Copula-based Stochastic Kernels for Abrupt Change Detection

Grégoire MERCIER*, Stéphane DERRODE†, Wojciech PIECZYNSKI‡, Jean-Marie NICOLAS§, Annabelle JOANNIC-CHARDIN‡, Jordi INGLADA¶

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

†EGIM / Institut Fresnel (CNRS UMR 6133) team GSM Domaine universitaire de Saint Jérôme, F-13013 Marseille Cedex 20 - France.

‡GET / INT (CNRS UMR 5157) dpt CITI 9, Rue Charles Fourier, 91011 Evry Cedex - France.

§GET / ENST (CNRS UMR 5141) dpt TSI 46, rue Barrault, 75634 Paris Cedex 13 - France.

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

***Abstract*— This paper shows how to obtain a binary change map from similarity measures of the local statistics of images *before* and *after* a disaster. The decision process is achieved by the use of a ν-SVM in which a stochastic kernel has been defined. Stochastic kernel includes two similarity measures, based on the local statistics, to detect changes from the images: 1) a distance between marginal probability density functions (pdfs) and 2) the mutual information between the two observations. Distance between marginal pdfs is evaluated by using a series expansion of the Kullback-Leibler distance. It is achieved by estimating cumulants up to order 4 from a sliding window of fixed size. Mutual information is estimated through a parametric model that is issued from the copulas theory. It is based on rank statistics and yields an analytic expression, that depends on the parameter of the copula only, to be evaluated to obtain the mutual information. Preliminary results are shown on a pair of Radarsat images acquire before and after a lava flow. A ground truth allows to show the accuracy of the stochastic kernels and the SVM decision.**

I. Introduction

Abrupt change detection in operational use requires, from the techniques that have been proposed in the literature, to be able to process data within a limited time of computation, while the conditions of acquisition may be different. Several papers have shown that the use of local statistics appears to be relevant for radar image and multi-sensor data processing [1, 2, 3, 4, 5].

From local statistics, change indicators have to be defined in order to detect or, at least, to give contrasted output on abrupt changes while remaining robust to normal changes between the two acquisitions. Normal changes may be induced by the normal evolution of landscape as well as by the different modalities of acquisition of the two observations. Moreover, no model may be applied for a parametric detection of those changes.

Two images I_1 and I_2, acquired before and after a disaster, are considered. Our goal is to detect abnormal changes between the two observations while remaining robust to *normal changes*. Local similarity measures are used as change indicator. Through a sliding window of fixed size, the neighborhood of the current pixel is associated to a random variable (RV): namely X_1 (resp. X_2) for pixels of I_1 (resp. I_2).

In this study, two measures have been investigated and used together: 1) distance between local pdfs that give better performances than the log ratio detector, as it is not limited to changes of the first order statistics only. The distance between distributions is evaluated by using cumulant expansion of the Kullback-Leibler divergence. It avoids local histograms estimation, which is computationally demanding and subject to variability when using small window size. 2) The distance between independence is used for its better point of view than the correlation measure, since no linear relationship may be found between SAR images when not acquired on interferometric conditions. In that case, a parametric model of dependency has been applied to yield a parametric expression of the local mutual information. Such a parametric model, which is based on the copulas theory [6], prevents from the use of large window size for parameter estimation.

A decision process has to be applied to yield a binary change map. Nevertheless, the change area may not be statistically representative in comparison to the *no change* class. That is why a Support Vector Machines (ν-SVM) [7] is applied here. The similarity measures have been included into the kernel, so-called stochastic kernels, to yield a change detection map.

II. Similarity Measure for change detection

A. Comparison between X_1 and X_2

Most of relevant change detection techniques are based on the difference operator or, the mean log ratio when using radar images [8, 5]. Nevertheless, on the case of multimodal change detection, some changes may appear through a shade of texture while the mean reflectivity remains similar. Instead of comparing the local means only, a Gaussian model may be used to compare the local pdf. The comparison between two Gaussians may be expressed easily by considering the Bhattacharyya distance [9]. The first two moments of X_1 and X_2 are required only for such a distance which make its estimation fast enough for operational use.

The results of [10] have shown that the Kullback-Leibler distance gave better results by considering the detection ($\mathrm{P_d}$) vs. false alarm ($\mathrm{P_{fa}}$) response. The Kullback-Leibler divergence from X_2 to X_1 is given by:

$$\mathcal{K}(X_2 \| X_1) = \int_{\mathbb{R}} \log \frac{f_{X_1}(x)}{f_{X_2}(x)} f_{X_1}(x) dx, \tag{1}$$

f_{X_1} (resp. f_{X_2}) being the pdf of X_1 (resp. X_2). A symmetric version of eq. (1), as referred to a distance, can be stated by writing:

$$\mathcal{K}(X_1, X_2) = \mathcal{K}(X_1 \| X_2) + \mathcal{K}(X_2 \| X_1). \tag{2}$$

In the general case, *i.e.* non Gaussian, eq. (1) requires the estimation of f_{X_1} and f_{X_2}. But histogram estimation remains a difficult task at local scale, whereas it is time consuming at larger scale. A cumulant-based approximation, up to order 4, is used instead, as stated in [10].

B. Dependence between X_1 and X_2

Correlation is often used to characterize relationships between two RVs. But in the case of radar images, as stated in [4], no linear dependency may be found; hence the correlation is being useless. Mutual information (*i.e.* distance to independence) is defined with the Kullback-Leibler divergence between joint distribution and the two marginal pdfs:

$$\mathcal{I}(X_1, X_2) = \iint_{\mathbb{R}^2} \log \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_1}(x_1) f_{X_2}(x_2)} f_{X_1,X_2}(x_1, x_2) dx_1 dx_2, \tag{3}$$

and is an interesting alternative for dependence characterization [3]. Unfortunately, eq. (3) is even more difficult to be estimated. 2D histogram estimation requires window of larger size than in 1D. Cumulant-based approximations of eq. (3) exist (often used, for instance, for Independent Component Analysis [11]) but are limited to order 4 for symmetric distributions, *not too far* from the Gaussian.

A parametric model of dependency is proposed to estimate eq. (3) through windows of limited size. This parametric model uses the copula theory.

A bivariate copula is any cumulative density function on the unit square with uniform marginal functions [6]:

$$C(u_1, u_2) = \Pr(U_1 \leqslant u_1, U_2 \leqslant u_2).$$

Such functions have the capability of giving an exhaustive description of the dependence between two RVs. Sklar has shown that the link between any continuous joint law F_{X_1,X_2} and its marginal laws F_{X_1}, F_{X_2} is achieved with a copula:

$$F_{X_1,X_2}(x_1, x_2) = C\Big(F_{X_1}(x_1), F_{X_2}(x_2)\Big).$$

Then copulas, also named *dependence functions*, act as a parametric model of the dependence between observations, whatever the marginal distributions [12]. By derivation, the density of the copula may be written as: $c(u_1, u_2) = \frac{\partial^2 C}{\partial u_1 \partial u_2}(u_1, u_2)$, it comes:

$$f_{X_1,X_2}(x_1, x_2) = c\Big((F_1(x_1), F_2(x_2)\Big) f_{X_1}(x_1) f_{X_2}(x_2). \tag{4}$$

Using eq. (3) and (4) together, it comes the interesting property that $\mathcal{I}$ is the entropy of the copula itself whatever the marginal pdfs:

$$\mathcal{I}(X_1, X_2) = \iint_{[0,1]^2} c(u_1, u_2) \log c(u_1, u_2) du_1 du_2. \tag{5}$$

As well as many parametric 1D pdfs exist: Gaussian, Gamma, K distribution, also the Pearson of Fisher system of distributions, many parametric copulas exist: namely Normal, Student's t, Frank's, Clayton's, Plackett's, Raftery's, Farlie-Gumbel-Morgenstern's, Fréchet's, Marchal-Olkin's [6]. The latter is used since it fits the best the *empirical copula* that characterized the non-parametric dependency between images *before* and *after*, comparatively to the others.

The Marchal-Olkin's copula is defined here with one parameter only by:

$$C(u_1, u_2) = \min\left(u_1^{1-\theta} u_2, u_1 u_2^{1-\theta}\right), \quad \theta \in [0, 1] \tag{6}$$

with is derivable on $[0, 1]^2$ but for $u_1 = u_2$:

$$c(u_1, u_2) = \begin{cases} (1-\theta) u_1^{-\theta} & \text{if } u_2 < u_1, \\ (1-\theta) u_2^{-\theta} & \text{if } u_1 < u_2. \end{cases}$$

As it is the case for many copulas, Marchal-Olkin's θ parameter depends on the *Kendall's* τ which is a concordance-discordance rank statistics:

$$\tau = \Pr\left((X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) > 0\right) - \Pr\left((X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) < 0\right),$$

where $(\tilde{X}_1, \tilde{X}_2)$ is a pair of RV of the same law but independent to (X_1, X_2). An empirical estimator of τ is:

$$\tau_{\text{empirical}} = \frac{\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} x_{1,ij} x_{2,ij}}{\binom{N}{2}} \tag{7}$$

with $x_{1,ij} = 1$ if $x_{1,i} \leqslant x_{1,j}$ or -1, and $x_{2,ij} = 1$ if $x_{2,i} \leqslant x_{2,j}$ or -1 elsewhere. When using copula, the Kendall's τ becomes:

$$\tau = 4 \iint_{[0,1]^2} C(u_1, u_2) dC(u_1, u_2) - 1 \text{ and then } \frac{\theta}{2-\theta}, \tag{8}$$

for the Marchal-Olkin's copula.

By using the Marchal-Olkin parametric model, eq. (5) becomes parametric also [13]:

$$\mathcal{I}(X_1, X_2) = 2\frac{1-\theta}{2-\theta} \log(1-\theta) - \frac{\theta}{2-\theta} + \frac{\theta^2}{(2-\theta)^2}. \tag{9}$$

This parameterization is much more accurate than the correlation parameter, while keeping the same level of computational complexity for small windows. Thanks to the copula theory, the estimation of the mutual information requires one parameter only which depends on the copula chosen.

III. Stochastic kernels for change detection

SVM have proved to be a relevant alternative for supervised clustering technique [14]. Kernels have to be defined in order to integrate the measures described previously (Distance between distribution through the Kullback-Leibler distance and Distance to independence estimated with a copula).

Kernels give an orthogonality point of view between two observations $\boldsymbol{x}$ and $\boldsymbol{y}$ projected into the feature space. $\boldsymbol{x}$ and $\boldsymbol{y} \in \mathbb{R}^2$ since they are two observations at two different positions of the images *before* and *after*. In our case, the feature space is defined through the distance between distributions and the distance to independence.

A. A kernel for the marginal pdf

This kernel is an contrast measure between two Kullback-Leibler distances. The kernel is simply defined as an RBF kernel, as:

$$k_m(\boldsymbol{x}, \boldsymbol{y}) = e^{-\gamma \sqrt{\mathcal{K}(X_1, X_2)\mathcal{K}(Y_1, Y_2)}}. \qquad (10)$$

It is not a Kullback-Leibler kernel by itself [15] since it achieves comparison between two Kullback-Leibler distances. It is easy to show that eq. (10) satisfy Mercer's eligibility conditions.

B. A kernel for the dependency

The comparison of dependency could be defined the same way as eq. (10) but it is more relevant to define a stochastic kernel that makes the comparison of the two copulas. It requires the assumption that the margins are identical, but if not, eq. (10), would informs the SVM optimization. By using the Bhattacharyya distance and the Marchal-Olkin model of eq. (6), it yields the simple expression:

$$\begin{aligned}
\mathcal{B}(f_{X_1,X_2}, f_{Y_1,Y_2}) \\
&= \iint_{\mathbb{R}^2} \sqrt{f_{X_1,X_2}(x_1,x_2) f_{Y_1,Y_2}(x_1,x_2)} dx_1 dx_2 \\
&= \iint_{[0,1]^2} \sqrt{c_{X_1,X_2}(u_1,u_2) c_{Y_1,Y_2}(u_1,u_2)} du_1 du_2 \\
&= \sqrt{(1-\theta_{X_1,X_2})(1-\theta_{Y_1,Y_2})} \\
&\qquad \left(\frac{1}{1-\Theta} + \frac{1}{2-\Theta} - \frac{1}{(1-\Theta)(2-\Theta)} \right),
\end{aligned}$$

with $\Theta = \frac{1}{2}(\theta_{X_1,X_2} + \theta_{Y_1,Y_2})$. Since this distance is bounded by 0 and 1, it can be easily integrated into a kernel, such as:

$$k_c(\boldsymbol{x}, \boldsymbol{y}) = e^{-\gamma \mathcal{B}(f_{X_1,X_2}, f_{Y_1,Y_2})}. \qquad (11)$$

The use of the Bhattacharyya distance between two copulas in eq. (11) justify *a posteriori* the use of the square root on eq. (10). This analogy gives also the best decision level, as a trade-off between $\mathrm{P_d}$ versus $\mathrm{P_{fa}}$.

C. Integration into ν-SVM

SVM classifications have been fully describe in previous studies for remote sensing image classification [16, 17]. But never for change detection from remote sensing data. The description remains the same but the kernel used, here, a mixture of the two previous kernels obtained by a linear combination:

$$k(\boldsymbol{x}, \boldsymbol{y}) = \mu k_m(\boldsymbol{x}, \boldsymbol{y}) + (1-\mu) k_c(\boldsymbol{x}, \boldsymbol{y}), \quad \mu \in [0,1]. \quad (12)$$

It is simply reminded here that the optimal separation between *change* and *no change* classes is performed by using the ν-SVM classifier [7]. Considering a set of training samples $\{\boldsymbol{x}_\ell\}_{0<\ell\leqslant L}$ in $\mathbb{R}^2$, associated to their labels $\{y_\ell\}_{0<\ell\leqslant L}$ of value ± 1, the problem is:

$$\min_{\lambda_1,\ldots,\lambda_L \in \mathbb{R}} \sum_{i,j=1}^{L} \lambda_i \lambda_j y_i y_j k(\boldsymbol{x}_i, \boldsymbol{x}_j) \qquad (13)$$

$$\text{subject to} \quad \forall \ell \in [1, L], \quad 0 \leqslant \lambda_\ell \leqslant 1/L$$

$$\sum_{\ell=1}^{L} \lambda_\ell \geqslant \nu \quad \text{and} \quad \sum_{\ell=1}^{L} y_\ell \lambda_\ell = 0.$$

The decision function becomes:

$$f(\boldsymbol{x}) = \operatorname{sgn}\left(\sum_{\ell=1}^{L} \lambda_\ell y_\ell k(\boldsymbol{x_\ell}, \boldsymbol{x}) + b \right),$$

but we will also evaluate the compete accuracy by considering the distance to the hyperplane as a new measure for change detection.

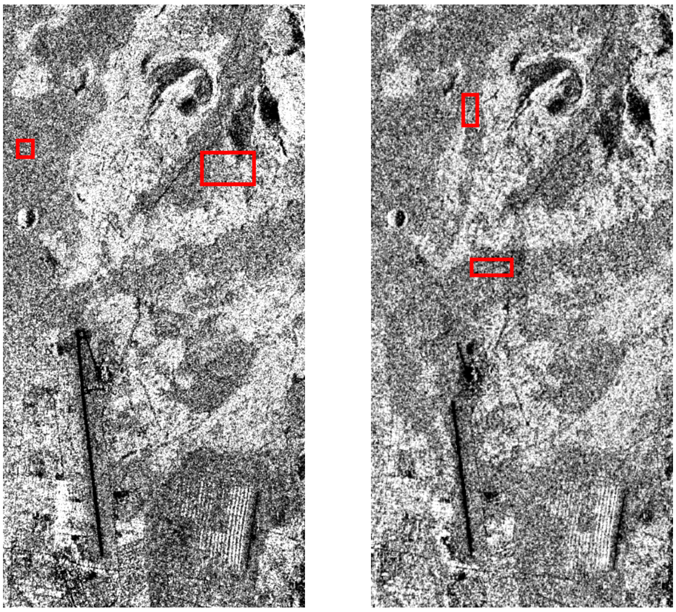
IV. Application

We show an example of application of this algorithm to a real case. A pair of F5 and F2 Radarsat images, acquired 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 with the Region of Interest (ROI) defined to train the support vectors. The images have a ground resolution of 10 m and cover an area of 4×8 km.

Results of the detection have to be seen on fig. 2. The ground truth has been superposed to the distance to hyperplane image for comparison. The area at the bottom right corner of the ground truth mask corresponds to an area where a severe mis-registration exists due to the lack of a proper digital terrain model. By using the distance to hyperplane, it is possible to apply several thresholds and to evaluate the compromise between good detection and false alarms ($\mathrm{P_d}$ versus $\mathrm{P_{fa}}$), since the ground truth is known. Such a result is presented on fig. 3.

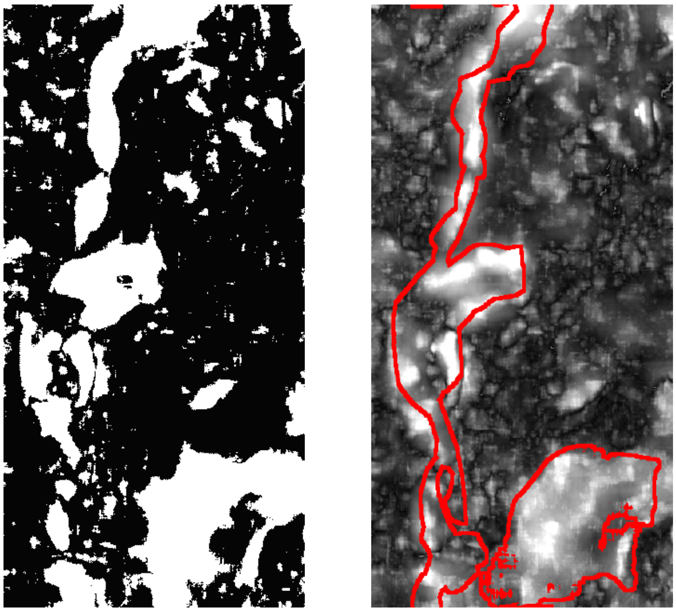
V. Conclusion

This paper has proposed the use of stochastic kernel for smart decision by SVM of similarity measures to detect changes between two images. This stochastic kernel integrates the distance between local pdf, estimated with an Edgeworth series expansion of the Kullback-Leibler distance, and also



before　　　after

Fig. 1: *Before* image on the left: Radarsat F5 mode data. *After* image on the right: Radarsat F2 data. Superposed on red: the training samples to define *no change* class shown on the image *before* (2623 samples), and the training samples of *change* class shown on the *after* image (1258 samples).



SVM change detection　　　ground truth superposed to the distance to hyperplane

Fig. 2: SVM classification on the left, and the distance to hyperplane on the right to be compared to the ground truth, superposed on red. Distance to hyperplane has to be seen as follows: white areas correspond to the *change* side of the hyperplane, grey is the transition area, dark corresponds to the *no change* side of the hyperplane in the feature space.

the distance between copulas as a contrast measure of the mutual information. First results has shown the availability of the technique since it yields a relevant binary change map. Other experiments intend to show its ability for multisensor change detection.

ACKNOWLEDGMENT

This work has been supported by the *Groupe des Écoles des Télécommunications* under the contract DéCA RE-022.

REFERENCES

[1] R. Touzi, A. Lopès, and P. Bousquet, "A statistical and geometrical edge detector for SAR images," *IEEE Trans. Geosci. Remote Sensing*, vol. 26, no. 6, pp. 764–773, Novembre 1988.
[2] J. Inglada, "Change detection on SAR images by using a parametric estimation of the Kullback-Leibler divergence," in *IGARSS*, vol. 6, July 21–25, 2003, pp. 4104–4106.
[3] P. Rogelj, S. Kovačič, and J. C. Gee, "Point similarity measures for non-rigid registration of multimodal data," *Computer Vision and Image Understanding*, vol. 92, pp. 112–140, 2003.
[4] J. Inglada and A. Giros, "On the possibility of automatic multisensor image registration," *IEEE Trans. Geosci. Remote Sensing*, vol. 42, no. 10, pp. 2104–2120, Oct. 2004.
[5] F. Bovolo and L. Bruzzone, "A wavelet-based change-detection technique for multitemporal SAR images," *IEEE Geosci. Remote Sensing Letters*, 2005.
[6] R. B. Nelsen, *An Introduction to Copulas*, ser. Lectures Notes in Statistics. New York: Springer Verlag, 1999, vol. 139.
[7] B. Schölkopf and A. J. Smola, *Learning with Kernels*. Cambridge, MA: The MIT Press, 2002.
[8] L. Bruzzone and D. F. Prieto, "Automatic Analysis of the Difference Image for Unsupervised Change Detection," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, no. 3, pp. 1171–1182, May 2000.
[9] G. Mercier and F. Girard-Ardhuin, "Oil slick detection by SAR imagery using Support Vector Machines," in *Proc. of the IEEE Oceans 2005 - Europe*, Brest, France, June 20–23, 2005.
[10] J. Inglada and G. Mercier, "Multi-scale abrupt change detection in remote sensing images: the multi-scale change profile," in *IGARSS*, Denver, Colorado, July 31–Aug. 04, 2006.
[11] J.-F. Cardoso, "High-order contrasts for Independent Component Analysis," *Neural Computation*, vol. 11, pp. 157–192, 1999.
[12] H. Joe, *Multivariate models and Dependence concepts*, ser. Monographs on Statistics and Applied Probability. Chapman & Hall / CRC, 2001, vol. 73.
[13] G. Mercier, "Mesures de dépendance entre images RSO," GET / ENST Bretagne, Tech. Rep. RR-2005003-ITI, 2005, (in french). [Online]. Available: http//:perso.enst-bretagne.fr/~mercierg
[14] V. N. Vapnick, *Statistical Learning Theory*. John Wiley and Sons Inc., 1998.
[15] P. Moreno, P. Ho, and N. Vasconcelos, "A Kullback-Leibler Divergence based Kernel for SVM Classification in Multimedia Applications," in *Proceedings of Neural Information Processing Systems*, Vancouver, Canada, 2003.
[16] F. Melgani and L. Bruzzone, "Support Vector Machines for Classification of Hyperspectral Remote Sensing Images," in *IGARSS*, 2002.
[17] G. Mercier and M. Lennon, "Support Vector Machines for hyperspectral image classification with spectral-based kernels," in *IGARSS*, 2003.

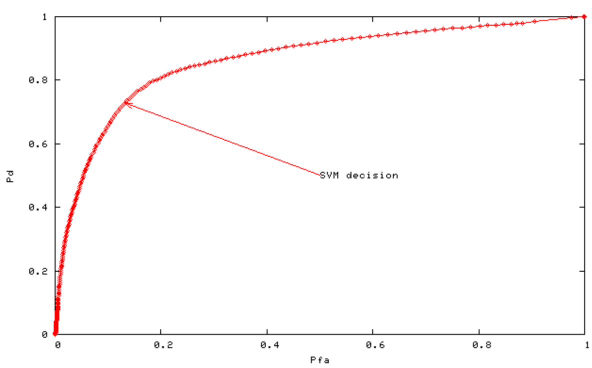


Fig. 3: ROC plots when considering distance to hyperplane as a new change detector. It proved the decision, performed by margin maximization, is almost optimal.