efficient framework for estimating the mixture model parameters.

### 3. Identifiability of the Proposed Mixture Model

This section discusses the property of identifiability of the GGMM. This property implies that GGMM can only be expressed by specific components. Obviously, this property is important for finite mixture model because it can guarantee the estimation procedures of the parameter set to be well defined [3, 22]. The property of identifiability is described as follows.

We defined the following set:

\[
F = \{ f : f_{\theta}(x) = h_1 p_1(x | v, k, \sigma) + h_2 p_2(x | u, \Sigma) \}.
\]

\( F \) is the family of proposed distributions, where \( v \neq 0, k > 0, \sigma > 0, h_1 + h_2 = 1 \). In this study, \( p_1(x | v, k, \sigma) \) is the generalized Gamma distribution and \( p_2(x | u, \Sigma) \) is the Gaussian distribution; the parameters of the proposed distribution are mutually independent. The set of the proposed mixture model with \( \pi_j \) satisfying (2) is

\[
H_F = \left\{ H : H(x) = \sum_{j=1}^{K} \pi_j f_{\theta_j}(x), \quad f_{\theta_i}(x) \in F \right\}.
\]

**Theorem 1.** The property of identifiability of \( H_F \) means that for any two mixture models \( H_{1,2} \in H_F \), that is,

\[
H_1 = \sum_{j=1}^{K_1} \pi_{1j} f_{\theta_{1j}}(x), \quad H_2 = \sum_{j=1}^{K_2} \pi_{2j} f_{\theta_{2j}}(x), \quad \pi_{1j} = \pi_{2j}, \quad f_{\theta_{1j}} = f_{\theta_{2j}}.
\]

**Proof.** According to [22], we prove that, for a linear transform, \( M : f_0(x) \rightarrow \phi_j \) with domain \( S(f) \). Let \( S_0(f) = \ldots \)
\{ m \in S(f) : \phi_i(m) \neq 0 \}. For a given point \( m_0 \) and any two proposed distributions \( f_1, f_2 \in F \), there exists a total order \( < \) on \( F \) that satisfies
\[
\frac{\phi_j(m)}{\phi_i(m)} \xrightarrow{m \to m_0} 0. 
\]

(14)

Given the expression of the linear transform \( M \) is as follows:
\[
M[f_0(x)] : \phi_f(m) = E(x^m) = \int_{-\infty}^{\infty} x^m f_0(x) \, dx, \quad \text{(15)}
\]

where \( f_0(x) \) is the proposed density function \( f_0(x) = h_1 p_1(x \mid v, k, \sigma) + h_2 p_2(x \mid u, \Sigma) \). Let \( g_{\theta_1}(x) = p_1(x \mid v, k, \sigma) \) and \( g_{\theta_2}(x) = p_2(x \mid u, \Sigma) \); then, \( f_0(x) = g_{\theta_1}(x) + h_2 g_{\theta_2}(x) \). Obviously, if \( g_1 < g_2 \) and \( h_1 < h_2 \), we can obtain \( f_1 < f_2 \). According to (15), we have
\[
\phi_f(m) = \sigma^m \Gamma(k + m/v) \Gamma(k), \quad m \in (-kv, +\infty), \quad \text{(16)}
\]

where \( S_0(g) = (-kv, +\infty) \) and \( m_0 = +\infty \). To facilitate the proof procedure, this study utilizes Stirling’s formula as follows:
\[
\Gamma(z + 1) \sim 2\pi z^{z} e^{-z}, \quad z \to +\infty. \quad \text{(17)}
\]

Thus, we have
\[
\phi_f(m) \sim \frac{\sigma^m}{\Gamma(k)} \sqrt{2\pi} \left( \frac{m}{v} \right)^{k + m/v - 1/2}
\]
\[
\cdot \left[ 1 + \frac{k \cdot v}{m} \right]^{k + m/v - 1/2} \exp \left\{ 1 - k - \frac{m}{v} \right\} \sim \frac{\sqrt{2\pi}}{\Gamma(k)} \exp \left\{ m \log \sigma - \frac{m}{v} \right\}
\]
\[
\cdot \exp \left\{ \left( k + \frac{m}{v} - \frac{1}{2} \right) \left( \log m - \log v \right) \right\}. \quad \text{(18)}
\]

The sign “\( \sim \)” indicates that the expressions on both sides are equivalent up to constant term when \( m \to +\infty \). Hence, for \( m \to m_0 \), we have
\[
\frac{\phi_{g_1}(m)}{\phi_{g_2}(m)} \sim C \exp \left\{ \left( \frac{1}{v_2} - \frac{1}{v_1} \right) m \log m \right. 
\]
\[
+ \left( \log \sigma_2 - \log \sigma_1 \right) - \left( \frac{1}{v_2} - \frac{1}{v_1} \right) 
\]
\[
+ \left( \frac{1}{v_2} \log \frac{1}{v_2} - \frac{1}{v_1} \log \frac{1}{v_1} \right) m 
\]
\[
+ \left( k_2 - k_1 \right) \log m + \left( k_2 - k_1 \right) \log m \right\}, \quad \text{(19)}
\]

where \( C \) is a constant. From (19), we can derive \( g_1 < g_2 \Leftrightarrow [v_2 > v_1], \) or \([v_2 = v_1, \sigma_2 < \sigma_1], \) or \([v_2 = v_1, \sigma_2 = \sigma_1, k_2 < k_1]\), which is apparently a total order. Analogously, we have
\[
\phi_{S_0}(m) = \exp \left\{ \mu m + \frac{1}{2} \Sigma m^2 \right\}, \quad \text{(20)}
\]

where \( S_0(h) = (-\infty, +\infty) \) and \( m_0 = +\infty \). For \( m \to m_0 \), we have
\[
\frac{\phi_{S_0}(m)}{\phi_{S_1}(m)} = \exp \left\{ \frac{1}{2} \left( \Sigma_2 - \Sigma_1 \right) m^2 + (u_2 - u_1) m \right\}. \quad \text{(21)}
\]

From (21), we can determine that \( h_1 < h_2 \Leftrightarrow [\Sigma_2 < \Sigma_1], \) or \([\Sigma_2 = \Sigma_1, u_2 < u_1]\), which is clearly a total order. Overall, we have \( f_1 < f_2 \); there exists a total order \( < \) on \( F \). Hence, we can draw the conclusion that the GGMMs are identifiable.

4. Parameter Learning

The main task of this section is to estimate the complete parameter set. In general, the EM algorithm provides an efficient scheme for unsupervised segmentation using iterative updating and guarantees that the log-likelihood function converges to a local maximum. Considering the complexity of (10), it is difficult to apply the EM algorithm directly for maximizing the log-likelihood function (10). Therefore, we employ Jensen’s inequality by defining the two hidden variables \( z_{ij} \) and \( y_{ijk} \), which are, respectively,
\[
z_{ij}^{(t)} = \frac{\pi_j [h_{j1} p_{jk} (x_i \mid \theta_{j1}) + h_{j2} p_{kj} (x_i \mid \theta_{j2})]}{\sum_{m=1}^{K} \pi_m [h_{m1} p_{mj} (x_i \mid \theta_{m1}) + h_{m2} p_{m2} (x_i \mid \theta_{m2})]},
\]
\[
y_{ijk}^{(t)} = \frac{h_{jk} p_{jk} (x_i \mid \theta_{jk})}{\sum_{m=1}^{K} h_{jm} p_{jm} (x_i \mid \theta_{jm})}. \quad \text{(22)}
\]

Clearly, \( z_{ij} \) and \( y_{ijk} \) satisfy the constraints \( \sum_{j=1}^{K} z_{ij} = 1 \) and \( \sum_{k=1}^{K} y_{ijk} = 1 \). Using Jensen’s inequality, one has
\[
\log(\sum_{j=1}^{K} z_{ij}) \geq \log(\sum_{j=1}^{K} z_{ij}) \geq \sum_{j=1}^{K} (z_{ij} \log e_{ij}); \text{thus, the log-likelihood function (10) can be rewritten as follows:}
\]
\[
L(X \mid \Theta) \geq \sum_{i=1}^{N} \sum_{j=1}^{K} z_{ij}^{(t)} \left\{ \log \pi_j 
\right.
\]
\[
+ \log \left[ \sum_{k=1}^{K} h_{jk} p_{jk} (x_i \mid \theta_{jk}) \right] \right\} - \log Z + \frac{1}{T} \quad \text{(23)}
\]
\[ \sum_{i=1}^{N} \sum_{j=1}^{K} G_{ij} \log \pi_j \geq \sum_{i=1}^{N} \sum_{j=1}^{K} z_{ij}^{(t)} \left( \log \pi_j \right) + \frac{1}{T} \sum_{i=1}^{N} \sum_{j=1}^{K} G_{ij} \log \pi_j. \]  
\[
+ \sum_{j=1}^{K} \left( \log h_{jk} + \log p_{jk} \left( x_j \mid \theta_{jk} \right) \right) - \log Z + \frac{1}{T} \sum_{i=1}^{N} \sum_{j=1}^{K} G_{ij} \log \pi_j
\]

Thus, we can define the following new objective function \( Q(X \mid \Theta) \) in terms of Jensen's inequality.

\[
Q(X \mid \Theta) = \sum_{i=1}^{N} \sum_{j=1}^{K} z_{ij}^{(t)} \left( \log \pi_j \right) + \sum_{j=1}^{K} \left( \log h_{jk} + \log p_{jk} \left( x_j \mid \theta_{jk} \right) \right) - \log Z + \frac{1}{T} \sum_{i=1}^{N} \sum_{j=1}^{K} G_{ij} \log \pi_j
\]

To realize clustering, we must maximize the log-likelihood function in (10), which is equivalent to maximizing the objective function in (25) with respect to \( \pi_j \), yielding

\[
\frac{\partial}{\partial \pi_j} \left[ Q(X \mid \Theta) - \lambda \left( \sum_{j=1}^{K} \pi_j - 1 \right) \right] = 0,
\]

where \( \lambda \) is the Lagrange multiplier in consideration of the constraint \( \sum_{j=1}^{K} \pi_j = 1 \); we have

\[
\pi_j^{(t+1)} = \frac{z_{ij}^{(t)} + G_{ij}^{(t)}}{\sum_{m=1}^{K} z_{im}^{(t)} + G_{im}^{(t)}}.
\]

Similarly, to estimate the weighting factor \( h_{jk} \), we take the partial derivative of the objective function in (25) with respect to \( h_{jk} \)

\[
\frac{\partial}{\partial h_{jk}} \left[ Q(X \mid \Theta) - \tau \left( \sum_{k=1}^{K} h_{jk} - 1 \right) \right] = 0,
\]

where \( \tau \) is the Lagrange multiplier in consideration of the constraint \( \sum_{k=1}^{K} h_{jk} = 1 \); we have

\[
h_{jk}^{(t+1)} = \frac{\sum_{i=1}^{N} z_{ij}^{(t)} y_{ijk}^{(t)}}{\sum_{i=1}^{N} z_{ij}^{(t)}}.
\]

In the following, we estimate the power parameter \( v_j \). We calculate the partial derivative of the objective function (25) with this power parameter \( v_j \) as follows:

\[
\frac{\partial Q(X \mid \Theta)}{\partial v_j} = \frac{N}{\sum_{i=1}^{N} \left[ z_{ij}^{(t)} y_{ij1}^{(t)} \left( k_j - \frac{x_j}{\sigma_j} \right)^v \log \left( \frac{x_j}{\sigma_j} \right) + \frac{1}{v_j} \right]}
\]

The solution of \( \frac{\partial Q(X \mid \Theta)}{\partial v_j} = 0 \) yields the estimates of \( v_j \) as follows:

\[
v_j^{(t+1)} = \frac{\sum_{i=1}^{N} \left[ z_{ij}^{(t)} y_{ij1}^{(t)} \left( v_j^{(t)} \right)^2 x_j / \sigma_j^{(t)} \right] \log \left( x_j / \sigma_j^{(t)} \right) + 1}{\sum_{i=1}^{N} \left[ z_{ij}^{(t)} y_{ij1}^{(t)} \left( v_j^{(t)} \right)^2 - k_j^{(t)} \right] \log \left( x_j / \sigma_j^{(t)} \right)},
\]

where \( v_j^{(t)} = v_j^{(t)} \log (x_j / \sigma_j^{(t)}) + 1 \). Then, to derive the solution of the shape parameter \( k_j \), we must calculate the partial derivative \( \frac{\partial Q(X \mid \Theta)}{\partial k_j} \) with respect to it. We have

\[
\frac{\partial Q(X \mid \Theta)}{\partial k_j} = \frac{N}{\sum_{i=1}^{N} \left[ z_{ij}^{(t)} y_{ij1}^{(t)} \left( v_j \right) \log \left( x_j / \sigma_j^{(t)} \right) - \phi_0 (k_j) \right]}
\]

It is clear to see that the solution \( \frac{\partial Q(X \mid \Theta)}{\partial k_j} = 0 \) yields the updates for shape parameter \( k_j \) by

\[
\phi_0 (k_j) = \frac{v_j^{(t+1)} \sum_{i=1}^{N} \left[ z_{ij}^{(t)} y_{ij1}^{(t)} \log \left( x_j / \sigma_j^{(t)} \right) \right]}{\sum_{i=1}^{N} \left[ z_{ij}^{(t)} y_{ij1}^{(t)} \right]}
\]

where \( \phi_0 (\cdot) \) is the Digamma function; \( k_j^{(t+1)} \) can be calculated by solving (33) via the bisection method [23]. In the same fashion, to obtain the estimate of the scale parameter \( \sigma_j \), we must derive the partial derivative of \( Q(X \mid \Theta) \) over it.
Equating $\partial Q(X | \Theta)/\partial \sigma_j$ to zero, we can obtain the update formulas for scale parameter $\sigma_j$ by

$$
\sigma_j^{(t+1)} = \left[ \frac{\sum_{i=1}^{N} z_{ij}^{(t)} y_{ij}^{2} (x_i - u_j)^2}{\sum_{i=1}^{N} z_{ij}^{(t)} y_{ij}^2} \right]^{1/2}.
$$

(35)

By calculating the partial derivative of the objective function in (25) with parameter set $\theta_j = \{u_j, \Sigma_j\}$, we can obtain the estimation of the parameters mean $u_j$ and covariance $\Sigma_j$:

$$
\frac{\partial Q(X | \Theta)}{\partial u_j} = \frac{N}{\sum_{i=1}^{N} z_{ij}^{(t)} y_{ij}^2} \left( x_i - u_j \right),
$$

$$
\frac{\partial Q(X | \Theta)}{\partial \Sigma_j} = \sum_{i=1}^{N} z_{ij}^{(t)} y_{ij}^2 \left( 1 - \frac{1}{2 \Sigma_j} + \frac{(x_i - u_j)^2}{2 \Sigma_j^2} \right).
$$

(36)

Eventually, the final updates for these two parameters can be obtained by

$$
u_j^{(t+1)} = \frac{\sum_{i=1}^{N} z_{ij}^{(t)} y_{ij}^2 x_i}{\sum_{i=1}^{N} z_{ij}^{(t)} y_{ij}^2},
$$

$$
\Sigma_j^{(t+1)} = \frac{\sum_{i=1}^{N} z_{ij}^{(t)} y_{ij}^2 (x_i - u_j)^2}{\sum_{i=1}^{N} z_{ij}^{(t)} y_{ij}^2}.
$$

(37)

(38)

At this point, the parameter learning procedure is complete.

### 5. Ordering Method for Likelihood Trimming

For observed data with heavy outliers, it is preferred to discard the outliers and to estimate the parameters of the proposed mixture model using the remaining data. Assume that $X$ is a sample with $N$ observations, $\eta$ is the trimming fraction, and $X_M$ is the subsample with a size $M = N(1 - \eta)$. Theoretically, the trimming fraction $\eta$ should be higher than the real outlier fraction value. After cutting the outliers, we estimate the model parameters by maximizing the objective function $Q(X_M | \Theta)$ in subsample $X_M$. The most important step is to discard the outliers and select the subsample. This requires a specific ordering for all of the observations in the sample. Typically, the number of outliers is unpredictable. Thus, it is important for the proposed model to avoid allowing observed data that belongs to the labels with a small number of observations falling into outliers. This study presents an effective component-based confidence level ordering method. In the proposed GGMM-TLE, we do not calculate the density function value as in FAST-TLE [20], for every single observation. Rather, we only utilize the concept of confidence level for these observations to eliminate the effects of mixture weights and sample scales. Combined with the posterior probability $z_{ij}^{(t)}$ in (22), we can order the observations that belong to the same group separately. Thus, it is more reasonable for ordering the observation with GGMM-TLE. Specifically, we derive the following increasing inequality based on component-based confidence level ordering:

$$
\int_{\Omega_{j}(x_{d_{jk}})_{k=1}}^{K_j} h_{jk} P_{jk}(\omega | \theta_{jk}) \, d\omega \leq \cdots \leq \int_{\Omega_{j}(x_{d_{jk}})_{k=1}}^{K_j} h_{jk} P_{jk}(\omega | \theta_{jk}) \, d\omega \leq \cdots
$$

(39)

$$
\int_{\Omega_{j}(x_{d_{jk}})_{k=1}}^{K_j} h_{kk} P_{kk}(\omega | \theta_{kk}) \, d\omega \leq \cdots \leq \int_{\Omega_{j}(x_{d_{jk}})_{k=1}}^{K_j} h_{kk} P_{kk}(\omega | \theta_{kk}) \, d\omega,
$$

where $\Omega_{j}(x_{d_{jk}}) = \{ \omega \in \Omega_j : \sum_{k=1}^{K_j} h_{jk} P_{jk}(\omega | \theta_{jk}) \geq \sum_{k=1}^{K_j} h_{kk} P_{kk}(x_{d_{jk}} | \theta_{jk}); j = (1, \ldots, K); \Omega_j$ is the $j$th component determined by the posterior probability. $N_j$ is the number of observations belonging to the $j$th component. Clearly, $N_j$ satisfies $\sum_{j=1}^{K_j} N_j = N$. $d_j = (d_{j1}, d_{j2}, \ldots, d_{jN_j})$ is the ordering of sample indices of the $j$th component. By sorting and discarding each component individually with the same trimming fraction $\eta$, we can obtain the subsample of each component $X_{M_j} = \{x_{d_{jm}}\}_{m=1}^{M_j}$, where $M_j = N_j(1 - \eta)$. Hence, the total subsample $X_M$ can be expressed by the union of the subsample of each component $X_{M_j}$,

$$
X_M = X_{M_1} \cup X_{M_2} \cup \cdots \cup X_{M_K}.
$$

(40)

Finally, the parameters of the proposed mixture model can be estimated with subsample $X_M$ and objective function $Q(X_M | \Theta)$. In the proposed GGMM-TLE, by evaluating the interval integral rather than log-likelihood value of the observations, we can obtain superior performance compared with the classical FAST-TLE. This is because we can order the observations within every individual label to retain the consistency of trimming proportion of each label. Therefore, regardless of the mixture weights and sample scales, all observations of each label are equally considered. Finally, combined with the steps of GGMM in Section 2, we summarize the procedures of GGMM-TLE as follows.

**Step 1.** Input the trimming fraction $\eta$. Initialize the parameter set $\Theta = \{\pi_j, h_{j1}, \theta_{j1}, h_{j2}, \theta_{j2}\}$, where $\theta_{j1} = \{v_j, k_j, \sigma_j\}$ and $\theta_{j2} = \{u_j, \Sigma_j\}$.

**Step 2.** Based on the current parameter set $\Theta^{(t)}$, evaluate the posterior probability $z_{ij}$ using (22), compute the variable $y_{ij}$ using (23), and classify the observations.

**Step 3.** Perform component-based confidence level ordering using (39) to obtain subsample $X_M^{(t+1)}$. 

Step 4. Compute the objective function \( Q(X_M^{(t+1)} \mid \Theta^{(t)}) \) in terms of (25). If \( Q(X_M^{(t+1)} \mid \Theta^{(t)}) \geq Q(X_M^{(t)} \mid \Theta^{(t)}) \), continue to Step 5; else, increase the value of the trimming fraction below the predefined threshold and obtain new subsample \( X_M^{(t+1)} \) until the following condition is satisfied: \( Q(X_M^{(t+1)} \mid \Theta^{(t)}) \geq Q(X_M^{(t)} \mid \Theta^{(t)}) \); set \( M = M^* \). If the condition is not satisfied, terminate the procedure.

Step 5. Update the prior probability \( \pi_j \) and weighting factor \( h_{jk} \) using (27) and (29), respectively. Compute the power parameter \( \nu_j \), shape parameter \( k_j \), scale parameter \( \sigma_j \), mean parameter \( u_j \), and the covariance parameter \( \Sigma_j \) by solving (31), (33), (35), (37), and (38), respectively.

Step 6. Maximize the objective function \( Q(X_M^{(t+1)} \mid \Theta^{(t)}) \) using (25) and obtain the new parameter set \( \Theta^{(t+1)} \). If the termination condition is satisfied, end the iterations. Otherwise, set \( t = t + 1 \) and return to Step 2.

6. Experimental Results

This section experimentally evaluates the proposed GGMM-TLE by considering the problem of real-world image segmentation and compares GGMM-TLE with other related algorithms. All algorithms are initialized using \( k \)-means. The experiments were developed in MATLAB R2012b and were executed on a personal computer with Intel(R) Core(TM) i7-6500U CPU @ 2.5 GHz, 8 GB RAM, 64-bit. To obtain an objective evaluation of the proposed method, this paper uses two measure criteria: the misclassification ratio (MCR) [24] and Dice similarity coefficient (DSC) [25]. The former has the following form:

\[
\text{MCR} = \frac{\text{number of mis-segmented pixels}}{\text{total number of pixels}},
\]

MCR is widely used in the literature to evaluate segmentation performance. For MCR, the smaller the value of MCR, the higher the accuracy of the segmentation. The popular overlap-based metric DSC is also employed to evaluate the proposed mixture model.

\[
\text{DSC}(S_a, S_m) = \frac{2 |S_a \cap S_m|}{|S_a| + |S_m|},
\]

where \( S_a \) denotes the shape of the automatic segmentation and \( S_m \) indicates the shape of the manual segmentation obtained from the algorithm output. The range of DSC is from zero to one, with one denoting ideal segmentation and zero indicating poor segmentation.

6.1. Test of the Proposed Trimming Approach. The first experiment presented herein validates the behaviour of the proposed GGMM-TLE. For this purpose, we generated three labels of inlier observations and one label of outliers. The inlier observations consisted of 10,000 points from a 3-component bivariate GMM with prior probability \( \pi_1 = \pi_2 = \pi_3 = 1/3 \). The means and variances of this bivariate GMM are

\[
\begin{align*}
\mu_1 &= (0, 4)^T, \\
\mu_2 &= (4, 0)^T, \\
\mu_3 &= (-4, 4)^T, \\
\Sigma_1 &= \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}, \\
\Sigma_2 &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \\
\Sigma_3 &= \begin{pmatrix} 1 & -0.5 \\ -0.5 & 3 \end{pmatrix}.
\end{align*}
\]

Labels 1, 2, and 3 have 2,500, 3,500, and 4,000 points, respectively. Random noise with 1,000 points was added in the ±10 rectangle. This random noise is considered as outliers. For comparison, apart from the proposed approach, we also included the performance of FAST-TLE [20] and CLO-TLE [21]. The trimming fraction \( \eta \) varied in the range [0.1, 0.5]. The segmentation results of the different methods are presented in Figure 1. From Figure 1, we can observe that the classical GMM was sensitive to outliers resulting in unstable performance in terms of visual interpretation. It is clear that the performance of the FAST-TLE method was less influenced by the outliers. However, it was also unstable, especially when the trimming fraction was high. The CLO-TLE ordering strategy exhibited a higher stability versus the previous two algorithms with the use of confidence level ordering. However, there remained misclassified outliers demonstrating that the fitting influence of CLO-TLE was not ideal. Conversely, we determined that the best performance was with the proposed GGMM-TLE, this is because every observation of the individual groups was considered. This figure indicates that the GGMM-TLE can extract clear point accumulations from noise data.

6.2. Segmentation of Noise-Degraded Images. To demonstrate the feasibility of GGMM-TLE, the following experiment used four real-world images (“Boat,” “Cow,” “House,” and “Man”) from the semantic boundaries dataset (SBD) [26] for comparison. These images were segmented into three labels. All these images were contaminated by salt and Pepper noise with intensity 5%. Figure 2 presents the visualization of the segmentation task with trimming fraction 0.2, where the second, third, and fourth columns correspond to the FAST-TLE, CLO-TLE, and GGMM-TLE algorithms, respectively. Figure 2 shows the detailed parts of the corresponding segmentation results using different approaches. Note that the GGMM-TLE eliminates obviously the noise as predicated. We demonstrate the log-likelihood function versus the number of iteration under trimming fractions 0.2 for the different test images in Figure 3. It can be clearly observed that the log-likelihood functions of FAST-TLE and CLO-TLE are similar.
to that of the proposed method. However, a closer inspection of the iteration ranges [5, 15] indicates that the GGMM-TLE method can moderately improve the convergence rate. When the iterations is low ($t \leq 5$), the convergence rate with GGMM-TLE is the biggest one. In the general case, the GGMM-TLE method converges after five iterations. In Figure 4, the MCR plots of each test image against different trimming fractions are displayed. This figure implies that the proposed scheme achieved superior segmentation accuracy as the basic scheme because the MCR of the GGMM-TLE method was the least of all test images.

The proposed algorithm was also assessed on a clinical MR image to label the white mass (WM) and grey mass (GM). For this purpose, a real MR image, slice 42 of IBSR2 from the IBSR dataset [27], was randomly selected to evaluate the performance of the proposed GGMM-TLE against FAST-TLE and CLO-TLE. Salt and Pepper noise with intensity 5% and 10% was considered in our experiment. Figure 5 presents the performance of these methods under 5% Salt and Pepper noise and different trimming fractions. It is clear that FAST-TLE did not demonstrate improved results for heavier outliers in the segmentation task. The CLO-TLE
tended to achieve superior performance with an increase of the trimming fraction and could maintain its stability and effectiveness. A closer inspection of Figure 5 indicates that the segmentation accuracy of GGMM-TLE was visually higher than the other methods. It is due to the fact that the proposed GGMM-TLE utilizes the advantages of confidence level for these observations so that the effects of mixture weights and sample scales are eliminated. Therefore, with the increasing of the trimming fractions, GGMM-TLE exhibits better stability to outliers than CLO-TLE. As shown in Figure 5, this is especially apparent in GGMM-TLE with high trimming fractions. Figure 6 displays the evaluation results using the MCR metric. It can be observed that the GGMM-TLE had the lowest MCR value; thus, its segmentation results were superior to those of FAST-TLE and CLO-TLE.

Further, we executed these algorithms 20 times, each time with different initialization. Then, we computed the (average) performance in terms of the number of correctly classified data points and the DSC for this MR image, including white matter and grey matter. Table 1 lists the mean values and the
standard deviations of the DSC obtained from 20 executions. The experiment results demonstrate that the accuracy was moderately improved compared with the other methods.

To assess the robustness of the proposed GGMM-TLE at different levels of noise, a set of real-world images from the Berkeley image dataset [28] was considered to compare the performance of GMM [29], SMM [30], GTMM [31], NSMM [12], and ACAP [8]. The ground-truth information was freely obtained from the website [31]. This was used for algorithm performance evaluation. The experiment was performed with noisy version images by adding Gaussian noise (zero mean, 0.01 variance) and Salt and Pepper noise (3%) to the images, as indicated in the first row of Figures 7 and 8. The evaluated algorithms were initialized using the $k$-means algorithm. The number of label $K$ was set according to human visual inspection. Figures 7 and 8 exhibit the results of image segmentation using the different methods. Owing to the application of a mean filter, we can observe that the performance of ACAP was superior to GMM, SMM, GTMM, and NSMM. The results generated by the ACAP achieved
Figure 4: Plot of MCR of test images against different trimming fractions. (a) Test image "Boat," (b) test image "Cow," (c) test image "House," and (d) test image "Man."

Table 1: Average DSC of different methods with diverse trimming fractions. Original MR image with 5% Salt and Pepper noise (mean ± standard deviation).

<table>
<thead>
<tr>
<th>Methods</th>
<th>Trimming fraction (WM)</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>FAST-TLE</td>
<td>0.7153 ± 0.4191</td>
<td>0.7269 ± 0.4785</td>
<td>0.7081 ± 0.2083</td>
<td>0.6771 ± 0.4762</td>
<td>0.6743 ± 0.5216</td>
<td></td>
</tr>
<tr>
<td>CLO-TLE</td>
<td>0.8181 ± 0.2721</td>
<td>0.7968 ± 0.2814</td>
<td>0.7966 ± 0.3612</td>
<td>0.7748 ± 0.3264</td>
<td>0.7484 ± 0.3673</td>
<td></td>
</tr>
<tr>
<td>GGMM-TLE</td>
<td>0.8994 ± 0.1023</td>
<td>0.8380 ± 0.1571</td>
<td>0.8177 ± 0.1238</td>
<td>0.8040 ± 0.0779</td>
<td>0.7803 ± 0.2317</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Methods</th>
<th>Trimming fraction (GM)</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>FAST-TLE</td>
<td>0.8367 ± 0.4020</td>
<td>0.8378 ± 0.4060</td>
<td>0.8313 ± 0.3553</td>
<td>0.8015 ± 0.3723</td>
<td>0.7829 ± 0.4062</td>
<td></td>
</tr>
<tr>
<td>CLO-TLE</td>
<td>0.9107 ± 0.3841</td>
<td>0.8931 ± 0.2379</td>
<td>0.8944 ± 0.1540</td>
<td>0.8803 ± 0.2783</td>
<td>0.8593 ± 0.1761</td>
<td></td>
</tr>
<tr>
<td>GGMM-TLE</td>
<td>0.9561 ± 0.1262</td>
<td>0.9227 ± 0.1040</td>
<td>0.9095 ± 0.1235</td>
<td>0.9002 ± 0.1156</td>
<td>0.8837 ± 0.1240</td>
<td></td>
</tr>
</tbody>
</table>
Figure 5: (a) to (c) display the segmentation results of MR image (slice 42 of IBSR2) using FAST-TLE, CLO-TLE, and GGMM-TLE, respectively. The trimming fractions from left to right are 0.1, 0.2, 0.3, 0.4, and 0.5.

Figure 6: Trimming fractions versus classification accuracy under noise environment, MCR of different methods for MR image (slice 42 of IBSR2). (a) 5% Salt and Pepper noise and (b) 10% Salt and Pepper noise.
similar results to GGMM-TLE; however, its performance was impaired when there was an abundance of rich details, for example, in test image 241004 (the sixth row of Figure 7). The GGMM-TLE provided a moderately improved performance under different noisy conditions and eliminated the influence of widely spread noise data. This characteristic is endemic to MRF and trimmed likelihood estimator. The resulting DSC is reported in Tables 2 and 3, providing a quantitative comparison among the algorithms. The DSC and standard deviation indicate that the proposed method outperformed the other methods by preserving the highest DSC.

To further demonstrate the goodness of GGMM-TLE against different noise, in Figure 9, we display the mean values and standard deviations of the MCR obtained from twenty runs on two Berkeley test images (24063 and 35010) under different noise environments. Considering the MCR, on average, the ACAP effectively eliminated the effects of noise during the segmentation processing and demonstrated acceptable segmentation results. We determined that classical GMM, SMM, and GΓMM were severely influenced by Gaussian noise and could not accurately separate a region from the background. In the majority of cases, the NSMM
7. Concluding Remarks

In this paper, a robust estimation of the proposed GGMM-TLE using a trimmed likelihood estimator for real-world image segmentation was proposed. GGMM-TLE with MRF implements a mixture of generalized Gamma and Gaussian distributions.

The main contribution of this paper is the presentation of an asymmetric finite model GGMM-TLE based on MRF. With this model, we have high flexibility to fit different shapes of observed data. Further, this study discussed the property of identifiability of the proposed mixture model, guaranteeing that the estimation procedure for the parameters was correctly developed. Then, to ensure that GGMM-TLE was robust against heavy outliers, the paper offered an effective method to discard the outliers in advance, and

approach was superior to SMM and GMM, yet continued to be influenced by varying degrees of Salt and Pepper and Gaussian noise. As expected, compared to the other algorithms, the GGMM-TLE was stable and achieved the best segmentation results according to the quantitative criterion.
Figure 9: Average MCR for different methods on two test images under noisy environment (Berkeley Dataset). ((a) and (b)) Test image 24063; ((c) and (d)) test image 35010.

therefore, GGMM-TLE demonstrated superior performance under modelling with samples contaminated with unknown outliers. Finally, combined with MRF, GGMM-TLE considered the spatial relationship between neighbouring pixels and demonstrated a stronger ability to resist different noise. The segmentation results of synthetic data and real-world images confirmed that the proposed method demonstrated superior competitiveness. The main limitation of this algorithm is that the segmentation task requires component-based confidence level ordering, which increases the computational cost.

As a future work, one direction is to obtain other finite mixture models by testing different probability density functions. Another possible direction is to extend the presented method to a higher dimension in a straightforward manner such as fMRI time-series clustering. We plan to address these topics in a separate paper.
Table 2: Average DSC of different methods for Berkeley test images with zero mean Gaussian noise, variance 0.01 (mean ± standard deviation).

<table>
<thead>
<tr>
<th>Images</th>
<th>Label</th>
<th>GMM</th>
<th>SMM</th>
<th>G-MM</th>
</tr>
</thead>
<tbody>
<tr>
<td>24063</td>
<td>3</td>
<td>0.86005 ± 0.32173</td>
<td>0.89173 ± 0.20533</td>
<td>0.88071 ± 0.20472</td>
</tr>
<tr>
<td>8068</td>
<td>3</td>
<td>0.85737 ± 0.28634</td>
<td>0.87439 ± 0.27448</td>
<td>0.86005 ± 0.31348</td>
</tr>
<tr>
<td>241004</td>
<td>4</td>
<td>0.85388 ± 0.32745</td>
<td>0.89018 ± 0.24834</td>
<td>0.87803 ± 0.20895</td>
</tr>
<tr>
<td>55067</td>
<td>4</td>
<td>0.87931 ± 0.21733</td>
<td>0.88342 ± 0.21631</td>
<td>0.88013 ± 0.30556</td>
</tr>
<tr>
<td>35010</td>
<td>4</td>
<td>0.84904 ± 0.33850</td>
<td>0.88463 ± 0.23632</td>
<td>0.87991 ± 0.18584</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Images</th>
<th>Label</th>
<th>NSMM</th>
<th>ACAP</th>
<th>GGMM-TLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>24063</td>
<td>3</td>
<td>0.91825 ± 0.15734</td>
<td>0.95004 ± 0.14848</td>
<td>0.97122 ± 0.14891</td>
</tr>
<tr>
<td>8068</td>
<td>3</td>
<td>0.89173 ± 0.28604</td>
<td>0.94485 ± 0.16055</td>
<td>0.96108 ± 0.10723</td>
</tr>
<tr>
<td>241004</td>
<td>4</td>
<td>0.92006 ± 0.21634</td>
<td>0.94670 ± 0.13114</td>
<td>0.96108 ± 0.14826</td>
</tr>
<tr>
<td>55067</td>
<td>4</td>
<td>0.92601 ± 0.17203</td>
<td>0.96893 ± 0.12051</td>
<td>0.97175 ± 0.13876</td>
</tr>
<tr>
<td>35010</td>
<td>4</td>
<td>0.92183 ± 0.20847</td>
<td>0.95380 ± 0.13954</td>
<td>0.96219 ± 0.15664</td>
</tr>
</tbody>
</table>

Table 3: Average DSC of different methods for Berkeley test images with 3% Salt and Pepper noise (mean ± standard deviation).

<table>
<thead>
<tr>
<th>Images</th>
<th>Label</th>
<th>GMM</th>
<th>SMM</th>
<th>G-MM</th>
</tr>
</thead>
<tbody>
<tr>
<td>24063</td>
<td>3</td>
<td>0.87126 ± 0.30566</td>
<td>0.90972 ± 0.23385</td>
<td>0.88482 ± 0.18808</td>
</tr>
<tr>
<td>8068</td>
<td>3</td>
<td>0.85234 ± 0.32531</td>
<td>0.88082 ± 0.30318</td>
<td>0.87627 ± 0.33601</td>
</tr>
<tr>
<td>241004</td>
<td>4</td>
<td>0.87294 ± 0.35061</td>
<td>0.88385 ± 0.30458</td>
<td>0.89714 ± 0.25054</td>
</tr>
<tr>
<td>55067</td>
<td>4</td>
<td>0.88099 ± 0.32392</td>
<td>0.90400 ± 0.25767</td>
<td>0.88517 ± 0.21092</td>
</tr>
<tr>
<td>35010</td>
<td>4</td>
<td>0.85337 ± 0.25011</td>
<td>0.89075 ± 0.21839</td>
<td>0.87630 ± 0.16432</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Images</th>
<th>Label</th>
<th>NSMM</th>
<th>ACAP</th>
<th>GGMM-TLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>24063</td>
<td>3</td>
<td>0.92079 ± 0.16049</td>
<td>0.96103 ± 0.23214</td>
<td>0.98394 ± 0.07514</td>
</tr>
<tr>
<td>8068</td>
<td>3</td>
<td>0.90213 ± 0.25321</td>
<td>0.95329 ± 0.11539</td>
<td>0.97328 ± 0.11482</td>
</tr>
<tr>
<td>241004</td>
<td>4</td>
<td>0.93487 ± 0.16806</td>
<td>0.96271 ± 0.12378</td>
<td>0.97117 ± 0.13683</td>
</tr>
<tr>
<td>55067</td>
<td>4</td>
<td>0.93247 ± 0.18579</td>
<td>0.98076 ± 0.10690</td>
<td>0.98587 ± 0.06655</td>
</tr>
<tr>
<td>35010</td>
<td>4</td>
<td>0.93752 ± 0.17804</td>
<td>0.96053 ± 0.10842</td>
<td>0.97420 ± 0.12772</td>
</tr>
</tbody>
</table>

Conflicts of Interest
The authors declare that they have no conflicts of interest.

Acknowledgments
This work was supported by the National Natural Science Foundation of China under Grant no. 61371150.

References


