A REDUCED RANK REGRESSION MIXTURE MODEL FOR CHANGE VALIDATION IN AERIAL IMAGES

FERNANDO PÉREZ NAVA

Departamento de Estadística, Investigación Operativa y Computación.
Universidad de la Laguna. Islas Canarias. 38271 Spain.
fdoperez@ull.es

JOSE MANUEL GÁLVEZ LAMOLDA

Departamento de Física Fundamental y Experimental, Electrónica y Sistemas
Universidad de La Laguna. Islas Canarias. 38271. Spain
jmgalvez@ull.es

Change detection is an important part of image interpretation and automated geographical data collection. In this paper we show a reduced rank regression mixture model for the verification of image changes detected by a human operator. Maximum likelihood estimators are used to learn the operator behaviour. Then, the operator uses the trained system to validate the image changes found. Computational results are given with real image data that show the performance of the system.

1. Introduction

Change detection on the terrestrial surface from aerial images is an important problem in the update and maintenance of digital cartographic maps. One way to detect these changes is the use of pairs of images of the same area at different times. A human operator then detects changes by means of the visual analysis of images. Human visual analysis is favoured by the implicit incorporation of intelligent processing associated to the operator. However, and also due to the explicit intervention of the operator, this analysis is not free of disadvantages due to operator fatigue, operator inconsistencies or operator costs. To achieve an automatization of this process several techniques have been developed: Image differencing [1], Image regression [2] or Principal component Analysis (PCA) [3], [4]. In this paper, we present a quality control system for the verification of the changes detected by a human operator in pairs of images. The systems checks if each change can be explained by the operator global behavior and eventually select a subset of the changes for revision. The learning system is

based on a reduced rank regression mixture model [5] whose parameters are estimated from the changes detected by the operator. This model is simpler and computationally more efficient than the full regression model.

2. A Probabilistic Model of Image Changes

In this section we propose a generative model of operator's response for two patches x from image I_1 captured at time t_1 and y from image I_2 captured at time t_2. These patches will be represented as vectors and are small enough so that they approximately belong to the 'Change' (CH) class or 'No Change' (NC) class. The proposed model shown in Figure 1 states that intensity levels on patch y depend on the intensity levels patch x, some uncontrolled factors (illumination changes or others) and whether there is change or not as expressed in the True Class (TC) variable. The model also states that once the true class is inferred there may be a random disturbance (due to operator's fatigue or other causes) that gives the actual response of the operator (Operator's Class or OC).
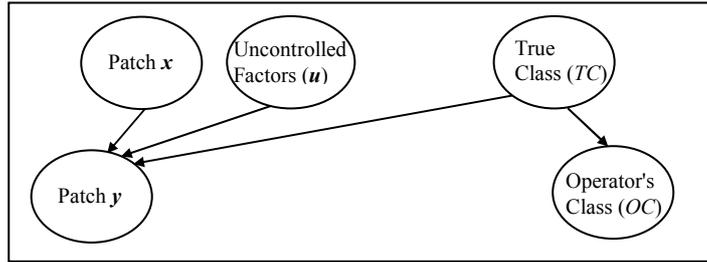

Figure 1. Probabilistic Model for image changes

To validate the operator's behaviour we are interested in the evaluating the probability P(TC| OC, x, y). Following the independence assumptions in the Bayesian network in Figure 1 we have:

$$P(TC|\ OC, x, y) \propto p(y|\ x, u,\ TC)\ P(TC)\ P(OC\ |\ TC). \tag{1}$$

Then, to complete the model we need to define the three terms in the right side of equation (1). The proposed model for $p(y|\ x, u,\ TC)$ is:

$$y = \mu_{TC} + A_{TC}t + C_{TC}u + \varepsilon, \tag{2}$$

$$t = B_{TC}x, \quad u \sim N(0, I), \quad \varepsilon \sim N(0, \sigma_{TC}^2 I), \quad TC = CH, NC \quad ,$$

where t is an r_{TC} dimensional random variable, u is q_{TC} dimensional random variable and t, u, ε are independently distributed. The dimension of matrix A_{TC} is $d \times r_{TC}$, the dimension of matrix B_{TC} is $r_{TC} \times d$, and both matrices have rank r_{TC}. The dimension of matrix C_{TC} is $d \times q_{TC}$ and its rank is q_{TC}.

Therefore we can express the d dimensional variable y in terms of two latent variables t and u of dimensions r_{TC} and q_{TC} where usually $d \gg r_{TC} + q_{TC}$. The proposed model then states that image patch y is related to:

- Latent variable t, that linearly depends on image patch x. The term $A_{TC}t$ $=A_{TC}B_{TC}x$ can be interpreted as a reduced rank regression [5] of y over x.
- Latent variable u that accounts for those aspects in y not explained by t. This term is related to probabilistic principal component analysis (PPCA) [6].
- An overall mean μ_{TC}
- Random variable ε, that expresses an isotropic noise model.

We can simplify the model in Figure 1 integrating out variable u. Then:

$$y \mid x, TC \sim N(\mu_{TC} + A_{TC}B_{TC}x, M_{TC}), \quad M_{TC} = C_{TC}C_{TC}^{\mathrm{T}} + \sigma_{TC}^2 I. \tag{3}$$

The probability distribution for P($OC|TC$) and P(TC) are defined from Bernoulli random variables.

2.1 The Learning Set

Before we show how to estimate the parameters in the model we will describe the learning set. This set is given by the operator's work and consist in a set of point locations in the image where she observes a change. Then, for each detected change, an image patch around this location is generated. This provides the learning set for the 'Change' class. After an operator processes an image, there are no examples for the 'No Change' class. To provide a set of examples for this class, we randomly sample locations from the image except in a neighborhood of the locations in the 'Change' class. The image patches generated from these locations are the learning set for the 'No Change' class. In Figure 2, we can see two examples from the learning set.

Figure 2. Sample image patches from the learning set of the 'Change' class (left) and 'No Change' class (right).

2.2 Model parameter estimation

We will estimate the parameters of the model, after eliminating variable ***u*,** by Maximum Likelihood Estimation in the joint probability distribution of the observed variables $p(\mathbf{y}, OC | \mathbf{x})$:

$$p(\mathbf{y}, OC | \mathbf{x}) = \sum_{TC} p(\mathbf{y} | \mathbf{x}, TC) P(OC|TC) P(TC). \tag{4}$$

It is simpler, however, to use the alternative factorisation:

$$p(\mathbf{y}, OC | \mathbf{x}) = \sum_{TC} p(\mathbf{y} | \mathbf{x}, TC) P(TC|OC) P(OC). \tag{5}$$

We will denote $P(TC|OC)$ as $\alpha_{TC|OC}$, $P(OC)$ as π_{OC}, $P(OC|TC)$ as $\alpha_{OC|TC}$ and $P(TC)$ as π_{TC}. These probability distributions are related by:

$$\pi_{TC} = \sum_{OC} \alpha_{TC|OC} \pi_{OC}, \quad \alpha_{OC|TC} = \frac{\alpha_{TC|OC} \pi_{OC}}{\sum_{OC} \alpha_{TC|OC} \pi_{OC}} \tag{6}$$

Then, given a learning set $\mathbf{H} = \{(\mathbf{y}_1, \mathbf{x}_1, OC_1), (\mathbf{y}_2, \mathbf{x}_2, OC_2), \dots, (\mathbf{y}_N, \mathbf{x}_N, OC_N)\}$, the Maximum Likelihood Estimators (MLE) for the model are evaluated iteratively with the Expectation Maximization (EM) algorithm [7] as follows:

For each sample in $\mathbf{H}$ initialise its resposibilities $R_{n,TC}$ for each of the *CH*, *NC* classes.

repeat {
Evaluate the parameters in the model from:

$$\tilde{\pi}_{OC} = \frac{1}{N} \sum_{i=1}^{N} [OC_n = OC], \quad [OC_n = OC] = \delta_{OC_n = OC}, \quad OC = NC, CH$$

$$\tilde{\alpha}_{TC|OC} = \frac{\sum_{n=1}^{N} R_{n,TC} [OC_n = OC]}{N \tilde{\pi}_{OC}}, \quad OC = NC, CH \tag{7}$$

$$\widetilde{\boldsymbol{\mu}}_{TC} = \left(\widetilde{\boldsymbol{\mu}}_{y,TC} - \widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{B}}\widetilde{\boldsymbol{\mu}}_{x,TC}\right), \quad \widetilde{\boldsymbol{\mu}}_{y,TC} = \frac{\sum_{n=1}^{N} R_{n,TC}\boldsymbol{y}_n}{\widetilde{N}_{TC}}, \quad \widetilde{\boldsymbol{\mu}}_{x,TC} = \frac{\sum_{n=1}^{N} R_{n,TC}\boldsymbol{x}_n}{\widetilde{N}_{TC}},$$

$$\widetilde{N}_{TC} = N\sum_{OC}\widetilde{\pi}_{OC}\widetilde{\alpha}_{TC|OC}$$

$$\widetilde{\boldsymbol{S}}_{yx,TC} = \frac{\sum_{n=1}^{N} R_{n,TC}\left(\boldsymbol{y}_n - \widetilde{\boldsymbol{\mu}}_{y,TC}\right)\left(\boldsymbol{x}_n - \widetilde{\boldsymbol{\mu}}_{x,TC}\right)^{\mathrm{T}}}{\widetilde{N}_{TC}},$$

$$\widetilde{\boldsymbol{S}}_{xx,TC} = \frac{\sum_{n=1}^{N} R_{n,TC}\left(\boldsymbol{x}_n - \widetilde{\boldsymbol{\mu}}_{x,TC}\right)\left(\boldsymbol{x}_n - \widetilde{\boldsymbol{\mu}}_{x,TC}\right)^{\mathrm{T}}}{\widetilde{N}_{TC}}$$

(7)

$$\widetilde{\boldsymbol{A}}_{TC} = \boldsymbol{V}_{r_{TC},TC}, \quad \widetilde{\boldsymbol{B}} = \boldsymbol{V}_{r_{TC},TC}{}^{\mathrm{T}}\widetilde{\boldsymbol{S}}_{yx,TC}\widetilde{\boldsymbol{S}}_{xx,TC}^{-1},$$

where the r_{TC} column vectors in the $d \times r_{TC}$ matrix $\boldsymbol{V}_{r_{TC},TC}$ are the eigenvectors corresponding to the r_{TC} largest eigenvalues of $\widetilde{\boldsymbol{S}}_{yx,TC}\widetilde{\boldsymbol{S}}_{xx,TC}^{-1}\widetilde{\boldsymbol{S}}_{xy,TC}$, $TC = NC, CH$.

$$\widetilde{\boldsymbol{S}}_{TC} = \frac{\sum_{n=1}^{N} R_{n,TC}\left(\boldsymbol{y}_n - \widetilde{\boldsymbol{\mu}}_{TC} - \widetilde{\boldsymbol{A}}_{TC}\widetilde{\boldsymbol{B}}_{TC}\boldsymbol{x}_n\right)\left(\boldsymbol{y}_n - \widetilde{\boldsymbol{\mu}}_{TC} - \widetilde{\boldsymbol{A}}_{TC}\widetilde{\boldsymbol{B}}_{TC}\boldsymbol{x}_n\right)^{\mathrm{T}}}{\widetilde{N}_{TC}},$$

$$\widetilde{\boldsymbol{C}}_{TC} = \boldsymbol{U}_{q_{TC},TC}(\boldsymbol{\Lambda}_{q_{TC},TC} - \widetilde{\sigma}_{TC}^2\boldsymbol{I})^{1/2},$$

where the q_{TC} column vectors in the $d \times q_{TC}$ matrix $\boldsymbol{U}_{q_{TC},TC}$ are the eigenvectors of $\widetilde{\boldsymbol{S}}_{TC}$, with corresponding eigenvalues in order of decreasing magnitude in the $q_{TC} \times q_{TC}$ diagonal matrix $\boldsymbol{\Lambda}_{q_{TC},TC}$, and $\widetilde{\sigma}_{TC}^2$ is defined as:

$$\widetilde{\sigma}_{TC}^2 = \frac{1}{d - q_{TC}} \sum_{j=q_{TC}+1}^{d} \lambda_{j,TC},$$

where $\lambda_{q_{TC}+1,TC}, \ldots, \lambda_{d,TC}$ are the smallest eigenvalues of $\widetilde{\boldsymbol{S}}_{TC}$.

For each sample, reevaluate the responsibilities:

$$R_{n,TC} = \frac{p(\boldsymbol{y}_n \mid \boldsymbol{x}_n, TC)\alpha_{TC|OC_n}}{\sum_{TC} p(\boldsymbol{y}_n \mid \boldsymbol{x}_n, TC)\alpha_{TC|OC_n}}, \quad TC = NC, CH$$

(8)

}
until likelihood converges

To classify as *NC* or *CH* a new pair of image patches $\boldsymbol{x}$ and $\boldsymbol{y}$, we apply the optimal Bayesian rule [7]:

$$\text{P}(TC = NC \mid \boldsymbol{x}, \boldsymbol{y}, OC) / \text{P}(TC = CH \mid \boldsymbol{x}, \boldsymbol{y}, OC) = \tag{9}$$
$$p(\boldsymbol{y} \mid \boldsymbol{x}, TC = NC)\alpha_{TC=NC|OC} / p(\boldsymbol{y} \mid \boldsymbol{x}, TC = CH)\alpha_{TC=CH|OC} > (<)\ bias$$

To test an arbitrary patch we can set the *OC* variable as missing obtaining:

$$\text{P}(TC = NC \mid \boldsymbol{x}, \boldsymbol{y}) / \text{P}(TC = CH \mid \boldsymbol{x}, \boldsymbol{y}) = \tag{10}$$
$$p(\boldsymbol{y} \mid \boldsymbol{x}, TC = NC)\pi_{TC=NC} / p(\boldsymbol{y} \mid \boldsymbol{x}, TC = CH)\pi_{TC=CH} > (<)\ bias$$

The *bias* variable is usually set to 1 (to minimize the probability of error) but other values may be used if other criteria are employed.

2.3 Model Selection

Once the estimation problem is solved we have to select the complexity of the model in (**1**). It depends on the rank parameters: r_{CH}, r_{NC}, q_{CH}, q_{NC}. To study the optimal values for the parameters we have made 5 replications of two fold cross-validation (5x2 cv) in several images. Results show that for most images the optimal values are: r_{CH}=0, q_{CH}=1, q_{NC}=1, and r_{NC} in the range [25,35]. For example, we show in Figure 3 the average negative log-likelihood (the smaller, the better) of the training and validation data after 5x2 cv in terms of the r_{NC} for the image pair partially reproduced in Figure 4, after fixing r_{CH}=0, q_{CH}=1, q_{NC}=1.
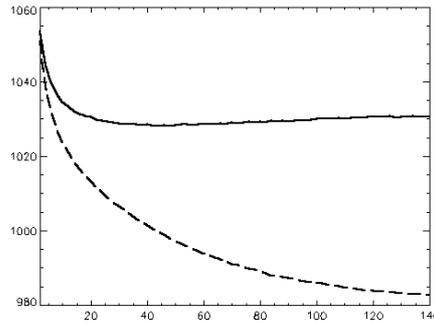

Figure 3 Average negative log-likelihood of the validation set (continuous line) and learning set (discontinuous line) after 5 replications of 2 fold cross-validation.

The minimum value for the average negative log-likelihood of the validation set is obtained for r_{NC} =48 and its value is equal to $l(r_{NC})$ = 1028.24 with a standard deviation of 1.13. To select the optimal value of r_{NC} we searched for

the smallest value of r_{NC} whose difference with that minimum is not significatively different. To obtain this value, the combined 5x2 cv test [8] was taken with significance level α =0.05. The obtained value is r_{NC} =28 with likelihood $l(r_{NC})$ =1029.09 and standard deviation of 1.64. Also the reduced rank model is significatively better than the full regression model.

It can be shown that the uncontrolled factor found by the model (q_{CH}=1, q_{NC} =1) corresponds to local changes of illumination and disappears if all patches are preset to a constant mean of illumination. Therefore, the model can be understood as a mixture of reduced rank regressions with rank r_{CH} = 0 for the *CH* class (no prediction is possible) and rank r_{NC} << d for the *NC* class.

3. Computational Results

We now present some computational results from the mixture model. In Figure 4 we show a partial reproduction (350x350 pixels) of an aerial image from the island of Gran Canaria (Spain). The learning set extracted from the full image (1250x1250 pixels) comprised 493 changes detected by the operator. Image patches were size 15×15 pixels, hence d =225. After a 5x2 cv step, rank values were r_{CH} = 0, q_{CH} = 1, r_{NC} = 28, q_{CH} = 1. In the left image of Figure 5 coincidences between the operator and the system are shown with white squares and differences with black circles. In the right image of Figure 5 the difference of log-likelihood between the *CH* class and *NC* is shown. Whiter regions denote a greater likelihood of change. This likelihood is also shown in 3D in Figure 6. The changes in Figure 5 (left) not validated by the system are shown in Figure 7.

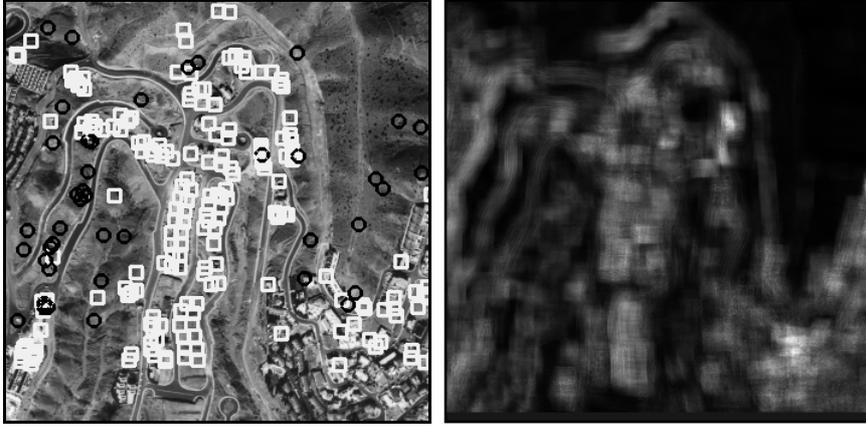Figure 4 Detail of one pair of the image from the experiments

Figure 5 Left: Coincidences of the operator and the system are denoted with white squares and differences with black circles. Right: Log-likelihood of change.
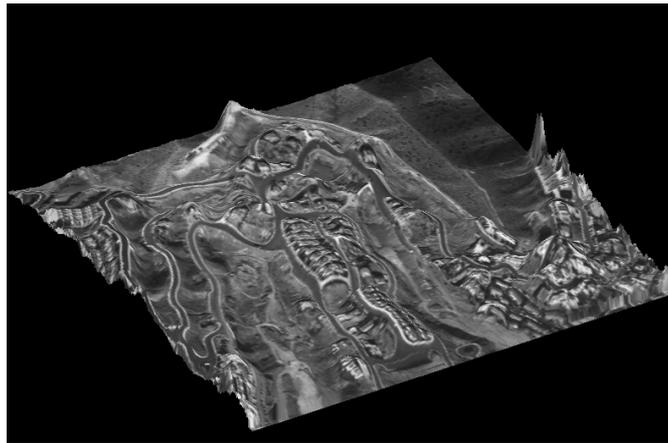

Figure 6 Likelihood of change represented in 3D over the right image in Figure 4. (the higher the more likelihood of change)
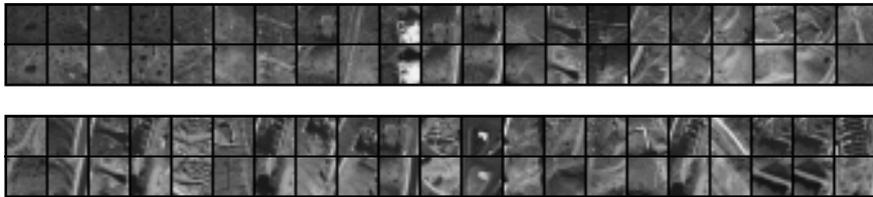

Figure 7 Changes not validated by the system in Figure 5 (left) are shown from top left to down right sorted by the differences of log-likelihood between the two classes (that is from less likelihood of change to more).

4. Conclusions

Change detection is an important part of image interpretation and automated data collection for Geographical Information Systems. In this paper, we have presented a probabilistic model for change validation in aerial images. This model uses a mixtute of reduced rank regressors whose parameters are estimated by Maximum Likelihood using the Expectation Maximization algorithm. Computational results are given that show the validity of the results.

References

1. D.L. Williams y M.L. Stauffer (1978). Monitoring Gypsy Moth Defoliation by Applying Change Detection Techniques to Landsat Imagery. Proc. of the Symp. for Vegetation Damage, Am. Soc. for Photogrammetry , pp. 221-229.
2. K. Ingram, E. Knapp and J.W. Robinson (1981). Change detection technique development for improved urbanized area delineation, Technical memorandum CSC/TM-81/6087, Comp. Sci. Corp, Maryland, USA.
3. G.F. Byrne, P.F. Crapper, and K.K. Mayo (1980). Monitoring Land Cover Change by Principal Component Analysis of Multitemporal Landsat Data. Remote Sensing of Environment, Vol. 10, pp. 175-184.
4. R. Wiemker, A. Speck, D. Kulbach, H. Spitzer and J. Bienlein (1997).Unsupervised Robust Change Detection on Multispectral Imagery Using Spectral and Spatial Features. Proceedings of the Third International Airborne Remote Sensing Conference and Exhibition, Copenhague
5. G. C. Reinsel and R.P. Velu. (1998) Multivariate Reduced Rank Regression: Theory and Applications. New- York. Springer-Verlag.
6. M. Tipping and Christopher Bishop. (1997). Mixtures of probabilistic principal component analyzers. Technical Report NCRG/97/003, Neural Computing Research Group, Aston University.
7. A.P. Dempster, N.M. Laird and D.B. Rubin. (1977), Maximum Likelihood from Incomplete Data via The EM Algorithm, Journal of Royal Statistical Society, Vol. 39, pp. 1-38, 1977.
8. E. Alpaydin (1999), Combined 5x2 cv F Test for Comparing Supervised Classification Learning Algorithms Neural Computation, 11(8), 1885-1982

Acknowledgments