Can we make remote sensing image classification fast enough?

Rodrigo José Pisani[1]
Rodrigo Mizobe[2]
Paulina Setti Riedel[1]
Célia Regina Lopes Zimback[3]
João Paulo Papa[2]

[1]UNESP – Univ Estadual Paulista - IGCE/UNESP
PO Box 176 - 13506900 – Rio Claro - SP, Brazil
{pisani,psriedel}@rc.unesp.br,

[2] UNESP – Univ Estadual Paulista - FC/UNESP
PO Box 473 - 17033360 - Bauru - SP, Brazil
m1z0b3@gmail.com,papa@fc.unesp.br

[3] UNESP – Univ Estadual Paulista - FCA/UNESP
PO Box 237 - 18610307 - Botucatu - SP, Brazil
czimback@fca.unesp.br

Abstract. In this paper we would like to shed light over the problem of efficiency and effectiveness of image classification in large datasets. As the amount of data to be processed and further classified has increased in the last years, there is a need for faster and more precise pattern recognition algorithms in order to perform online and offline training and classification procedures. We deal here with the problem of land use classification in middle resolution satellite images in a fast manner. Experimental results using Optimum-Path Forest and its training set pruning algorithm are also provided and discussed.

Keywords: land use classification, optimum-path forest, remote sensing, classificação de uso da terra, floresta de caminhos ótimos, sensoriamento remoto.

1. Introduction

Automatic classification in large collection of images has been a challenge in the last years. The exponential growing of embedded technologies in digital cameras and satellite onboard systems has introduced a new concept in the image classification research field: can we ally efficiency and effectiveness in object recognition?

A common non-professional digital camera may produce images with millions of pixels to be classified, leading us to face a new paradigm that involves classification in massive datasets. Several techniques have been developed in order to overcome such problem, such as LASVM (Bordes et al., 2005), SVMTorch (Collobert and Bengio, 2001), LibLINEAR (Fan et al., 2008), and LibOPF (Papa et al., 2009a). The former approaches were designed to adapt the well succeed Support Vector Machines (SVMs) for large datasets, and the latter technique, Optimum-Path Forest (OPF) was proposed aiming to provide graph-based pattern recognition algorithms with reduced training time (Papa et al., 2009b).

Although OPF has demonstrated to be similar to SVMs regarding accuracy, it performs training much faster (Papa et al., 2009b). Moreover, it is often desirable to handle the pattern classification as fast as possible. Imagine a situation in which we have an interactive remote sensing image classification system. The user may want to mark some samples, which will be used to train a classifier. The remaining pixels will be classified

according to the classifier designed over the training samples. After that, the user may refine the results marking another set of samples, and even so unmarking misclassified ones. This process can be repeated over several iterations until the user be satisfied. Note that it is desirable to have a fast training and classification processes in order to provide a user-friendly framework, mainly in applications that tackle high resolution and multispectral images.

Aiming to speed up the OPF classification time, Papa et al. (2009c) proposed a learning algorithm to identify the irrelevant samples of the training set, i.e., the ones that did not participate from any classification process, for further remove them. This training set pruning algorithm was first validated in the context of rainfall occurrence estimation in satellite images, and has demonstrated to be very efficient with less affecting the accuracy over the test set in some applications. In that approach, when a sample from the evaluating set is classified, all training nodes responsible for that classification are marked. At the final of the process, the unmarked samples are then removed from the training set and discarded.

In such a way, we would like to shed light here the importance of having efficient and effectiveness classifiers, which will bound the amount of information that can be handled in a nearby future. In this paper we evaluated how much we can penalize the OPF classifier effectiveness by increasing its efficiency. The remainder of this paper is organized as follows. Section 2 presents OPF and its training set pruning algorithm. Section 3 discusses the experimental results, and Section 4 states conclusions.

2. Optimum-Path Forest

Given a training set with samples from distinct classes, we wish to design a pattern classifier which can assign the true class label to any new sample. Each sample is represented by a set of features and a distance function measures their dissimilarity in the feature space. The training samples are then interpreted as the nodes of a graph, whose arcs are defined by a given adjacency relation and weighted by the distance function. It is expected that samples from a same class/cluster are connected by a path of nearby samples. Therefore, the degree of connectedness for any given path is measured by a connectivity (path-value) function, which exploits the distances along the path. In supervised learning, the true label of the training samples is known and so it is exploited to identify key samples (prototypes) in each class. Optimum paths are computed from the prototypes to each training sample, such that each prototype becomes root of an optimum-path tree composed by its most strongly connected samples. The labels of these samples are assumed to be the same of their root. In unsupervised learning, each cluster is represented by an optimum-path tree rooted at a single prototype but we do not know the class label of the training samples. Therefore, we expect that each cluster contains only samples of a same class and some other information about the application is needed to complete classification. The basic idea is then to specify an adjacency relation and a path-value function, compute prototypes and reduce the problem into an optimum-path forest computation in the underlying graph. The training forest becomes a classifier which can assign to any new sample the label of its most strongly connected root.

Papa et al. (2009b) presented a first method for supervised classification using a complete graph (implicit representation) and the maximum arc weight along a path as connectivity function. The prototypes were chosen as samples that share an arc between distinct classes in a minimum spanning tree of the training set. This OPF classifier has been widely used in several applications. Another supervised learning method was proposed by Papa and Falcão (Papa and Falcão, 2008). In this case, the arcs connect *k*-nearest neighbors (*k*-nn) in the feature space. The distances between adjacent nodes are used to estimate a probability density value of each node and optimum paths are computed from the maxima of this probability density function. For large datasets, we usually use a smaller training set and a

much larger evaluation set to learn the most representative samples from the classification errors in the evaluation set. This considerably improves classification accuracy of new samples. This strategy was assessed with *k*-nn graphs by (Papa and Falcão, 2009). The accuracy results can be better than using similar strategy with complete graph for some situations, but the latter is still preferred because it is faster and does not require the optimization of the parameter *k*.

Although OPF be faster than SVM for training, it is often desirable to design classifiers that can handle online training in massive datasets. Hence, any effort devoted to such applications will be always welcome. Papa et al. (2009c) proposed an algorithm that prunes the training set by identifying irrelevant samples in a learning process and further removing them. This approach has demonstrated to be very efficient and effective in some applications, speeding up the OPF training and classification times. Section 2.1 presents the OPF pruning algorithm.

2.1 Pruning training samples

Large datasets usually present redundancy, so at least in theory it should be possible to estimate a reduced training set (Z_1) with the most relevant patterns for classification. The use of training and an evaluation set has allowed OPF to learn relevant samples for training set from the classification errors in the evaluating one, by swapping misclassified samples of Z_2 and non-prototype samples of Z_1 during a few iterations (Papa et al., 2009b). In this learning strategy, Z_1 remains with the same size and the classifier instance with the highest accuracy is selected to be tested in the unseen set Z_3.

Further, Papa et al. (2009c) proposed an algorithm that aimed to identify the most relevant samples from Z_1 by classifying Z_2 and marking each sample from the training set until its root in the optimum-path tree that classified some node from Z_2. This process is repeated until some convergence criterion is satisfied, and at the final the unmarked samples are removed from the training set. Therefore, this approach aims to learn the relevant samples from the training set and also to reduce its size. As the OPF computational complexity is proportional to the number of training samples, we can speed up it by decreasing the training set size. Obviously, the accuracy over the test set may be affected, but in some applications this phenomenon has not been observed (Papa et al., 2009c).

The OPF pruning algorithm can be summarized in Figure 1a-d. Figure 1a shows an optimum-path forest generated in the training phase, composed by two optimum-path trees (a blue and a red one) and two prototypes (bounded nodes). In Figure 1b, one can see the classification process from a sample *p* (yellow node) of the evaluating set: *p* is connected to all nodes of Z_1, and it is evaluated the node *t* from Z_1 that offered the optimum-path in Figure 1c. For sake of simplicity, some arcs were not represented in Figure 1b. Finally, in Figure 1d all training nodes that participate from the classification process of *p* are marked with the black color. At the final of the whole classification step, the unmarked training samples are discarded from the training set.

The pruning process is repeated until some criterion is satisfied. The user may define a number of iterations, or even so can set a lower bound to the absolute difference between the accuracy over the original evaluating set and the pruned one (this threshold is called *MLoss*). In this case, when this absolute difference is reached, the algorithm stops its execution. We use this implementation for the experiments in this paper. This is the first time that OPF pruning algorithm is used to assess its robustness regarding different training set sizes in the context of land use classification.

Although the reader may argue that the pruning step complexity can be increased with respect to simple training, this procedure can be used once, or only when it is required by

user. If the resolution of the images to be classified increase, the pruning algorithm may be applied in order to reduce the training set size again. Note that is expected a lower bound to the training set size, and the user needs take into account this point when it adjusts the parameters of pruning algorithm, aiming the best trade-off between good recognition rates in the classification set and speed for training patterns.

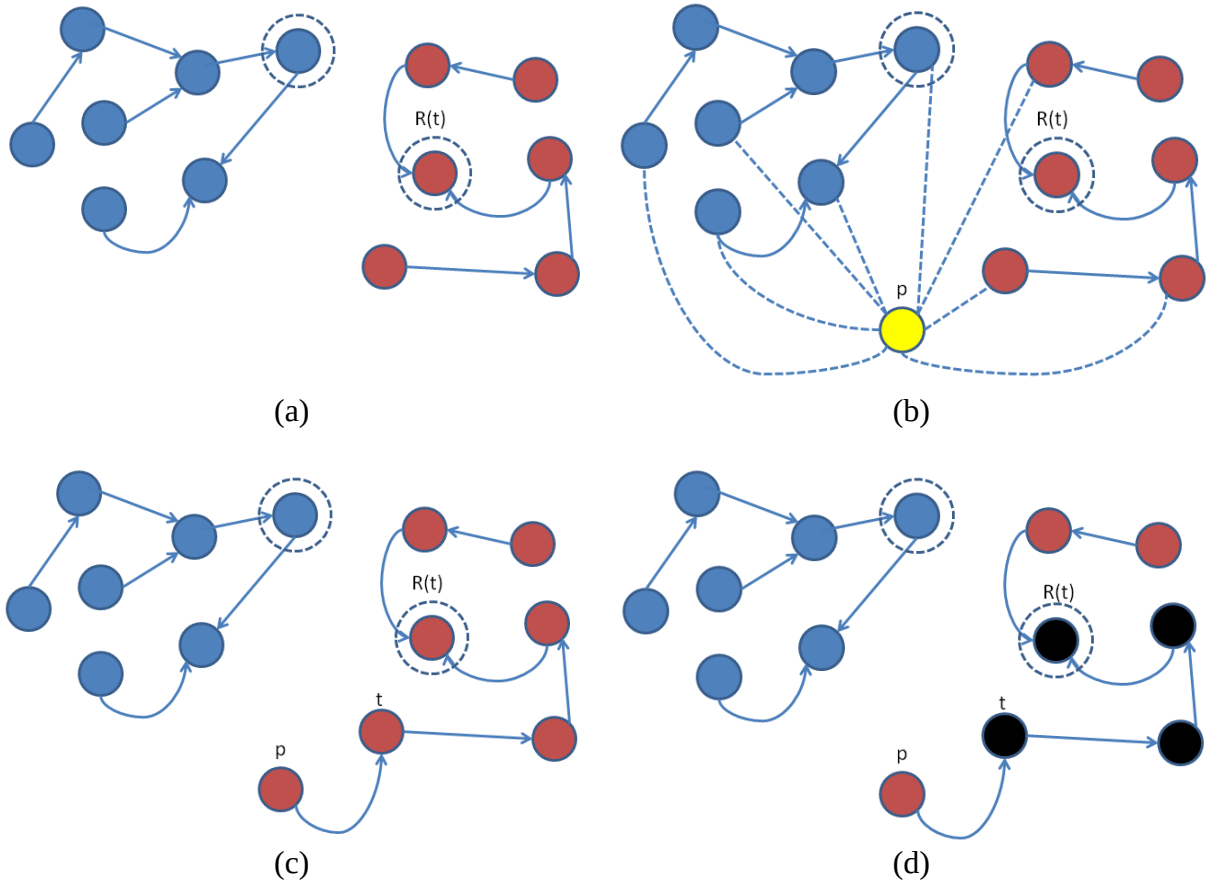


Figure 1. OPF training set pruning algorithm proposed by Papa et. Al (2009c): (a) optimum-path forest generated over Z_1, (b) classification of a sample *p* from Z_2, (c) *p* is conquered by *t* and receives the red label, (d) all nodes from the optimum-path used to conquer *p* are marked with the black label.

3. Experiments

We conducted the experiments in two phases: in the former we evaluated the Optimum-Path Forest accuracy over two images covering the area of Distrito do Lobo, Itatinga – SP, obtained from CBERS-2B and Landsat 5 satellites, and in the former we evaluated the pruning algorithm in order to speed up classification phase. Figure 2 shows the CBERS-2B and Landsat 5 images used in this work.

In the first round of experiments (Section 3.1) we used 70% of the image for training and the whole one for testing. In this case, we applied this procedure for CBERS-2B (Figure 2a) and Landsat 5 (Figure 2b) images. In the second round we used 70% for training, 10% for the evaluating set and the whole image for testing. The training and evaluating set are used to the learning process involved in the pruning algorithm (Section 3.2).

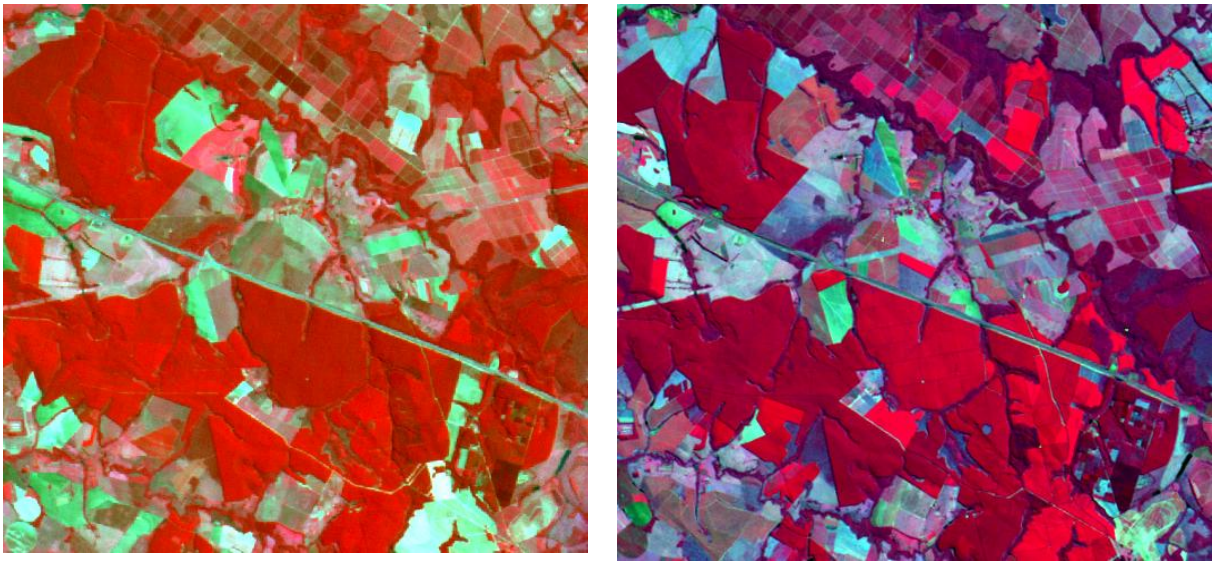


Figure 2. Satellite images used in the experiments covering the area of Distrito do Lobo, Itatinga – SP, Itapetininga-SP: (a) CBERS-2B CCD (B2G3R4 composition) and (b) Landsat 5 TM (B5G3R4 composition).

3.1 Land use classification through Optimum-Path Forest

In this section, we evaluate the accuracy of OPF considering the land use classification task. Since that Pisani et al. (2009) already demonstrated that OPF outperformed Support Vector Machines and Artificial Neural Networks with Multilayer Perceptrons to this task, we opted to not show here these experiments. To compose the feature vector for the pixel-based classification, we used 21 features described as follows:

- 3 color features composed by the gray levels in the Red, Green and Blue channels.
- 18 texture features obtained through the convolution between the original image and a Gabor filter in 6 different orientations ($\boldsymbol{\theta} = 0°, 45°, 90°, 135°, 225°$ and $315°$) and 3 spatial resolutions ($\boldsymbol{\lambda} =$ 2.5, 3 and 3.5). For each one of the $\boldsymbol{\lambda}$ values, we applied a different value for $\boldsymbol{\sigma}$, say that $\boldsymbol{\sigma} =$ 1.96, 1.40 and 1.68.

Equation 1 below describes the mathematical formulation of the Gabor filter applied in this paper:

$$\boldsymbol{f(x, y, \theta, \gamma, \sigma, \lambda, \psi) = e^{-\frac{x'^2 + y'^2 \delta^2}{2\sigma^2}} \cos\left(2\pi \frac{x'}{\lambda} + \psi\right)}, \quad (1)$$

where $\boldsymbol{x' = x\cos(\theta) + y\sin(\theta)}$ e $\boldsymbol{y' = x\sin(\theta) + y\cos(\theta)}$. In the above equation, λ means the sinusoidal factor, θ represents the orientation angle, ψ is the phase offset, σ is the Gaussian variance and γ is the aspect spatial ratio. Regarding the remaining variables, we used the following values: $\boldsymbol{\psi = 0}$ and $\boldsymbol{\gamma = 1}$. Tables 1 and 2 display the kappa results for different training set size percentages for CBERS-2B and Landsat 5 satellites, respectively. The training and classification times in seconds are also shown.

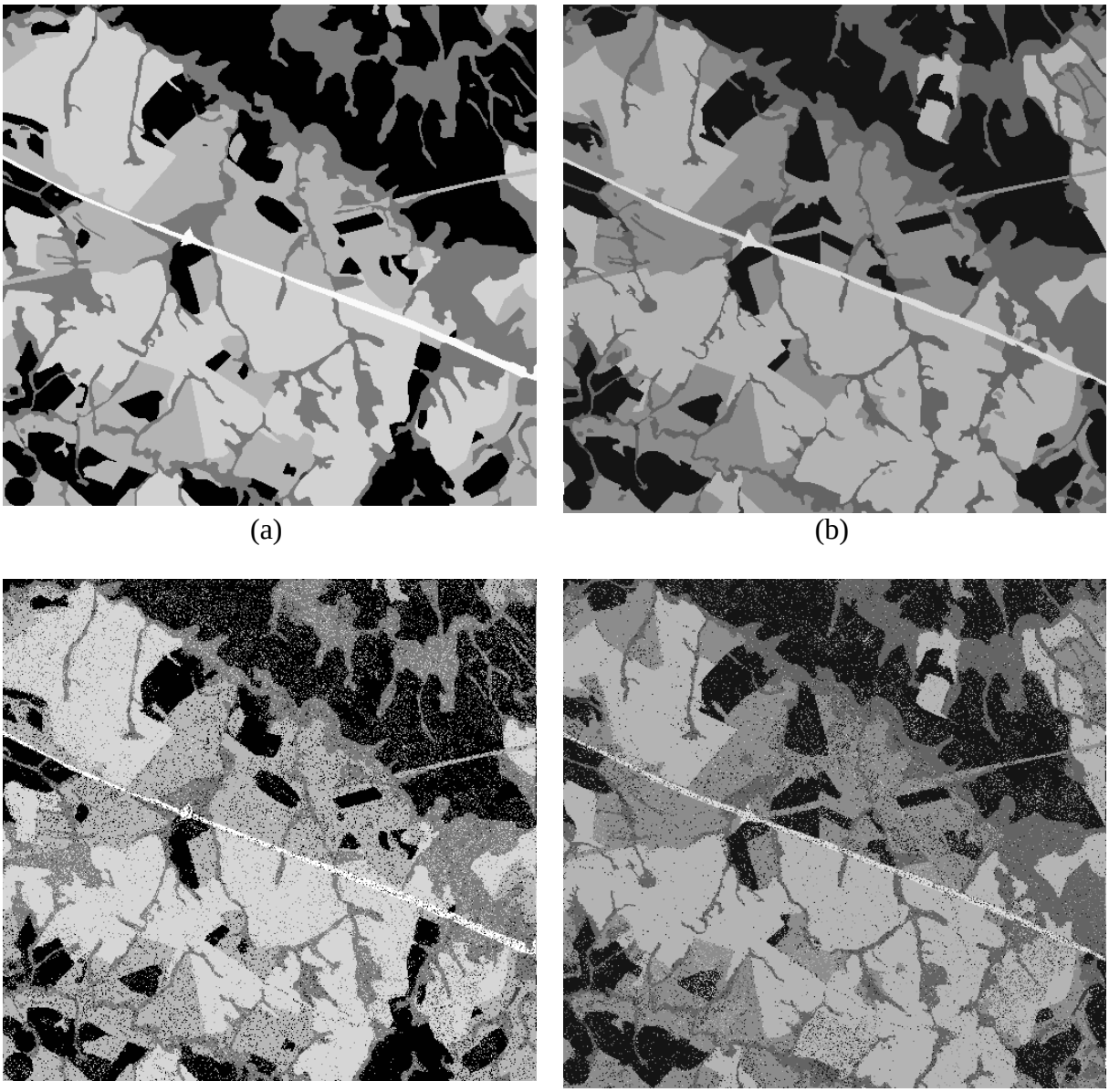
Table 1. Quantitative results for OPF classification in the CBERS-2B image (Figure 2a).

Training set percentage	Kappa	Training time [s]	Classification time [s]
30%	0.71	539.95	1680.77
40%	0.75	960.55	2301.72
50%	0.79	1525.25	2832.67
60%	0.83	2204.03	3303.75
70%	0.87	3005.87	3695.98

Table 2. Quantitative results for OPF classification in the Landsat image (Figure 2b).

Training set percentage	Kappa	Training time [s]	Classification time [s]
30%	0.77	561.73	1882.24
40%	0.81	1009.91	2611.51
50%	0.84	1586.87	3004.15
60%	0.87	2291.99	3537.68
70%	0.90	3123.68	4015.71

We can see that OPF can get good results, which are also displayed by the classified images in Figure 3. The ground truth images with respect to the ones displayed in Figure 2 were also shown. Note that land use classification is not a usual task, mainly because of the proximity of pixels from distinct classes, which share similar properties with different meaning. In the ground truth images (Figures 3b and 3d) there are six gray levels, which denote the following classes: cultures, dams, reforesting, grasslands, bushes and road.



(a) (b)

(c) (d)

Figure 3. Satellite images used in the experiments: (a) and (c) are the ground truth and classified images obtained by CBERS-2B, respectively, and (b) and (d) are the ground truth and classified images obtained by Landsat, respectively.

3.2 Pruning samples from training set

In this section we evaluate how much the images are affected after a training set size reduction. As aforementioned, we used 70% for training, 10% for the evaluating set and the whole image for testing. Table 3 displays the results.

Table 3. Quantitative results for OPF pruning.

Satellite	Kappa	Classification time [s]	Pruning rate
CBERS-2B	0.59	260.92	94.24%
Landsat 5	0.78	337.49	92.90%

As one can see, although the performance of OPF was quite degraded, mainly in CBERS-2B case, the classification time was speeded up about 14.16 and 11.89 times for CBERS-2B and Landsat 5, respectively. However, we believe that we can make the accuracy much better adopting other values of the threshold for pruning, since that here we used $MLoss = 0.06$. This parameter is used as the convergence criteria, as explained in Section 2.1. Let Acc_O be the OPF accuracy over the evaluating set after training in the original training set, i.e., without pruning. Let Acc_P be the OPF accuracy over the evaluating set after training in the actual pruned training set. Hence, $MLoss$ can be defined as $MLoss = |Acc_O - Acc_P|$. This means that the pruning algorithm will stop only when $MLoss > 0.06$. Therefore, our future works will be dedicated to find other values of $MLoss$ that may less degrade the accuracy after pruning

4. Conclusions

In this paper we shed light the importance of having efficiency and effectiveness in image classification tasks, mainly in applications that require millions of pixels to be recognized. This problem has been found in interactive systems of image segmentation and classification, in which user-friendly environment are often desired.

We conducted two rounds of experiments: in the first one we deal with the problem of land-use classification in images acquired by CBERS-2B CCD and Landsat 5 TM sensors. The experiments showed good recognition rates, which leaded us to classified images with interesting visual results. In the former round, we demonstrated that we can speed up the classification time, but the accuracy may be degraded up to a certain level, mainly because of some training samples pruning. In this case, we have obtained up to 95% of training set reduction. However, this pruning strategy may be hard in the sense of pruning all samples of the optimum-path that originated in a given sample that did not participate of the classification process over the evaluating set. Nowadays, we are looking to develop soft strategies for pruning, which may less affect the accuracy, but may be also slightly slower than the implementation used here, but still faster than not using pruning.

References

Bordes A., Ertekin S., Weston J., Bottou L. Fast Kernel Classifiers with Online and Active Learning. **Journal of Machine Learning Research**, v.6, p.1579-1619, 2005.

Collobert R., Bengio S. SVMTorch: Support Vector Machines for Large-Scale Regression Problems. **Journal of Machine Learning Research**, v.1, p.143-160, 2001.

Fan R. E., Chang K. W., Hsieh C. J., Wang X. R., Lin C. J. LIBLINEAR: A library for large linear classification. **Journal of Machine Learning Research**, v. 9, p.1871-1874, 2008.

Papa, J. P., Suzuki, C. T. N., Falcão, A. X., **LibOPF: Library For The Design Of Optimum-Path Forest Classifiers.** Campinas: IC/Unicamp. Software version 2.0. 2009. Disponível em: ttp://www.ic.unicamp.br/~afalcao/LibOPF.

Papa, J. P., Falcão, A. X., Suzuki, C. T. N., Supervised Pattern Classification Based on Optimum-Path Forest**. International Journal of Imaging Systems and Technology**, v.19, no. 2, p.120-131, 2009.

Papa, J. P., Falcão, A. X., Freitas, G. M., Ávila, A. M. H., Robust Pruning of Training Patterns for Optimum-Path Forest Classification Applied to Satellite-Based Rainfall Occurrence Estimation. **Geoscience and Remote Sensing Letters,** v.7, no. 2, p.396-400, 2009.

Papa, J. P., Falcão, A. X., A New Variant of the Optimum-Path Forest Classifier**. International Symposium on Visual Computing**, v.5358, no. 2, p.935-944, 2008.

Papa, J. P., Falcão, A. X., On the Training Patterns Pruning for Optimum-Path Forest Classifier. **International Conference on Image Analysis and Processing**, v.5716, no. 2, p.259-268, 2009.

Pisani, R. J., Papa, J. P., Zimback, C. R. L. Falcão, A. X., Barbosa, A. P. Land use Classification Using Optimum-Path Forest**. XIV Simpósio Brasileiro de Sensoriamento Remoto**, p.7063-7070, 2009.