In *Proceedings of International Geoscience and Remote Sensing Symposium, July 1998, Seattle, WA, USA*

Feature Selection for ERS-1/2 InSAR Classification: High Dimensionality Case

Reinhold Huber$^{(1)}$ and Luciano V. Dutra$^{(2)}$
(1) Aero–Sensing Radarsysteme GmbH
c/o DLR Oberpfaffenhofen, 82234 Weßling, Germany
Fax: +49–8153–281543 / E–mail: reinhold.huber@dlr.de

(2) Instituto Nacional de Pesquisas Espaciais
12227.010 São José dos Campos, Brazil
Phone: +55–12–345–6480 / Fax: +55–12–345–6468 / E–mail: dutra@dpi.inpe.br

A systematic way of selection and assessment of the performance of a large number of texture features extracted from spaceborne interferometric SAR data and classified with different types of classifiers is presented. Multi-seasonal ERS–1 and ERS–2 SAR data of the Czech Republic is used to classify into four different land–cover classes. A multistage search method in the space of all possible feature subsets taken from local statistics, fractal analysis and co–occurrence matrices is proposed and tested. In the early stages of the method, features are ranked according to its discriminatory power measured by a ranking coefficient based on subset performance measured by Jeffreis–Matusita–distance. Best ranked features are chosen and a new set is formed and evaluated using the hold–out method employing maximum–likelihood, nearest neighbor and multilayer perceptron classifiers.

INTRODUCTION

Feature selection for SAR and InSAR data, i.e. the selection of a subset of features providing the most discriminative power out of the numerous possible features, is addressed in this paper. We formulate the problem of feature selection as finding a mapping from the initial, possibly very high–dimensional, feature space of dimension D to a m–dimensional subspace, where $m \leq D$. Considering all possible subsets of an initial feature space of dimension D requires 2^D evaluations of feature subsets. This can be done more efficiently using search techniques [1]. Firstly, such methods do not guarantee to find the optimal solution, especially with the given InSAR data set, so we are left with evaluation of the performance measure for all possible subsets. Secondly, we are interested on classifier performance on new data. The hold–out method using a data set which is independent from that used for training accomplishes this [2].

The study area is located in the Czech Republic near the city of Olomouc, used data sets are ERS–1/2 tandem data from December 1995 and ERS–2 data from July 1996. The data has been geocoded to grid size of 25m and covers 30×20 km. The four basic land–cover classes to be identified from the SAR and InSAR data are water, forest, built-up area and open area.

EXTRACTION OF FEATURES

We extracted 36 different features, three of which we call primary features. The primary features are f_1 the mean backscatter of the ERS–1/2 tandem mission from December 1995, f_2 the backscatter of the ERS-2 pass in July 1996 and the interferometric coherence f_3 taken again from the winter tandem pair. The extracted secondary features are measures of fractalness of the digital elevation model (DEM), features of first order local statistics and co–occurrence matrix features of f_2 and f_3. All three single–look SAR images are multi–look processed and scaled to a dynamic range of 8 Bit and a speckle filtering algorithm was applied [3].

Grey-level co-occurrence matrices GLCM are derived from pair-wise pixel intensity statistics [4]. Each entry g_{ij} of a GLCM is derived from the grey-level image as the expectation for the probability for two pixels having grey-value i and j and being d pixels separated in angle direction α. We use a single GLCM for α of $0, 45, 90$ and 135 degrees and $d = 1$ and extract the following features from this undirected GLCM (see [5] for an exhaustive list and definition of GLCM features):

f_4 Energy	f_9 Correlation
f_5 Entropy	f_{10} Cluster shade
f_6 Maximum Probability	f_{11} Cluster prominence
f_7 Contrast	f_{12} Information correlation I
f_8 Homogeneity	f_{13} Information correlation II

The features f_{14} to f_{23} are derived from f_3 as f_4 to f_{13} are extracted from f_2.

Features derived from local statistics describe the textural appearance of the surrounding of a pixel by calculating first–order statistical parameters in a small estimation window centered at the pixel p_{xy}. For $p_{ij} \in f_2$ we extract the following features (see [6] for definition):

f_{24} Local mean	f_{27} Kurtosis
f_{25} Coefficient of variation	f_{28} Contrast
f_{26} Skewness	f_{29} Homogeneity

Table 1: Multi-stage feature selection by JMD and hold–out performance ranking.

stage	group	selected feature and rank									
1	F_C	f_7 0.77	f_8 0.71	f_{18} 0.71	f_{20} 0.74	f_{23} 0.75					
	F_L	f_{24} 0.99	f_{25} 0.68	f_{30} 0.85	f_{31} 0.63						
2	F_A	f_1 0.74	f_2 0.54	f_7 0.75	f_8 0.85	f_{18} 0.56	f_{23} 0.54	f_{24} 0.57	f_{25} 0.63	f_{30} 0.90	f_{31} 0.59
3	MLP	f_1 0.70	f_7 0.91	f_{23} 0.62	f_{25} 0.67	f_{30} 0.78					
	ML	f_1 0.68	f_7 0.92	f_{18} 0.63	f_{25} 0.70	f_{30} 0.77					
	NN	f_1 0.63	f_2 0.62	f_7 0.88	f_{18} 0.62	f_{24} 0.73	f_{30} 0.87				

The features f_{30} to f_{35} for coherence data are derived in the same fashion by setting $p_{ij} \in f_3$.

To describe the roughness of the observed land surface we employ the idea of fractal dimension. For local estimation of the Hurst parameter H the method presented in [7] is employed. The relationship between the fractal dimension D and the Hurst parameter H is defined as: $f_{36} = D = 3 - H$

SELECTION OF FEATURES

A widely used measure to measure the separability between two distributions is the Jeffreis–Matusita–distance (JMD) [8]. For multivariate Gaussian distributions the average JMD for N classes is given by:

$$JMD = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{i-1} J_{ik}, \qquad J_{ik} = 2(1 - e^{-B_{ik}}),$$

in which B_{ik} (Battacharyya–distance) is given by

$$B_{ik} = \frac{1}{8}(m_i - m_k)^t \Sigma^{-1}(m_i - m_k) + \frac{1}{2} \ln \left[\frac{|\sum|}{|\Sigma_i|^{\frac{1}{2}}\, |\Sigma_k|^{\frac{1}{2}}} \right]$$

$$\text{and} \qquad \Sigma = \frac{\sum_i + \sum_k}{2}.$$

One possible selection by the JMD criterion is selection of a m elements sub–set providing the largest average distance between pairs of classes. For any possible m normally all D features will be selected as having the highest average JMD, a property known as monotonicity. In the case that the best subset in terms of overall accuracy does not obey a monotonicity property suboptimal solutions based on stepwise selection or rejection of individual features depending on increase or decrease of JMD are employed [1].

We propose the use of the accumulated histogram of features present in sub-sets ranked from higher to lowest JMD as an indicator of discriminating power for features. Therefore, we define a ranking function $\text{rank}_i \in [0, \ldots, 1]$ for each feature i by

$$\text{rank}_i = \frac{1}{M/2} \sum_{j=1}^{M/2} \frac{h_{ij}}{j} \qquad \text{where} \quad M = 2^D,$$

$$h_{ij} = \sum_{k=1}^{j} f_{ik} \quad f_{ik} = \begin{cases} 1 & \text{if feature i in ranked subset k,} \\ 0 & \text{else.} \end{cases}$$

Most features used in this study are not Gaussian, but still JMD can be used as a tool based on second order properties of data and its ease of use. Despite of its computational efficiency, it's is practically impossible to apply exhaustive JMD computation to a feature space of $D = 36$. Therefore, a multistage selection scheme is proposed:

1. Select best feature subset F_C from co-occurence features $\{f_4, \ldots, f_{23}\}$ and F_L from local statistics features $\{f_{24}, \ldots, f_{35}\}$ by JMD ranking.
2. Select best feature subset F_A from $F_C \cup F_L \cup \{f_1, f_2, f_3, f_{36}\}$ by JMD ranking.
3. Select best feature subset from F_A by hold–out method.

RESULTS

Table 2: Classifier performance of MLP, ML and 1–NN.

Classifier	Overall accuracy	Kappa	Tau
MLP	0.84	0.66	0.78
ML	0.82	0.63	0.76
1–NN	0.80	0.60	0.73

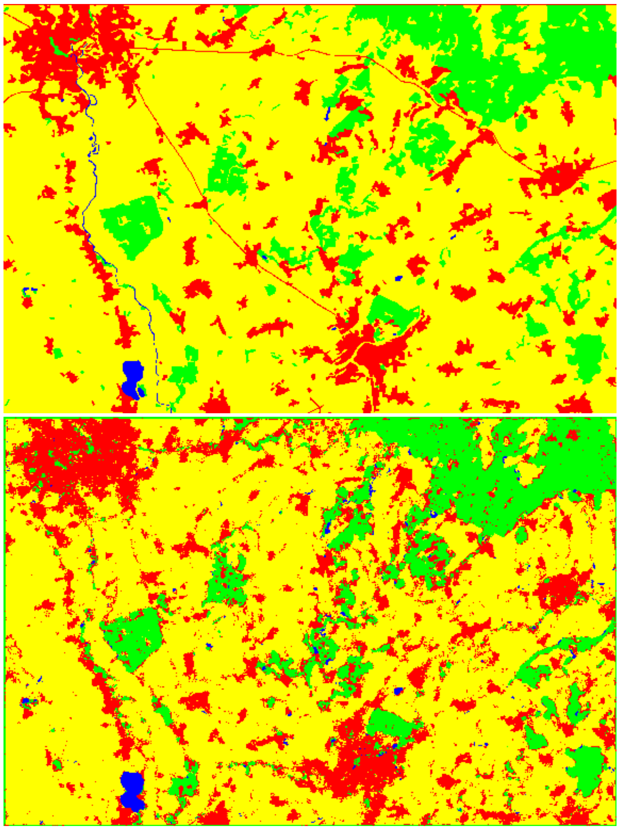

Figure 1: Ground–truth map and MLP classification

Table 1 summarizes the selected features for each group of features and stage together with the features rank. In the first stage we selected five features from F_C and four features from F_L. The input to the second stage consists of 13 features, 10 of which are selected as input to the third stage. The third stage takes the selected features and evaluates the performance of subsets on a specific classifier by the hold–out method. Results for the hold–out method using a one hidden layer multilayer perceptron (MLP) classifier [2] trained for 2000 epochs by resilient backpropagation (RPROP) [9] on a training set containing 2000 examples with equal class frequency are given. Results are also shown for a multivariate Gaussian maximum–likelihood (ML) [8] and a non–parametric nearest neighbor (1–NN) [8] classifier based on the same training set. The ranking function measures ranked overall accuracy on the validation set, see Table 1, stage 3. Validation set size is 2000 examples and equal class frequency. The result of the MLP classification is shown together with the ground–truth in Fig. 1, accuracy measures are given in Table 2.

DISCUSSION

The results of the proposed selection scheme showed that a small number of features is sufficient for the classification task, resulting in a speed–up in feature extraction and classification and that the selected features are different for different classifiers. Normally, the original backscatter data, one or two texture features and coherence are contained in the classifiers best features subset. The achieved classification accuracy might be further increased by contextual methods which are able to correct misclassifications by incorporation of prior class probabilities for specific spatial neighborhoods [10].

ACKOWLEDGMENTS

L.V. Dutra is grateful to Dr. João Moreira of Aero–Sensing and CNPq that partially funded this research under Grant 680.061/94-0 (GeoTec), sub-grant 480.115/97-4.

REFERENCES

[1] Wojciech Siedlecki and Jack Sklansky. On Automatic Feature Selection. *Int. J. of Pattern Recognition and Artificial Intelligence*, 2(2):197–220, 1988.

[2] Christopher M. Bishop. *Neural Networks for Pattern Recognition*. Clarendon Press, 1995.

[3] Jong-Sen Lee. Refined Filtering of Image Noise using Local Statistics. *Computer Graphics and Image Processing*, 15:380–389, 1981.

[4] Robert M. Haralick, K. Shanmugam, and Its'hak Dinstein. Textural Features for Image Classification. *IEEE Trans. on Systems, Man and Cybernetics*, 3(6):610–621, November 1973.

[5] D. DeKruger and B.R. Hunt. Image Processing and Neural Networks for Recognition of Cartographic Area Features. *Pattern Recognition*, 27(4):461–483, 1994.

[6] Anne H. Schistad Solberg and Anil K. Jain. Texture Fusion and Feature Selection Applied to SAR Imagery. *IEEE Trans. on Geoscience and Remote Sensing*, 35(2):475–479, March 1997.

[7] Maurizio Migliaccio and Daniele Riccio. Fractal Brownian Model for SAR Image Analysis: Edge Detection and Classification Issues. In *Proceedings of IGARSS*, pages 1605–1607, Tokyo, Japan, August 1993.

[8] Keinosuke Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, 1990.

[9] Martin Riedmiller and Heinrich Braun. A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm. In *Proceedings of ICNN*, pages 586–591, San Francisco, CA, USA, 1993.

[10] Luciano V. Dutra and Reinhold Huber. Feature Extraction and Selection for ERS–1/2 InSAR Classification. unpublished.