Multivariate Statistical Modeling for Multi-Temporal SAR Change Detection using Wavelet Transforms

Nizar Bouhlel, Guillaume Ginolhac, Eric Jolibois and Abdourrahmane Atto

LISTIC - Polytech'Annecy-Chambry,
BP 80439, 74944 Annecy le Vieux Cedex, France

24-07-2015

Table of contents

Table of contents

LISTIC

Introduction

- **Change detection**: is a process that analyzes multi-temporal remote sensing images acquired on the same geographical area for identifying changes occurred between the considered acquisition dates.
- **The goal** is a generation of a change detection map in which changed area are explicitly identified.
- **The number of images**: 2 or more
- **Application**: earth monitoring, earth observation, damage assessment and land cover dynamics

General Architecture (Stochastic approach)

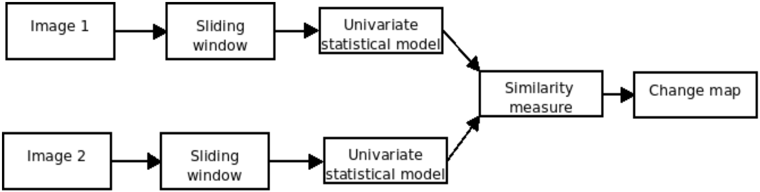


Figure : Block diagram for a classical change detection processing chain.

- For each sliding window, we model the data by an univariate statistical model
- Kullback-Leibler divergence between probability density functions of two sliding windows is used to generate a change map.

Our Architecture

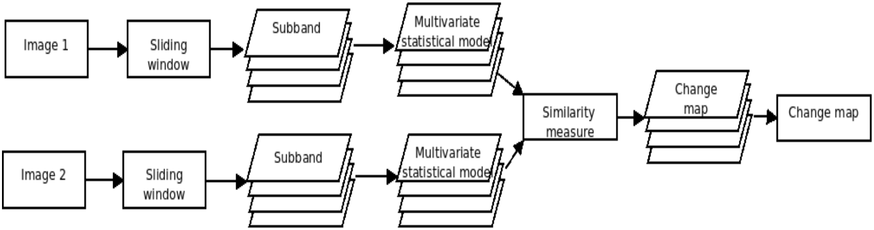


Figure : Block diagram for our change detection processing chain.

- The first step is to decompose a sliding window into multiple subands by wavelet transform.
- The probability density function of the sliding window coefficients of each subband is assumed to be **multivariate Gaussian distribution**
- **The Symmetric Kullback-Leibler divergence** is used as a similarity measure for change detection and used to generate a subband specific change map.
- The total Kullback-Leibler divergence is the sum of the Kullback-Leibler of each subband and it is used to obtain a final change map.

Why multivariate statistical model?

- The idea is to model the spatial interaction and information found in the sliding window by multivariate statistical distribution.
- When some Gaussianity are introduced into the data when the images were resampled and filtered during the pre-processing step, the Gaussian model can be accepted and can give a quite good approximation of the probability distributions.
- The existence of a closed-form expression of the KL.
- The calculation of KL is practical in real-time.

Why Wavelet decomposition?

- Implementation of our approach in a wavelet decomposition scheme can lead to performant texture modeling and change detection.
- Indeed, texture can be easily represented and discriminated in wavelet domain

LISTIC

Table of contents

LISTIC

Univariate statistical model: 1D Gaussian distribution

- In the community of radar image processing, a wide of statistical model distributions are used to characterize SAR images: Gamma distribution, generalized gamma distribution, K-distribution, etc.
- In our study, we simply use the Gaussian distribution which is given as follows

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-\mu)^2}{2\sigma^2}), \quad x \in \mathbb{R} \tag{1}$$

- The Gaussian model may be justified where the real data were subject to transformations during the pre-processing step.

LISTIC

Multivariate statistical model: kD multivariate Gaussian distribution

- The sliding window are modeled as the realization of a random vector $\mathbf{X} = (X_1, X_2, ..., X_k)^t$, where X_i are random variables.
- The joint density distribution function is given by $f_{\mathbf{X}}(\mathbf{x})$, where the vector $\mathbf{x} = (x_1, x_2, ..., x_k)^t$ is the realization of the random vector $\mathbf{X}$.
- The random k-vector $\mathbf{X}$ has the k-variate Gaussian distribution with mean k-vector $\boldsymbol{\mu}$ and positive-definite, symmetric $(k \times k)$ covariance matrix $\boldsymbol{\Sigma}$ and the density function is given by the

$$f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{\frac{-k}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})), \mathbf{x} \in \mathbb{R}^k \qquad (2)$$

Case (k=4)D

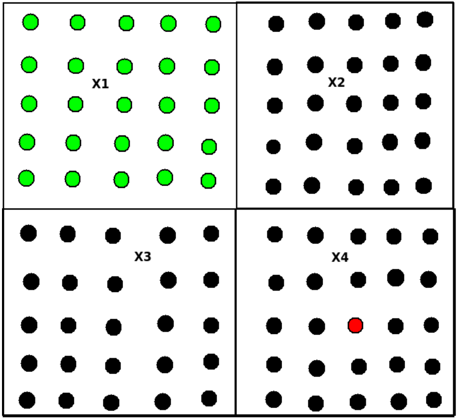


Figure : An example of sliding window organized in 2×2 blocks each one is constituted by $n \times n$ pixels (here 5×5). One block is a realization of vector component. The random vector $\mathbf{X}$ is composed here by $k = 4$ components (X_i, $i = 1..4$)

Case (k=9)D

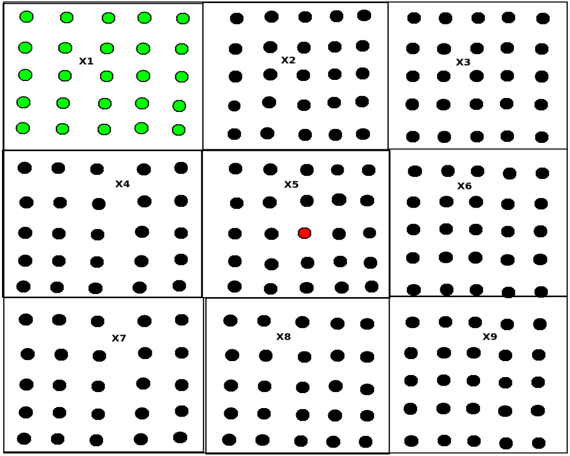


Figure : An example of sliding window organized in 3×3 blocks each one is constituted by $n \times n$ pixels (here 5×5). One block is a realization of vector component. The random vector $\mathbf{X}$ is composed here by $k = 9$ components $(X_i, i = 1..9)$

kD MGD in wavelet domain

- After wavelet decomposition, we finally have $3L+1$ images where L is the number of scales. These images correspond to the horizontal, vertical, diagonal details respectively at scale i and the approximation at scale L.
- The sliding windows of each wavelet subband coefficients are modeled using the kD multivariate Gaussian distribution.
- These sub-images are represented by

$$\boldsymbol{\mathcal{X}} = \{\mathbf{X_{H_i}}, \mathbf{X_{V_i}}, \mathbf{X_{D_i}}, \mathbf{X_{A_L}}\}, \quad i \in \{1, \ldots, L\} \tag{3}$$

and

$$\mathbf{X_{H_i}} = (X_{1,H_i}, X_{2,H_i} \ldots, X_{k,H_i})^t, \mathbf{X_{V_i}} = (X_{1,V_i}, X_{2,V_i} \ldots, X_{k,V_i})^t$$
$$\mathbf{X_{D_i}} = (X_{1,D_i}, X_{2,D_i} \ldots, X_{k,D_i})^t, \mathbf{X_{A_L}} = (X_{1,A_L}, X_{2,A_L} \ldots, X_{k,A_L})^t$$

are random k-vectors representing the sub-image horizontal, vertical and diagonal details at scale i respectively, and finally the sub-image approximation at scale L and are distributed according kD MGD.

LISTIC

Table of contents

LISTIC

- To quantify a change detection between two acquisition dates we need to analyze the modification of the statistics of each pixel's neighborhood.
- We choose to use the Kullback-Leibler divergence as a similarity measure. If the statistics of the two sliding windows are the same the symmetric Kullback-Leibler divergence is small.
- Let X^1 and X^2 be two random variables with probability density functions f_{X^1} and f_{X^2}. The Kullback-Leibler divergence from X^2 to X^1 is given by

$$KL(X^2||X^1) = \int \log\left(\frac{f_{X^1}(x)}{f_{X^2}(x)}\right) f_{X^1}(x)dx, \quad (4)$$

- **The symmetric KL similarity measure** between X^1 and X^2 is

$$KL(X^1, X^2) = KL(X^2||X^1) + KL(X^1||X^2). \quad (5)$$

LISTIC

- **Case of 1D GD** if the X^1 and X^2 are distributed according to a Gaussian distribution with mean μ_1 and μ_2 and variance σ_1 and σ_2, the symmetric version of the Kullback-Leibler divergence has the following form

$$KL(X^1, X^2) = \frac{\sigma_1^4 + \sigma_2^4 + (\mu_1 - \mu_2)^2(\sigma_1^2 + \sigma_2^2)}{2\sigma_1^2\sigma_2^2} \tag{6}$$

- **Case of kD MGD** if $\mathbf{X}^1$ and $\mathbf{X}^2$ are two random k-vectors with joint density functions $f_{\mathbf{X}^1}$ and $f_{\mathbf{X}^2}$, respectively, and are distributed according to the multivariate Gaussian distribution with k-dimensional mean vector $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ and $k \times k$ covariance matrix $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$, then the symmetric version of the Kullback-Leibler divergence has the following form

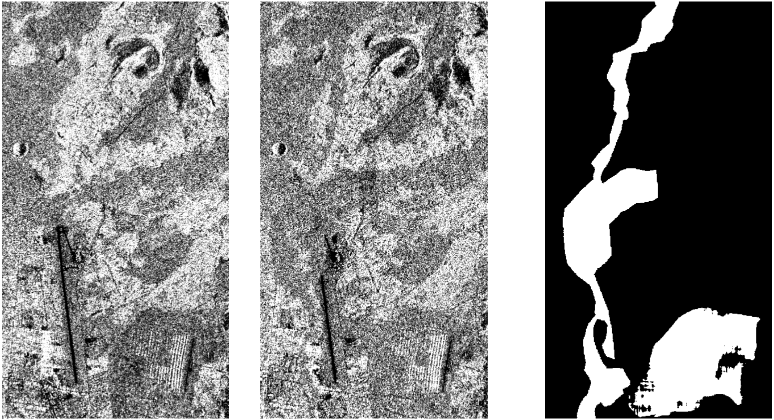
$$KL(\mathbf{X}^1, \mathbf{X}^2) = \frac{1}{2}[tr({\boldsymbol{\Sigma}_2}^{-1}\boldsymbol{\Sigma}_1) + tr({\boldsymbol{\Sigma}_1}^{-1}\boldsymbol{\Sigma}_2) - 2k + (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^t({\boldsymbol{\Sigma}_2}^{-1} + {\boldsymbol{\Sigma}_1}^{-1})(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)] \tag{7}$$

- **Case of kD MGD in wavelet domain** The subbands are assumed independent and the total similarity of two sliding windows are defined as the sum of similarity measures of each subband

$$KL(\boldsymbol{\mathcal{X}}^1, \boldsymbol{\mathcal{X}}^2) = KL(\mathbf{X}^1_{\mathbf{A_L}}, \mathbf{X}^2_{\mathbf{A_L}}) + \sum_{i=1}^{L} KL(\mathbf{X}^1_{\mathbf{H_i}}, \mathbf{X}^2_{\mathbf{H_i}}) + KL(\mathbf{X}^1_{\mathbf{D_i}}, \mathbf{X}^2_{\mathbf{D_i}}) + KL(\mathbf{X}^1_{\mathbf{V_i}}, \mathbf{X}^2_{\mathbf{V_i}}) \tag{8}$$

Where $KL(\mathbf{X}^1_{\mathbf{H_i}}, \mathbf{X}^2_{\mathbf{H_i}})$, $KL(\mathbf{X}^1_{\mathbf{D_i}}, \mathbf{X}^2_{\mathbf{D_i}})$, $KL(\mathbf{X}^1_{\mathbf{V_i}}, \mathbf{X}^2_{\mathbf{V_i}})$ and $KL(\mathbf{X}^1_{\mathbf{A_L}}, \mathbf{X}^2_{\mathbf{A_L}})$ are calculated using the Eq.(7)

Table of contents



Before change | After change | Binary change (ground truth)

Figure : Data and ground truth for the Nyiragongo volcanic eruption of January 2002.

Experiments

- The input image is decomposed into $L = (1, 2, 3)$ scales using discrete stationary wavelet transform (SWT) with a Daubechies filter bank.
- For each coefficient magnitude of each scale different sliding windows with size (33, 39, 45, 51, 57) are used for performance evaluation
- The two models both 1D GD and $(k = 9)$D MGD are estimated with different window sizes in spatial domain.
- The third method referred as SWT9D is applied at each scale of the wavelet domain
- The receiver operating characteristic (ROC) curve is used and the area under ROC curve (AUC) is computed as a performance measure.
- The ROC curve is the evolution of the true positive rate (TPR) as function of false positive rate (FPR)
- The area under curve (AUC) is a good indicator of change. The larger the area the better the performance

LISTIC

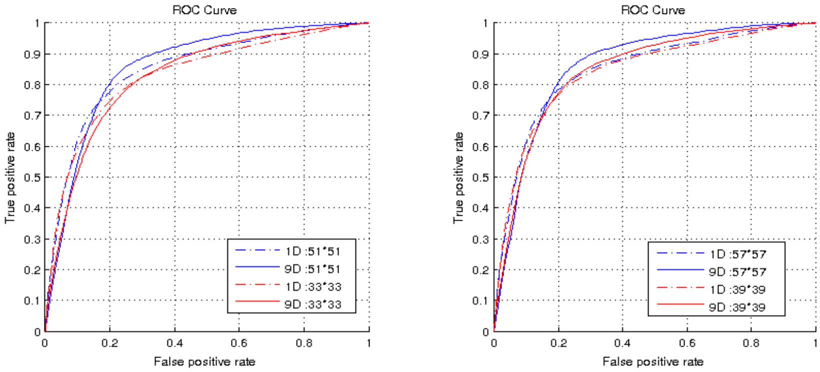


Figure : ROC curve comparison between 1D and 9D

- The ROC Curve of 1D is above the ROC curve of 9D for low FPR values and is below for important FPR values.
- This behavior is the same for all window sizes.
- The difference between the two curves is significant when the window size increases especially for important FPR values.

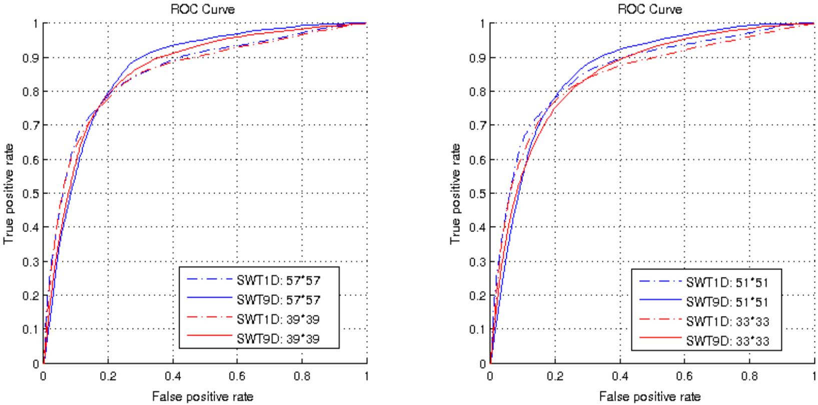


Figure : ROC curve comparison between SWT1D and SWT9D, $L = 3$

- The same behavior is shown in this figure comparing the ROC curve of SWT(k = 1)D and SWT9D

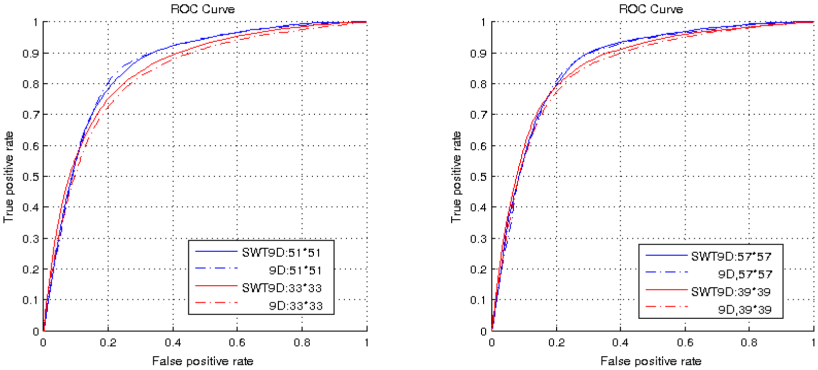


Figure : ROC curve comparison between SWT9D and 9D, $L = 3$

- The figure shows a comparison between SWT9D and 9D. As we can see, the difference is minimal for important sliding windows 51×51 and 57×57.
- For the other window sizes, the ROC curve of SWT9D is always above the ROC curve of 9D.

Scale	Method	27	33	39	45	51	57
1	SWT1D	0.8196	0.8374	0.8467	0.8509	0.8507	0.8484
	SWT9D	0.7927	0.8259	0.8491	0.8517	0.8601	0.8626
2	SWT1D	0.8209	0.8385	0.8487	0.8542	0.8552	0.8535
	SWT9D	0.8051	0.8358	0.8569	0.8575	0.8623	0.8651
3	SWT1D	0.8278	0.8418	0.8504	0.8560	0.8577	0.8566
	SWT9D	0.8222	0.8464	0.8597	0.8593	0.8589	0.8660
	1D	0.8160	0.8323	0.8426	0.8481	0.8499	0.8488
	9D	0.8044	0.8295	0.8478	0.8534	0.8593	0.8622

Table : The Area Under Curve (AUC) for different window size and different scales are measured for 1D, 9D, SWT1D and SWT9D. The best values are marked by red color and the worst by green color. Daubechies wavelets (db1) are used in this study.

- We can see clearly that the SWT($k = 9$)D is always the best for any window size. On the other hand, 1D gives the worst when the window size is bigger than 33×33.
- This can be explained by the fact that texture is better characterized in wavelet domain than in spatial domain.
- We can see that as the window size increases, the AUC always increases.
- For fixed window size, the AUC increases as the number of scales increases. The AUC of 9D is quite higher than that of 1D. This can be explained that spatial information given by texture is better characterized by a multivariate statistical model than an univariate statistical model.
- Based on this table, we conclude that the best window size and the best scale are 57×57 and $L = 3$, respectively.

Table of contents

LISTIC

Conclusion and perspective

- Change detection method in wavelet domain is proposed.
- Probability density function of each sliding windows of the coefficient magnitudes of each subband is assumed to be multivariate Gaussian distribution.
- The total Kullback-Leibler divergence is the sum of the Kullback-Leibler of each subband.
- Our approach is evaluated using different window sizes and different scales compared with the univariate Gaussian distribution.
- Through the study, the multivariate Gaussian distribution in wavelet domain shows promising results comparing to the conventional approach as the univariate Gaussian distribution.
- Improvement can be achieved by including other multivariate distributions as the multivariate Generalized Gamma distribution and multivariate generalized Gaussian distribution.