New statistical modeling of multi-sensor images with application to change detection

Jorge PRENDES

Supervisors:
Marie CHABERT, Frédéric PASCAL,
Alain GIROS, Jean-Yves TOURNERET

TéSA Telecommunications for Space and Aeronautics
SONDRA
Supélec
INP ENSEEIHT
cnes CENTRE NATIONAL D'ÉTUDES SPATIALES

October 22, 2015

Outline

Section 1

Introduction

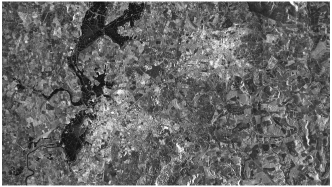
Change Detection for Remote Sensing

Remote sensing images are images of the Earth surface captured from a satellite or an airplane.

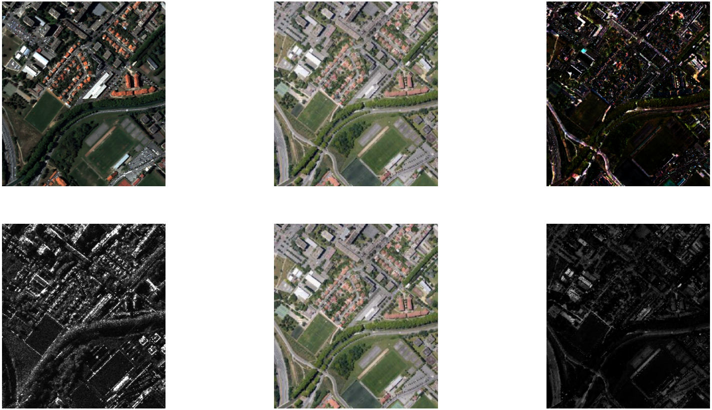
Multitemporal datasets are groups of images acquired at different times. We can detect changes on them!

Heterogeneous Sensors

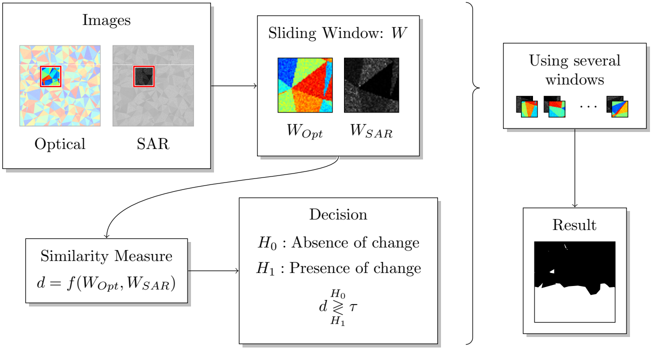
Optical images are not the only kind of images captured.
For instance, SAR images can be captured during the night, or with bad weather conditions.



Difference Image



Sliding window



Similarity measures

Statistical similarity measures

- Measure the dependency between pixel intensities
  - Correlation Coefficient
  - Mutual Information
- Others
  - KL-Divergence

Estimation of the joint pdf

- Non parametric computation
  - Histogram
  - Parzen windows
- Based on a parametric modeling
  - Bivariate gamma distribution [1]
  - Pearson distribution [2]
  - Copulas modeling [3]

[1] F. Chatelain et al. "Bivariate Gamma Distributions for Image Registration and Change Detection". In: IEEE Trans. Image Process. 16.7 (2007), pp. 1796–1806.
[2] M. Chabert and J.-Y. Tourneret. "Bivariate Pearson distributions for remote sensing images". In: Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS). Vancouver, Canada, July 2011, pp. 4038–4041.
[3] G. Mercier, G. Moser, and S. B. Serpico. "Conditional Copulas for Change Detection in Heterogeneous Remote Sensing Images". In: IEEE Trans. Geosci. Remote Sens. 46.5 (May 2008), pp. 1428–1441.

Section 2

Image model

Optical image

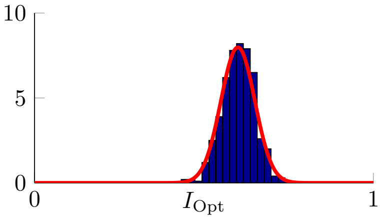
- Affected by additive Gaussian noise

$$I_{\text{Opt}} = T_{\text{Opt}}(P) + \nu_{\mathcal{N}(0,\sigma^2)}$$

$$I_{\text{Opt}} | P \sim \mathcal{N}\left[T_{\text{Opt}}(P), \sigma^2\right]$$

where

- $T_{\text{Opt}}(P)$ is how an object with physical properties P would be ideally seen by an optical sensor
- σ^2 is associated with the noise variance



Histogram of the normalized image

[1] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J.-Y. Tourneret, "A new multivariate statistical model for change detection in images acquired by homogeneous and heterogeneous sensors," IEEE Trans. Image Process., vol. 24, no. 3, pp. 799–812, March 2015.

SAR image

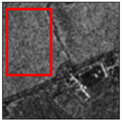
- Affected by multiplicative speckle noise (with gamma distribution)

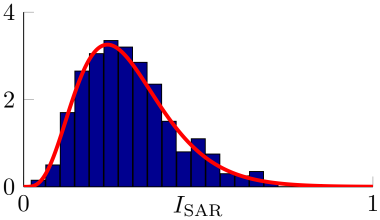
$$I_{\text{SAR}} = T_{\text{SAR}}(P) \times \nu_{\Gamma(L,\frac{1}{L})}$$

$$I_{\text{SAR}}|P \sim \Gamma\left[L, \frac{T_{\text{SAR}}(P)}{L}\right]$$

where

- $T_{\text{SAR}}(P)$ is how an object with physical properties P would be ideally seen by a SAR sensor
- L is the number of looks of the SAR sensor





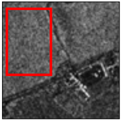
Histogram of the normalized image

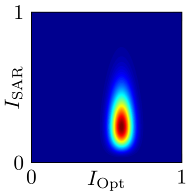
Joint distribution

- Independence assumption for the sensor noises

$$p(I_{\text{Opt}}, I_{\text{SAR}} | P) = p(I_{\text{Opt}} | P) \times p(I_{\text{SAR}} | P)$$

- *Conclusion*
  Statistical dependency (CC, MI) is not always an appropriate similarity measure





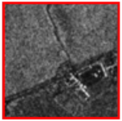
Sliding window

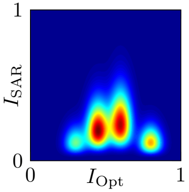
- Usually includes a finite number of objects, K
- Different values of P for each object

$$\Pr(P = P_k | W) = w_k$$

$$\mathrm{p}(I_{\text{Opt}}, I_{\text{SAR}} | W) = \sum_{k=1}^{K} w_k \mathrm{p}(I_{\text{Opt}}, I_{\text{SAR}} | P_k)$$

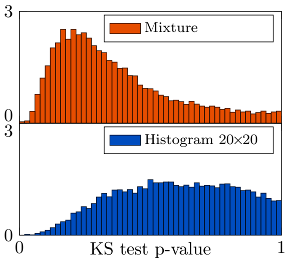
- Mixture distribution!



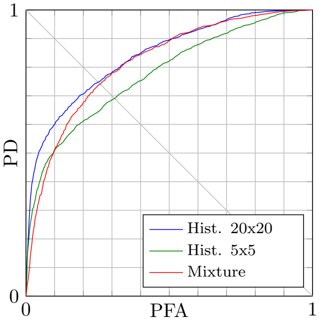


Resulting improvement

Goodness of fit of the image model

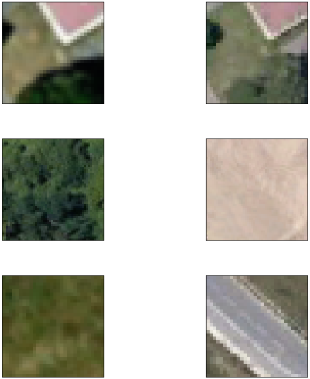


Performance for change detection



Limitation of dependency based measures

Correct detection



Incorrect detection

Section 3

Similarity measure

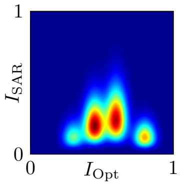
Motivation

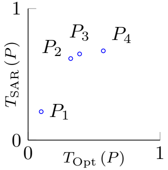
Parameters of the mixture distribution

- Can be used to derive $[T_{\text{Opt}}(P), T_{\text{SAR}}(P)]$ for each object

$$I_{\text{Opt}}|P \sim \mathcal{N}\left[T_{\text{Opt}}(P), \sigma^2\right]$$
$$I_{\text{SAR}}|P \sim \Gamma\left[L, \frac{T_{\text{SAR}}(P)}{L}\right]$$

- Related to P
- They are all related

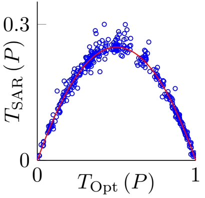




Distance measure

Unchanged regions

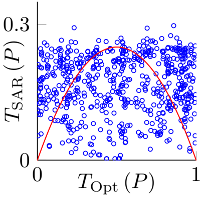
- Pixels belong to the same object
- P is the same for both images
- $\hat{v} = \left[\hat{T}_{\text{Opt}}(P), \hat{T}_{\text{SAR}}(P)\right]$



✓

Changed regions

- Pixels belong to different objects
- P changes from one image to another
- $\hat{v} = \left[\hat{T}_{\text{Opt}}(P_1), \hat{T}_{\text{SAR}}(P_2)\right]$

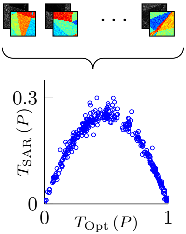


✗

Manifold

- For each unchanged window, $v(P) = [T_{\text{Opt}}(P), T_{\text{SAR}}(P)]$ can be considered as a point on a manifold
- The manifold is parametric on P
- Estimating $v(P)$ from pixels with different values of P will build the manifold

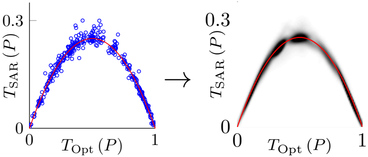
Several unchanged windows



[1] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J.-Y. Tourneret, "A new multivariate statistical model for change detection in images acquired by homogeneous and heterogeneous sensors," IEEE Trans. Image Process., vol. 24, no. 3, pp. 799–812, March 2015.

Manifold estimation

- The manifold is *a priori* unknown
- We must estimate the distance to the manifold
- PDF of $v(P)$
  - Good distance measure
  - Learned using training data from unchanged images

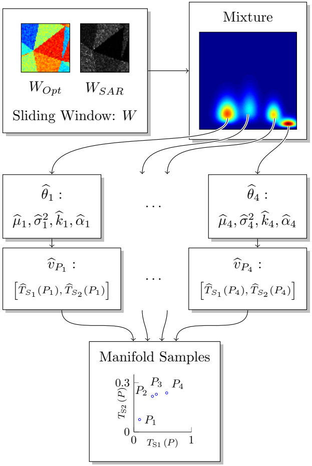


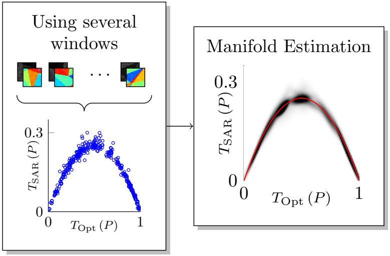
H_0 : Absence of change

H_1 : Presence of change

$$\hat{p}_{\mathbf{v}}(\hat{\mathbf{v}}) \underset{H_0}{\overset{H_1}{\lessgtr}} \tau \equiv \hat{p}_{\mathbf{v}}(\hat{\mathbf{v}})^{-1} \underset{H_1}{\overset{H_0}{\lessgtr}} \frac{1}{\tau}$$

Summary


W_{Opt}
W_{SAR}
Sliding Window: W
Mixture
$\widehat{\theta}_1$:
$\widehat{\mu}_1, \widehat{\sigma}_1^2, \widehat{k}_1, \widehat{\alpha}_1$
. . .
$\widehat{\theta}_4$:
$\widehat{\mu}_4, \widehat{\sigma}_4^2, \widehat{k}_4, \widehat{\alpha}_4$
$\widehat{v}_{P_1}$:
$[\widehat{T}_{S_1}(P_1), \widehat{T}_{S_2}(P_1)]$
. . .
$\widehat{v}_{P_4}$:
$[\widehat{T}_{S_1}(P_4), \widehat{T}_{S_2}(P_4)]$
Manifold Samples
0.3
0
$T_{S2}(P)$
P_1
P_2
P_3
P_4
0
$T_{S1}(P)$
1


Using several windows
. . .
0.3
0
$T_{SAR}(P)$
0
$T_{Opt}(P)$
1
Manifold Estimation
0.3
0
$T_{SAR}(P)$
0
$T_{Opt}(P)$
1

Section 4

Expectation maximization

Motivation

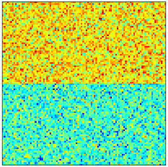
- To estimate $v(P)$ we must estimate the mixture parameters θ
- We can use a maximum likelihood estimator

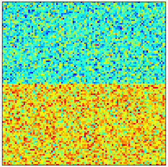
$$\hat{\theta} = \arg\max_{\theta} \mathrm{p}(I_{\mathrm{Opt}}, I_{\mathrm{SAR}} | \theta)$$

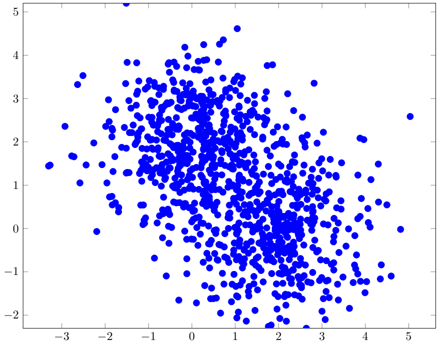
- EM algorithm: find local maxima of the likelihood function
- The value of K is fixed, or estimated heuristically[1]

[1] M. A. T. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 3, pp. 381–396, March 2002.

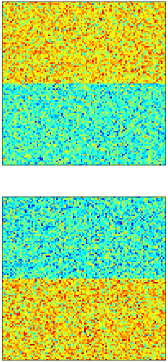
Example

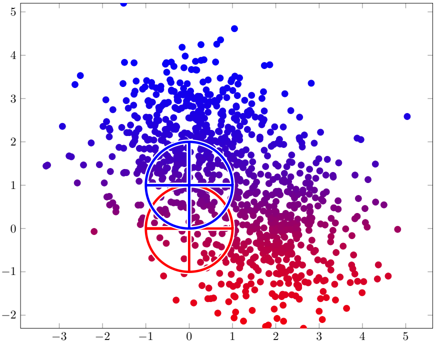




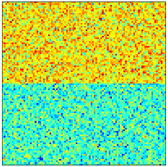


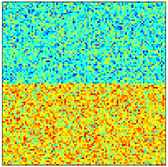
Example

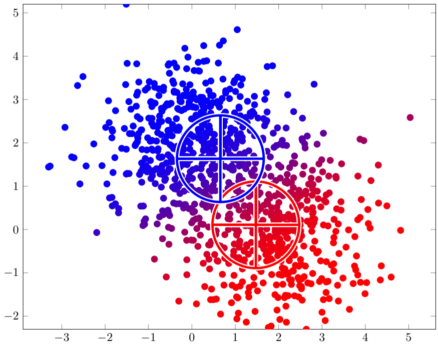




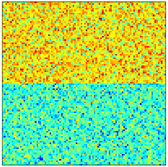
Example

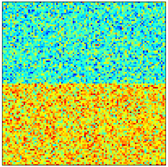


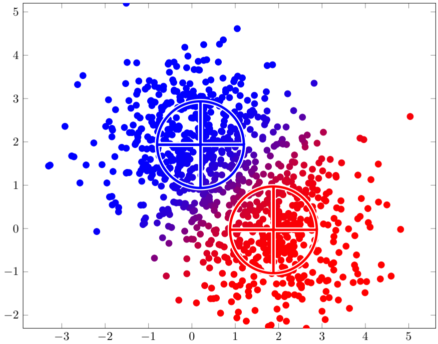




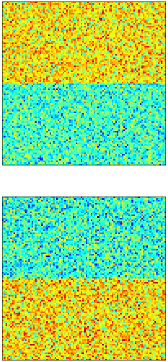
Example

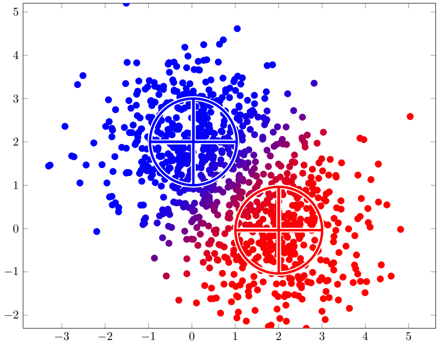




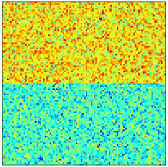


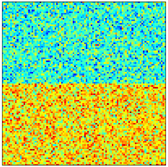
Example

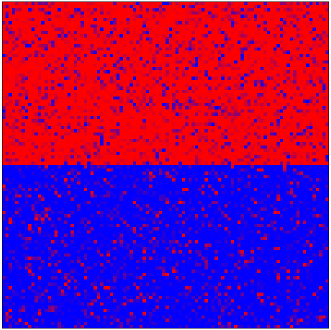




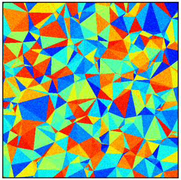
Example

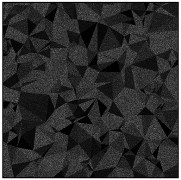






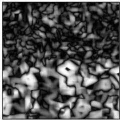
Results – Synthetic Optical and SAR Images

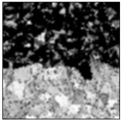

Synthetic optical image

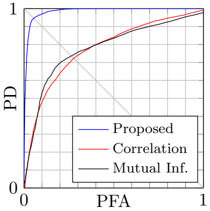

Synthetic SAR image


Change mask


Mutual Information


Correlation Coefficient

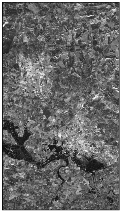

Proposed Method

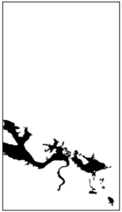


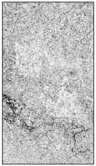
Performance – ROC

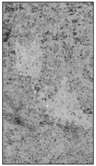
Results – Real Optical and SAR Images

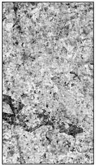

Optical image before the flooding

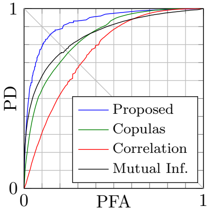

SAR image during the flooding


Change mask


Mutual Information


Conditional Copulas [1]

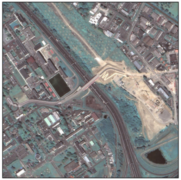

Proposed Method



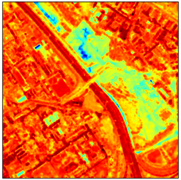
Performance – ROC

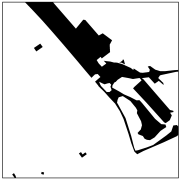
[1] G. Mercier, G. Moser, and S. B. Serpico, "Conditional copulas for change detection in heterogeneous remote sensing images," IEEE Trans. Geosci. and Remote Sensing, vol. 46, no. 5, pp. 1428–1441, May 2008.

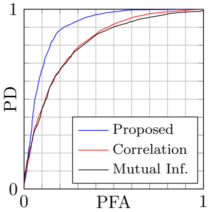
Results – Pléiades Images


Pléiades – May 2012


Pléiades – Sept. 2013


Change map


Change mask

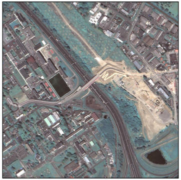


Performance – ROC

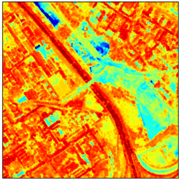
Special thanks to CNES for providing the Pléiades images

[1] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J.-Y. Tourneret, "Performance assessment of a recent change detection method for homogeneous and heterogeneous images", Revue Française de Photogrammétrie et de Télédétection, vol. 209, pp. 23– 29, January 2015.

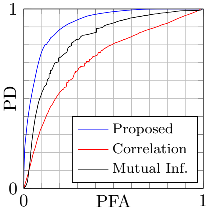
Results – Pléiades and Google Earth Images


Pléiades – May 2012


Google Earth – July 2013


Change map


Change mask



Performance – ROC

Section 5

Bayesian non parametric

Motivation

- Unknown number of objects in an image
- High variability in the expected number of objects (urban vs rural)
- Spatial correlation in images

Proposed solution

- Dirichlet Process Mixture
  - Chinese Restaurant Process prior on the labels
- Markov Random Field prior on the labels
- Jeffreys Prior on the concentration parameter
- Implemented through a Collapsed Gibbs Sampler

[1] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J.-Y. Tourneret, "A Bayesian nonparametric model coupled with a Markov random field for change detection in heterogeneous remote sensing images".

Classic mixture

- Introduce a Bayesian framework into the labels: K is not fixed
- Classic mixture model

$$\boldsymbol{i}_n \big| \boldsymbol{v}_n \sim \mathcal{F}(\boldsymbol{v}_n)$$

$$\boldsymbol{v}_n \big| \boldsymbol{V}' \sim \sum_{k=1}^{K} w_k \delta\left(\boldsymbol{v}_n - \boldsymbol{v}'_k\right)$$

$\boldsymbol{i}_n = \left[i_{\text{Opt},n}, i_{\text{SAR},n}\right]$, and $\mathcal{F}$ is a distribution family which is application dependent, i.e., a bivariate Normal-Gamma distribution.

Bayesian approach

- Prior in the mixture parameters

$$\boldsymbol{v}'_k \sim \mathcal{V}_0$$
$$\boldsymbol{w} \sim \mathrm{Dir}_K(\alpha)$$

- Now make $K \to \infty$
  - $\boldsymbol{v}_n$ will still present clustering behavior
  - There is an infinite number of parameters for the prior of $\boldsymbol{v}_n$

$\mathrm{Dir}_K(\alpha)$ is a K dimensional Dirichlet distribution, with concentration parameter α.

Bayesian non parametric

Dirichlet Process

$$\boldsymbol{i}_n | \boldsymbol{v}_n \sim \mathcal{F}(\boldsymbol{v}_n)$$
$$\boldsymbol{v}_n \sim \mathcal{V}$$
$$\mathcal{V} \sim \mathrm{DP}(\mathcal{V}_0, \alpha).$$

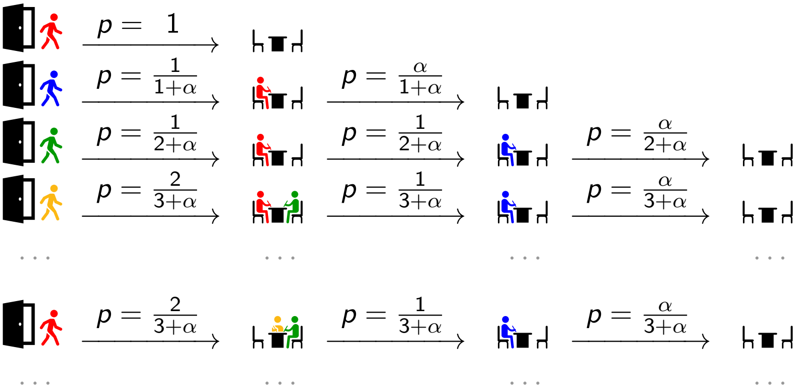
Chinese Restaurant Process

$$\boldsymbol{i}_n | z_n \sim \mathcal{F}\left(\boldsymbol{v}'_{z_n}\right)$$
$$\boldsymbol{z} \sim \mathrm{CRP}(\alpha)$$
$$\boldsymbol{v}'_k \sim \mathcal{V}_0.$$

$$\mathrm{p}_{\mathrm{BNP}}\left(z_n \middle| \boldsymbol{i}_n, \mathcal{V}_0, \boldsymbol{V}'\right) \propto \begin{cases} \alpha \; \mathrm{p}(\boldsymbol{i}_n | \mathcal{V}_0) & \text{if } z_n \text{ is new label} \\ N'_{z_n} \mathrm{p}\left(\boldsymbol{i}_n \middle| \boldsymbol{v}'_{z_n}\right) & \text{if } z_n \text{ is existing label} \end{cases}$$

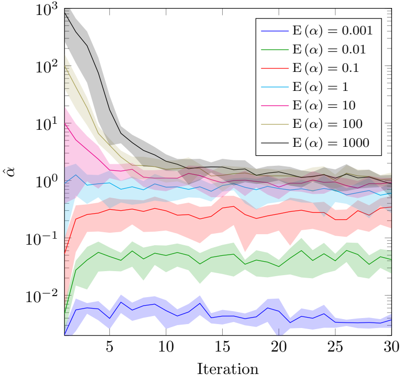
$$\mathrm{p}_{\mathrm{BNP}}\left(z_n \middle| \boldsymbol{z}_{\backslash n}, \boldsymbol{I}, \mathcal{V}_0\right) \propto \begin{cases} \alpha \; \mathrm{p}(\boldsymbol{i}_n | \mathcal{V}_0) & \text{if } z_n \text{ is new label} \\ N'_{z_n} \dfrac{\mathrm{p}\left(\boldsymbol{I}_{\{z_n\}} \middle| \mathcal{V}_0\right)}{\mathrm{p}\left(\boldsymbol{I}_{\{z_n\}\backslash n} \middle| \mathcal{V}_0\right)} & \text{if } z_n \text{ is existing label} \end{cases}$$

Bayesian non parametric



Concentration Parameter

α with Gamma prior proposed in (Escobar 1995, Antoniak 1974)



Concentration Parameter

- Method to define uninformative priors
- α non informative w.r.t. K

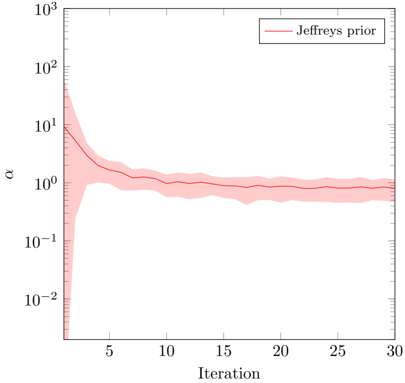
$$p(\alpha|N) \propto \sqrt{E_K\left[\left(\frac{d}{d\alpha}\log p(K|\alpha,N)\right)^2\right]}$$

$$p(\alpha|N) \propto \sqrt{\frac{\Delta\Psi_N^{(0)}(\alpha)}{\alpha} + \Delta\Psi_N^{(1)}(\alpha)}$$

$$\Delta\Psi_N^{(i)}(\alpha) = \Psi^{(i)}(N+\alpha) - \Psi^{(i)}(1+\alpha)$$

- $p(\alpha|K,N)$ rejection sampling from $\text{Gamma}\left(K+\frac{1}{2}, -\frac{1}{\log t}\right)$

Concentration Parameter



α with Jeffreys prior

Markov random fields

- Markov random fields are a common tool to capture spatial correlation
- We would like to define

$$\mathrm{p}\left(z_n \middle| \boldsymbol{z}_{\setminus n}\right) = \mathrm{p}\left(z_n \middle| \boldsymbol{z}_{\delta(n)}\right)$$

- MRF define the constraints to define a joint distribution $\mathrm{p}(Z)$

Markov random fields

- We will define the joint distribution as

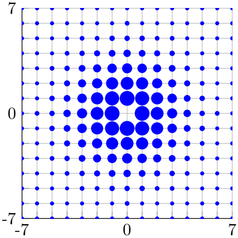
$$p\left(z_n \middle| \boldsymbol{z}_{\backslash n}\right) \propto \exp\left[H\left(z_n \middle| \boldsymbol{z}_{\backslash n}\right)\right]$$

$$H\left(z_n \middle| \boldsymbol{z}_{\backslash n}\right) = H_n(z_n) + \sum_{m \in \delta(n)} \omega_{nm} \mathbf{1}_{z_n}(z_m) = H_n(z_n) + \sum_{\substack{m \in \delta(n) \\ z_n = z_m}} \omega_{nm}$$

- The trick is to take $H_n(z_n) = \log p\left(z_n \middle| I_n, \boldsymbol{V}', \mathcal{V}_0\right)$

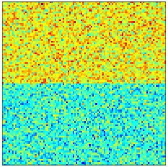
Markov random fields

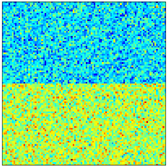
$$p\left(z_n \middle| \boldsymbol{z}_{\backslash n}, \boldsymbol{I}, \mathcal{V}_0\right) \propto \begin{cases} \alpha\ p\left(\boldsymbol{i}_n \middle| \mathcal{V}_0\right) & \text{if } z_n \text{ is new label} \\ N'_{z_n} \frac{p\left(\boldsymbol{I}_{\{z_n\}} \middle| \mathcal{V}_0\right)}{p\left(\boldsymbol{I}_{\{z_n\}\backslash n} \middle| \mathcal{V}_0\right)} \prod_{\substack{m \in \delta(n) \\ z_n = z_m}} e^{\omega_{nm}} & \text{if } z_n \text{ is existing label} \end{cases}$$

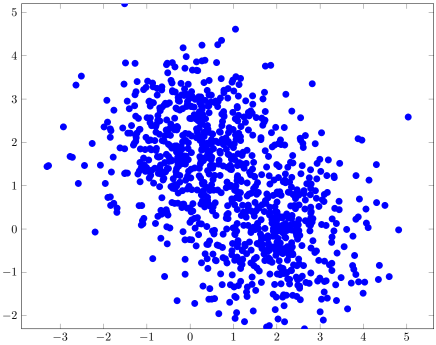


Representation of ω_{nm}

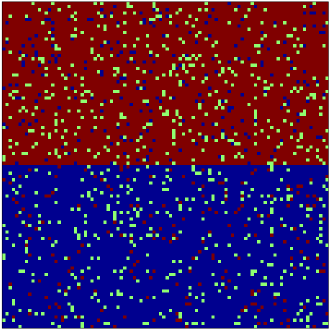
Example

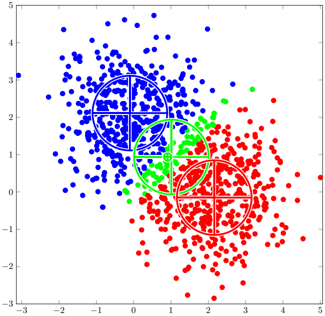




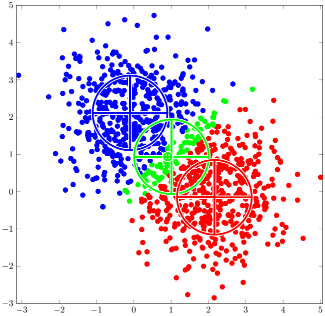


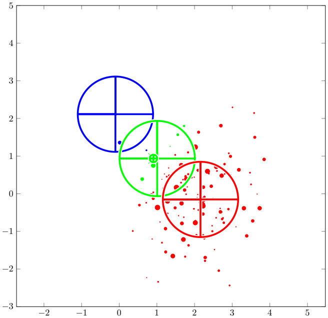
Example



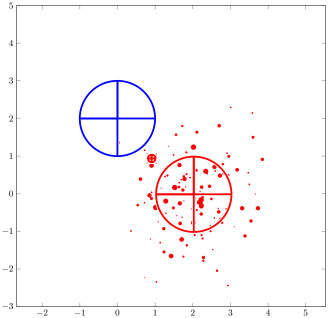


Example



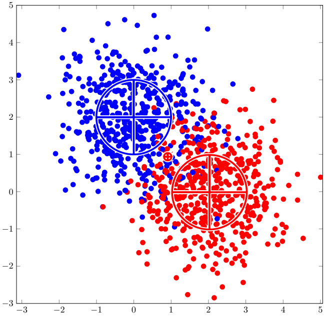


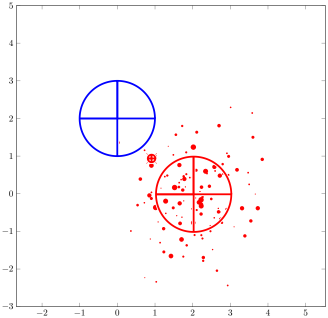
Example



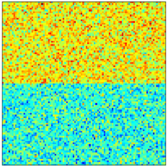
$$N'_{z_n} \frac{\mathrm{p}\left(\boldsymbol{I}_{\{z_n\}} \middle| \mathcal{V}_0\right)}{\mathrm{p}\left(\boldsymbol{I}_{\{z_n\}\backslash n} \middle| \mathcal{V}_0\right)} \prod_{\substack{m \in \delta(n) \\ z_n = z_m}} e^{\omega_{nm}}$$

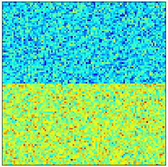
Example

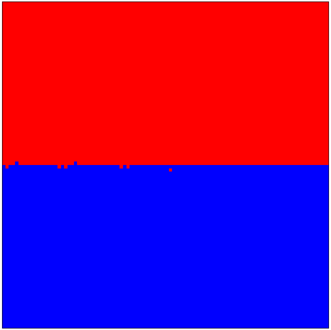




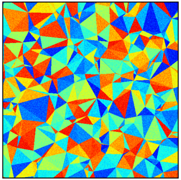
Example

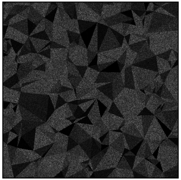


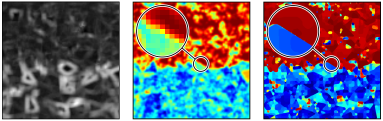


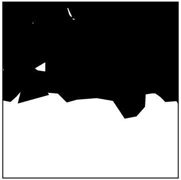


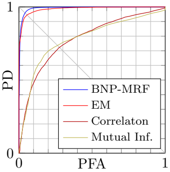
Results – Synthetic Optical and SAR Images


Synthetic optical image


Synthetic SAR image


Mutual Information EM BNP


Change mask

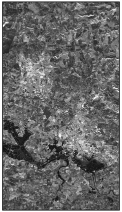


Performance – ROC

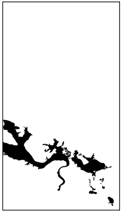
[1] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J.-Y. Tourneret, "Change detection for optical and radar images using a Bayesian nonparametric model coupled with a Markov random field", in Proc. IEEE ICASSP, Brisbane, Australia, April 2015.

Results – Real Optical and SAR Images

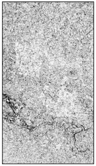
Optical image before the flooding



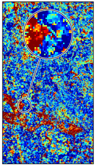
SAR image during the flooding



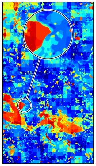
Change mask



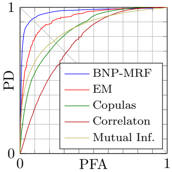
Mutual Information



EM

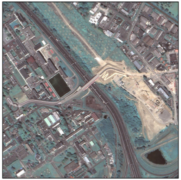


BNP



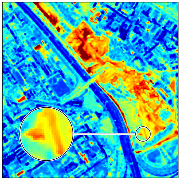
Performance – ROC

Results – Pléiades Images

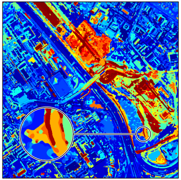


Pléiades – May 2012

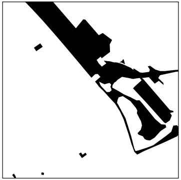
Pléiades – Sept. 2013



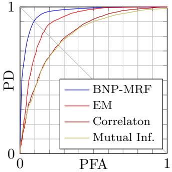
EM



BNP



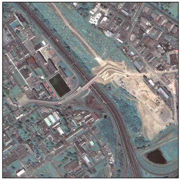
Change mask



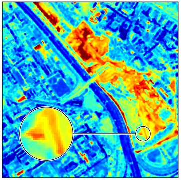
Performance – ROC

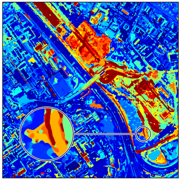
Special thanks to CNES for providing the Pléiades images

Results – Pléiades and Google Earth Images

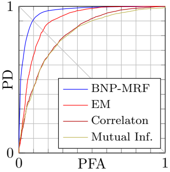

Pléiades – May 2012


Google Earth – July 2013


EM


BNP


Change mask



Performance – ROC

Section 6

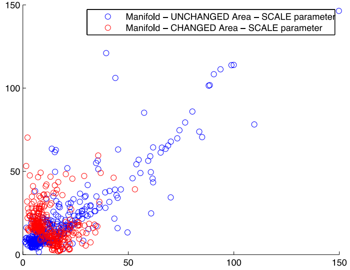
Conclusions

Conclusions

- New statistical model to describe heterogeneous images
- New similarity measure showing encouraging results for homogeneous and heterogeneous sensors
- Interesting for many applications
  - Change detection – local similarity measure
  - Classification
  - Registration – global similarity measure

Future Work

- Study the method performance for different image features (wavelets, gradient, texture coefficients)
  - Homogenize the parametrization for different image modalities
  - Wavelets coefficients: Generalized Gaussian distribution



Future Work

- Consider a robust estimation of the mixture parameters
  - M-Estimators [1]
  - Using noise sparsity approaches [2]
- Consider intra-object dependency of the pixel intensities
  - i.e., in the case of pansharpened images
- Estimate parameters using empirical likelihood methods [3]
  - Overcomes the need to propose a particular statistical model

[1] P. J. Huber. Robust Statistics. Wiley Series in Probability and Statistics. Wiley, 2004
[2] J. Wright et al. "Robust Face Recognition via Sparse Representation". In: IEEE Trans. Pattern Anal. Mach. Intell. 31.2 (Feb. 2009), pp. 210–227
[3] A. B. Owen. Empirical Likelihood. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press, 2001

Future Work

- Add a prior on the spatial parameter of the MRF
- Speed-up the BNP-MRF algorithm with a smart initialization
  - i.e., initialize the algorithm with the output of mean-shift [4]
  - Preliminary results: 10x reduction in the number of iterations

[4] D. Comaniciu and P. Meer. "Mean shift: a robust approach toward feature space analysis". In: IEEE Trans. Pattern Anal. Mach. Intell. 24.5 (May 2002), pp. 603–619

Thank you for your attention

Jorge Prendes
jorge.prendes@tesa.prd.fr