

# The Multiscale Change Profile: a Statistical Similarity Measure for Change Detection in Multitemporal SAR Images

Jordi INGLADA\* and Grégoire MERCIER†

\*CNES - DCT/SI/AP - BPI 1219  
18, avenue Edouard Belin,  
31401 Toulouse Cedex 09 – France.

†GET/ENST Bretagne, (CNRS UMR 2872) dpt ITI,  
Technopôle Brest-Iroise, CS 83818,  
F-29 238 Brest Cedex – France.

**Abstract**—In this paper, we present a new similarity measure for automatic change detection in multitemporal SAR images. This measure is based on the evolution of the local statistics of the image between two dates. The local statistics are estimated using a cumulant-based series expansion which approximates the probability density functions in the neighborhood of each image pixel. The degree of evolution of the local statistics is measured using the Kullback-Leibler divergence. An analytical expression for this detector is given allowing a simple computation which depends only on the 4 first statistical moments of the pixels inside the analysis window.

The concept of multiscale change profile (MCP) is also introduced and its optimized implementation is presented. MCP yields change information on a wide range of scales and better characterizes the appropriate scale to be used for the detection. Two simple examples of application show that the MCP allows the design of change indicators which provide better results than a monoscale analysis.

## I. INTRODUCTION

Remote sensing imagery is a precious tool for rapid mapping applications. In this context, one of the main uses of the remote sensing is the detection of changes occurred after a natural or anthropic disaster. Since they are abrupt and seldom predictable, these events can not be well observed by the polar orbit satellites which provide the medium, high and very high resolution imagery needed for an accurate analysis of the land cover. Therefore, the rapid mapping is often produced by detecting the changes between an acquisition after the event and available archive data.

This change detection procedure is made difficult due to the time constraints imposed by the emergency context. Indeed, the first available acquisition, after the event, is more likely to be a radar one, due to weather constraints. Moreover, the changes of interest are all mixed up with *normal changes*, which can be majority if the time gap between the two acquisitions is too wide.

In the case of radar acquisitions, the standard detector is based on the ratio of local means [1]. This detector is robust to speckle noise, but it is limited to the comparison of first order statistics. More information may be extracted from the comparison of the local probability density functions (pdfs). The estimation of pdfs can be made with different approaches, but a high resolution change map requires small analysis

window sizes. This excludes the use of histogram computation. Several eligible approaches for this estimation are presented, based on the local statistics up to order 4.

Once the pdfs are estimated, their comparison can also be performed using different criteria. It appears that the Kullback-Leibler (KL) divergence is superior to the classical detector when the pdfs are correctly estimated.

Therefore, these measures will be based on the comparison of local neighborhoods where an analysis window for the computation of the local estimation of probabilities is used. The main point of the problem is how to choose the largest window size which robustly detects the changes but which is small enough to preserve the resolution of the final map without miss-detections.

We propose to use multiscale change profiles, which are defined as the change indicator for each pixel in the image as a function of the analyzing window size. We present here an efficient recursive computation method which is usable on operational context.

## II. PROBLEM FORMULATION

Let us consider two co-registered SAR intensity images  $I_X$  and  $I_Y$  acquired at two different dates  $t_X$  and  $t_Y$  respectively. Our objective is to produce a map representing the changes occurred in the scene between  $t_X$  and  $t_Y$ . The final goal of a change detection analysis is to produce a binary map corresponding to the two classes: *change* and *no change*. The problem can be decomposed into two steps: the generation of a change image and the thresholding of the change image in order to produce the binary change map.

The overall performances of the detection system will depend on both, the quality of the change image and the quality of the thresholding. In this work, we choose to focus on the first step of the procedure, that is, the generation of an indicator of change for each pixel in the image.

## III. DISTANCE BETWEEN PROBABILITY DENSITIES

The main difficulty in the multitemporal analysis of SAR images is the presence of speckle noise. When moving away from interferometric configurations, the speckle is different from one image to the other and it can induce a high number of

false alarms in the change detection procedure. Because of the multiplicative nature of speckle, the classical approach consists in using the ratio of the local means in the neighborhood of each pair of homologous pixels. The Mean Ratio Detector, MRD may be defined as:

$$r_{MRD}(X, Y) = 1 - \min \left\{ \frac{\mu_X}{\mu_Y}, \frac{\mu_Y}{\mu_X} \right\}, \quad (1)$$

where  $\mu_X$  and  $\mu_Y$  are the local mean values of the images before and after the event of interest. The logarithm of eq. (1) is often used also. This detector assumes that a change in the scene will appear as a modification of the local mean value of the image. If the change preserves the mean value but modifies the local texture, it will not be detected.

#### A. Kullback-Leibler divergence

Let  $f_X$  and  $f_Y$  be the pdfs of the random variables  $X$  and  $Y$ . The KL divergence from  $Y$  to  $X$  is given by:

$$K(Y|X) = \int \log \frac{f_X(x)}{f_Y(x)} f_X(x) dx. \quad (2)$$

The measure  $\log \frac{f_X(x)}{f_Y(x)}$  can be thought of as the log-likelihood ratio in a *on-line* change detection process between  $f_X$  and  $f_Y$ . This divergence appears to be an appropriate tool to detect changes as soon as we consider that changes on ground induce different shape of the local pdf. This measure is positive but not symmetric:  $K(Y|X) \neq K(X|Y)$ . One can define a symmetric version by writing:  $D(X, Y) = D(Y, X) = K(Y|X) + K(X|Y)$ , that we will call KL *distance*, KLD.

In order to estimate the KLD, we need to know the pdfs of the two variables to be compared. Several approaches are investigated. They allow the estimation of the pdfs by using a small set of samples as a trade-off between the quality of estimation and the high-resolution change map which is required.

#### B. KLD using the Pearson system

It is now accepted that the statistics of SAR images can be well modeled by the family of probability distributions known as the Pearson system [2], [3]. It is composed of eight types of distributions among which one can find, for instance, the Gaussian and the Gamma distributions. The Pearson system is very easy to use since the type of distribution can be inferred from the following parameters:  $\beta_{X;1} = \frac{\mu_{X;3}^2}{\mu_{X;2}^3}$  and  $\beta_{X;2} = \frac{\mu_{X;4}}{\mu_{X;2}^2}$ , where  $\mu_{X;i}$  is the centered moment of order  $i$  of variable  $X$ .

The Pearson-based KL detector (PKLD) was originally introduced in [4]. It does not have a unique analytic expression, since the type of pdf has to be estimated for each pixel in the image. Therefore, 64 different possibilities for the couples of pdf exist. Once the couple of pdfs is identified, the detection can be performed by numerical integration:

$$r_{PKLD}(X, Y) = \int \left[ \log \left( \frac{f_X^{\beta_{X;1}, \beta_{X;2}}(x)}{f_Y^{\beta_{Y;1}, \beta_{Y;2}}(x)} \right) f_X^{\beta_{X;1}, \beta_{X;2}}(x) + \log \left( \frac{f_Y^{\beta_{Y;1}, \beta_{Y;2}}(x)}{f_X^{\beta_{X;1}, \beta_{X;2}}(x)} \right) f_Y^{\beta_{Y;1}, \beta_{Y;2}}(x) \right] dx. \quad (3)$$

Nevertheless, the Pearson system can be restricted to the Gaussian case. It yields a Gaussian KL detector (GKLD), which takes an expression that depends on the mean ( $\mu$ ) and variance ( $\sigma^2$ ) only:

$$r_{GKLD}(X, Y) = \frac{\sigma_X^4 + \sigma_Y^4 + (\mu_X - \mu_Y)^2(\sigma_X^2 + \sigma_Y^2)}{2\sigma_X^2\sigma_Y^2} - 1. \quad (4)$$

#### C. Cumulant-based KL approximation

Instead of considering a parameterization of a given density, or set of densities, it may be of interest to *describe* the shape of the distribution. The cumulants are interesting to *describe* such a shape: third order ( $\kappa_3$ ) is linked to the symmetry (*i.e.* skewness), while the fourth ( $\kappa_4$ ) to the flatness (*i.e.* kurtosis).

Let us assume that the density to be approximated is not *too far* from a Gaussian pdf (denoted as  $\mathcal{G}_X$  to underline the fact that it has the same mean and variance as  $X$ ). The Edgeworth series gives an approximation of a pdf by using the cumulants and the Chebyshev-Hermite polynomials of order  $r$ , here up to 6 [5]:

$$f_X(x) = \left( 1 + \frac{\kappa_{X;3}}{6} H_3(x) + \frac{\kappa_{X;4}}{24} H_4(x) + \frac{\kappa_{X;5}}{120} H_5(x) + \frac{\kappa_{X;6} + 10\kappa_{X;3}^2}{720} H_6(x) \right) \mathcal{G}_X(x). \quad (5)$$

The Edgeworth series expansion of the two pdfs  $f_X$  and  $f_Y$  may be introduced into the KL divergence (eq. (2)). It yields an approximation of the KL divergence by Edgeworth series, truncated at a given order. In [6], such an approximation has been given up to order 4.  $KL_{\text{Edgeworth}}(X, Y)$  depends on the first four cumulants of  $X$  and  $Y$  only. Finally, the cumulant-based KL detector (CKLD) between two observations  $X$  and  $Y$  is written as:

$$r_{CKLD}(X, Y) = KL_{\text{Edgeworth}}(X, Y) + KL_{\text{Edgeworth}}(Y, X). \quad (6)$$

The reader may pay attention to the fact that, as for the Pearson-based detector, the moments up to order 4 have to be computed only.

## IV. MULTISCALE CHANGE PROFILE

Scale plays a strategic role in image analysis in general and for change detection purposes in particular. Instead of applying a multiscale analysis of the change image, we propose here to produce a multiscale change indicator, that we will call the multiscale change profile, MCP.

The multiscale term refers here to the size of the analyzing window. The MCP will therefore consist in computing the change indicator for a pixel using neighborhoods of increasing sizes. The so-called profile corresponds to the sequence of change measures as a function of the scale.

#### A. Optimized computation of the MCP

Let's consider the following problem: how to update the moments when a  $N + 1^{\text{th}}$  observation  $x_{N+1}$  is added to a set of observations  $\{x_1, x_2, \dots, x_N\}$  already considered. When considering raw moments of order  $r$ , the formulation comes

easily as:  $\mu'_{r,[N+1]} = \frac{N}{N+1}\mu'_{r,[N]} + \frac{1}{N+1}x_{N+1}^r \cdot \mu'_{r,[N]}$  (resp.  $\mu'_{r,[N+1]}$ ) is the raw moment of order  $r$  estimated with  $N$  samples (resp.  $N + 1$  samples). Since the analyzing window may contain textured areas, the mean value itself may be modified by the increase of the number of samples. Therefore, by using simple binomial properties, it can be shown that central moments may be characterized by:

$$\begin{aligned} \mu_{1,[N]} &= \frac{1}{N} s_{1,[N]} \\ \mu_{r,[N]} &= \frac{1}{N} \sum_{\ell=0}^r \binom{r}{\ell} (-\mu_{1,[N]})^{r-\ell} s_{\ell,[N]}, \end{aligned} \quad (7)$$

where the notation  $s_{r,[N]} = \sum_{i=1}^N x_i^r$  has been used.

Hence, when considering a new sample  $x_{N+1}$ , each moment may be updated directly by using updates of  $s_{1,[N+1]}$  and  $s_{r,[N+1]}$  for increasing value of order  $r$ . Then, Edgeworth series is updated also by transforming moments to cumulants to be introduced in eq. (6).

In fact, the availability of updating the estimation of the distance between distributions from windows of any size without re-processing the overall data is the most interesting point for multiscale change detection purpose. This on-line multiscale moment estimation is the key for the usability of the MCP concept.

For example, the computation of  $r_{CKLD}$  for windows with sizes ranging from  $5 \times 5$  pixels to  $51 \times 51$  pixels (22 different scales) takes only 42% additional time with respect to the computation of a single detection with a window of size  $29 \times 29$  pixels (300 s. versus 210 s. for a  $800 \times 400$  pixel image).

### B. MCP exploitation

The MCP computation produces a multichannel image (one scale per channel). Exploitation of the MCP is illustrated by following two approaches: the first one consists in choosing the *best* scale for each image pixel (*i.e.* the scale which gives the highest distance value). The second one consists in *fusing* the information available at all scales in order to provide a single change value (here, the first principal component of the MCP multichannel image is selected).

## V. EXPERIMENTAL RESULTS

We show here an example of application of these algorithms to a real case. A pair of Radarsat images before and after the eruption of the Nyiragongo volcano (D.R. of Congo), which occurred in January 2002, have been used. Fig. 1 shows the two images to be compared and a change map produced using ground measures. The area at the bottom right corner of the ground truth mask corresponds to an area where a severe misregistration exists due to the lack of a proper digital terrain model.

### A. Measure comparison

Fig. 2 shows the result obtained with the classical image intensity ratio approach and the methods based on the KL distance. Fig. 3 gives the ROC plots using the ground truth of fig. 1(c). It shows that the use of KL approximation by

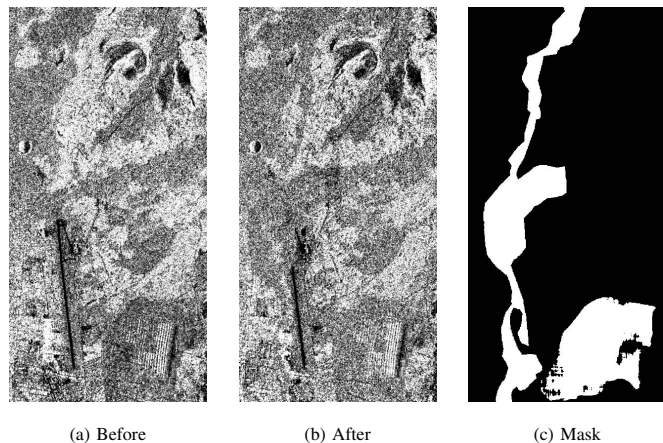


Fig. 1. Data and ground truth for the Nyiragongo volcanic eruption of January 2002.

the Edgeworth series outperforms any other methods such as model-based (Gaussian-based or Pearson-based) KL distance, or the ratio measure. The ratio criterion (which uses only the mean pixel values) is not always worse than pdf-based criteria. That proves that a density model has to suit to the data in order to yield relevant results.

This point confirms that it is more interesting, for operational use, to consider a more flexible pdf approximation by using Edgeworth series instead of a pdf parameterization. The cumulant-based approximation may give equivalent results as the Pearson-based approximation if the estimated cumulants correspond to a pdf belonging to the Pearson's system of distributions. If cumulants of order 3 and 4 vanish, the Edgeworth series is equivalent to a Gaussian model. If the variance of  $X$  and  $Y$  are equivalent, the Edgeworth series yields the same results as the ratio measure. However, when the local observations  $X$  and  $Y$  to be compared do not fit an *a priori* model, Edgeworth series becomes the only usable tool.

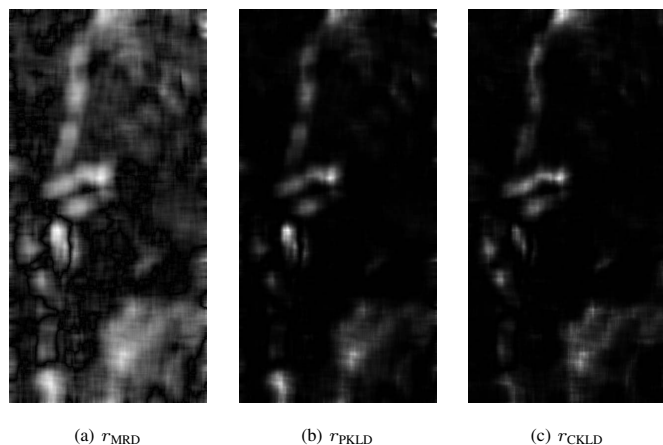


Fig. 2. Change detection: comparison between the image different change indicators using the same window size ( $35 \times 35$  pixels).

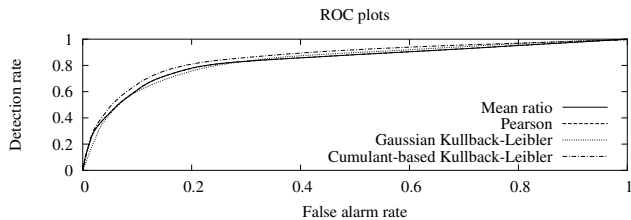


Fig. 3. ROC plots for the different detectors. The cumulant-based KL detector outperforms all other detectors. The Pearson-based detector gives identical results as the classical mean ratio. The Gaussian-based detector shows the worse behavior.

### B. Multiscale change profile

1) *Supervised analysis of the MCPs*: fig. 4 presents some collected multiscale change profiles issued from  $r_{CKLD}$  of eq. (6) applied to our data set. It appears that the detection is well contrasted when the *change* area is larger than the analysis window, whatever its size. There is a high contrast between numerical values of the *change* profile and the *no change* profile, at each scale. When, the *change* area becomes smaller to the analyzing window at some scale, the change profile shows a local maximum that correspond the most appropriate window size to Analyse the local change. Some times, a global maximum may be found on the smallest scales. This corresponds to a bad estimation of the moments due to a too small window. This best scale is signal dependent and varies from the smaller window size (here  $5 \times 5$  to  $50 \times 50$ ). Nevertheless, the shape of the multiscale change profile may help to reduce some false alarms.

2) *Multiscale change indicators*: The interests of MCP are illustrated with window sizes ranging from  $29 \times 29$  to  $51 \times 51$ . In order to select the appropriate analysis window for each pixel in the image, we will choose the maximum of the MCP. It can be shown that the appropriate analysis window can be found at any size from  $29 \times 29$  to  $51 \times 51$ . The ROC plots of fig. 5 show that this simple strategy improves the results with respect to the case where the  $35 \times 35$  window was used.

As an approach to multiscale fusion, we propose here to use

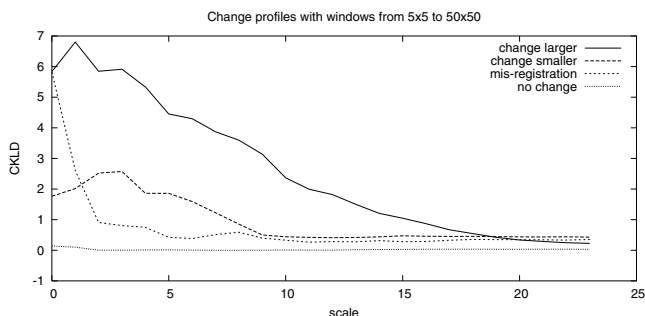


Fig. 4. Typical examples of multiscale change profiles issued from the Edgeworth approximation of the KL distance by using windows of size  $5 \times 5$  to  $50 \times 50$ .

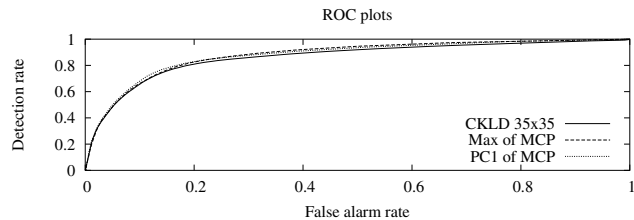


Fig. 5. ROC plots for two possibilities of MCP exploitation: the maximum and the first principal component. They outperform the CKLD for a fixed window size of  $35 \times 35$ .

the first principal component of the stack of multiscale detection images. The ROC plot of figure 5 shows that this approach provides also better performances than the monoscale detector.

## VI. CONCLUSION

In this paper a new similarity measure between images has been introduced in the context of multitemporal SAR image change detection. This measure is based on the use of the cumulant-based series expansion of the local image statistics combined with the KL divergence. The concept of multiscale change profile has been developed and a fast and efficient implementation has been proposed. Finally, two simple approaches for the production of change images containing multiscale information have been presented. The first one is based on the selection of the scale which gives the highest change indicator, and the second one uses the first principal component of the multiscale change image stack.

The proposed original cumulant-based detector has been showed to outperform all other detectors in terms of receiving operator characteristics. The two simple schemes for the exploitation of the multiscale change profile provide better performances than the monoscale detector.

The main advantages of the proposed approach are the following: our detector needs only the computation of the first 4 statistical moments and can deal with a great variety of pdfs; the multiscale change profile provides change information over a wide range of scales at a very small computation cost.

## REFERENCES

- [1] E. J. M. Rignot and J. J. van Zyl, "Change Detection Techniques for ERS-1 SAR Data," *IEEE Trans. Geosci. Remote Sensing*, vol. 31, no. 4, pp. 896–906, July 1993.
- [2] Y. Delignon, R. Garello, and A. Hillion, "Statistical modelling of ocean SAR images," *IEE Proc. on Radar, Sonar and Navig.*, vol. 44, no. 66, pp. 348–354, 1997.
- [3] N. Johnson and S. Kotz, *Distributions in statistics: continuous univariate distributions*. Wiley Interscience, 1969.
- [4] J. Inglada, "Change detection on SAR images by using a parametric estimation of the Kullback-Leibler divergence," in *IGARSS*, Toulouse, France, July, 21–25 2003.
- [5] A. Stuart and J. K. Ord, *Kendall's advanced theory of Statistics*, 5<sup>th</sup> ed. Edward Arnold, 1991.
- [6] J. Lin, N. Saito, and R. Levine, "Edgeworth approximation of the Kullback-Leibler distance towards problems in image analysis," University of California, Davis, Tech. Rep., 1999, <http://www.math.ucdavis.edu/~saito>.