Mining Satellite Image Time Series: Statistical Modeling and Evolution Analysis

Shiyong Cui and Mihai Datcu
Remote Sensing Technology Institute (IMF)
German Aerospace Center (DLR)
Oberpfaffenhofen D-82234, Germany
shiyong.cui@dlr.de, mihai.datcu@dlr.de

***Abstract*—Due to the short revisit time of high resolution satellites, huge amount of high resolution satellite images can be acquired every few days even few hours. It promotes the construction of Satellite Image Time Series (SITS), which contain valuable spatio-temporal information. Therefore, it is strongly needed to develop methods to explore such huge data to provide useful information in the context of earth observation. To address this issue, a patch based method for mining satellite image time series is proposed, consisting of statistical modeling and evolution analysis. Many statistical models has been proposed for Synthetic Aperture Radar (SAR) image modeling. Among them, $\mathcal{G}$ distribution has been proved efficient in modeling extremely heterogenous area especially for urban areas. In this paper, it is used to estimate the marginal distribution of SAR images by second-kind statistics. For the purpose of joint distribution modeling given the marginal distributions, optimal copula function is selected from a set of copulas by a Bayesian method and estimated using Kendall's τ. Based on the statistical model and the optimal copula, mixed information is computed between two neighboring patches along time for evolution analysis of the SITS. A ν-support vector machine is applied for evolution classification. Performance of both estimation and classification are evaluated using our database produced by iterative classification.**

I. Introduction

In recent years, due to rapid development of high resolution satellite sensors, huge amount of high resolution satellite images have been acquired. Furthermore, short revisit time of high resolution satellites lead to Satellite Image Time Series (SITS), which contain a lot of detailed and valuable spatio-temporal information. Traditionally, multitemporal image analysis is performed on two images for change detection [1] and classification [2]. Nowadays, with a long SITS, more information could be extracted and exploited, such as the dynamic evolution of object and the scene. Furthermore, we can predict the developing and potential trends of the object evolution in the scene. However, what information can be extracted from SITS and how to represent and extract this information are still unanswered, which is the main goal of image time series analysis. In [3], trajectory modeling of dynamic clusters is performed based hierarchical Bayesian modeling. The hierarchy is composed of two steps: first is unsupervised modeling of dynamic clusters resulting in a graph representation of the trajectory and the second is interactive learning based on the trajectory graph leading to the semantic labeling of spatio-temporal patterns. Relevant information extraction from SITS based on information-bottleneck principle is proposed in [4]. Information contents in SITS are characterized by Gaussain-Markov random filed and autobinomial random filed models. Unsupervised learning is carried out on the basis of information-bottleneck principle. These two outstanding works both contribute to the content discovery in SITS. In this paper, an alternative approach for mining SITS is proposed based on statistical models for evolution classifidcation.

II. Density Estimation

In the literature, many statistical models have been proposed for SAR image modeling, e.g., $\mathcal{K}$ distribution, $\mathcal{F}$ distribution and $\mathcal{G}$ distribution. The parameters of these models are usually needed to be estimated and used as features for further processing. Among these models, $\mathcal{G}$ model [5], estimated using moments $m_{1/2}$ and $m_{1/4}$, has been proved be able to model extremely heterogeneous region, especially for urban area. Parameters of the $\mathcal{G}$ distribution estimated using mothod of moments $m_{1/2}$ and m_1, was used to derive features that, in turn, were used as the input of Gaussian maximum likelihood. classification [6]. An unsupervised classification approach of SAR images using Markov Random Fields and $\mathcal{G}$ Model is proposed in [7] and it is shown that $\mathcal{G}$ distribution is superior to $\mathcal{K}$ distribution. However, the estimation of the model can be only addressed by maximum likelihood or Method of Moment (MoM). MoM can not estimate the number of looks, therefore the accuracy of data fitting is decreasing especially for urban areas. An estimation method based on second kind statistics is proposed. The $\mathcal{G}$ model of the SAR intensity is defined as

$$P(x) = \frac{n^n \Gamma(n-\alpha)}{\gamma^\alpha \Gamma(n)\Gamma(-\alpha)} \frac{x^{n-1}}{(\gamma+nx)^{n-\alpha}} \tag{1}$$

where x is the SAR intensity value, n is the number of looks, γ is scale parameter and α is shape parameter. We propose to use second kind statistics [8] for estimation, which is based on Mellin transform. The first second-kind characteristic function is defined as the Mellin transform of the $\mathcal{G}$ model and given as.

$$\phi(x) = \left(\frac{\gamma}{n}\right)^{x-1} \frac{\Gamma(n+x-1)\Gamma(-\alpha-x-1)}{\Gamma(n)\Gamma(-\alpha)} \tag{2}$$

The second second-kind characteristic function is given by the logarithm of the first second-kind characteristic function, defined as

$$\xi(x) = (s-1)\ln(\frac{\gamma}{n}) + \ln\Gamma(n+x-1) \\ + \ln\Gamma(-\alpha-x+1) - \ln\Gamma(n) - \ln\Gamma(-\alpha) \quad (3)$$

Second kind log-cumulants of $\mathcal{G}$ model is given as follows by the derivative of the second second-kind characteristic functions.

$$\begin{aligned} K_1 &= \ln(\gamma/n) + \Psi(n) - \Psi(-\alpha) \\ K_i &= \Psi(i-1,n) + (-1)^i\Psi(i-1,-\alpha) \end{aligned} \quad (4)$$

where $\Psi(x)$ is the Digamma function and $\Psi(i-1,x)$ is the $i-1$ order Polygamma function. If assume the sample log-cumulants $\hat{K_i}$ are equal to second kind log-cumulants K_i, the following equations can be derived

$$\begin{aligned} &\ln(\hat{\gamma}/\hat{n}) + \Psi(\hat{n}) - \Psi(-(\hat{\alpha})) = \hat{K_1} \\ &\Psi(1,\hat{n}) + \Psi(1,-\hat{\alpha}) = \hat{K_2} \\ &\Psi(2,\hat{n}) - \Psi(2,-\hat{\alpha}) = \hat{K_3} \end{aligned} \quad (5)$$

These nonlinear equations can be solved for parameters by numerical methods, such as Newton-Raphson Method and Trust region algorithms. It is worth to note that it may not converge rarely if the initial value of the parameters are not appropriate. In this case, the parameters can be derived by interpolation using the neighboring values.

III. COPULA AND MODEL SELECTION

In this paper, we focus on bivariate copula function. It can be extended to multivariate case by using multivariate copula function.

A. Bivariate copula

A bivariate copula function is a joint cumulative distribution of two uniform random variables X_1 and X_2, defined as,

$$C(x_1,x_2) = Pr(X_1 < x_1, X_2 < x_2) \quad (6)$$

where $X_i \sim U(0,1)$ for $i=1,2$. Based on Sklar's theorem [9], joint cumulative distribution of two random variables X_1 and X_2 is given by the copula function at $F_1(x_1), F_2(x_2)$. Therefore, the joint probability density can be given by the derivative of the copula function as follows,

$$\begin{aligned} f(x_1,x_2) &= \frac{\partial^2 C}{\partial x_1 \partial x_2}(x_1,x_2) \\ &= c(F_1(x_1),F_2(x_2))f_1(x_1)f(x_2) \end{aligned} \quad (7)$$

where $f_i(x_i), i=1,2$ is the marginal distribution and $c(u,v)$ is the copula density given by the derivative of the copula function. In practice for SAR images, distribution functions $f_i(x_i)$ can be derived by the statistical model presented in previous section. The remaining problem is to choose and estimate a appropriate copula function. In this paper, we consider four copulas (Clayton, Ali-Mikhail-Haq, Gumbel and Frank) [9] belonging to archimedean family as shown in table I.

B. Copula selection

To select a appropriate copula from these four options, Bayesian model selection method [10] is applied to attribute a weight to each one. The one with maximum weight would be selected to model the joint distribution. Let $\mathcal{M}$ represents the copula family. we need to select a best one from this copula family. Suppose the data is composed of mutually independent pairs of quantiles (u_i, v_i), using Bayesian theorem, the posterior probability of each copula is given as

$$P(M_i|D) = \frac{P(D|M_i)P(M_i)}{P(D)} \quad (8)$$

where $P(D|M_i)$ is the likelihood, $P(M_i)$ is the prior of the copula and $P(D)$ is the normalization constant. As all copulas depend on a parameter θ and to parameterize the copulas using a common parameter, Kendall's τ is introduced as a nuisance variable based on its relationship $\tau = g(\theta)$ with the parameter θ. Therefore, the posterior can be written as

$$\begin{aligned} P(M_i|D) &= \int_{-1}^{1} P(M_i,\tau|D)d\tau \\ &= \int_{-1}^{1} \frac{P(D|M_i,\tau)P(M_i|\tau)P(\tau)d\tau}{P(D)} \end{aligned} \quad (9)$$

where $P(\tau)$ is the prior on τ, $P(M_i|\tau)$ is the prior on the copula. The likelihood of the data can be calculated from the copula density as follows,

$$P(D|M_i,\tau) = \prod_{i=1}^{n} c(u_i,v_i|g_i^{-1}(\tau)) \quad (10)$$

As there is no knowledge about the prior on τ, uniform distribution is assumed for the prior

$$P(\tau) = \frac{1}{w(\Lambda)} \quad (11)$$

where Λ is the interval of correlation between two random variables, w is the width of the Λ. In the case of no information about correlation available, Λ is simply $[-1,1]$. Since each copula is equally probable with respect to a give τ, $P(M_i|\tau)$ is assumed to be proportional to $\mathbb{1}(\tau \in \Omega_i)$. If some knowledge about the priors is available, certain distributions can be chosen to reflect the prior. The weight attributes to ith copula is as follows,

$$W_i = \frac{1}{w_i}\int_{\Omega_i \bigcap \Lambda} \prod_{i=1}^{n} c(u_i,v_i|g_i^{-1}(\tau))d\tau \quad (12)$$

This integral can be computed numerically using Gaussian-Legendre quadrature. The copula with maximum weight is chosen for joint distribution construction. After selecting the optimal copula, the remaining problem is to estimate the copula parameter θ. For the sake of simplicity, Kendall's τ is used to estimate copula parameter $\theta = g^{-1}(\tau)$. Kendall's τ for sample is defined as

$$\tau = \frac{c-d}{\frac{1}{2}n(n+1)} \quad (13)$$

c is the number of concordant pairs and d is the number of discordant pairs.

TABLE I
COPULAS CONSIDERED: CLAYTON, ALI-MIKHAIL-HAQ (AMH), GUMBEL AND FRANK

Copula	$C(u,v)$	$\theta(\tau)$	$\tau = g(\theta)$	$\tau \in \Omega$
Clayton	$\left(u^{-\theta} + v^{-\theta} - 1\right)^{-1/\theta}$	$(0, \infty)$	$1 - \frac{2}{2+\theta}$	$(0, 1]$
AMH	$\frac{uv}{1-\theta(1-u)(1-v)}$	$[-1, 1)$	$1 - \frac{2}{3} \frac{(\theta-)^2 \ln(1-\theta)+\theta}{\theta^2}$	$[-0.181726, 1/3]$
Gumbel	$\exp\{-[(-\ln u)^{\theta} + (-\ln v)^{\theta}]^{1/\theta}\}$	$[1, \infty)$	$1 - \theta^{-1}$	$[0, 1]$
Frank	$-\frac{1}{\theta} \ln(1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-}\theta - 1})$	$[-1, 0) \bigcup (0, 1]$	$1 - \frac{4}{\theta}(1 - \frac{1}{\theta} \int_0^{\theta} \frac{t}{e^t - 1} dt)$	$[-1, 0) \bigcup (0, 1]$

C. *Similarity features*

Based on the optimal copula, the joint distribution can be computed by equation (7). It has been shown that mixed information [11] between two neighboring patches along time can be used as a similarity metric and computed as

$$I_\alpha = \iint_D f(x,y) \log \frac{f(x,y)^{1+\alpha}}{f(x)f(y)} dxdy \tag{14}$$

Suppose we have n patches acquired at different times in specific scene, a feature vector of mixed information composed of $n-1$ elements $(I_\alpha^1, ..., I_\alpha^{n-1})$ can be computed using any neighboring pairs of patches. This feature vector will be used for evolution classification by a ν-SVM classifier.

IV. EXPERIMENTS AND EVALUATION

A dataset consisting of 12 TerraSAR-X images covering Vâlcea county in Romania is used to create the test database. These images are acquired every 11 days since angst 5, 2010 with incidence angle around 35.93^{o} and average height about 384m. The images have a ground resolution of 2.9m. Each image is cut into 34×41 patches with patch size 100×100 pixels. Image alignment is performed by normalized correlation coefficients as there is only translation along azimuth and range direction. In most cases, the average translation is 2 pixels. The reference data for change detection and evolution classification based on this database is produced respectively by iterative classification using support vector machine. For the sake of simplicity, we assume in this scenario there are five evolution classes in our database, e.g., building, forest, agriculture, grassland and water as shown Fig.1.

Symmetric Kullback-Leibler divergence is used to quantitatively evaluate the accuracy of estimation in both marginal distribution and joint distribution. Test patches from each class are randomly selected for accuracy estimation as shown in Fig.2. The patches in the first column are used for assessing the accuracy of marginal distribution. Patches in both the first and the second column are used for assessing the accuracy of joint distribution. As an example, marginal density and the joint density of agriculture field are shown in Fig.2. For the sake of comparison, Gamma distribution is also estimated. The accuracy is shown in table III. As expected, $\mathcal{G}$ distribution is promising in modeling high heterogenous region, where the accuracy is 0.0023 for urban region. On the contrary, the accuracy of Γ distribution in modeling urban region is 0.0098, which is 4 times lower than $\mathcal{G}$ distribution. For homogenous region, such as grassland in this test, Γ distribution is better than $\mathcal{G}$ distribution. However,with the increasing resolution of SAR image. $\mathcal{G}$ model is becoming superior. In this experiment, the prior on τ is assume to be normal distribution $N \sim (0.5, 3)$. It is shown that in this test, Clayton and Gumbel copulas are more suitable for dependence modeling among the four investigated copulas.

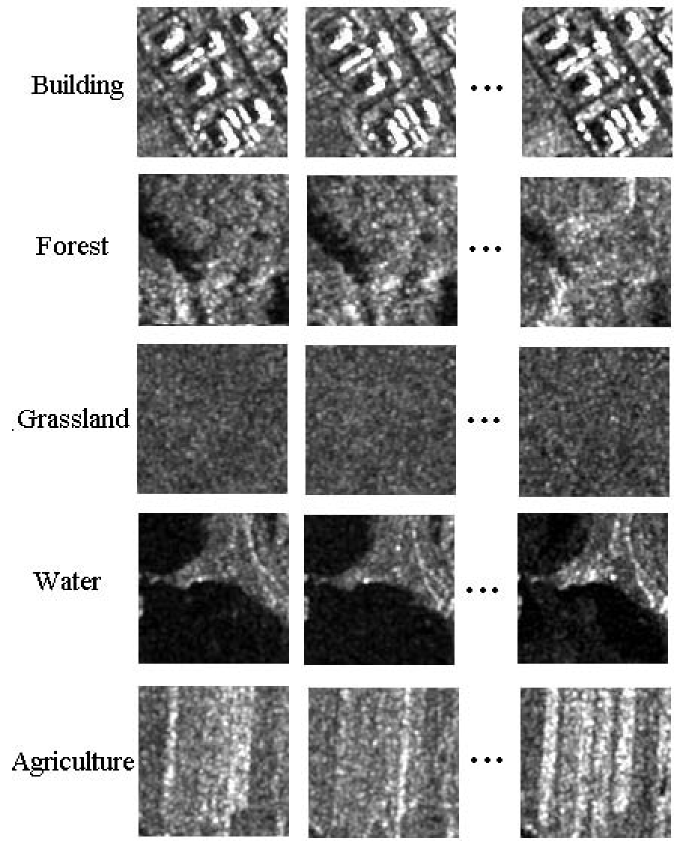


Fig. 1. Example of different evolution patterns from left to right.

In principle, each physical class should associate with one evolution pattern. Based ont the feature vector $(I_\alpha^1, ..., I_\alpha^{n-1})$, a ν-SVM classifier is applied to classify evolution patterns. Classification result of evolution patterns is shown in table II. The overall accuracy is 83.29%. The experimental results show potential ability for mining SITS.

V. CONCLUSION

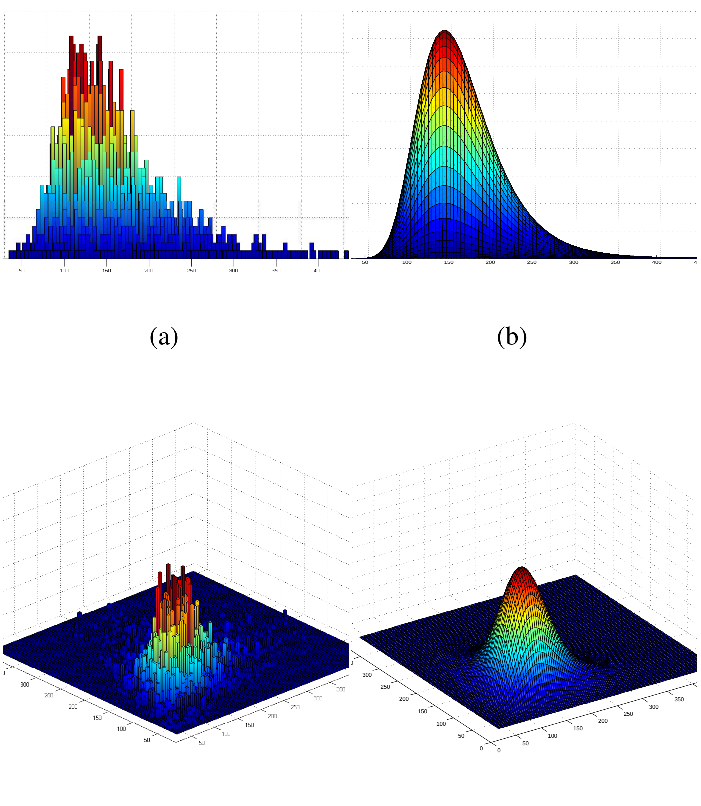
In this paper, an alternative approach for mining SITS is proposed and evaluated using TerraSAR-X images. $\mathcal{G}$ distribution

TABLE II
CONFUSION MATRIX OF EVOLUTION CLASSIFICATION

Ref. / *Pred.*	Forest	Building	Grassland	Water	Agriculture
Forest	86.51%	10.81%	6.41%	50.00%	11.15%
Building	3.82%	75.00%	6.41%	0.00%	1.44%
Grassland	5.09%	6.52%	79.06%	0.00%	6.47%
Water	0.38%	1.09%	0.85%	50.00%	0.00%
Agriculture	4.20%	6.52%	7.26%	0.00%	80.94%

TABLE III
ACCURACY ASSESSMENT OF ESTIMATION USING THE PATCHES SHOWN IN FIG.1.

Class / *KLD*	Forest	Building	Grassland	Water	Agriculture
$\mathcal{G}$ distribution	0.0211	0.0023	0.1413	0.1419	0.0066
Γ distribution	0.0250	0.0098	0.0456	0.1601	0.0140
Joint distribution	0.0243	0.0017	0.0567	0.2132	0.0323
Estimated copula	Clayton	Clayton	Gumbel	Clayton	Gumbel



(a) (b) (c) (d)

Fig. 2. Example of marginal and joint density estimation: (a) Histogram of agriculture; (b) Estimated pdf of agriculture; (c) Joint histogram; (d) Estimated joint pdf.

has been estimated using second-kind statistics for modeling SAR image intensity. To model the joint distribution, optimal copula is selected from a set of copulas using Bayesian model selection and estimated using Kendall's τ. Based on the $\mathcal{G}$ distribution and optimal copla function, features composed of similarity metric are computed and used for classification of evolution patterns by a ν-SVM Classifier. It is shown that Clayton and Gumbel copulas are more appropriate for modeling joint distribution of SAR images from this test. As there are less methods in the literature for modeling joint distribution of SAR images, this could be a choice for multitemporal SAR image analysis.

ACKNOWLEDGMENT

The author would like to thank the German Academic Exchange Service (DAAD) to provide finical support to carry out this research in the German Aerospace Center (DLR).

REFERENCES

[1] L. Bruzzone and D. F. Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1171–1182, 2000.

[2] F. Bovolo, L. Bruzzone, and L. Carlin, "A novel technique for subpixel image classification based on support vector machine," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2983–2999, 2010.

[3] P. Heas and M. Datcu, "Modeling trajectory of dynamic clusters in image time-series for spatio-temporal reasoning," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 7, pp. 1635–1647, 2005.

[4] L. Gueguen and M. Datcu, "Image time-series data mining based on the information-bottleneck principle," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 4, pp. 827–838, 2007.

[5] A. C. Frery, H.-J. Muller, C. C. F. Yanasse, and S. J. S. Sant'Anna, "A model for extremely heterogeneous clutter," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 3, pp. 648–659, 1997.

[6] M. E. Mejail, J. C. Jacobo-Berlles, A. C. Frery, and O. H. Bustos, "Classification of sar images using a general and tractable multiplicative model," *Int. J. Remote Sens.*, vol. 24, pp. 3565–3582, 2003.

[7] M. Picco and G. Palacio, "Unsupervised classification of sar images using markov random fields and $\mathcal{G}_I^0$ model," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 2, pp. 350–353, 2011.

[8] L. Bombrun and J.-M. Beaulieu, "Fisher distribution for texture modeling of polarimetric sar data," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 512–516, 2008.

[9] R. B. Nelsen, *An Introduction to Copulas*, Springer, New-York, 2nd ed edition, 2007.

[10] D. Huard, G. Évin, and A. Favre, "Bayesian copula selection," *Computational Statistics & Data Analysis*, vol. 51, no. 2, pp. 809–822, 2006.

[11] L. Gueguen, S. Cui, G. Schwarz, and M. Datcu, "Multitemporal analysis of multisensor data: Information theoretical approaches," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, 2010, pp. 2559–2562.