

Article

SAR Images Statistical Modeling and Classification Based on the Mixture of Alpha-Stable Distributions

Yijin Peng ¹, Jiayu Chen ¹, Xin Xu ^{1,*} and Fangling Pu ²

¹ Signal Processing Laboratory, School of Electronic Information, Wuhan University, Wuhan 430072, China; E-Mails: pengyijin@whu.edu.cn (Y.P.); jiayu.chen@ieee.org (J.C.)

² Wireless Communication and Sensor Network Laboratory, School of Electronic Information, Wuhan University, Wuhan 430072, China; E-Mail: flpu@whu.edu.cn (F.P.)

* Author to whom correspondence should be addressed; E-Mail: xinxu@whu.edu.cn; Tel.: +86-27-6875-2836.

Received: 14 March 2013; in revised form: 23 April 2013 / Accepted: 24 April 2013 /

Published: 3 May 2013

Abstract: This paper proposes the mixture of Alpha-stable (MAS) distributions for modeling statistical property of Synthetic Aperture Radar (SAR) images in a supervised Markovian classification algorithm. Our work is motivated by the fact that natural scenes consist of various reflectors with different types that are typically concentrated within a small area, and SAR images generally exhibit sharp peaks, heavy tails, and even multimodal statistical property, especially at high resolution. Unimodal distributions do not fit such statistical property well, and thus a multimodal approach is necessary. Driven by the multimodality and impulsiveness of high resolution SAR images histogram, we utilize the mixture of Alpha-stable distributions to describe such characteristics. A pseudo-simulated annealing (PSA) estimator based on Markov chain Monte Carlo (MCMC) is present to efficiently estimate model parameters of the mixture of Alpha-stable distributions. To validate the proposed PSA estimator, we apply it to simulated data and compare its performance to that of a state-of-the-art estimator. Finally, we exploit the MAS distributions and a Markovian context for SAR images classification. The effectiveness of the proposed classifier is demonstrated by experiments on TerraSAR-X images, which verifies the validity of the MAS distributions for modeling and classification of SAR images.

Keywords: synthetic aperture radar; mixture of Alpha-stable distributions; parameter estimation; statistical modeling; classification

1. Introduction

Synthetic aperture radar (SAR) works in the microwave band and has the ability to operate under all weather conditions. SAR images have been increasingly used in different areas [1–3]. One fundamental technique for interpretation of SAR images is classification, which is of great importance for land planning, natural risk prevention, target recognition, *etc.* However, SAR image classification is still a tough work due to the existence of intense speckle. Different classification approaches have been considered in both supervised and unsupervised contexts, based on Markov random field [4], neural networks [5] and other methods [6]. We focus on supervised classification of SAR images based on its statistical property in Markovian framework, in which a proper statistical distribution must be adopted to model SAR images data.

SAR images generally exhibit heavy-tailed and sharp-peaked statistical properties. Many distributions, e.g., Gamma ([7]; Chapter 4), Weibull [8,9], Log-normal ([7]; Chapter 5), [10], K [11,12], generalized Gamma [13], Fisher [14], G^0 [15], Alpha-stable and heavy-tailed Rayleigh [16,17], and recently proposed HG^0 [18], have been applied to the statistical modeling of SAR images. Gao [19] summarizes the development history and the current researching state of statistical modeling and discusses relevant issues in his work. However, as the resolution of the images increases, multimodal statistical properties become apparent due to the different types of backscatter from a single class. For instance, in an SAR image, building areas contain strong reflectors, textures, and shadows. In this case, the previously mentioned unimodal distributions fail to capture such multimodal statistical properties. Thus, a model with the ability to describe these attributes, as well as a heavy tail and sharp peak, is necessary. Alpha-stable distribution is a flexible model that can fit many other distributions into it. The appeal of Alpha-stable distribution also derives from theoretical reason, which states that stable processes arise as limiting processes of sums of independent identically distributed (i.i.d.) random variables via the generalized central limit theorem. Actually, the only possible nontrivial limit of normalized sums of i.i.d. terms is Alpha-stable. Moreover, Alpha-stable distribution has been used as a successful alternative for modeling non-Gaussian data and has also been applied to understand SAR images, e.g., for image restoration [17], object detection [20,21], image classification [22], and image fusion [23]. Applications of Alpha-stable distribution in other areas such as watermark detection, network traffic and stock returns have also been reported [24–26]. Driven by the multimodal and impulsive histogram of high resolution SAR image, as well as the theoretical appeal and successful applications of Alpha-stable family itself, we take the mixture of Alpha-stable distributions that can fit any situation for the statistical modeling of high resolution SAR images.

The main challenge in applying MAS distributions comes from the parameter estimation, as there is no analytical expression for the probability density function. Work on MAS distributions in the literature is very limited: Salas-Gonzalez and Kuruoglu [27] proposed a reversible jump Markov chain Monte Carlo (MCMC) method for MAS distributions with an unknown number of components in the Bayesian frame, but their estimations suffered from large deviations. The same authors introduced an estimator for a mixture of symmetric alpha-stable distributions with identical characteristic exponent for each component [28], and also an estimator for a mixture of skewed alpha-stable distributions with identical shape parameters for each component [29]. In both cases, they considered the number of

components to be unknown. Using Buckle's work [30] on the estimation of distribution parameters, Casarin [31] introduced a random auxiliary vector for MAS distributions with a fixed number of components. However, the algorithm necessary for this approach makes it difficult to draw samples for the auxiliary variable and is time consuming.

In this paper, we introduce a pseudo-simulated annealing (PSA) estimator for the MAS distributions. Experimental results on simulated data show that the PSA estimator obtains accurate parameter values for MAS distributions. We utilize the MAS distributions estimated by PSA for the statistical modeling and classification of SAR images. To verify the effectiveness of the mixture model, classification experiments are carried out using TerraSAR-X data. The supervised Bayesian maximum a posteriori (MAP) classifier is selected over MAS distributions in a Markov random field (MRF) framework. SAR images are represented by MAS distributions that are trained via the PSA estimator and appear as likelihood probabilities. In order to preserve spatial coherence, the MRF introduces prior information, which builds a restraint between pixels and their neighborhoods. The classification is then the solution to the maximization problem of the posterior probability. To solve this problem, we use the Graph Cuts method [32–34] to obtain a global optimization. The proposed classifier is validated using TerraSAR-X images.

The remainder of this paper is organized as follows. Section 2 introduces MAS distributions and the PSA estimator. Section 3 gives an explanation of the statistical modeling of SAR images, and Section 4 introduces the MAS-based MRF classification algorithm. Our experiments and results are presented in Section 5, and we summarize the conclusions in Section 6.

2. Mixture of Alpha-Stable Distributions

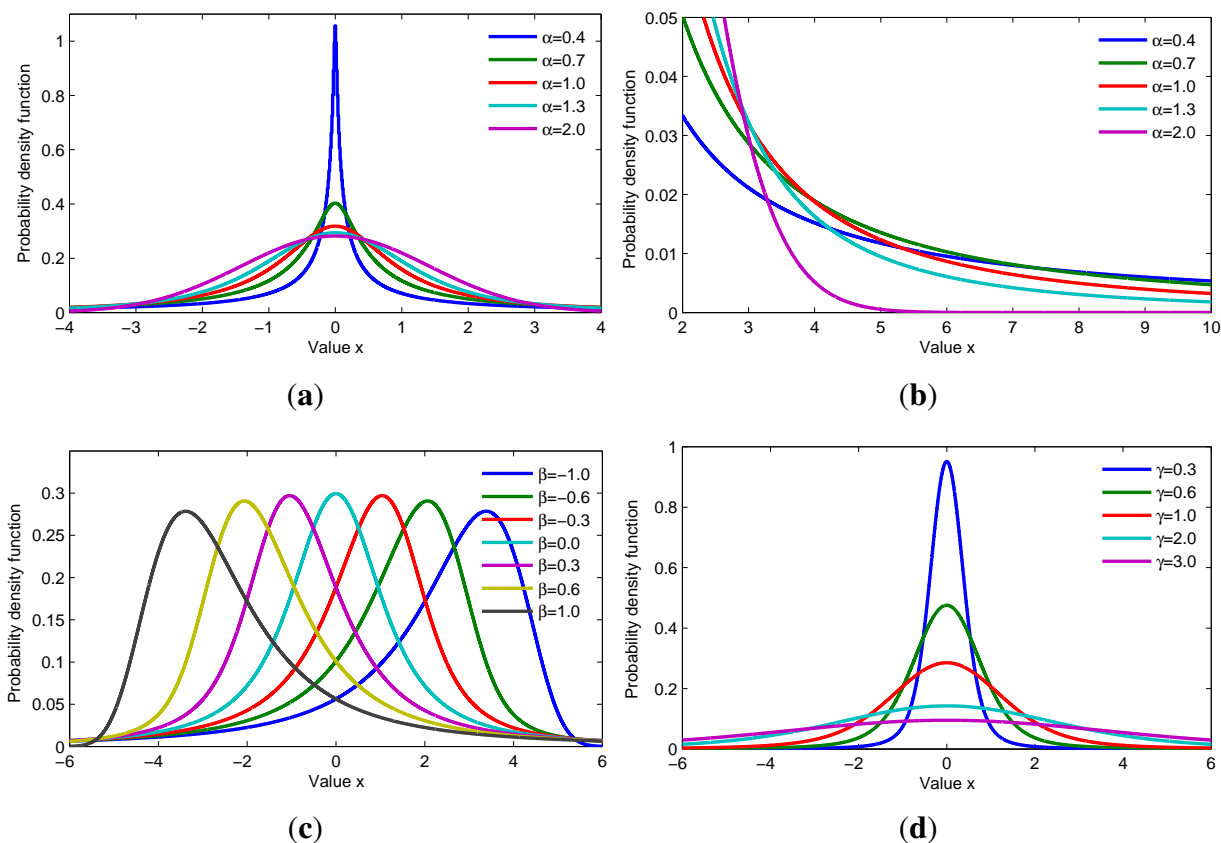
2.1. Alpha-Stable Distribution

The notion of stability was introduced by Levy [35] in his study of normalized sums of independent and identically distributed terms. Due to the lack of an analytical expression for the probability density function (PDF), Alpha-stable distributions are usually described in terms of the characteristic function as

$$\phi(\omega) = \begin{cases} \exp\{j\mu\omega - |\gamma\omega|^\alpha[1 - j \operatorname{sgn}(\omega)\beta \tan(\frac{\pi\alpha}{2})]\}, & \alpha \neq 1 \\ \exp\{j\mu\omega - |\gamma\omega|[1 + j \operatorname{sgn}(\omega)\beta \frac{2}{\pi} \ln(\omega)]\}, & \alpha = 1 \end{cases} \quad (1)$$

where $j = \sqrt{-1}$, “sgn” is the sign function. The four parameters of Alpha-stable distribution are $\alpha \in (0, 2]$, $\beta \in [-1, 1]$, $\gamma \in \mathbb{R}^+$ and $\mu \in \mathbb{R}$, representing the characteristic exponent, skewness parameter, dispersion parameter, and location parameter, respectively. The parameter α sets the level of impulsiveness (smaller α gives a more peaked PDF and a heavier tail); the parameter β controls the skewness of the PDF (symmetric if $\beta = 0$, negatively skewed if $\beta = -1$, and positively skewed if $\beta = 1$); the dispersion parameter and location parameter are similar to the variance and mean, respectively, in a Gaussian distribution. There are three special cases: the distribution reduces to a Gaussian distribution when $\alpha = 2$, a Cauchy distribution when $\alpha = 1$, $\beta = 0$, and a Levy distribution when $\alpha = 0.5$, $\beta = 1$. Figure 1 illustrates the Alpha-stable distributions given by various parameter values.

Figure 1. PDF of Alpha-stable distributions ($\mu = 0$). (a) Symmetric Alpha-stable case: $\alpha = [0.4, 0.7, 1.0, 1.3, 2.0]$, $\beta = 0.0$, $\gamma = 1.0$; (b) Enlargement of the tails in (a); (c) Skewness parameter: $\alpha = 1.2$, $\beta = [0.0, 0.2, 0.5, 0.8, 1.0]$, $\gamma = 1.0$; (d) Dispersion parameter: $\alpha = 1.6$, $\beta = 0.0$, $\gamma = [0.3, 0.6, 1.0, 2.0, 3.0]$.



Although there is no closed-form expression for the PDF of an Alpha-stable distribution, it is possible to obtain the PDF by applying a Fourier transformation to the characteristic function:

$$S_{\alpha}(x|\gamma, \beta, \mu) = \int_{-\infty}^{+\infty} \phi(\omega) \exp(-j\omega x) d\omega \quad (2)$$

However, it is difficult to compute Equation (2) because the characteristic function is complex and the interval of integration is infinite. Therefore, Equation (2) does not admit an analytical solution except in a few special cases. In this paper, we utilize a FFT-based approach [36] to obtain the PDF of the Alpha-stable distribution.

2.2. Mixture of Alpha-Stable (MAS) Distributions

Mixture models allow us to describe and estimate complex multimodal data by considering them as sampled from different subpopulations. The density of MAS distributions is given by

$$p(x) = \sum_{k=1}^K \omega_k S_{\alpha_k}(x|\gamma_k, \beta_k, \mu_k) \quad (3)$$

$$\forall k, 0 \leq \omega_k \leq 1, \text{ and } \sum_{k=1}^K \omega_k = 1$$

where ω_k is the weight of the k^{th} component. In a Bayesian scheme, we can estimate the distribution parameters via prior information and observations using the Bayesian rule

$$p(\theta|X) = \frac{p(X|\theta)p(\theta)}{p(X)} \propto p(X|\theta)p(\theta) \quad (4)$$

where $p(\theta|X)$ denotes the posterior probability of θ given an observation $X = (x_1, \dots, x_j, \dots, x_M)$, $p(\theta)$ is the prior probability, $p(X|\theta)$ is the likelihood of X given θ , and $p(X)$ is the prior probability of X .

2.3. PSA Estimator for MAS Distributions

The Bayesian approach is an efficient technique for the estimation of MAS distribution parameters. This technique has received a lot of attention recently thanks to the evolution of MCMC computational tools. We name this the PSA estimator.

When estimating its parameters, it is convenient to consider the mixture model as a missing data problem. We assume that the data vector $X = (x_1, \dots, x_j, \dots, x_M)$ has been randomly drawn from K subpopulations. For each variable x_j , let z_j be an allocation variable for the unobserved or missing variable that indicates to which component x_j belongs. Thus, z_j will be equal to 1 if x_j belongs to the k^{th} subpopulation, or 0 otherwise.

Each observation x_j is considered to be drawn from the k^{th} component described by $\theta_k = (\alpha_k, \beta_k, \gamma_k, \mu_k)$ with probability

$$p(z_j = k|x_j) = \frac{\omega_k S_{\alpha_k}(x_j|\gamma_k, \beta_k, \mu_k)}{\sum_{m=1}^K \omega_m S_{\alpha_m}(x_j|\gamma_m, \beta_m, \mu_m)} \quad (5)$$

The allocation variable z_j can be defined as follows

$$z_j = \arg \max_k p(k|x_j) \quad (6)$$

Parameters then can be updated based on $\{z_j\}$. The candidate parameter $\theta_k^{new} = (\alpha_k^{new}, \beta_k^{new}, \gamma_k^{new}, \mu_k^{new})$ is sampled from a proposal distribution $q(\cdot|\cdot)$, and is accepted with probability $A_{\theta_k^{new}}$, where

$$A_{\theta_k^{new}} = \min \left(1, \left(\prod_{\substack{j=1 \\ z_j=k}}^M \frac{S_{\alpha_k^{new}}(x_j|\gamma_k^{new}, \beta_k^{new}, \mu_k^{new})}{S_{\alpha_k^{old}}(x_j|\gamma_k^{old}, \beta_k^{old}, \mu_k^{old})} \right)^{\frac{1}{T(t)}} \times \frac{p(\theta_k^{new})q(\theta_k^{old}|\theta_k^{new})}{p(\theta_k^{old})q(\theta_k^{new}|\theta_k^{old})} \right) \quad (7)$$

In Equation (7), $T(t)$ is the temperature function, which cools as the iteration proceeds. The introduction of $T(t)$ can prevent the chain from trapping in small area, especially in case the target distribution may have multiple peaks. $p(\theta)$ is the prior function for parameter vector θ . In the case of an independent prior for each parameter, $p(\theta)$ can be expressed in terms of product $p(\theta) = p(\alpha)p(\beta)p(\gamma)p(\mu)$. We take the prior distributions of α , β , γ and μ to be uniform, uniform,

inverse gamma ($IG(\cdot|\cdot)$) and normal ($N(\cdot|\cdot)$), respectively. In addition, we use a symmetric proposal satisfying $q(\theta_k^{new}|\theta_k^{old}) = q(\theta_k^{old}|\theta_k^{new})$ for simplicity. The acceptance probability $A_{\theta_k^{new}}$ can then be simplified as

$$A_{\theta_k^{new}} = \min \left(1, \left(\prod_{\substack{j=1 \\ z_j=k}}^M \frac{S_{\alpha_k^{new}}(x_j|\gamma_k^{new}, \beta_k^{new}, \mu_k^{new})}{S_{\alpha_k^{old}}(x_j|\gamma_k^{old}, \beta_k^{old}, \mu_k^{old})} \right)^{\frac{1}{T(t)}} \times \frac{IG(\gamma_k^{new}|a_0, b_0)N(\mu_k^{new}|\xi, \kappa)}{IG(\gamma_k^{old}|a_0, b_0)N(\mu_k^{old}|\xi, \kappa)} \right) \quad (8)$$

where a_0 and b_0 are parameters of the inverse gamma prior for γ , and κ and ξ are parameters of the normal prior for μ . Moreover, a normal proposal distribution $q(\cdot|\cdot) = N(\cdot|\delta, \sigma)$ is utilized in our algorithm, and its parameters are adaptively updated using estimations obtained in previous iterations: δ is set to the estimation of the previous iteration, and σ is set to the standard deviation of the previous L estimations. Adaptively updating the parameters of the proposal distribution ensures that new candidates can properly explore the entire parameter space, which further ensures that estimations rapidly converge to the true values.

The PSA estimator is described in Algorithm 1. The input includes observation $X = (x_1, x_2, \dots, x_M)$, the maximum iteration number $Iter$, the number of components K , starting temperature $T(0)$, the value of L for updating the parameters of the proposal distribution, and the initial parameters θ_k^0, ω_k^0 for each component.

Algorithm 1 PSA estimator for parameters of MAS distributions

- 1: **Input:**
 $X = \{x_j\}, \theta_k^0 = (\alpha_k^0, \beta_k^0, \gamma_k^0, \mu_k^0), \omega_k^0, k = 1, 2, \dots, K, L, T(0), Iter$
 - 2: **for** each $t < Iter$ **do**
 - 3: Decrease temperature $T(t) = \frac{T(0)}{\lg(t+1)}$
 - 4: Assign initial parameters $\theta_k^{old} = \theta_k^{t-1}, \omega_k^{old} = \omega_k^{t-1}, k = 1, 2, \dots, K$
 - 5: For each data sample x_j , obtain allocation variable z_j using Equation (6)
 - 6: Update parameters of the proposal distribution $q(\cdot|\cdot) = N(\cdot|\delta, \sigma)$: set δ to value of the previous iteration and set σ to the standard deviation of the previous L estimations if $t > L$, otherwise set σ to its initialization value
 - 7: Sample new candidates $\theta_k^{new} = (\alpha_k^{new}, \beta_k^{new}, \gamma_k^{new}, \mu_k^{new})$ from proposal distribution $q(\cdot|\cdot) = N(\cdot|\delta, \sigma)$ for each component
 - 8: Accept θ_k^{new} according to Equation (8) and set $\theta_k^t = \theta_k^{new}$, otherwise set $\theta_k^t = \theta_k^{t-1}$
 - 9: Obtain weight $\omega = (\omega_1, \dots, \omega_k, \dots, \omega_K)$ of each component as in [27] by drawing samples from distribution $\omega \sim D$, where $D(\zeta + n_1, \dots, \zeta + n_k, \dots, \zeta + n_K)$ is the Dirichlet distribution with $\zeta > 0$, and n_k is the number of samples assigned to the k^{th} component
 - 10: **end for**
 - 11: **Output:**
 $\theta_k = (\alpha_k, \beta_k, \gamma_k, \mu_k), \omega_k, k = 1, 2, \dots, K$
-

The critical differences between PSA estimator and Salas-Gonzalez's method [27] are twofold. Firstly, PSA takes a cooling schedule, introducing $T(t) = \frac{T(0)}{\lg(t+1)}$ over the acceptance probability. This

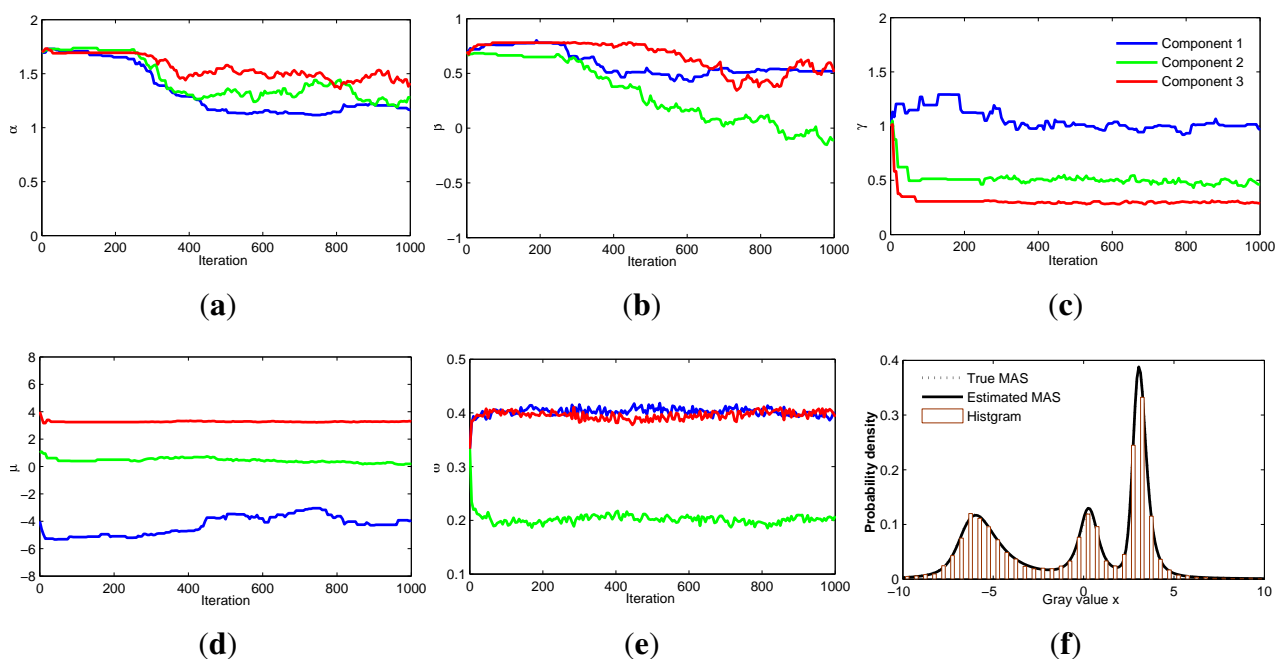
ensures a reasonable probability of a downhill move, thus preventing the estimation from being trapped in small region of the parameter space for long periods of time. When the chain is trapped in small region, it is worthless to the estimation but will slow down the global convergence of the estimation. Secondly, in PSA estimator, parameters of the proposal distribution are adaptively updated by using the former estimations, which makes the proposal distribution increasingly closer to the distribution of the true parameter as the estimation procedure proceeds.

2.4. Simulation Result for PSA Estimator on MAS Distributions

To illustrate the effectiveness of the PSA estimator, we conduct an experiment using the MAS model shown in Equation (9). Random samples are generated using the algorithm proposed in [37].

$$\begin{aligned} p(x) &= \omega_1 S_{\alpha_1}(x|\gamma_1, \beta_1, \mu_1) + \omega_2 S_{\alpha_2}(x|\gamma_2, \beta_2, \mu_2) + \omega_3 S_{\alpha_3}(x|\gamma_3, \beta_3, \mu_3) \\ &= 0.4S_{1.2}(x|1.0, 0.5, -4.25) + 0.2S_{1.2}(x|0.5, 0.0, 0.3) + 0.4S_{1.5}(x|0.3, 0.5, 3.25) \end{aligned} \quad (9)$$

Figure 2. Simulation results for the PSA estimator. (a)–(e) Evolution procedure of the three components for parameters α , β , γ , μ and ω , respectively (red: component 1, green: component 2, blue: component 3); (f) Fitting results of the estimated MAS distributions to simulated data histogram and the PDF curve obtained by the true parameters.



As stated in Section 2.3, the priors are uniform for α and β , inverse Gamma for γ , and normal for μ . Initial values of the MAS distribution parameters are empirically chosen as follows: $\alpha_{1\sim 3} = [1.7, 1.7, 1.7]$, $\beta_{1\sim 3} = [0.7, 0.7, 0.7]$, $\gamma_{1\sim 3} = [1.0, 1.0, 1.0]$, $\mu_{1\sim 3} = [-4.0, 1.0, 4.0]$. The parameters of inverse gamma prior for γ are $a_0 = 2$ and $b_0 = 3$, and those for the normal prior for μ are $\xi = 0$ and $\kappa = 6$. $T(0)$ and L are set to 15 and 200, respectively. Experimental results for the PSA estimation with $Iter = 1000$ iterations are displayed in Figure 2(a–e). These show that the PSA estimator has a good convergence procedure. Figure 2(f) displays the discrete histogram of the originally simulated data, the

PDF curve of the estimated MAS distributions with three components and the PDF curve of the true MAS distributions. This illustrates that the estimated MAS distributions fit the simulated data well.

To quantitatively evaluate the performance of the PSA estimator, we list in Table 1 true values, starting values, estimated values, and standard deviation (std) by PSA estimator. Results of Salas-Gonzalez's method [27] are also included for comparison. Details of the initial conditions and implementation for Salas-Gonzalez's method were reported in his work [27].

Table 1. Simulation results for the PSA estimator on MAS distributions.

Parameter		PSA			Salas-Gonzalez	
Name	True	Starting	Estimated	std	Estimated	std
α_1	1.20	1.70	1.20	0.02	1.27	0.09
β_1	0.50	0.70	0.53	0.01	0.65	0.08
γ_1	1.00	1.00	1.01	0.02	0.98	0.06
μ_1	−4.25	−4.00	−4.11	0.19	−4.30	0.60
ω_1	0.40	0.33	0.40	0.01	0.40	0.02
α_2	1.20	1.70	1.28	0.07	1.30	0.17
β_2	0.00	0.70	−0.03	0.07	0.04	0.30
γ_2	0.50	1.00	0.47	0.02	0.45	0.05
μ_2	0.30	1.00	0.25	0.07	0.40	0.30
ω_2	0.20	0.33	0.20	0.01	0.20	0.02
α_3	1.50	1.70	1.45	0.04	1.37	0.12
β_3	0.50	0.70	0.51	0.08	0.34	0.20
γ_3	0.30	1.00	0.30	0.01	0.30	0.02
μ_3	3.25	4.00	3.28	0.02	3.24	0.06
ω_3	0.40	0.33	0.40	0.01	0.40	0.02

Table 1 shows that 9 of the 15 values estimated by PSA are closer to true values than the estimates given by Salas-Gonzalez's method. The estimations of the two methods for ω_1 , ω_2 , ω_3 and γ_3 are similar and the estimations of PSA for two other parameters μ_1 and μ_3 are slightly worse than those of Salas-Gonzalez's method. In addition, the standard deviation obtained by the proposed estimator is smaller than those of Salas-Gonzalez's method. These results demonstrate the effectiveness of the proposed PSA estimator. Therefore, it can be concluded that the proposed PSA estimator is accurate for MAS distributions.

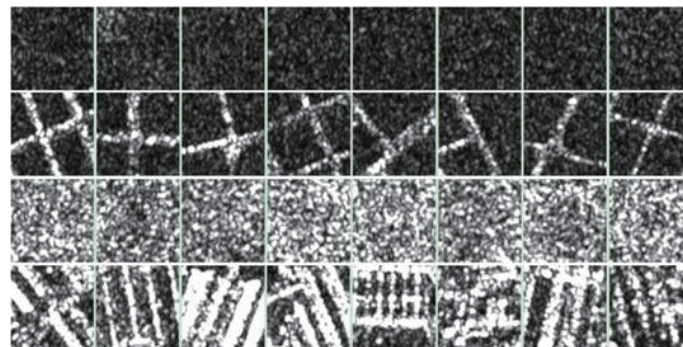
3. MAS-Based Statistical Modeling of SAR Images

A statistical model offers a description of SAR images' statistical properties. The quality of classification depends on how well the statistical model agrees with the statistical properties. As mentioned in Section 1, many different models for SAR images exist, but these do not provide a good fit. The high probability of high intensities, produced by a large number of strong reflectors, leads to a heavy tail with a rather slow discrepancy in the SAR image histogram. In addition, each class of SAR image, especially at high resolutions, is usually a combination of different objects. For example, building areas

generally contain strong reflectors, textures, and shadows. Such images present multimodal statistical properties; therefore, we expect that a multimodal distribution will be required to describe them.

We first analyze the performance of existing statistical models using TerraSAR-X image samples representing four different classes (river, marsh, farmland, and buildings). Each class contains 600 samples of $64 \text{ pixels} \times 64 \text{ pixels}$. Figure 3 displays eight samples from each class.

Figure 3. Eight samples from each class (from top to bottom: river, marsh, farmland and building).



Kolmogorov–Smirnov (KS) test is applied to evaluate whether a particular model is able to describe the statistical properties of the SAR images. We calculate the KS distance (KSD) between the empirical distribution function (EDF) and the cumulative distribution function (CDF) of the reference distribution. The KSD is defined by Equation (10):

$$KSD = \max_n |F(x_n) - \hat{F}(x_n)| \quad (10)$$

where $F(x_n)$ is the CDF and $\hat{F}(x_n)$ is the EDF. In addition, we also considered the acceptance probability, which is defined as the ratio of the number of samples that the KS test accepts for a distribution (at the 5% significance level) to the total number of samples in a class. The evaluation results are shown in Tables 2 and 3.

Table 2 shows that some models may be suitable for one class but not for another. Gamma distributions are not good for modeling marsh areas and building areas, because the corresponding average KSD s are particularly large. The Weibull and K distributions do not fit marsh areas well, whereas the G^0 distribution can model marsh and building, but not river or farmland. This indicates that G^0 distribution is suitable for modeling heterogeneous but not for homogeneous images. The Alpha-stable distribution is robust for various SAR image classes, as its average KSD s are not larger than 0.046, and it provides the best model for marsh areas. In addition, the maximum KSD s of the Alpha-stable distribution are the smallest of the different distributions tested. However, the average KSD s of the Alpha-stable distribution are still large. Similar results are also displayed in Table 3, which shows that Gamma distribution and Weibull distribution have a high probability of acceptance for modeling river and farmland, whereas they are always rejected for modeling the other two classes. The K distribution is always rejected when modeling marsh areas. The Alpha-stable distribution has a high probability of acceptance, but is not always accepted for modeling all four classes. These results

indicate that unimodal distributions are not sufficient for modeling high-resolution SAR images due to the multimodal statistical properties inherent in such images.

Table 2. Kolmogorov–Smirnov distance (KSD) of different statistical models.

Model	Average KSD (Maximum KSD)			
	River	Marsh	Farmland	Building
Gamma	0.042 (0.213)	0.205 (0.427)	0.044 (0.172)	0.280 (0.590)
Weibull	0.058 (0.207)	0.109 (0.171)	0.047 (0.164)	0.085 (0.152)
G^0	0.118 (0.915)	0.051 (0.100)	0.145 (0.806)	0.035 (0.287)
K	0.024 (0.133)	0.081 (0.164)	0.014 (0.069)	0.064 (0.287)
Alpha-stable	0.046 (0.079)	0.027 (0.071)	0.042 (0.072)	0.039 (0.100)

Table 3. Acceptance probability of different statistical models.

Model	Acceptance Probability at 5% Significance Level (Percentage)			
	River	Marsh	Farmland	Building
Gamma	81.17	0.17	71.67	0.00
Weibull	55.33	0.00	70.33	0.67
G^0	83.83	45.44	81.69	83.33
K	97.00	4.17	99.83	43.83
Alpha-stable	68.83	96.50	70.83	67.50

To intuitively illustrate the multimodal statistical property of SAR image, we consider a 512 pixels \times 768 pixels image of a building area from TerraSAR-X data. Figure 4 displays the original SAR image (Figure 4(a)), the fitting results (Figure 4(b)) of different models (including Gamma, Alpha-stable and MAS), and the estimated MAS distributions with three components (Figure 4(c)). From Figure 4(a), we can observe that many bright regions exist, caused by the existence of many strong reflectors in building area. Shadows of darkness are also observed, a result of the imaging mechanism of the SAR system, and few scatter waves are received by the SAR system in this kind of area. These factors lead to the complicated multimodal statistics of SAR images. In this case, none of unimodal distributions has the ability to model it well. Therefore, a multimodal distribution is necessary for accurate modeling of SAR images. MAS distributions can capture this essential statistical property pretty well.

We further evaluate the performance of MAS distributions with different number of components for modeling SAR image samples used previously. The results are shown in Tables 4 and 5.

From Tables 4 and 5, it can be concluded that MAS distributions substantially improve the modeling performances for all four classes. As we increase the number of components, the average KSD decreases and the acceptance probability increases for each class. The results also indicate that MAS distributions with five components are sufficient, as the average KSD and acceptance probability for each class change only slightly when the number of components increase to seven. Therefore, we choose the MAS distributions with five components to model the statistics of SAR images in the proposed classification algorithm.

Figure 4. Selected building area and fitting results. (a) Building area in SAR image; (b) Fitting results by different models; (c) Three components of the estimated MAS distributions.

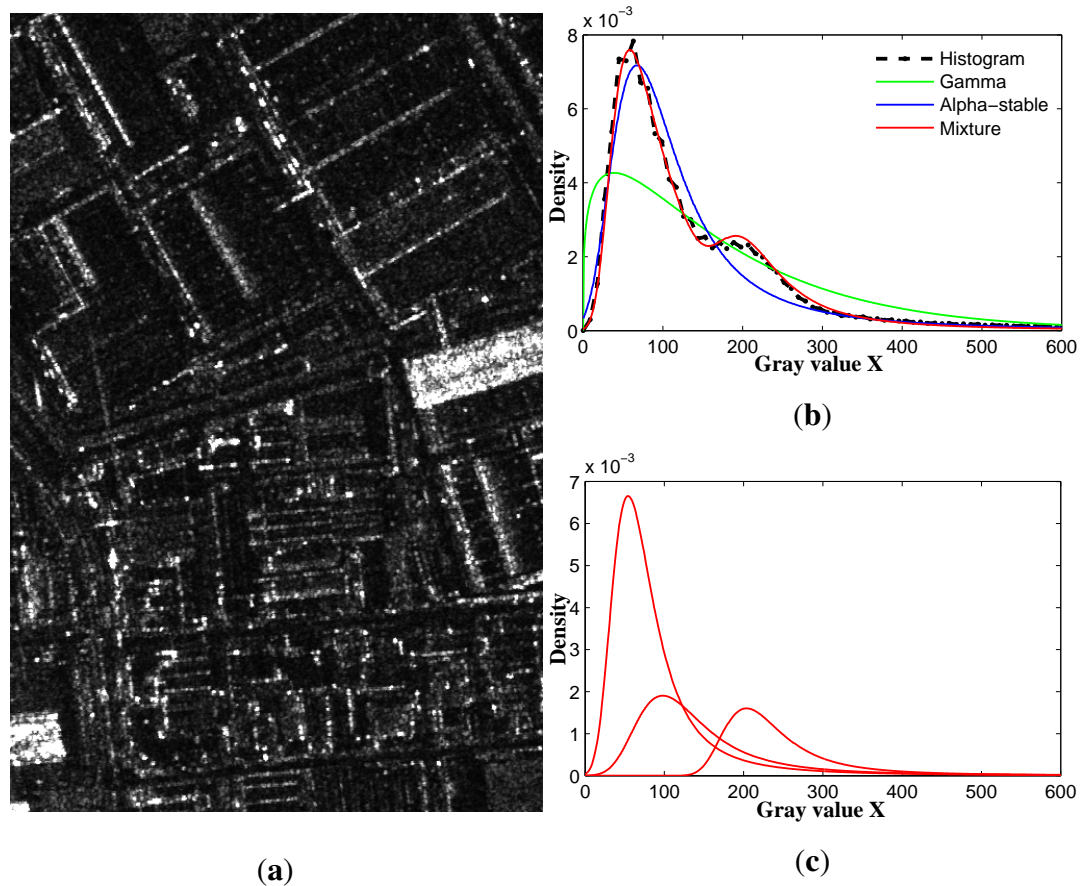


Table 4. Mixture of Alpha-stable distributions for modeling various SAR images.

Components	Average <i>KSD</i> (Maximum <i>KSD</i>)			
	River	Marsh	Farmland	Building
1	0.046 (0.079)	0.027 (0.071)	0.042 (0.072)	0.039 (0.100)
3	0.025 (0.050)	0.024 (0.088)	0.012 (0.049)	0.026 (0.099)
5	0.021 (0.041)	0.023 (0.113)	0.011 (0.025)	0.024 (0.115)
7	0.021 (0.074)	0.022 (0.110)	0.010 (0.023)	0.023 (0.115)

Table 5. Acceptance probability of mixture models with different number of components.

Components	Acceptance Probability at 5% Significance Level (Percentage)			
	River	Marsh	Farmland	Building
1	68.83	96.50	70.83	67.50
3	99.83	96.83	100.00	87.17
5	100.00	97.17	100.00	88.67
7	100.00	98.00	100.00	89.33

4. MAS-Based MRF Classification Algorithm

We propose the use of MAS distributions in a Bayesian classification scheme. MAS distributions are utilized to model the statistical properties of SAR images, and the spatial context is introduced by an MRF framework. A Graph Cuts-based algorithm is applied to obtain a global optimization.

4.1. MRF-Based Segmentation

Let $X = (x_1, \dots, x_j, \dots, x_M)$ be a random observation and $Y = (y_1, \dots, y_j, \dots, y_M)$ be the expected labeling result, with $y_j \in \{1, 2, \dots, K\}$. In the Bayesian estimation framework, the SAR image classification could be described as a MAP problem, $\hat{Y} = \arg \max_Y P(Y|X)$. Based on the Bayesian theory, $P(Y|X) \propto P(X|Y)P(Y)$, it is equivalent with $\hat{Y} = \arg \max_Y P(X|Y)P(Y)$. The restoration method proposed in [38] assumes that the conditional probability of labeling $Y = y$, given an observation $X = x$, can be modeled by a Markov random field. Using the Hammersley–Clifford theorem, this probability can be written under Gibbs field formalism as Equation (11):

$$P(Y = y|X = x) = \frac{\exp\{-U(y|x)\}}{Z} \quad (11)$$

where Z is a normalization coefficient, and U is the energy function given by the sum of a data term and a regularization term Equation (12):

$$U(y|x) = - \sum_i \ln p(x_j|y_j) + \sum_{c \in C_y} V_c(y) \quad (12)$$

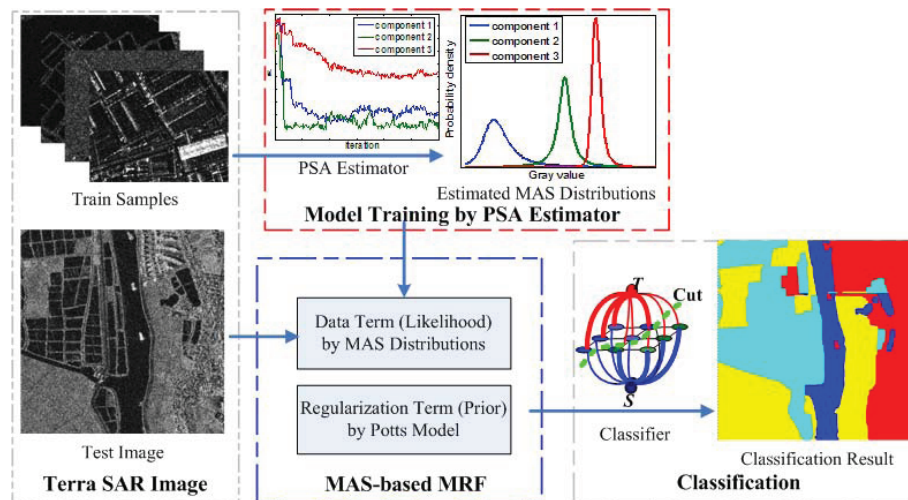
where C_y is the set of cliques of the selected neighborhood, and V_c is the potential of label configuration. In this paper, we focus solely on the data term, which is evaluated by mixture of Alpha-stable distributions. The second term is simply modeled by the Potts model [39], which can be written as Equation (13):

$$V_{C_y=(s,t)}(y_s, y_t) = \begin{cases} \lambda, & y_s \neq y_t \\ -\lambda, & y_s = y_t \end{cases} \quad (13)$$

with s and t being in the same clique. Result of the Markovian segmentation is the label field $Y = y$ that minimizes U (or equivalently maximizes P). In the energy function U of random field, mixture of Alpha-stable distributions are exploited to drive the data term.

4.2. MAS-Based MRF Classification Algorithm

We propose a supervised Bayesian classification algorithm by combining MAS distributions with MRF. Samples are required for each class to estimate the MAS distribution parameters via the PSA estimator. The first term of Equation (12) implies that the conditional probabilities are then represented by the MAS distributions. The regularization term is introduced via MRF, which builds a labeling restraint between the current pixel and its neighborhood. The classification turns out to be a posterior probability maximization problem, which is equivalent to minimizing the energy. To solve the energy minimization problem, we utilize a Graph Cuts optimization, as this is fast and obtains near-global optimization. The proposed MAS-based MRF classification algorithm is described in Figure 5.

Figure 5. Flowchart of the MAS-based MRF classification algorithm.

5. Experiment and Result Analysis

The MAS-based MRF classifier was applied to the TerraSAR-X data. We compared the results from the MAS-based MRF classifier (MAS + MRF) with three other classifiers: Alpha-stable (AS)-based MRF classifier (AS + MRF), Gamma (GAM)-based MRF classifier (GAM + MRF), and mixture Gamma (MGAM)-based MRF classifier (MGAM + MRF). To enable a quantitative analysis, the classification accuracy for each class was calculated as the ratio of the number of correctly classified pixels to the total number of pixels in the class, given as a percentage with reference to the visually interpreted map. The regularity parameter for the Graph Cuts algorithm was empirically set to $\lambda = 8$.

5.1. Experimental Data

5.1.1. Wuhan Data

The TerraSAR-X data for Wuhan in Hubei, China, was acquired on 28 September 2008, with a spatial resolution of approximately 3.0 meters in range and azimuth (Stripmap mode, multilook ground range detected). The image, shown in Figure 6(a), has a size of 2,000 pixels \times 1,800 pixels. Figure 6(b) is the optical image from Google Earth (© Google 2013). Figure 6(c) is the visually interpreted map. A total of four land covers were identified, namely river, marsh, farmland, and buildings. The white areas labeled 'None' in Figure 6(c) could not be defined. These regions were not included in the accuracy assessment step.

5.1.2. Foshan Dataset

The TerraSAR-X data for Foshan in Guangdong, China, was acquired on 24 May 2008, with a spatial resolution of approximately 1.25 meters in both range and azimuth (Stripmap mode, multilook ground range detected). The image, shown in Figure 7(a), has a size of 7,000 pixels \times 5,000 pixels. Figure 7(b) is the optical image from Google Earth (© Google 2013). Figure 7(c) is the visually interpreted map.

Again, four land covers were identified, *i.e.*, river, marsh, farmland, and buildings. The black areas labeled ‘None’ in Figure 7(c) could not be defined. These regions were not included in the accuracy assessment step.

Figure 6. TerraSAR-X data of Wuhan city. (a) SAR image of Wuhan, Hubei, China; (b) Optical image from Google Earth (©Google 2013); (c) Visually interpreted map; (d) Classification result of MAS + MRF classifier.

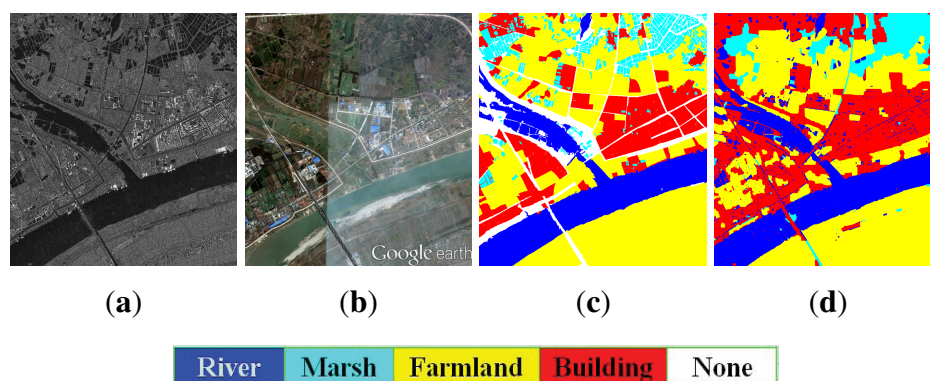
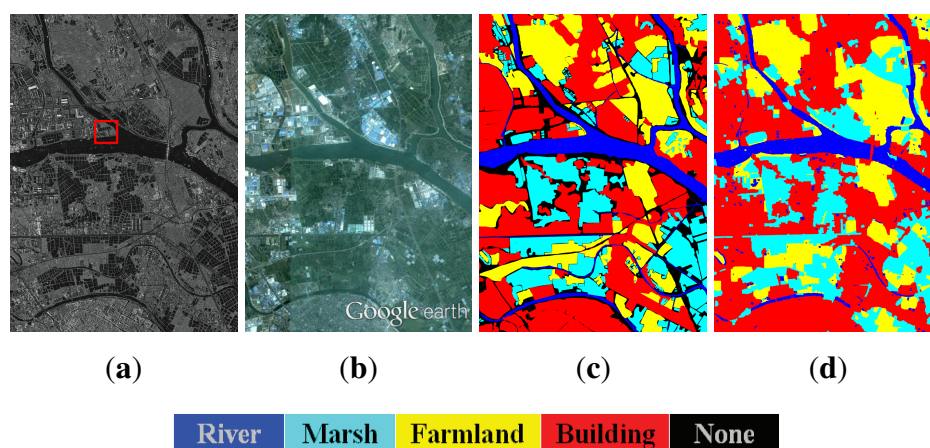


Figure 7. TerraSAR-X data of Foshan city. (a) SAR image of Foshan, Guangdong, China; (b) Optical image from Google Earth (©Google 2013); (c) Visually interpreted map; (d) Classification result of MAS + MRF classifier.



5.2. Classification Result

In this experiment, samples were selected from each class in order to learn the parameters of the MAS distributions. The Potts model was applied to introduce the spatial contextual information, which behaves as the regularization term. Various classifiers were selected to compare their classification performance. The results from the MAS + MRF classifier for the Wuhan and Foshan data are displayed in Figures 6(d) and 7(d), respectively. These both show that the global ground cover acquired by the MAS + MRF classifier was well recognized. In particular, the region edges were accurate. In our experiment, we found that the AS + MRF classifier and the MGAM + MRF classifier also obtained

acceptable classification results, but those of the GAM + MRF classifier were poor, especially for marsh areas, which were probably mislabeled as rivers.

5.3. Result Analysis

The classification accuracy for each class, as well as the average accuracy, was calculated for different classifiers. Confusion matrices for the results obtained by the MAS+MRF classifier on Wuhan data and Foshan data are displayed in Tables 6 and 7. In addition, the classification accuracy for each class and the overall accuracy of different classifiers are illustrated in Figures 8 and 9.

Table 6. Confusion matrix for the MAS + MRF result (percentage).

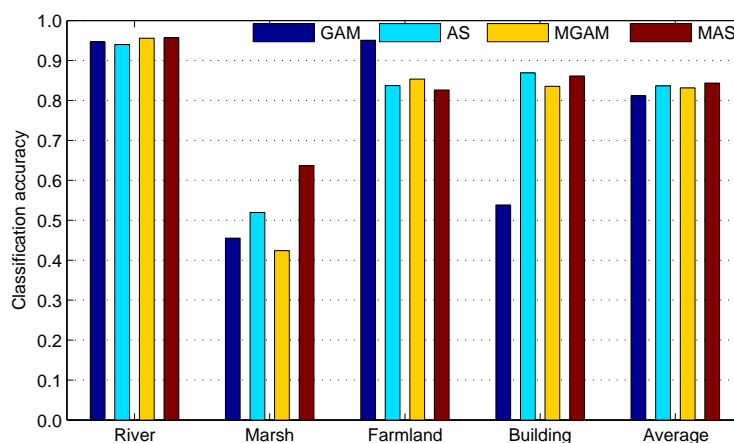
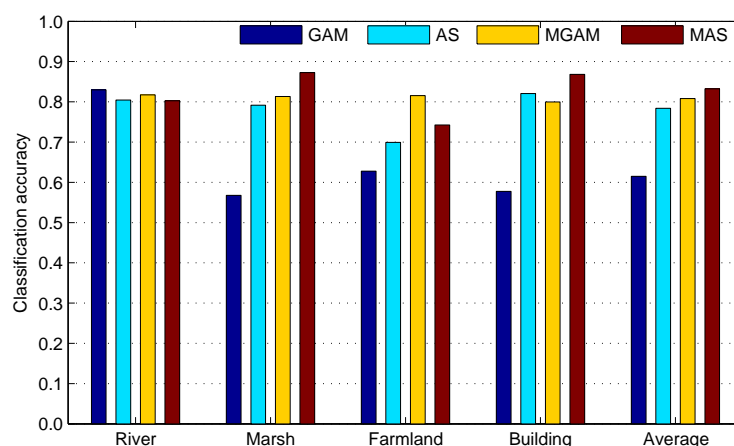
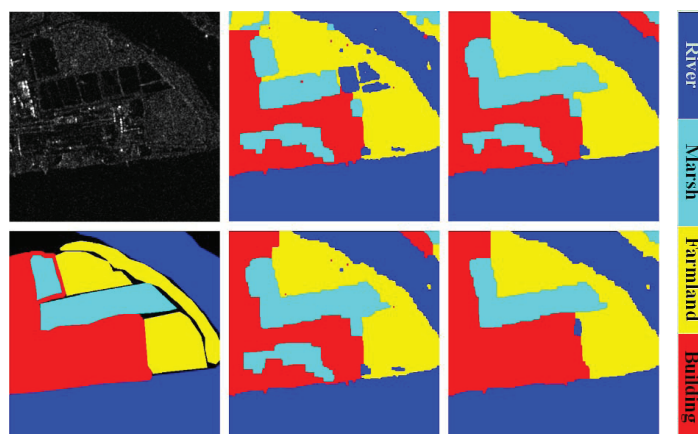
Dataset	Wuhan data (Average accuracy: 84.35)			
Class	River	Marsh	Farmland	Building
River	95.7	0.7	1.2	2.4
Marsh	22.2	63.7	2.5	11.6
Farmland	0.6	1.3	82.6	15.5
Building	3.6	0.7	9.6	86.1

Table 7. Confusion matrix for the MAS + MRF result (percentage).

Dataset	Foshan data (Average accuracy: 83.22)			
Class	River	Marsh	Farmland	Building
River	80.3	11.2	2.0	6.5
Marsh	4.0	87.3	3.4	5.3
Farmland	0.3	0.9	74.3	24.5
Building	0.7	10.0	2.5	86.8

Tables 6 and 7 show that the MAS + MRF classifier obtains a high accuracy for each class on both Wuhan data and Foshan data, with an average accuracy of 84.35% and 83.22%, respectively.

Figure 8 illustrates that the MAS + MRF classifier attains an average accuracy of 84.35% for the Wuhan data, which is higher than that of the other classifiers by 0.69%–3.21%. The average accuracies are 83.12%, 83.66% and 81.14% for MGAM + MRF, AS + MRF and GAM + MRF, respectively. Similarly, Figure 9 shows that the MAS+MRF classifier is the most accurate for marsh and building with the Foshan data, and obtains good results for river and farmland. The average accuracy obtained by MAS + MRF is higher than that of the other classifiers by 2.41%–21.74%. The average accuracies are 61.48%, 78.40%, 80.81% and 83.22% for GAM + MRF, AS + MRF and MGAM + MRF, respectively. This leads us to the conclusion that the MAS + MRF classifier outperforms the other classifiers. In addition, it can also be concluded that a classifier based on a mixture of distributions obtains better classification accuracy than a single distribution classifier. The improvement given by the mixture model is obvious. The high accuracy of classification by the MAS-based MRF classifier can be explained by the accurate modeling of the multimodal statistical properties of SAR images.

Figure 8. Classification accuracies of different classifiers on Wuhan dataset.**Figure 9.** Classification accuracies of different classifiers on Foshan dataset.**Figure 10.** Classifications of localized area (**top**, left–right: SAR image, result from GAM + MRF, result from AS + MRF; **bottom**, left–right: Visually interpreted map, result from MGAM + MRF, result from MAS + MRF).

To evaluate the performance of the various classifiers in a localized area, we studied the $480 \text{ pixels} \times 480 \text{ pixels}$ image indicated by the red box in Figure 7(a). The visually interpreted map and classifications are displayed in Figure 10. The GAM + MRF classifier result is not good, as it

misclassifies large areas of marsh as river. The AS + MRF classifier and MGAM + MRF classifier obtain slightly better classification results than GAM + MRF. However, all three classifiers misclassify large areas of buildings as marsh. Fortunately, the proposed MAS + MRF classifier correctly labels these areas. It can be observed that different regions are perfectly distinguished, and the edges are also accurate. These results further illustrate the efficiency of the proposed MAS + MRF classifier.

6. Conclusions

This paper proposed the mixture of Alpha-stable distributions, the first application in image processing, to model the multimodal and impulsive statistical properties of SAR images especially at high resolution, and introduced a PSA estimator to determine the distribution parameters. Simulation results on simulated data validated the effectiveness of the PSA estimator, and the modeling performance over TerraSAR-X samples verified the validity of MAS distributions for SAR images. Finally, a Bayesian classification algorithm that combines MAS distributions and MRF was presented for SAR images. The classification performances over TerraSAR-X images demonstrated that the proposed mixture model is a promising candidate for the modeling and classification of SAR images and can be potentially useful for SAR images analysis, interpretation, and other important applications.

Acknowledgements

This work was supported in part by the National Basic Research Program of China (973) (No. 2011CB707102), in part by the National Natural Science Foundation of China (No. 40871199), and in part by the Fundamental Research Funds for the Central Universities (No. 20102120201000128).

The authors would like to thank the editor and all the reviewers for their helpful suggestions and comments on this paper.

Conflict of Interest

The authors declare no conflict of interest.

References

1. Perko, R.; Raggam, H.; Deutscher, J.; Gutjahr, K.; Schardt, M. Forest assessment using high resolution SAR data in X-band. *Remote Sens.* **2011**, *3*, 792–815.
2. Alberga, V. Similarity measures of remotely sensed multi-sensor images for change detection applications. *Remote Sens.* **2009**, *1*, 122–143.
3. Hasager, C.B.; Badger, M.; Pena, A.; Larsén, X.G.; Bingol, F. SAR-based wind resource statistics in the Baltic Sea. *Remote Sens.* **2011**, *3*, 117–144.
4. Tison, C.; Nicolas, J.M.; Tupin, F.; Maitre, H. A new statistical model for Markovian classification of urban areas in high-resolution SAR images. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 2046–2057.

5. Jacob, A.M.; Hemmerly, E.M.; Fernandes, D. SAR Image Classification Using a Neural Classifier Based on Fisher Criterion. In Proceedings of the 7th Brazilian Symposium on Neural Networks (SBRN02), Recife, Brazil, 11–14 November 2002; pp. 168–172.
6. Chamundeeswari, V.V.; Singh, D.; Singh, K. Unsupervised Land Cover Classification of SAR Images by Contour Tracing. In Proceedings of Geoscience and Remote Sensing Symposium (IGARSS 2007), Barcelona, Spain, 23–27 July 2007; pp. 547–550.
7. Oliver, C.J. *Understanding Synthetic Aperture Radar Images*; Artech House: Boston, MA, USA/London, UK, 1998.
8. Oliver, C. Optimum texture estimators for SAR clutter. *J. Phys. D Appl. Phys.* **1993**, *26*, 1824–1835.
9. Menon, M. Estimation of the shape and scale parameters of the Weibull distribution. *Technometrics* **1963**, *5*, 175–182.
10. Szajnowski, W. Estimators of log-normal distribution parameters. *IEEE Trans. Aerosp Electron. Syst.* **1977**, *AES-13*, 533–536.
11. Jakeman, E.; Pusey, P. A model for non-Rayleigh sea echo. *IEEE Trans. Antennas Propagat.* **1976**, *24*, 806–814.
12. Jao, J.K. Amplitude distribution of composite terrain radar clutter and the K-distribution. *IEEE Trans. Antennas Propag.* **1984**, *AP-32*, 1049–1062.
13. Li, H.C.; Hong, W.; Wu, Y.R.; Fan, P.Z. On the empirical-statistical modeling of SAR images with generalized gamma distribution. *IEEE J. Sel. Top. Signal Process.* **2011**, *5*, 386–397.
14. Galland, F.; Nicolas, J.M.; Sportouche, H.; Roche, M.; Tupin, F.; Refregier, P. Unsupervised synthetic aperture radar image segmentation using fisher distributions. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 2966–2972.
15. Frery, A.; Muller, H.-J.; Yanasse, C.; Sant’Anna, S. A model for extremely heterogeneous clutter. *IEEE Trans. Geosci. Remote Sens.* **1997**, *35*, 648–659.
16. Kuruoglu, E.E.; Zerubia, J. Modeling SAR images with a generalization of the Rayleigh distribution. *IEEE Trans. Image Process.* **2004**, *13*, 527–533.
17. Achim, A.; Tsakalides, P.; Bezerianos, A. SAR image denoising via bayesian wavelet shrinkage based on heavy-tailed modeling. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1773–1784.
18. Gao, G.; Shi, G.T.; Zou, H.X.; Zhou, S.L. Characterizing the statistical properties of SAR clutter by using an empirical distribution. *Int. J. Antennas Propag.* **2013**, 109145:1–109145:8, doi:10.1155/2013/109145.
19. Gao, G. Statistical modeling of SAR images: A survey. *Sensors* **2010**, *10*, 775–795.
20. Liao, M.S.; Wang, C.C.; Wang, Y.; Jiang, L.M. Using SAR images to detect ships from sea clutter. *IEEE Geosci. Remote Sens. Lett.* **2008**, *5*, 194–198.
21. Chen, J.Y.; Sun, H. Multi-Resolution Edge Detection Based on Alpha-Stable Model in SAR Images Using Translation-Invariance Contourlet Transform. In Proceedings of Signal Processing and Information Technology (ISSPIT 2006), Vancouver, BC, Canada, 27–30 August 2006; pp. 264–270.
22. Peng, Y.J.; Xu, X.; Zhou, W.B.; Zhao, Y. SAR image classification based on alpha-stable distribution. *Remote Sens. Lett.* **2011**, *2*, 51–59.

23. Wan, T.; Canagarajah, N.; Achim, A. Segmentation-driven image fusion based on alpha-stable modeling of wavelet coefficients. *IEEE Trans. Multimed.* **2009**, *11*, 624–633.
24. Deng, C.Z.; Zhu, H.S.; Wang, S.Q. Curvelet Domain Watermark Detection Using Alpha-Stable Models. In Proceedings of Information Assurance and Security (ICIAS 2009), XI'an, China, 18–20 August 2009; pp. 313–316.
25. Ge, X.H.; Zhu, G.X.; Zhu, Y.T. On the testing for alpha-stable distributions of network traffic. *Comput. Commun.* **2004**, *5*, 447–457.
26. Xu, W.D.; Wu, C.F.; Dong, Y.C. Modeling Chinese stock returns with stable distribution. *Math. Comput. Modell.* **2011**, *54*, 610–617.
27. Salas-Gonzalez, D.; Kuruoglu, E.E.; Ruiz, D.P. Finite mixture of alpha-stable distributions. *Digital Signal Process.* **2009**, *19*, 250–264.
28. Salas-Gonzalez, D.; Kuruoglu, E.E.; Ruiz, D.P. Modelling with mixture of symmetric stable distributions using Gibbs sampling. *Signal Process.* **2009**, *90*, 774–783.
29. Salas-Gonzalez, D.; Kuruoglu, E.E.; Ruiz, D.P. Bayesian Estimation of Mixtures of Skewed Alpha-Stable Distributions with an Unknown Number of Components. In Proceedings of European Signal Processing, EUSIPCO 2006, Florence, Italy, 4–8 September 2006.
30. Buckle, D.J. Bayesian inference for stable distribution. *J. Am. Stat. Assoc.* **1995**, *90*, 605–613.
31. Casarin, R. *Bayesian Inference for Mixtures of Stable Distributions*; Working Paper No. 0428; CEREMADE, University Paris IX: Paris, France, 2004.
32. Boykov, Y.; Veksler, O.; Zabih, R. Fast approximate energy minimization via Graph Cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 1222–1239.
33. Kolmogorov, V.; Zabih, R. What energy functions can be minimized via Graph Cuts? *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 147–159.
34. Boykov, Y.; Kolmogorov, V. An experimental comparison of Min-Cut/Max-Flow algorithms for energy minimization in vision. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 1124–1137.
35. Levy, P. Theorie des erreurs : La loi de Gauss et les lois exceptionnelles. *Bull. Socit Mathematique France* **1924**, *52*, 49–85.
36. Belov, I.A. On the Computation of the Probability Density Function of Alpha-Stable Distributions. In Proceedings of the 10th International Conference on Mathematical Modelling and Analysis (ICMMA 2005), Trakai, Lithuania, 1–5 June 2005; pp. 333–341.
37. Chambers, J.; Mallows, C.; Stuck, B. A method for simulating stable random variables. *J. Am. Stat. Assoc.* **1976**, *71*, 340–344.
38. Geman, S.; Geman, D. Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **1984**, *PAMI-6*, 721–741.
39. Wu, F. The Potts model. *Rev. Mod. Phys.* **1982**, *54*, 235–268.