CONDITIONAL MIXED-STATE MODEL FOR STRUCTURAL CHANGE ANALYSIS FROM VERY HIGH RESOLUTION OPTICAL IMAGES

Benjamin Belmudez[1,3,4]*, Véronique Prinet*[1]*, Jian-Feng Yao*[2]*, Patrick Bouthemy*[3]*, Xavier Descombes*[4]

[1] NLPR & LIAMA, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, P.R. China
[2] IRMAR, Univ. Rennes 1, 35042 Rennes Cedex, France
[3] INRIA, Centre Rennes, Campus de Beaulieu, 35042 Rennes Cedex, France
[4] INRIA, Centre Sophia-Antipolis, 06902 Cedex, France

ABSTRACT

The present work concerns the analysis of dynamic scenes from earth observation images. We are interested in building a map which, on one hand locates places of change, on the other hand, reconstructs a unique visual information of the non-change areas. We show in this paper that such a problem can naturally be takled with conditional mixed-state random field modeling (mixed-state CRF), where the ”mixed state” refers to the symbolic or continous nature of the unknown variable. The maximum a posteriori (MAP) estimation of the CRF is, through the Hammersley-Clifford theorem, turned into an energy minimisation problem. We tested the model on several Quickbird images and illustrate the quality of the results.

Index Terms— Image analysis, Conditional random fields, Mixed-state model, Change detection, Remote sensing.

1. INTRODUCTION

The advent of very high resolution optical satellite images has opened, since nearly one decade already, a whole range of new possibilities in application domains like digital cartography, urban planning or land survey. The variety and the amount of information available in these images make nevertheless classical techniques of classification and segmentation inadequate. High accuracy comes with high level of details (car, road lines,..), which are not necessarily relevant information and contribute to the highly correlated noise in the image. The challenge is then to get rid of this 'geometric noise', while exploiting valid and accurate information.

The present work concerns the analysis of dynamic scenes from earth observation images. More precisely, we are interested in building a map which, on one hand locates places of change, on the other hand, reconstructs a unique visual information of the non-change areas. To this end, we propose a new approach based on conditional mixed-state random field modeling (mixed-state CRF). The so-called "mixed-state" stems from the mixed nature of the unknown random variable: the symbolic value is a binary indicator function of change v.s. non change, while the continuous value provides an estimation of the ”background image” where no change happened. The maximum a posteriori (MAP) estimation of the CRF is, through the Hammersley-Clifford theorem, turned into an energy minimisation problem.

The fundamentals of mixed-state CRF, the design of the energy functional, the preprocessing step and the optimisation process are described in sections 3 to 7. Results and conclusions are given in sections 8 and 9. We first briefly introduce in the following section some recent works on change detection.

2. RELATED WORK

Recent advances in digital change detection involve multi-resolution approaches. In [1] Carvalho & *al.* applied wavelet transform to multi-temporal and multiresolution Landsat (TM and MSS) data with the objective to fuse multi-resolution data while reducing radiometric and geometric mis-registration. In [2], the authors generate a multiscale dataset using object-specific analysis (OSA) and object-specific up-scaling (OSU); they detect features using marker-controlled watershed segmentation (MCS) and finally proceed to change detection by differencing the features images.

L. Bruzzone [3] proposes a segmentation framework in which images are decomposed in homogeneous connected regions and some features are computed for each region. The change detection map is produced by thresholding the features difference. In [4], a probabilistic approach inspired from [5] is proposed; it consists in detecting man-made structures using a Discriminative Random Field model. DRF models are directly derived from Conditional Random Fields [6], but introduce a data term which is a discriminative classifier.

Markov Random Field models have been extensively used for various segmentation and labelling applications in vision. The modeling task consists in defining a joint probability distribution $p(\mathbf{x}, \mathbf{y})$ over observation variables $\mathbf{x}$ and unknown variables $\mathbf{y}$. Typically, in a generative framework and using the Bayes rule, the joint probability decomposes into an observation model (or likelihood) $p(\mathbf{x}|\mathbf{y})$ and a prior model $p(\mathbf{y})$. A contrario, Conditional Random Fields (CRFs) [6] directly model the posterior probability of the unknown variable given the observation, $P(\mathbf{y}|\mathbf{x})$. We will, in the next sections, develop a new model belonging to the family of CRFs.

3. OVERVIEW

The overall motivation of this work stems from the following considerations:

- High resolution remote sensing images exhibit strong structural pattern/organisation; for this reason, robust processing of VHR images requires, in particular, to model *long range dependencies of the observations* (note that higher-order MRF can account for long-range interaction between *labels* only); hence the choice of *Conditional Random Fields models (CRF)*;
- We aim at fully exploiting the bi-modal nature of the problem, *i.e.* not only to generate a map of changes, but also to retrieve the 'optimal' visual image of non-changed background; these two distinct aspects are naturally modeled by *mixed-state variables*, which simultaneously account for symbolic and real variables and are estimated without additional computational step.
- CRFs or MRFs do not allow to retrieve an exact optimum, but under specific conditions [7]. In this context, a close-to-the-solution initialisation before optimising the CRF model is important. We thus derived a robust initialisation procedure.

The rest of the paper describes the model and each step of the processing. Details can be found in [8].

4. DEFINITIONS AND FUNDAMENTALS ON MIXED-STATE CONDITIONAL RANDOM FIELDS

Let us first introduce a few notations; we define a graph $G(V, E)$, where $V = \{i \,|\, i \in [1, ..., N]\}$ are the vertices (or nodes) and $E = \{e_{i,j} \,|\, i \neq j;\, i, j \in [1, ..., N]^2\}$ are edges linking two nodes. I^1 and I^2 correspond to images acquired at time t^1 and t^2 respectively; we assume that pairs of images are co-registered. $\mathbf{y} = \{y_i\}$ is the unknown random variable; the observation variable $\mathbf{x} = \{x_i\}$ is a feature vector of the two images (points of same coordinates are coupled to form a pair). Vertices and egdges define cliques of size (order) 1 and 2, $\mathcal{C}_o$ and $\mathcal{C}_1$ respectively.

CRF starts from the paradigm that we can define the posterior distribution directly, without explicitly modeling the data:

$$p(\mathbf{y}|\mathbf{x}) = \frac{e^{-E(\mathbf{x},\mathbf{y})}}{\sum_y e^{-E(\mathbf{x},\mathbf{y})}} = Z^{-1} e^{-\sum_C E_c(\mathbf{x},\mathbf{y})} \qquad (1)$$

where $E()$ is the total energy functional of the model, $E_c()$ are clique-wise energy functions. Z is the partition function, which normalises the distribution ; it is a constant over $\mathbf{y}$ and therefore do not play any role during the task of inference.

CRF models until now were developed for either continuous or discrete random variables [6, 5]. Conversely, inspired from [9], we are here interested in modelling a energy functional $E(\mathbf{x}, \mathbf{y})$, over $\mathbf{y} \in S^N$, where S is a *mixed space* defined by $S = \Omega \cup \mathbb{R}$, with $\mathbb{R}$ the space of reals and Ω the space of symbolic concepts. The notion of "mixed-state" comes then from the bi-modal nature of the unknown $\mathbf{y}$, either continuous (in $\mathbb{R}$), either symbolic ($\omega \in \Omega$).

The mixed state nature of the random variable compels us to define properly a density function that is associated to it (see [9, 10]). Let us first define, for a given element y_i, a mixed measure : $m(dy_i) = \delta_\omega(dy_i) + \lambda(dy_i)$ where δ_ω is the Dirac measure at $\omega \in \Omega$ and λ the Lebesgue measure over $\mathbb{R}$. We note $\delta^*_\omega = 1 - \delta_\omega(y_i)$. The variable $y_i \in S$ is defined such as:

1. $y_i = \omega$ with a probability $P_i \in [0, 1]$,
2. with a probability $(1 - P_i)$, y_i follows a continous distribution with a density function d.

Consequently, we define the probability density function of a mixed variable $y_i \in S$ associated to the measure $m(dy_i)$ by:

$$f(y_i) = P\,\delta_\omega(y_i) \;+\; (1 - P)\,d(y_i)\,\delta^*_\omega(y_i)$$

that indeed verifies $\int_{-\infty}^{\infty} f(y_i)m(dy_i) = 1$.

We can now design a mixed-state energy functional expressed as the sum of two terms, associated respectively with the symbolic value and the continuous value of $\mathbf{y}$.

5. RADIOMETRIC FEATURES AND REVISITED ITERATIVE PCA

The first step of our approach is to classify pixels pairs into two *rough* classes, "Change" and "No-Change", respectively l_1, l_0. The main principle is (see [11] or [8] for details) to first compute the intensity-level linear change between two images, then to evaluate, after linear correction, the deviation of the intensity level in the second image. The computation is based on a iterative computation of principal component analysis.

Let us consider a bi-dimensional feature space defined by the orthogonal axes (oI^1) and (oI^2) —namely, a *bi-temporal* feature space. A site i is thus represented in this feature space by its coordinates $(I^1(i), I^2(i))$. We can reasonably expect all unchanged pixels to lie in a narrow elongated cluster along a principal axis (which tangent would be 1 if there were no illumination change at all between the two acquisitions). On the other hand, the pixels of which spectral appearance has changed are expected to lie far away from this axis. In other words, the magnitude of change can be quantified by the following inner product : $c(i) = g\,.\,(I - \mu)$, with $g = [g^1\; g^2]^T$ the second eigenvector of the covariance matrix (it indicates the direction along which there is little variation of the intensity level) , and $\mu = [\mu^1 \;\, \mu^2]^T$ the mean vector of intensity levels $I = [I^1 \;\, I^2]$ in each image. An obvious problem with principal component based change detection is that the covariance matrix is computed from all the pixels including those which have experienced change. Thus the computation of the second principal axis g is obviously biased.

The iterative approach aims at decreasing progressively the influence of outliers (*i.e.* the pixels of change), by computing the empirical expectation : $\bar{\mu} = \sum_i I(i)\, p(l_0|i)$. Computing $p(l|i) \simeq p(l)p(i|l)$, $l = \{l_0, l_1\}$, boils down to estimate the likelihoods $p(i|l)$ and the priors $p(l)$. The latter is evaluated as the ratio between the pixels belonging to class l and the total number of pixels. The likelihood is assumed to follow a (zero-centred) Gaussian distribution $N(c(i); 0, \sigma_l)$, which parameters σ_l are estimated by fitting the empirical distribution of $c(i)$ and re-estimated at each iteration (see Figure 1).

Iterations converge (*i.e.* $\Delta\bar{\mu} < \epsilon$) within 10 iterations. Finally, a rough map of change is obtained by maximising the posterior at each node: $l_i = \arg\max_{l \in \{l_0, l_1\}} p(l|i)$.

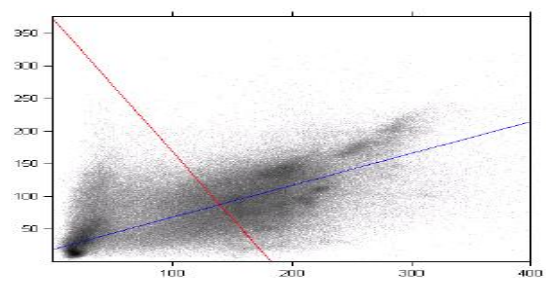

Fig. 1. Bi-temporal feature space and results of iterative PCA (g orientation in red, and principal axis in blue).

6. CHANGE ANALYSIS FROM CONDITIONAL MIXED STATE MODEL

6.1. Overview

We can re-write equation 1 such as to highlight the mixed-state nature of the model. Taking the negative log-likelihood, and incorporating the variable $\mathbf{y} \in S$, we have

$$\begin{aligned} L = -\log\left(p(\mathbf{y}|\mathbf{x})\right) &= E(\mathbf{x}, \mathbf{y}) + log(Z) \\ &\sim E_{symb}(\mathbf{x}, \mathbf{y}) + \beta\, E_{cont}(\mathbf{x}, \mathbf{y}) \end{aligned} \tag{2}$$

β is a weight balancing the contributions of the symbolic and continuous terms, E_{symb} and E_{cont} respectively. (Z being constant, it is neglected in the rest of the paper). Both components are in turn decomposed into a $data$ term (or likelihood), and a $pairwise$ term (or regularisation).

6.2. Data energy term

We define

$$E_{data} = \sum_{i \in \mathcal{C}_0} \; E_{data,cont}(i) + \alpha\, E_{data,symb}(i)$$

with

$$\begin{aligned} E_{data,symb}(i) &= -\; \mathbb{I}_{y_i = \omega}\, \log\left(p(l_0|i)\right) \\ E_{data,cont}(i) &= \mathbb{I}_{y_i \neq \omega}\, \rho\left(\frac{(x_i - y_i)^2}{\sigma_d^2}\right) \end{aligned} \tag{3}$$

The symbolic part is the negative log of the probability of change, $p(l_0|i)$, as defined in section 5. If $p(l_0|i)$ is close to one then this term is close to zero. If there is no change, then the energy term is positive and large. As for the continuous part, it tells us that, if y_i takes value in $\mathbb{R}$, then its value should be close to $x(i) = \min(I^1(i), I^2(i))$ (we could have chosen the mean function rather than the minimum, it is an arbitrary choice). We set $\rho(t) = \tanh(t) = \frac{e^t - e^{-t}}{e^t + e^{-t}}$; its role is to make the energy terms vary on comparable intervals and, when the dynamic of the variations is too tight, to stretch it. σ_d is the variance computed on a neighbour around node i (usually 7 x 7 or 9 x 9 pixels): it enables a local normalisation.

6.3. Pairwise energy terms

We define the pairwise/regularisation term such that

$$E_{reg} = \sum_{i \in \mathcal{C}_0} E_{reg,cont}(i) \; + \; \gamma\, E_{reg,symb}(i)$$

with

$$\begin{aligned} E_{reg,symb}(i) &= \mathbb{I}_{y_i = \omega}\left(1 - \frac{\sum_{j \in \mathcal{C}_{i,1}} \mathbb{I}_{y_j = \omega}}{|\mathcal{C}_{i,1}|}\right) \\ E_{reg,cont}(i) &= \mathbb{I}_{y_i \neq \omega} \sum_{j \in \mathcal{C}_{i,1}} \mathbb{I}_{y_i \neq \omega} \;\; \rho\left(\phi(i,j)\frac{(y_i - y_j)^2}{\sigma_r^2}\right) \end{aligned} \tag{4}$$

$|\mathcal{C}_{i,1}|$ is the number of $\mathcal{C}_1$ cliques around node i (typically we choose a 4-connectivity neighbourhood, hence $|\mathcal{C}_1| = 4$). The symbolic part of the spatial regularisation is inspired from the Ising model : it favours the presence of symbolic labels in the neighbourhood of i. The continuous term has for effect to prevent large intensity gradients between two neighbourhood pixels labelled "non-changed", except at edges. Function $\phi(i,j)$ is used to reduce smoothing around edges; its expression is : $\phi(i,j) = \frac{1}{\max(\Delta I^1, \Delta I^2, \xi)}$, where $\Delta I^k(i,j) = \left|I^k(i) - I^k(j)\right|$, for $k = 1, 2$, index of each of the two images, and ξ is a constant, typically small enough. The variance σ_r plays similar role as above. Note that this pairwise term depends on the observation variable.

7. OPTIMISATION FRAMEWORK

We seek for the optimal configuration $\mathbf{y}$ that minimises L, given the observation $\mathbf{x}$. We are not dealing with conventional random fields and need to develop an algorithm that is adapted to the mixed-state nature of the energy function.

For its simplicity of implementation, we adapted the Iterated Conditional Model (ICM) (it would be worth comparing with global optimisation methods, such as graph-cuts or simulated annealing [12, 7]). At each iteration, at each node, we compute the local energy for the two possible cases (*i.e.* y_i is real —no-change–, or symbolic –change) and retain the value of y_i which minimises the energy. Note that, when minimising the energy for $y_i \in \mathbb{R}$, y_i can be estimated analytically: if one develops the expression $E_{reg,cont} + E_{data,cont}$, it results in a polynomial of degree two, whose minima can be computed without difficulty.

8. EXPERIMENTAL RESULTS

Experiments were performed on several pairs of Quickbird panchromatic images (0.6m resolution), covering the same geographical site (Beijing area). Pairs are assumed to be registered (via geo-coding in our case). While selecting test images, in purpose, we chose difficult cases which contain true structural changes (that we want to detect) and visual changes (cars, vegetation, projective effects) associated to "noise".

Figure 2 illustrates some results. Top images show the two input data and their associated ground truth (in red). The results from mixed-state CRF is given in bottom right : gray level values indicate the reconstructed image associated to the unchanged regions of the scene, while in red colour are pixels detected as change. More results are given in [8]. It appears clearly that mixed-state CRF model performs better than a simple IPCA. Remaining false positive detections are mainly due to projective effects of high buildings, and could be overcome by a multi-scale approach or a better fitted data term.

9. CONCLUSION

We have described a mixed-state conditional random field model for change analysis problems. We designed an energy functional robust to visual (appearance) variations and geometric noise inherent to VHR images. We have shown that the mixed-nature of the random variable enables to naturally retrieve a bi-modal solution, mapping the areas of changes and simultaneously reconstructing the visual background image where no changes happened. The model could further be exploited to multi-category change analysis.

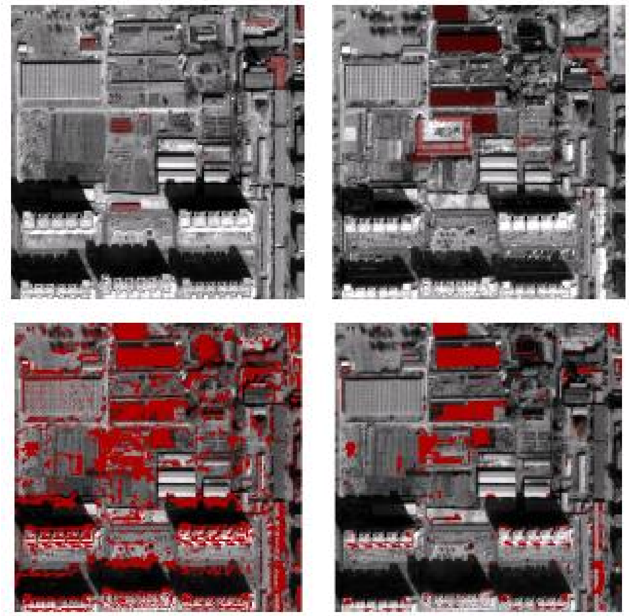

Fig. 2. Change detection from Mixed-State CRF. Top: Pair of Quickbird images (2001 and 2003) and manual labeling of structural changes (in red). Bottom- Left: preprocessing from IPCA ; Bottom-Right: Change map (in red) and reconstructed visual image of non-change (background image).

10. REFERENCES

[1] L. Carvalho and L. Fonseca, "Digital change detection with the aid of multiresolution wavelet analysis," *Int. J. of Remote Sensing*, vol. 22(18), pp. 38713876, 2001.

[2] O. Hall and G.J. Hay, "A multiscale object-specic approach to digital change detection," *Int. J. of Applied Earth Observation and Geoinformation*, vol. 4(4), pp. 311327, 2003.

[3] L. Bruzzone and D. F. Prieto, "An adaptative parcel-based technique for unsuper- vised change detection," *Int. J. of Remote Sensing*, vol. 21(4), pp. 817822, 2000.

[4] W. Liu and V. Prinet, "Probabilistic modeling for structural change inference," in *Proc. Asian Conf. on Computer Vision*, Hyderabad, India, 2006.

[5] S. Kumar and M. Herbert, "Discriminative random fields," *Int. J. of Computer Vision*, vol. 68(2), pp. 179–201, 2006.

[6] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proc. Int. Conf. on Machine Learning*, Massachusetts, USA, 2001.

[7] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts," *IEEE PAMI*, vol. 26(2), pp. 147–159, 2004.

[8] Benjamin Belmudez, "Structural change detection on urban areas from high-resolution satellite images," M.S. thesis, ENSPS, 2007.

[9] P. Bouthemy, C. Hardouin, G. Piriou, and J. Yao, "Mixed-state automodels and motion texture modeling," *J. of Mathematical Imaging and Vision*, vol. 25(3), pp. 387–402, 2006.

[10] T. Crivelli, B. Cernuschi-Frias, P. Bouthemy, and J.-F. Yao, "Mixed-state markov random fields for motion texture modeling and segmentation," in *Proc. IEEE Int. Conf. on Image Processing*, Atlanta, USA, 2006.

[11] R. Wiemker, A. Speck, D. Kulbach, H. Spitzer, and J. Beinlein, "Unsupervised robust change detection on multispectral imagery using spectral and spatial features," in *Proc. Int. Airborne Remote Sensing Conference*, Copenhagen, Denmark, 1997.

[12] Julian Besag, "Spatial interaction and the statistical analysis oflattice systems," *J. R. Statist. Soc.*, vol. B 36, pp. 192–236, 1974.