MVA2015 IAPR International Conference on Machine Vision Applications, May 18-22, 2015, Tokyo, JAPAN

VHR Satellite Image segmentation based on topological unsupervised learning

Nistor Grozavu[*], Nicoleta Rogovschi [**], Guénaël Cabanes[*],
Andrés Troya-Galvis[***], and Pierre Gançarski[***]

[*]LIPN UMR CNRS 7030, Paris 13 University, 99 av. J.B. Clément, 93430 Villetaneuse, France,
nistor.grozavu,guenael.cabanes@lipn.univ-paris13.fr
[**]LIPADE, University Paris Descartes, 45, rue de Saintes Pères, 75005 Paris, France,
{*nicoleta.rogovschi*}*@parisdescartes.fr*
[***] ICube UMR CNRS 7357, Strasbourg University, 300 bd Sébastien Brant, Illkirch Cedex, France,
{*troyagalvis,gancarski*}*@unistra.fr*

Abstract

High spatial resolution satellite imagery has become an important source of information for geospatial applications. Automatic segmentation of high-resolution satellite imagery is useful for obtaining more timely and accurate information. In this paper we introduce a new approach for automatic image segmentation into different regions (corresponding to various features of texture, intensity, and color) based on topological unsupervised learning. Three types of methods were studied in this work: matrix factorization, self-organizing maps and probabilistic models. The approaches were applied on a real Very High Resolution (VHR) image of the French city of Strasbourg. The obtained segmentation results were validated using internal and external clustering validation indexes.

1 Introduction

Many real applications require Image Segmentation methods to identify regions of interest or to annotate the image. These methods can be categorized into three types: region-based segmentation, edge-based segmentation and data clustering [2]. Region-based segmentation includes the seeded and unseeded region growing algorithms (e.g. JSEG: J measure based SEGmentation), and the Fast Scanning algorithm. They expand each region of the image, pixel by pixel, based on their value so as to obtain a high positional relation between clusters. The edge-based segmentation methods usually use edge detection (i.e. the watershed algorithm). Finally, data clustering is used to group automatically similar pixels and/or regions of the image.

In the exploratory data analysis of high dimensional data, like Very High Resolution (VHR) satellite images, one of the main tasks is the formation of a simplified, usually visual, overview of the dataset. This can be achieved through simplified description or summaries, which provide the possibility to discover most relevant features or patterns. Clustering and projection are among the examples of useful methods to achieve this task. Classical clustering algorithms produce groups of similar data (e.g. different area or pixels of the image) according to a chosen criterion. Projection methods, on the other hand, represent the data in a lower dimensional space in such a way that the clusters and the metric relations of the data items are preserved as faithfully as possible. In this field, most algorithms use similarity measures based on Euclidean distance. However there are several types of data where the use of this measure is not adequate as in VHR images.

Topological learning is a recent direction in Machine Learning which aims to develop methods grounded on statistics to recover the topological invariants from the observed data points. Most of the existed topological learning approaches are based on graph theory or graph-based clustering methods.

In this work, we present new clustering approaches for VHR image segmentation based on two recently proposed topological learning approaches: TPNMF (Topographic Projective Matrix Factorization) [5] and lwo-SOM (local weighting observation Self Organizing Maps) [3]. Both are compared to GTM (Generative Topographic Mappings) [4].

The rest of the paper is organized as follows: a quick presentation of the three topological clustering models (TPNMF, lwo-SOM and GTM) is given in Section 2 after a short introduction in Section 1. In Sections 3, we present the validation of the proposed approach on the VHR satellite image of the city of Strasbourg (France). Finally the paper ends with a conclusion and some future works.

2 Topological Clustering

Topological learning is one of the most known technique to obtain a clustering and a visualization simultaneously. At the end of the topographic learning, the data points will be regrouped in different clusters based on their similarity. These clusters can be represented is a concise way using their gravity center or different statistical moments. This information is easier to manipulate than the original data points.

2.1 The GTM model

GTM was proposed by Bishop et al. [4] as a probabilistic counterpart to the Self-organizing maps (SOM) [6]. GTM is defined as a mapping from a low dimensional latent space onto the observed data space. The mapping is carried through by a set of basis functions generating a constrained mixture density distribution. It is defined as a generalized linear regression model:

$$y = y(z, W) = W\Phi(z) \quad (1)$$

where y is a prototype vector in the D-dimensional data space, Φ is a matrix consisting of M basis functions $(\phi_1(z), \ldots, \phi_M(z))$, introducing the non-linearity, W is a $D \times M$ matrix of adaptive weights w_{dm} that defines the mapping, and z is a point in latent space. The standard definition of GTM considers spherically symmetric Gaussians as basis functions, defined as:

$$\phi_m(x) = \exp\left\{-\frac{\|x - \mu_m\|^2}{2\sigma^2}\right\} \quad (2)$$

where μ_m represents the centers of the basis functions and σ - their common width. Let $\mathcal{D} = (x_1, \ldots, x_N)$ be the data set of N data points. A probability distribution of a data point $x_n \in \Re^D$ is then defined as an isotropic Gaussian noise distribution with a single common inverse variance β:

$$p(x_n|z, W, \beta) = \left(\frac{\beta}{2\pi}\right)^{D/2} \exp\left\{-\frac{\beta}{2}\|x_n - y(z, W)\|^2\right\} \quad (3)$$

The distribution in x-space, for a given value of W, is then obtained by integration over the z-distribution

$$p(x|W, \beta) = \int p(x|z, W, \beta)p(z)\,dz \quad (4)$$

and this integral can be approximated defining $p(z)$ as a set of K equally weighted delta functions on a regular grid,

$$p(z) = \frac{1}{K}\sum_{i=1}^{K}\delta(z - z_k) \quad (5)$$

So, equation (4) becomes

$$p(x|W, \beta) = \frac{1}{K}\sum_{i=1}^{K} p(x|z_i, W, \beta) \quad (6)$$

For the data set $\mathcal{D}$, we can determine the parameter matrix W, and the inverse variance β, using maximum likelihood. In practice it is convenient to maximize the log likelihood, given by:

$$\mathcal{L}(W, \beta) = \sum_{n=1}^{N} \ln\left\{\frac{1}{K}\sum_{i=1}^{K} p(x_n|z_i, W, \beta)\right\} \quad (7)$$

In this paper, we propose to use two topographic learning methods to compare with the GTM algorithm.

2.2 TPNMF: Topographic projective NMF

The TPNMF [5] incorporates neighborhood connections between PNMF basis functions arranged on a 2D topographic map, the objective to optimize becomes

$$||A - \mathbf{RHR}^T A||^2, \quad (8)$$

where A, R are a non-negative input matrix and a non-negative coefficient matrix, respectively, as in the original NMF. The new term $\mathbf{H} = (\mathbf{h}_{rs})$ is a $K \times K$ non-negative dimensional matrix that defines neighborhood connections between K basis functions. Choosing $\mathbf{H}$ as the identity matrix reduces the TPNMF to the PNMF. We arranged basis functions on a two-dimensional square-lattice topographic map, and set neighborhood connection weights to be normal distribution (Gaussian) functions on the map. The model consists of a discrete set $\mathcal{C}$ of cells called "map". This map has a discrete topology defined by an undirected graph, which usually is a regular grid in two dimensions. For each pair of cells (r,s) on the map, the distance $\delta(r, s)$ is defined as the length of the shortest chain linking cells r and s on the grid. For each cell this distance defines a neighbor cell; in order to control the neighborhood area, we introduce a kernel positive function $\mathbf{h}$ ($\mathbf{h} \geq 0$ and $\lim_{|y|\to\infty} \mathbf{h}(y) = 0$). We define the mutual influence of two cells r and s by $\mathbf{h}_{r,s}$. In practice, as for traditional topological maps we use a smooth function to control the size of the neighborhood as:

$$\mathbf{H} = (\mathbf{h}_{r,s}) = \exp\left(\frac{-\delta(r, s)}{T}\right).$$

Using this kernel function, T becomes a parameter of the model and as in the Kohonen [6] algorithm, we decrease T from an initial value T_{max} to a final value T_{min}. The minimization of (8) leads to the following update rule

$$\mathbf{R} \leftarrow \mathbf{R} \odot \frac{2AA^T\mathbf{RH}}{\mathbf{RHR}^T AA^T\mathbf{RH} + AA^T\mathbf{RHR}^T\mathbf{RH}}. \quad (9)$$

Hereafter, the pseudo code of the proposed algorithm.

Algorithm 1 TPNMF

Input: data $A \in R^{m\times n}$ and $K \leq min(m, n)$
Output: $\mathbf{R}, \mathbf{H}$
Initialize: select random nonnegative $\mathbf{R} \in \mathbb{R}_+^{N\times K}$ and $\mathbf{H} \in \mathbb{R}_+^{K\times K}$. Choose T_{max}, T_{min} and N_{iter}.
repeat

$$\mathbf{H} = (\mathbf{h}_{r,s}) = \exp\left(\frac{-\delta(r, s)}{T}\right) \quad (10)$$

$$\mathbf{R} \leftarrow \mathbf{R} \odot \frac{2AA^T\mathbf{RH}}{\mathbf{RHR}^T AA^T\mathbf{RH} + AA^T\mathbf{RHR}^T\mathbf{RH}} \quad (11)$$

until stabilization of R ($t \leq N_{iter}$)**.**
Classification-step: For $i = 1, \ldots, N$ **each** $\mathbf{a_i}$ **is assigned to the k*th* cluster, according to:**

$$k =_{k'} \mathbf{R}_{ik'}, k' = 1, \ldots, K$$

2.3 Local Weighting Observations : *lwo*-SOM

The method *lwo*-SOM proposed by [3] is an adaptation of the classical SOM by adding a weighting parameter in the objective function which allows to weight the features during the learning process and to obtain better results.

Indeed, the proposed clustering algorithm and feature weighting aims to select the optimal prototypes, observations and feature weights at the same time. Each prototype $\mathbf{w}_j = (w_j^1, w_j^2, ..., w_j^m)$ corresponding to cell j is allowed to have its own set of local features weights $\pi_j^{(o)} = (\pi_j^{(o)1}, \pi_j^{(o)2}, ..., \pi_j^{(o)m})$ and its own set of local distance weights $\pi_j^{(d)} = (\pi_j^{(d)1}, \pi_j^{(d)2}, ..., \pi_j^{(d)m})$ respectively. We denote the set of weight vectors ($|\Pi| = |W|$) by $\Pi = \{\pi_j, \pi_j \in \Re^m\}_{j=1}^{|\Pi|}$ for both observation and distance weighting.

The *lwo*-SOM has the following objective function:

$$R_{lwo}(\chi, \mathcal{W}, \Pi) = \sum_{i=1}^{|E|} \sum_{j=1}^{|\mathcal{W}|} \mathcal{K}_{j,\chi(\mathbf{x}_i)} \|\pi_j \mathbf{x}_i - \mathbf{w}_j\|^2 \quad (12)$$

where π_j are the observations weights.

According to this function, the prototype's vectors are updated using the following expression:

$$\mathbf{w}_j(t+1) = \mathbf{w}_j(t) + \epsilon(t)\mathcal{K}_{j,\chi(\mathbf{x}_i)} \left(\pi_j \mathbf{x}_i - \mathbf{w}_j(t)\right)$$

As in the traditional stochastic learning algorithm of Kohonen, we denote by $\epsilon(t)$ the learning rate at time t. The training is usually performed in two phases. In the first phase, a large initial learning rate $\epsilon(0)$ and a large neighborhood radius T_{max} are used .

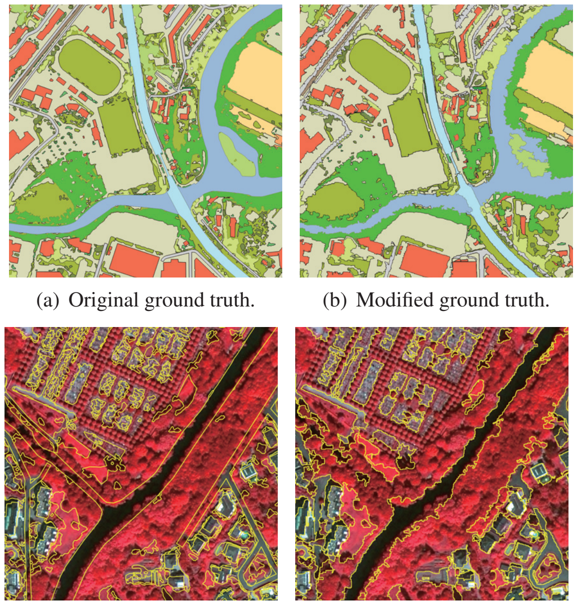
These two methods (TPNMF and lwo-SOM) were already tested by the authors for differents types of data and the experimental results shows that these models outperforms classical clustering methods [5] [3].

3 Experimental results

In this section we present the results obtained using the three topological learning methods on the Strasbourg images data (Figure 1).

Figure 1. A small part of the VHR Strasbourg image

The image has first been preprocessed in order to create a data set made of 187058 superpixels, each of them described by 27 attributes either geometrical or radiometrical [1]. These attributes include the geographic position of the superpixel, the surface of the area covered by the superpixel, the mean RBG values, the contrast compared to neighbor pixels and superpixels, the brightness, and the standard deviations, among others. In order to validate our results, we compared the obtained segmentations with maps of the area made by expert geographers (cf. Figure 2(a)). These maps were produced by a hybrid methodology, which mixes data from topographic databases for roads and buildings, supervised classification for different types of water and vegetation, as well as further manual refinement in order to reduce classification errors. However, there is an unknown amount of uncertainty associated to these maps and they are not fully reliable to be used as ground truth as they are. Indeed, topographic data are taken *in situ* and do not align well with the image data, because of the satellite capture angle, as well as further geometrical corrections applied to the image. This problem is easily observable by a closer inspection of the maps (cf. Figure 2(c)).



(a) Original ground truth. (b) Modified ground truth.

(c) Zoomed boundary representation of the original ground truth maps. (d) Zoomed boundary representation of the modified ground truth.

Figure 2. Extract of the original and improved ground-truth images

To validate the approaches, we used two types of validations : visual validation and numerical validation.

Visual validation

To test the quality of the obtained segmentation, we first performed a visual evaluation of the segmentation results (Figures 3 and 4) on the VHR Strasbourg image dataset. These figures represent a portion of the automatic segmentation of the image by the TPNMF and lwo-SOM models. It is easy to see that the water cluster (the Rhine river), the roads and the vegetation are clearly detected by both models. The oval building situated on the left of the image corresponds to the European Parliament, and we can note that, for this part of the image, the TPNMF model give better results compared to lwo-SOM model.

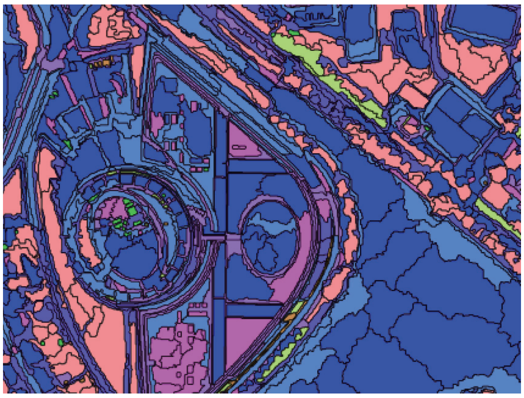


Figure 3. Image segmentation using lwo-SOM

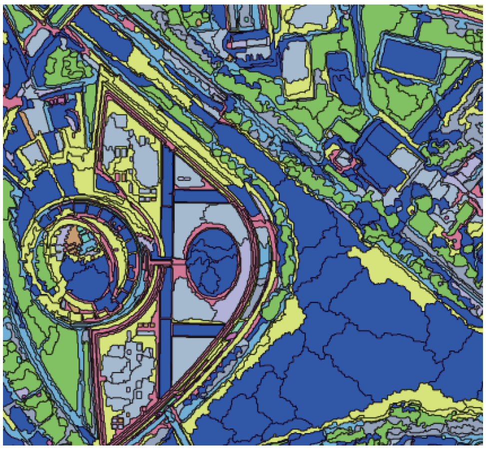


Figure 4. Image segmentation using TPNMF

Table 1. Quality Index results

Method	RI	GCE	VI	DB
TPNMF	0.7793	0.7326	5.1868	31.745
lwo-SOM	0.7752	0.7293	5.1674	34.249
GTM	0.7712	0.7223	4.8618	36.017

Numerical Validation

The comparison is based on four quantitative performance measures [7], [8]:
1. The Probabilistic Rand Index (PRI) counts the fraction of pairs of superpixels whose labellings are consistent between the computed segmentation and the ground truth, averaging across multiple ground truth segmentations to account for scale variation in human perception.
2. The Variation of Information (VoI) metric defines the distance between two segmentations as the average conditional entropy of one segmentation given the other, and it measures the amount of randomness in one segmentation which cannot be explained by the other. It is used to compare the segmentation with the ground truth.
3. The Global Consistency Error (GCE) measures the extent to which one segmentation can be viewed as a refinement of the other. Segmentations which are related in this way are considered to be consistent, since they could represent the same natural image segmented at different scales. This index is used to compare the computed segmentation with the ground truth.
4. The DB_{nc} is the average similarity between each cluster $c_i, i = 1, ..., n_c$ and the most similar other. We seek clusterings that minimize the DB, and thus have minimum possible similarity between the clusters.

As we can see in Table 1, the quality of the the segmentation is good for this kind of image, which are know to be difficult to segment accurately due to their complexity. The obtained image segmentations are relatively close to the ground truth for the three methods (the RI indexes are above 0.77, a perfect agreement would be 1) and the new approaches outperform the GTM model for the four quality indexes.

We can note that the fusion of the three segmentations gives a better result than lwo-SOM and GTM alone (RI = 0.778), but does not outperform TPNMF.

4 Conclusions

In this paper, we proposed a framework for Very High Resolution image segmentation, based on topological learning. The algorithms described in this paper provide clustering of small regions from a VHR satellite image and the experimental results have shown promising performance. Several perspectives can be considered for this work: currently, we are investigating a "collaborative learning" approach between the presented methods in order to increase the quality of the final segmentation.

Acknowledgments

This work has been supported by the ANR Project COCLICO, ANR-12-MONU-0001, France.

References

[1] S. Rougier, and A. Puissant: "Improvements of urban vegetation segmentation and classification using multi-temporal Pleiades images" *5th Inter. Conf. on Geographic Object-Based Image Analysis*, p.6, 2014.

[2] Yu Hsiang Wang Tutorial: Image Segmentation Graduate Institute of Communication Engineering National Taiwan University, Taipei, Taiwan, ROC.

[3] Nistor Grozavu, Youns Bennani, Mustapha Lebbah, "From variable weighting to cluster characterization in topographic unsupervised learning". Intern. Joint Conference on Neural Networks, 2009: 1005-1010

[4] Christopher M. Bishop and Markus Svensn and Christopher K. I. Williams, "GTM: The generative topographic mapping", Journal of Neural Computation, 1998, volume 10, pages 215-234.

[5] Nicoleta Rogovschi, Lazhar Labiod, Mohamed Nadif, "A topographical nonnegative matrix factorization algorithm", International Joint Conference on Neural Networks, IJCNN 2013: 1-6.

[6] T. Kohonen, Self-organizing Maps. Springer Berlin, 2001.

[7] C. Pantofaru and M. Hebert. A comparison of image segmentation algorithms. Technical Report CMU-RI-TR-05-40, CMU, 2005. 2, 9

[8] M. Meila. Comparing clusterings: an axiomatic view. In Proc. International Conference on Machine Learning, pages 577584, 2005.