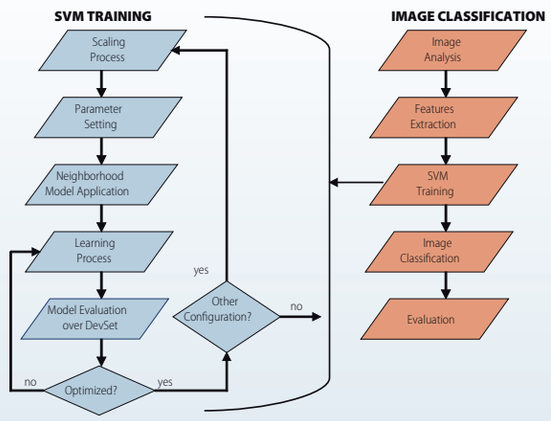
A COMPARATIVE ANALYSIS OF KERNEL-BASED METHODS FOR THE CLASSIFICATION OF LAND COVER MAPS IN SATELLITE IMAGERY

R. Basili, F. Del Frate, M. Luciani, F. Mesiano, F. Pacifici, R. Rossi
Earth Observation Laboratory and Artificial Intelligence Laboratory
Dept. of Computer Science, Systems and Production, Tor Vergata University, Rome, Italy

ABSTRACT

This paper studies the impact of several learning issues in an image classification task with SVMs, such as rich feature based representations, optimization and sensitivity to novelty in the test data sets. The employed imagery refers to the city of Rome, Italy and is acquired in different years and seasons by the European Remote Sensing Satellites ERS-1 and ERS-1/2 tandem mission. A comprehensive evaluation according to varying training conditions is reported, showing that SVMs provide robust and largely applicable tools.

THE CLASSIFICATION PROCESS



01 · DATASET

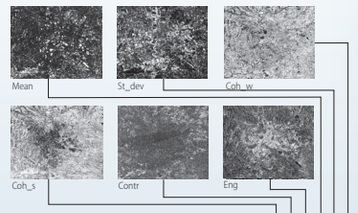
Acquisition Date	Satellite
January 25,31 1994	ERS 1
March 26,29 1994	ERS 1
July 13, 1994	ERS 1
February 24, 1996	ERS 1
February 25, 1996	ERS 2
March 30, 1996	ERS 1
March 31, 1996	ERS 2
July 14, 1996	ERS 2
February 13, 1999	ERS 1
February 14, 1999	ERS 2
March 20, 1999	ERS 1
March 21, 1999	ERS 2

Output classes			
1 Asphalt	608	1,338	730
2 Forest	1,955	6,128	4,173
3 High density	706	1,632	926
4 Isolated buildings	585	1,140	555
5 Low density	9,020	20,050	11,030
6 Vegetation	8,994	18,940	9,946
7 Water	1,066	2,340	1,274
Total ROIs	22,934	51,568	28,634
Training ROIs	8,182	21,382	13,200
Validation ROIs	14,752	14,752	15,434

02 · FEATURE DESCRIPTION

- Pieces of information embedded in both the amplitude and phase of the radar return have to be exploited.
- The classification scheme exploits six different inputs parameters such as average and standard deviations of backscattering intensity, coherence and textural features
- Two features related to the backscattering mechanisms, Contrast and Energy ,and two features related to the phase information.
- The information contained in the backscattering has been exploited by considering the time-average and the standard deviation values.
- To increase the separability among classes, the features were computed over moving square boxes of 7x7 pixels, and considering a 16 grey levels quantization.
- The phase information has been included into two degree of interferometric coherence values, one calculated over two winter tandem acquisitions, the other one over two late spring tandem acquisitions.

List of Input Parameters:
- Mean Int,
- Int. St. Dev.
- Winter Coh,
- Spring Coh,
- Contrast,
- Energy

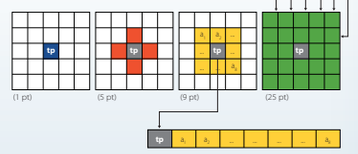


03 · OPTIMIZATION

- This paper studies the impact of several learning The optimization process of the SVMs has been carried out by setting five different parameters: the kernel type, the value for data scaling, the parameters C and γ employed in the SVM learning and the adopted vector representation of the target pixel.
- Parameter setting is computed over a randomly chosen 20% of the training data points named development set.
- The optimal configuration that achieves the best accuracy in the development set is obtained through iterative and it is applied to every representation chosen and every kernel function.
- The optimal configuration of SVM is then retrained over the entire training set.
- Radial kernel increases the accuracy of the default SVM configuration by 40% and it is two times faster than the polynomial case.

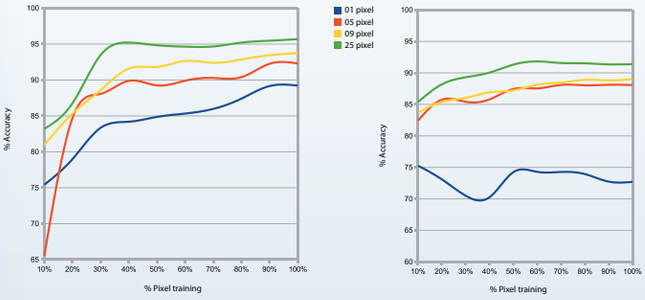
04 · NEIGHBORHOOD MODEL

- Neighborhood model: inclusion in the vector representing a target pixel the (surface) information characterizing the surrounding pixels (i.e. the tp neighbors).
- (5pt) includes the 4 points adjacent to the pixel along the two major dimensions.
- (9pt) and (25pt) respectively, represent the full square of 9 and 25 pixels centered on the tp.
- The individual dimensions increase the expressiveness of the resulting kernel function. Neighborhood model doesn't changes the amount of training information.
- Significantly higher accuracies are observed and no relevant boost in the computational load is observed.



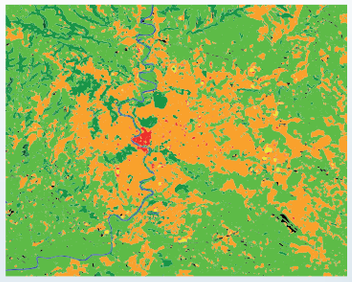
05 · LEARNING CURVES

- The learning curves of classifiers have been derived by employing training subsets of increasing size during the SVM learning.
- Accuracy over all subsets is measured, from a randomly chosen 10% to the full training set.
- All the SVMs reach significant levels of accuracy after a limited set of examples is made available.
- The best model (i.e. the 25 point adjacency grid), has the higher learning rate and it reached very high accuracy values when only the 25% of the total training set is considered.
- More expressive representations have a strong impact on the applicability of SVMs: they require much less manually tagged material than simpler feature vectors.



06 · CLASSIFICATION MAP

The classification exercise has been carried out for ERS data set acquired in 1994, 1996 and 1999, thus obtaining three different land cover maps. A visual inspection suggests that the main large built area has been identified with good accuracy, as well as some specific structures such as the parks inside the city, the Tiber river, the compact old section of the city and the Ciampino airport. In spite of the decametric resolution of the SAR acquisition, several features, even of relatively small dimensions like the trees along the river and the squares with lawns and plants, have been captured.



07 · CONCLUSIONS

An extensive evaluation of SVM-based land cover mapping has been carried out over multitemporal SAR imagery acquired over the Rome area. Different feature representations have been suggested that make the resulting kernels more accurate and robust with respect to the amount of novelty and lacks in the test data. The proposed neighborhood representation offers a simple model of the image textural properties around the target pixel, by capitalizing relevant information local to each classification case. They nicely exploit the high flexibility offered by the SVM paradigm, with practically no impact on the parametrization and learning complexity.

The tests run over the SVMs classifiers suggest that a significant reduction in the learning requirements can be achieved: significant performances are obtained even on small amounts of training data (about 35% is sufficient to closely approximate the upper bound). This confirms the high applicability of the SVM methods in real applications characterized by a strong heterogeneity in the sources, by a significant novelty in the target data and corresponding poorer training conditions. The work here described enabled a specific study on the combination of the neural and SVM classifiers over the same data sets. It will be the near-future target of this research.

Comparison across differentt neighborhood												
Image year	1994				1996				1999			
Adjacent grid	1p	5p	9p	25p	1p	5p	9p	25p	1p	5p	9p	25p
Max value	1	0.5	0.5	0.1	0.01	0.5	0.5	0.1	1	5	5	0.1
C	5	50	10	5	500	50	100	1	5	1	1	10
Gamma	50	10	10	50	50	0.1	0.1	10	10	0.1	0.1	50
Accuracy %	89.22	92.30	93.77	95.72	72.55	77.13	78.94	89.81	72.70	88.07	89.01	93.12