LOGISTIC REGRESSION FOR DETECTING CHANGES BETWEEN DATABASES AND REMOTE SENSING IMAGES

M. Chabert[(1)], *J.-Y. Tourneret*[(1)], *V. Poulain*[(2)] *and J. Inglada*[(2)]

[(1)] University of Toulouse, IRIT-ENSEEIHT-TéSA, Toulouse, France
[(2)] Centre National d'Études Spatiales (CNES), Toulouse, France
{jean-yves.tourneret,marie.chabert}@enseeiht.fr, vincent.poulain@these.cnes.fr, jordi.inglada@cnes.fr

1. INTRODUCTION

Cartographic database updating is a key application of high resolution remote sensing imagery. Classically, the updating is performed by comparing particular objects detected in the image to the information contained in the database [1], [2]. This paper proposes a change detection strategy based either on the optical and SAR (Synthetic Aperture Radar) images or on appropriate features extracted from the detected objects. The relationship between the database and the images or features are described using a logistic regression model. Logistic regression is an efficient tool for comparing continuous and discrete random variables [3]. The database can be transformed into a discrete (generally binary) random variable (e.g., modeling the absence/presence of buildings). Conversely, the images and associated features can be modeled as continuous random variables [4]. The proposed approach estimates the logistic regression model parameters either between the database and the image for each pixel (pixel approach) or between the database and the extracted features for each detected object (object approach). The estimation is performed using the maximum likelihood (ML) method. The logistic regression parameter estimates provide a measure of similarity between the database and the available images. Binary hypothesis tests are finally constructed to detect changes between the optical/radar images and the database using these estimates.

2. LOGISTIC REGRESSION MODEL

The basic problem addressed in this paper is the detection of changes between an optical and/or SAR image and a database. The comparison can be performed directly on the image using a sliding window centered on a given pixel. The change detection can also use features derived from the optical and SAR images for each detected object [5]. In the case of building database updating, these features are defined as follows:

- **edges**: mean distance between building borders and optical image nearest edges,
- **shadow**: mean value of the shadow mask near building walls not oriented toward the sun,
- **vegetation**: percentage of vegetation pixels located inside the buildings on the optical image,
- **line segments**: percentage of building wall pixels containing an extracted segment in their vicinity,
- **contrast**: ratio between the layover region and shadow region mean values on the SAR image.

The optical and SAR images as well as each extracted feature can be modeled as continuous random variables whereas each database element is transformed into a binary random variable. Let $\mathbf{U_i} = [U_i(1), \ldots, U_i(n)]^T$, for $i = 1, \ldots, N$, denote n pixels of the sliding window for the i^{th} image or n independent measurement of the i^{th} extracted feature. Denote as $\mathbf{M} = \{M(j), j = 1, \ldots, n\}$ the associated database elements. For a given pixel or a given object, the logistic regression model is defined by

$$\pi_1 = P\left[M(j) = 1 | U_1(j) = u_1, U_2(j) = u_2, ..., U_N(j) = u_N\right] = \frac{\exp(\beta_0 + \beta_1 u_1 + \beta_2 u_2 + ... + \beta_N u_N)}{1 + \exp(\beta_0 + \beta_1 u_1 + \beta_2 u_2 + ... + \beta_N u_N)}$$

and

$$\pi_0 = 1 - \pi_1 = P\left[M(j) = 0 | U_1(j) = u_1, U_2(j) = u_2, ..., U_N(j) = u_N\right]$$

where $\beta_i, i = 1, \ldots, N$ are the regression model parameters to be estimated and π_1 is the probability to observe a building in the database from the image pixels or the corresponding features.

3. CHANGE DETECTION

We propose to estimate the regression parameters for each pixel (image approach) or for each detected object (feature approach). This provides an estimate of the conditional probability π_1 that can be compared to a threshold for building detection. The change detection is then performed by comparison to the database. Another strategy consists of comparing the logistic regression parameters to a reference value available from the previous database updating, for a one-step change detection strategy.

3.1. Maximum Likelihood Estimation

The ML method estimates the unknown parameter vector $\boldsymbol{\beta} = [\beta_0, ..., \beta_N]$ by maximizing the joint likelihood of $M_1, ..., M_n$ given $U_i(1) = u_i(1), ..., U_i(n) = u_i(n)$ for $i = 1, \ldots, N$ on a set of n independent measurements (the n pixels of the sliding window or the features extracted for n detected objects). In this case, the joint likelihood is

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \pi_1^{m_i} \left[1 - \pi_1\right]^{(1-m_i)}.$$

Maximizing the joint likelihood is equivalent to minimize the negative log-likelihood

$$l(\boldsymbol{\beta}) = -\ln L(\boldsymbol{\beta}) = -\sum_{i=1}^{n} m_i \ln(\pi_1) + (1 - m_i) \ln(1 - \pi_1).$$

Differentiating the log-likelihood with respect to β_0 and β_i, $i = 1, \ldots, N$ leads to the following equations

$$\sum_{j=1}^{n} [m_j - \pi_1] = 0 \qquad \text{and} \qquad \sum_{j=1}^{n} u_i [m_j - \pi_1] = 0, i = 1, \ldots, N. \tag{1}$$

The parameter vector $\boldsymbol{\beta}$ can then be estimated using a numerical optimization procedure. The closed-form expressions of the partial derivative given in (1) can be used for the optimization procedure that was conducted using the unconstrained Nelder-Mead simplex method [6].

3.2. Binary hypothesis Tests

The detection of changes between the database and the images can be achieved by using the estimated logistic regression parameters. A binary hypothesis test allows one to decide whether the observed pixel (image approach) or the detected object (feature approach) corresponds to a building (hypothesis H_0) or not (hypothesis H_1). For that purpose, the conditional probability is then compared to a given threshold. The binary hypothesis testing problem can then be expressed as

$$H_0 : \text{ no building} \qquad \text{and} \qquad H_1 \text{ building.}$$

For each pixel on the image or for each detected object, an estimate of π_1 is computed by using the regression parameter estimates. The decision is made according to the following rule:

$$T = \widehat{\pi_1} = \frac{\exp(\widehat{\beta_0} + \widehat{\beta_1} u_1 + \widehat{\beta_2} u_2 + ... + \widehat{\beta_N} u_N)}{1 + \exp(\widehat{\beta_0} + \widehat{\beta_1} u_1 + \widehat{\beta_2} u_2 ... + \widehat{\beta_N} u_N)} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} S_{\text{PFA}}. \tag{2}$$

The second strategy directly uses the regression parameter estimates as a test statistics for change detection. Assume that a reference image (associated to the last database updating) is available with $\boldsymbol{\beta} = \boldsymbol{\beta}_{H_0} = \boldsymbol{\beta}_{\text{ref}}$. The hypothesis testing can then be expressed as

$$H_0 \text{ (no change)} : \boldsymbol{\beta} = \boldsymbol{\beta}_{\text{ref}}, \qquad \text{and} \qquad H_1 \text{ (change)} : \boldsymbol{\beta} \neq \boldsymbol{\beta}_{\text{ref}}.$$

The decision is then made according to the following rule $T = \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{\text{ref}}\| \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} S_{\text{PFA}}$.

In both cases, T is the test statistics and S_{PFA} is the threshold depending on the probability of false alarm (PFA), i.e. the probability of deciding that hypothesis H_1 is true when it is actually not true [7, p. 38].

4. SIMULATION RESULTS

First, the adequation of the logistic regression model has been evaluated for an optical image and the associated building database. The number of pixels m_i associated to each value $u_i \in \{0, ..., 255\}$ of the optical image has been derived. The estimated cumulative distribution is compared to the theoretical curve derived from the logistic regression model after estimation of $\boldsymbol{\beta}$ on the whole optical image. Figure 1(left) shows the perfect adequation with the model in the case of the optical image.

The next simulations have been obtained with synthetic data. The detection strategy uses the database and only one image with the pixel approach. The first experiment studies the mean square errors (MSEs) of the ML estimates (log scale) for parameters β_0 and β_1 for different sizes of the estimation window. The actual parameter vector is $\beta_0 = 1, \beta_0 = 0.2$ for this example. All MSEs have been computed using 1000 Monte Carlo runs. The MSEs of the ML estimates are compared with the corresponding Cramer-Rao bounds that can be computed for the proposed statistical model (more details will be provided in the final version of the paper). Figure 1(right) shows that the the MSEs of the ML estimator are very close to the Cramer-Rao bounds even for relatively small sample sizes.

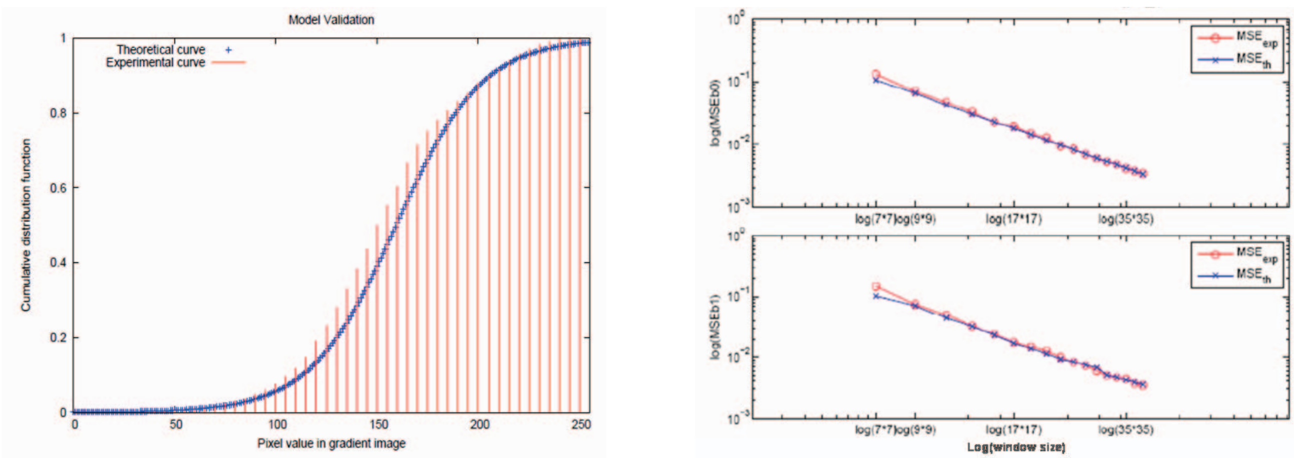

Fig. 1. (left) Logistic regression model validation. (right) Theoretical and empirical MSE for $\beta_0 = 1$ and $\beta_1 = 0.2$.

Real images have been provided by the CNES, the french spatial agency. Figure 2 shows the optical image acquired over a suburban area of Toulouse, France by the airborne PELICAN sensor (left) and the associated building database (right).

Fig. 2. Real optical image (left) Real database (right)

Figure 3 show the results of building detection from the conditional probability thresholding for two threshold values.

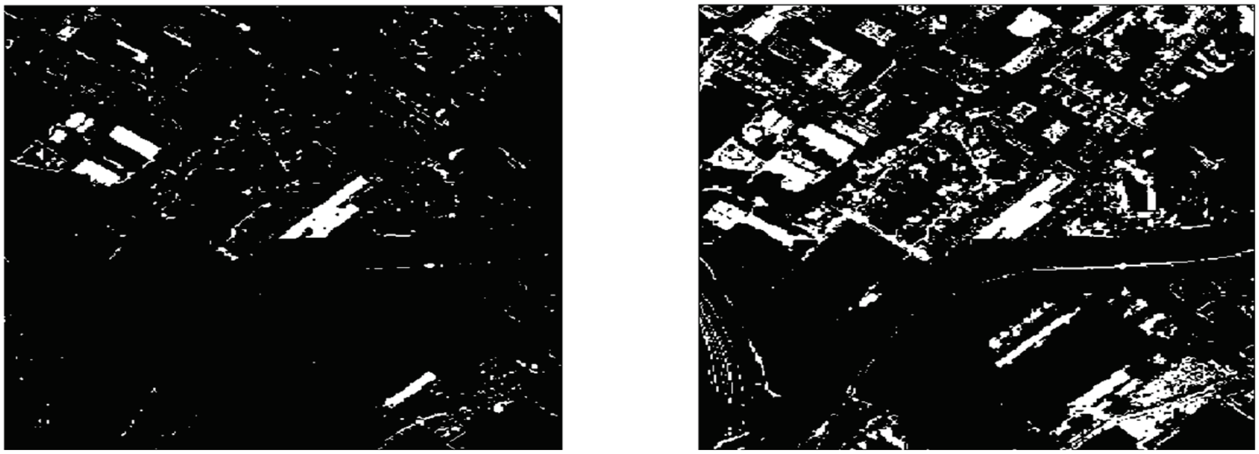

Fig. 3. Building detection by conditional probability thresholding (left) $S_{PFA} = 0.93$. (right) $S_{PFA} = 0.95$.

5. CONCLUSIONS

This paper studied a joint statistical model for a binary data basis, optical or SAR images and particular extracted features. The proposed logistic regression model has shown a good adequation with real data in the case of an optical image and the associated building database. The model parameters were estimated using a maximum likelihood approach on a sliding window or on particular objects to allow for localized change detection. The detection of changes between the two kinds of data is then performed using the logistic regression parameter estimates. More intensive simulations will be provided in the final version of the paper. In particular, the interest of using extracted features will be studied.

6. ACKNOWLEDGEMENT

The authors would like to thank the CNES for its financial support.

7. REFERENCES

[1] T. Bailloeul, V. Prinet, B. Serra, P. Marthon, P. Chen, and H. Zhang, "Urban building land use change mapping from high resolution satellite imagery, active contours and hough voting," in *Proc. 9th International Symposium on Physical Measurements and Signature in Remote Sensing (ISPMSRS)*, Beijing, China, Oct. 2005.

[2] J. Duan, V. Prinet, and H. Lu, "Building extraction in urban areas from satellite images using GIS data as prior information," in *Proc. IEEE International Geoscience and Remote Sensing Symposium*, Anchorage, Alaska, USA, Sep. 2004.

[3] D. W. Hosmer and S. Lemeshow, *Applied logistic regression, Second edition*, Wiley, New York, 2000.

[4] J.-Y. Tourneret, V. Poulain, M. Chabert, and J. Inglada, "Similarity measures between vector data bases and optical images for change detection," in *Proc. IEEE International Geoscience Remote Sensing Symposium (IGARSS'09)*, Cape Town, South Africa, Jul. 2009.

[5] V. Poulain, J. Inglada, M. Spigai, J.-Y. Tourneret, and P. Marthon, "Fusion of high resolution optical and sar images with vector data bases for change detection," in *Proc. IGARSS'09 Cape Town*, 2009.

[6] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Computer Journal*, vol. 7, pp. 308–313, 1965.

[7] Van Trees and Harry L., *Detection, Estimation, and Modulation Theory: Part I*, Wiley, New York, 1968.

[8] D. Borghys, M. Shimoni, and C. Perneel, "Supervised classification of hyperspectral images using a combination of spectral and spatial information," in *Proc. SPIE Image and Signal Processing for Remote Sensing XI, Brugge, Belgium*, Sept. 2005, vol. 5982.