

Contextual classification of remotely sensed images with integer linear programming

Manuel Lameiras Campagnolo

Dept. Matemática, Instituto Superior de Agronomia, Univ. Técnica de Lisboa and SQIG - IT

mlc@isa.utl.pt

Jorge Orestes Cerdeira

Dept. Matemática, Instituto Superior de Agronomia, Univ. Técnica de Lisboa

orestes@isa.utl.pt

Supervised classification techniques are commonly used to assign pixels of multispectral satellite imagery to a predefined set of classes in order to generate or update land use or land cover maps from remote sensed data. These techniques have a limited ability in expressing spatial relationships among pixels. We propose a new contextual approach to address this issue. In particular, we present an integer linear programming formulation which restricts the number of distinct objects in the classified image and we propose a heuristic for the resulting problem. We test it on one generated data set and one real data set.

1 INTRODUCTION

In a multispectral remote sensed image with n bands, each pixel of the image is described by a n -dimensional vector called the pixel's spectral signature. In image classification, one considers k distinct classes and looks for the “best” assignment of each pixel to one and only one class. Many techniques have been proposed to solve this problem (Landgrebe 2003) but the achieved solutions are frequently not satisfactory. This usually happens when classes are hard to discriminate in the space of spectral signatures and when the existing spatial relations among pixels in the remotely sensed images are not fully explored.

Formally, an assignment of pixels is a function y such that $y_i^c = 1$ if pixel i is assigned to class c and is 0 otherwise. For simplicity, we say that pixel i is of class c if $y_i^c = 1$. Supervised classification techniques typically look for a partition $R_1, \dots, R_k \subset \mathbb{R}^n$ of the space of spectral signatures such that the best decision rule is given by $y_i^c = 1$ if and only if $x^{(i)} \in R_c$ and $x^{(i)}$ is the spectral signature of pixel i . If d is an appropriate distance between the pixels' signatures and classes, this is equivalent to minimize the global function $D = \sum_{i,c} d_i^c y_i^c$, where d_i^c is the distance from pixel i to class c . Classifiers differ in the choice of function d . For instance, quadratic discriminant classifiers (usually known as “maximum likelihood classifiers” in the remote sensing literature) use

the Mahalanobis distance, which is a function of the spectral signature of each pixel and of the mean vector and variance-covariance matrix of each class. Alternatively, the m nearest neighbors non-parametric classifier estimates the probability of pixel i belonging to class c as the proportion of the closest m neighbors of i (in the spectral space) that belong to class c . Then, the class which is closest to pixel i is the one with the highest probability. Other classifiers like neural networks, support vector machines and decision trees can be described in the same framework.

However, something is missing in the approaches outlined above. Since d depends uniquely on the spectral signature of each individual pixel and on the description of the classes, then the spatial relations among pixels in the image are not accounted for. This per-pixel *non contextual classification* exhibits the well known “salt and pepper” effect. In opposition, classification techniques which take into account the spatial neighborhood relations among pixels are called “contextual”. As far as we know, currently used contextual classifiers only incorporate in the decision rule information about the *local* neighbors of each pixel. The most straightforward approach for local contextual classification is to revise pixel-to-class assignment *a posteriori*. A simple example is the application of a 3×3 mode filter to the image after classification. A popular and more realistic model of lo-

cal spatial context is given by Markov random fields in the framework of statistical Bayesian classification (Tso and Mather 1999).

While local decision rules may be suitable for certain cases, they may not apply to all spatial objects that can be identified in satellite imagery. For instance, if a patch of pixels of the same class is crossed by a one-pixel-wide feature of a different class (for instance, a road in a satellite image), which may not be clearly discriminated in the spectral space from the surrounding pixels, then a local rule would tend to remove this feature from the classified image.

2 THE CONCEPTUAL MODEL

We propose here a *global* contextual classifier. As before, and given any distance d between pixels and classes in the spectral space, we look for the assignment y that minimizes $D = \sum_{i,c} d_i^c y_i^c$. However, we add a restriction to the problem in order to prevent the “salt and pepper” effect derived from excessive number of small size patches of pixels. A patch is a maximal set of contiguous pixels belonging to the same class. One can view a patch as a spatial object in the classified image. Since objects in the image are somewhat spatially continuous, the key idea is to restrict the number of patches in the image, while preserving meaningful “narrow” objects. To turn this more precise consider the graph $G = (P, E)$, where P is the set of pixels and E is the set of pairs of pixels which are neighbors in the image. If y is any assignment of pixels to classes, a patch is a connected component (i.e., a maximal connected subgraph) of the subgraph of G induced by the set of pixels in a same class under y . Thus, different patches may be from the same class, but two neighbor patches (i.e., with a pair of neighbor pixels not both in the same patch) have to belong to different classes.

In this setting, the optimal assignment is given by the solution of

$$\min D = \sum_{i,c} d_i^c y_i^c \quad (1)$$

$$\sum_c y_i^c = 1, \text{ for every } i \quad (2)$$

$$y_i^c \in \{0, 1\}, \text{ for every } i, c \quad (3)$$

$$\text{number of patches} \leq N \quad (4)$$

where N is a suitable number of patches for the image to be processed.

3 ANALYSIS

We now show how to turn condition (4) into a linear description and we propose a heuristic for the resulting problem.

3.1 A linear model

The key idea is to define a spanning forest T (i.e., an acyclic subgraph spanning all vertices) of G with N components, and assign to the same class all pixels within the same component of T .

Let us therefore consider decision variables x_{ij} on every edge $[i, j]$ of G , and define $T = \{[i, j] \in E : x_{ij} = 1\}$ with

$$\sum_{i,j} x_{ij} = |P| - N \quad (5)$$

$$x_{ij} \in \{0, 1\}, \text{ for every } i, j \quad (6)$$

$$\sum_{i,j \in E(S)} x_{ij} \leq |S| - 1, \text{ for every } S \subset P \quad (7)$$

where $E(S)$ is the set of edges of G with both ends in S . Equation (5) ensures that forest T has $|P| - N$ edges, and inequalities (7) are the usual subtour elimination constraints preventing T from having cycles. Together, (5) and (7) imply that the forest has exactly N components.

We now have to guarantee that pixels within the same component of T will be assigned to the same class. The following simple observation allows us to guarantee this just from the knowledge of the edges of T . If two pixels from the same component of T are in different classes, then there will be some edge $[i, j]$ of T such that i and j are in different classes. Thus, the valid inequalities

$$1 - x_{ij} \geq y_i^c - y_j^c, \text{ for every } i, j, c \quad (8)$$

$$1 - x_{ij} \geq y_j^c - y_i^c, \text{ for every } i, j, c \quad (9)$$

ensures that whenever $[i, j]$ is in T , i.e., $x_{ij} = 1$, then i and j belong to the same class.

Note that distinct neighbor components of the N -component forest T may belong to the same class. However, if (y^*, x^*) is an optimal solution of (1)-(3),(5)-(9), then y^* is an optimal solution of (1)-(4), where the number of patches can be strictly smaller than N .

The use of an integer linear programming solver to deal with the formulation (1)-(3), (5)-(9) has to handle the huge number of constraints (7). One possible approach is to start with only (1)-(3), (5),(6), (8),(9), and successively add inequalities (7) that are violated by the current solution until a feasible solution is reached, which would be optimal. However, the number of constraints (7) to be added may make this procedure unpractical, even for a small data set (say, an image with 400 pixels like the one described in subsection 4.1).

To speed up this procedure classes of strong valid inequalities (i.e., cuts) for this classification problem have to be devised. This is the subject of polyhedral combinatorics theory: see for example (Pulleyblank

1983) or (Schrijver 1995). Nevertheless, any such integer cutting algorithm will certainly suffer from serious limitations regarding the size of the data sets. Hence, the development of accurate heuristics is a reasonable option for the problem in hand.

3.2 A heuristic approach

We design a heuristic for (1)-(3), (5)-(9) that can be viewed as a dynamic version of Kruskal’s algorithm for minimum spanning trees (Kruskal 1956).

The algorithm performs $|P| - N + 1$ iterations. In each iteration $t = 0, 1, \dots, |P| - N$ a spanning forest T_t with $|P| - t$ components is (implicitly) defined, and all the pixels within each component I of T_t are assigned to the same class c . Thus, $D_I = \sum_{v \in I} d_v^c$ is the contribution of the component I for the objective function (1), and $\sum_I D_I$ is the value of the current solution.

At iteration 0 the spanning forest T_0 has $|P|$ components, each consisting of a single pixel which is assigned to its closest class. This is the non contextual classification.

In iteration $t \geq 1$, for each pair of adjacent components I, J of T_{t-1} , the value

$$c_{IJ} = \min_{\{c\}} \left\{ \sum_{v \in I} d_v^c + \sum_{v \in J} d_v^c \right\} - (D_I + D_J) \quad (10)$$

is computed. The weight c_{IJ} on “edge” (I, J) is the minimum increment on the objective function (1) derived from merging components I and J within the same class. The “edge” (I^*, J^*) for which the minimum weight

$$c_{I^*J^*} = \min_{\{I, J\}} \{c_{IJ}\}$$

is attained is added to T_{t-1} , meaning that components I^* and J^* will be contracted into a single component of T_t , with every pixel of $I^* \cup J^*$ assigned to the same class (that one which was used to settle $c_{I^*J^*}$ in (10)).

The algorithm can be implemented with time and space complexities $O(k|E| + k|P|(|P| - N))$ and $O(k|P| + |E|)$, respectively.

4 EXAMPLES

To assess the quality of the solutions produced by the above heuristic algorithm, we use two data sets. The first is a small generated data set with 400 pixels and 2 classes. The second is a data set with 12513 pixels and 5 classes derived from SPOT-XS satellite imagery.

4.1 Generated data

This data set has two classes. Class 1 represents some linear features (e.g. roads) and class 2 represents an homogeneous background. The image has 20 rows and 20 columns. It was generated from the image in

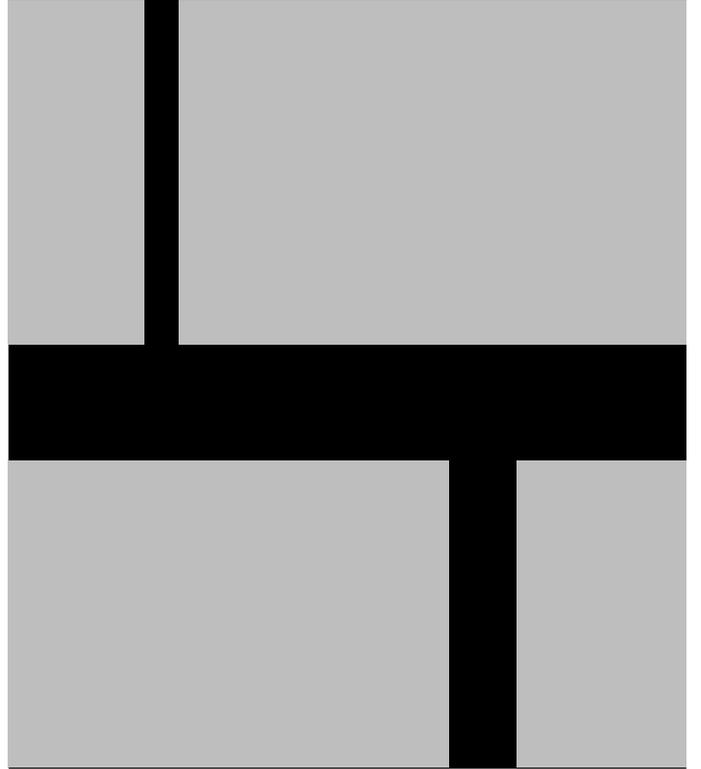


Figure 1: Reference image with two classes.

Figure 1, adding Gaussian noise to each class. Specifically, the amount of noise is such that the probability of misclassification for a pixel extracted from class 1 is 10%. For class 2, the corresponding probability is 20%. The resulting non contextual (per pixel) classification is shown in Figure 2. Since the number of patches in the reference image is 5, we run our algorithm up to that number of components. The resulting classification is shown in Figure 3.

4.2 Real data

The image covers the Monsanto area, in Lisbon, Portugal, with SPOT-XS data from the summer of 1990. The spatial resolution is 20 meters. It is a small area of 2580 meters by 1940 meters covered by 12513 pixels. We have only considered five classes: water, roads, urban areas, forested areas and agricultural areas. We used a random sample of 500 pixels (approximately 4% of the total) for estimating the spectral signatures of the classes. The distances d_i^c were computed using the k-nearest neighbors technique. The land cover map of that area is known from the photo-interpretation of aerial photographs of the same year (see Figure 4).

The resulting per pixel classification is shown in Figure 5. In this map, approximately 75% of the pixels are well classified according to the land cover map.

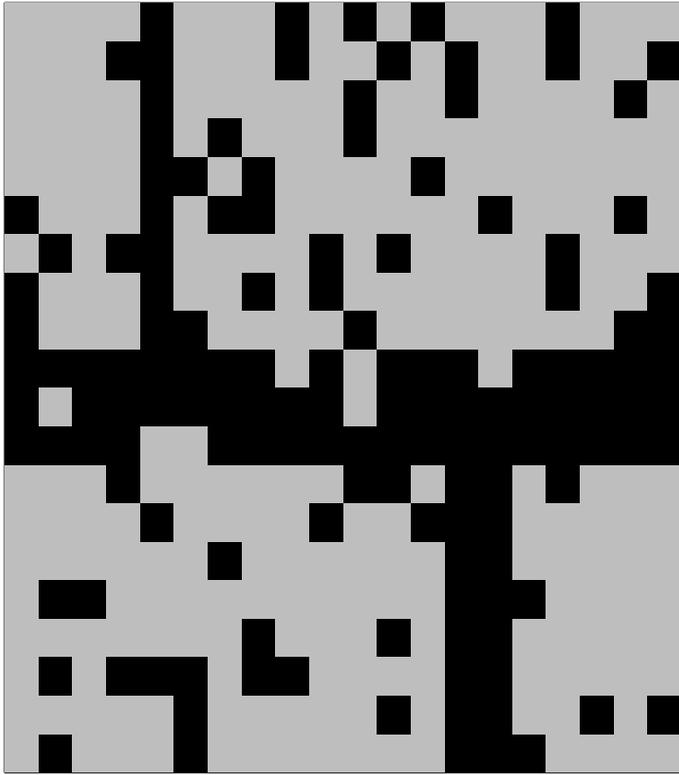


Figure 2: Non contextual classification after adding noise to Figure 1.

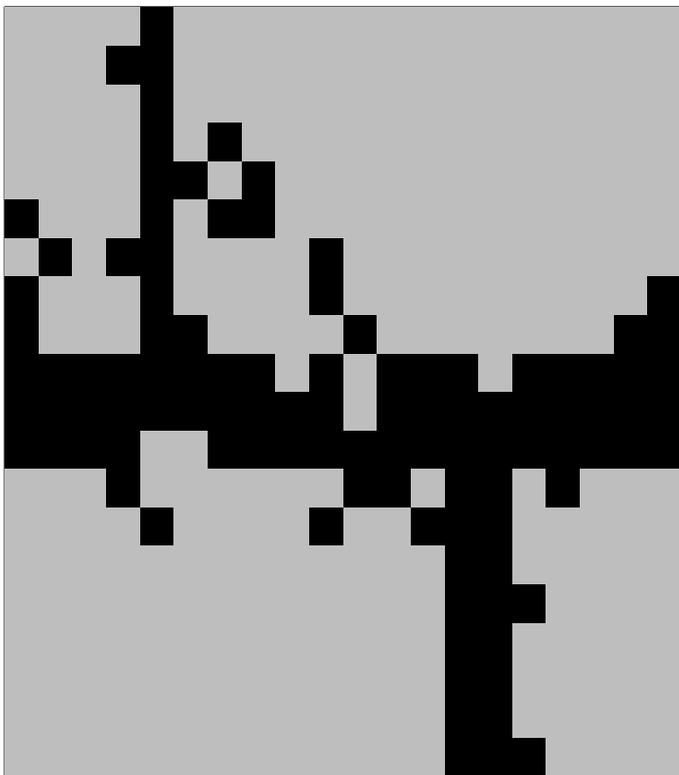


Figure 3: Contextual classification with 5 patches (heuristic solution).

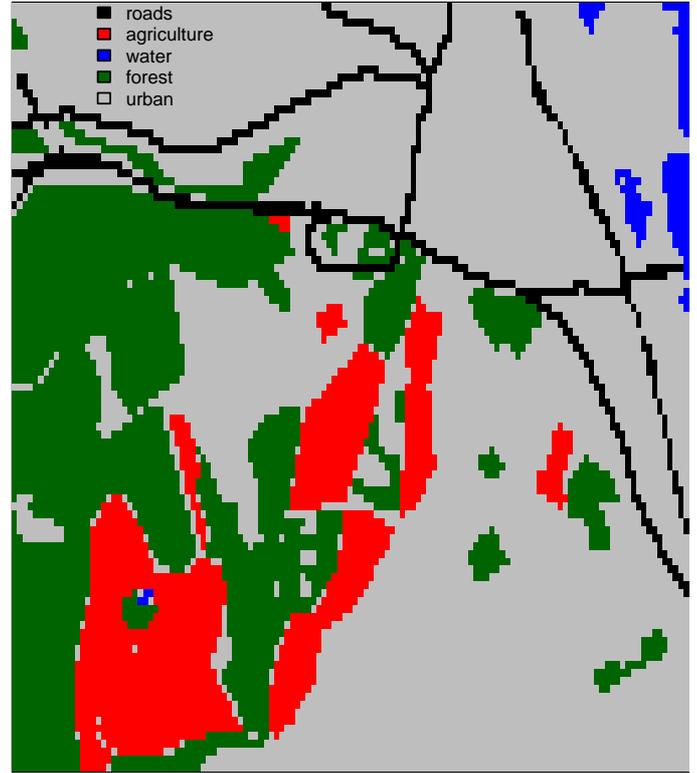


Figure 4: Land cover map obtained by photo-interpretation.

We computed the value of the objective function (1) for every iteration of the heuristic algorithm. The values are plotted in Figure 6. The objective function (for the per pixel classification) stays constant until the iteration $t = 12181$, meaning that the non contextual classification has $12513 - 12181 = 332$ patches. Note that for a number of components less than 48, i.e. for iteration $t > 12513 - 48$, the objective function is strictly increasing, which means that for less than 48 components all components of the solution are patches. We show in Figure 7 the resulting map with 15 patches. For this map, the percentage of pixels misclassified by the non contextual classifier that are correctly classified by the algorithm is 24.62 %. In opposition, the percentage of pixels correctly assigned in the non-contextual classification that are re-assigned to a different class is 4.17 %.

5 CONCLUSIONS

We propose a mathematical model to deal with global contextual classification of multispectral images. Furthermore we were able to devise a 0-1 linear formulation (1)-(3), (5)-(9), and developed a heuristic algorithm for the problem.

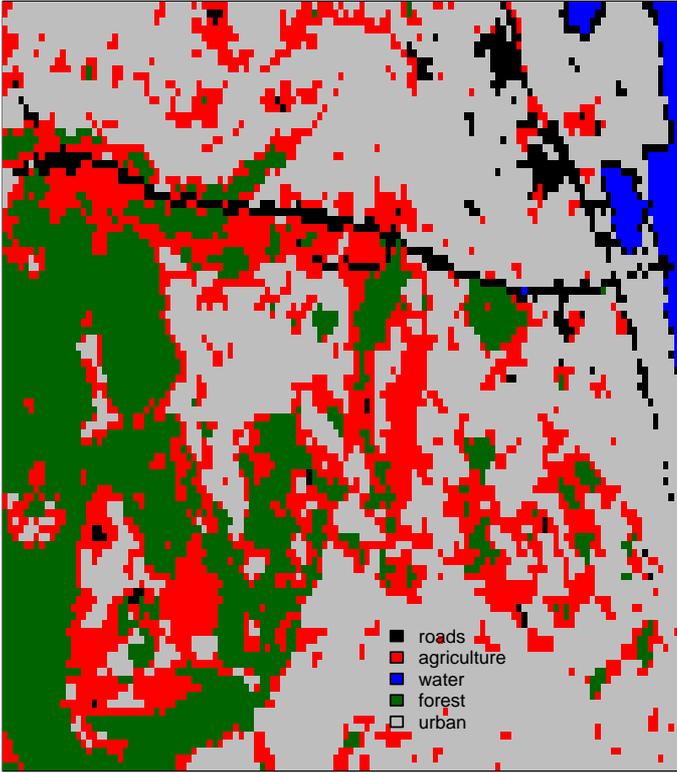


Figure 5: Non contextual classification of the SPOT-XS image.

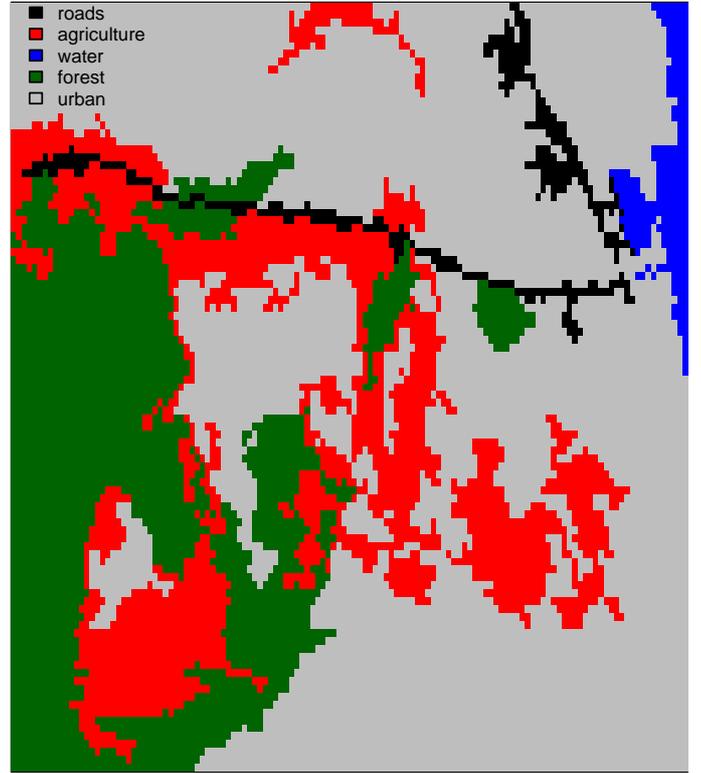


Figure 7: Contextual classification of the SPOT-XS image with 15 patches (heuristic solution).

The computational tests carried out allow us to conclude that the heuristic approach deals with some of the problems of the non contextual classification such as the “salt and pepper” effect. However, the tests we ran show that the algorithm behaves poorly regarding two kinds of features in the image. Firstly, it does not remove spurious “peninsulas” attached to relevant objects in the image (see Figure 3). Secondly, it is not successful in restoring linear features (e.g. roads) of the original image (compare Figures 5 and 7).

The first limitation is inherent to the model (1)-(4). We discuss below a possible way to overcome this limitation. The second is due to the particular heuristic we used. Since components defined by the algorithm are never broken apart, pixels in the background that are merged together in early iterations cannot be re-assigned to the linear features. However, the optimal solution of (1)-(4) is not contrived in this manner.

As we said above, we think that the current formalization can be further improved. We believe that the restrictions (2)-(4) are effective in accounting for the spatial relations among pixels of the image. However, the objective function (1) suffers from the lack of accuracy on the distances d_i^c , which are estimated from the data. For pixels whose spectral signatures are clearly related to one class, the estimates are reliable. However, for the remainder pixels, the random variability on the estimated values of d_i^c will affect significantly the results. A possible approach to overcome this could be to “randomize” (1) to take into account

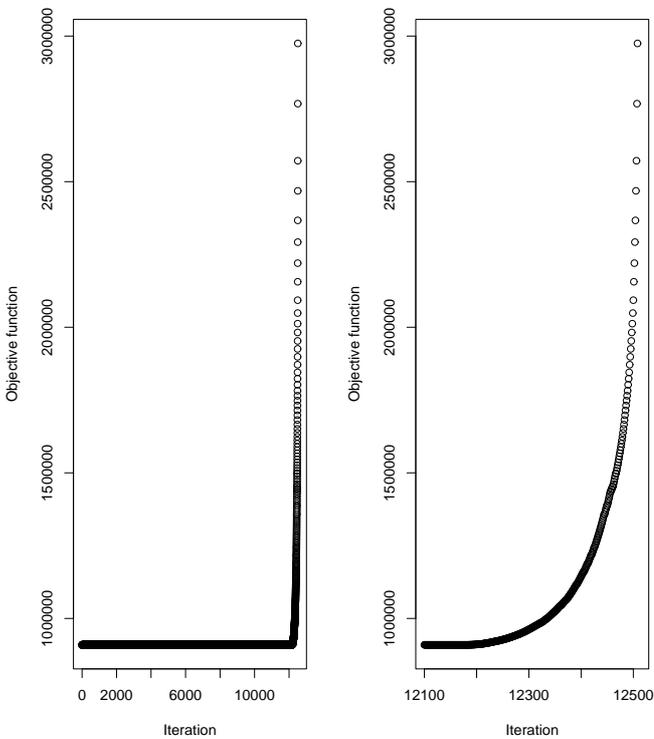


Figure 6: Objective function values for iterations 0 through 12508 of the algorithm. The figure on the right highlights iterations 12100 to 12508.

that phenomena. One possibility is to repeatedly run our algorithm on slightly perturbed values of the distances between pixels and classes. Then, one of the possible outcomes could be selected (for example, by an expert or by comparing their accuracy) as the solution of the problem. We intend to develop these ideas in a forthcoming paper.

The present work describes a new approach for contextual classification of remotely sensed images. Its global nature distinguishes it from usual techniques based on local post-classification filters or local random fields. Our formulation of the problem in a combinatorial optimization setting opens the way to explore many tools and techniques which are seldom applied to image classification. This paper is just a first step in that direction.

6 ACKNOWLEDGMENTS

This work was partially supported by FCT and EU FEDER, namely by CLC, grant SFRH/BPD/21012/2004 (MLC) and by POCTI Program from FCT (JOC). MLC's contribution was partly done while visiting the Department of Geography at UMD. We would like to thank Chris Justice from UMD and Leonor Pinto from ISEG for useful discussions on the subject of this paper.

REFERENCES

- Kruskal, J. B. (1956). On the shortest spanning-subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society* 7, 48–50.
- Landgrebe, D. (2003). *Signal Theory Methods in Multispectral Remote Sensing*. John Wiley and Sons.
- Pulleyblank, W. R. (1983). Polyhedral combinatorics. In *Mathematical Programming – The State of the Art*, pp. 312–345. Springer, Berlin.
- Schrijver, A. (1995). Polyhedral combinatorics. In *Handbook of Combinatorics*, Volume II, pp. 1649–1704. Elsevier, Amsterdam.
- Tso, B. and P. Mather (1999). Classification of multisource remote sensing images using a Markov random field. *IEEE Transactions on Geoscience and Remote Sensing* 37, 1255–1260.