Polarimetric SAR Image Classification on Urban Area using a Subset Selection Method

Maryam Soheili Majd1, Elisabeth Simonetto1 and Laurent Polidori1

GeF/L2G, ESGT, le Mans, France; elisabeth.simonetto(at)cnam.fr

Abstract. In this study, we consider the capability of a single look high-resolution PolSAR image for discriminating different surfaces in urban area. First, a basic framework is set up to extract different polarimetric descriptors from Sinclair matrix and coherency matrices, coherent and incoherent decomposition descriptors, and some discriminators. Then, we use a subset selection method (SSM) based on GA-MI (Genetic Algorithm – Mutual Information) and mRMR (minimum Redundancy Maximum Relevance) evaluation function to obtain an optimal subset of descriptors as input of the classification process ([1],[2],[3],[4]). To show the efficiency of this approach, we perform a comparison between the results of SVM classification obtained from the optimal feature subset and from different target decompositions based on different scattering mechanism assumptions, including the Pauli [5], Huynen [6], Cloude [7], Holm - Barnes [8], H/A/Alpha [7], Freeman [9,10], VanZyl [11], Krogager [12], Yamaguchi [13], or Touzi [14] methods. Experimental results on the optimal subset of SSM based on GA-MI show that the proposed method provides a meaningful selection in regards of the different scattering mechanisms and leads to better classification performance. Our experiments are performed on a single look PolSAR image acquired by the airborne RAMSES SAR sensor of ONERA over a suburban area.

Keywords. polarimetry, SAR, high-resolution, classification, genetic algorithm, mutual information, feature selection.

1. Introduction

The accuracy of a supervised classification in the urban context mainly depends on the SAR data characteristics such as spatial resolution, frequency, number of looks, look angle, polarization mode, etc. It is also related to the type of used polarimetric descriptors, the classification method, and its implementation. Moreover, an accurate result is obtained when integrating adequate ground knowledge.

In this paper, we focus on the minimum Redundancy and Maximum Relevance (mRMR) which are derived from MI to qualify all possible subsets of features derived from the radar data ([3],[4]). Then, we use GA in order to select the optimal subset of features. We present this method having in mind to find an optimal solution through different descriptors for considering the ability of PolSAR image for the classification of urban surfaces.

Further details of the proposed method and relative results are described in the next sections.
2. Methods

2.1. Data preparation:

In our work, we use one single full polarimetry X-band image of Toulouse which was acquired in 2006 by the RAMSES airborne sensor of ONERA, with 60° look angle and 35 cm pixel size (figure 4 (a)). The spatial resolution is around 50 cm. Our classes consist in highways, lawns, trees, flat roofs and sloped roofs. We also consider two classes that result from radar imagery effects: shadows and bright pixels. The training samples and the control samples are extracted based on the observation of an optical aerial photo and the PoLSAR image.

2.2. Feature extraction:

We consider several attributes extracted from PoLSAR data in two ways: features directly derived from the scattering matrix or using decomposition methods and some discriminators [15]. The complete list of features is presented in the Table 1. The subscripts used in this table represent the related element in the corresponding category. For example, \((T_{13-\text{Phase}})\) refers to the phase parameter in the first row and third column of coherency matrix, or \([\text{Barnes I}]_{\text{phase}}\), is referring to the intensity (in dB) and phase of the distinct elements in the Barnes-I coherency matrix and so on.

Table 1. Considered polarimetric descriptors. 

<table>
<thead>
<tr>
<th>Descriptor extraction technique</th>
<th>Features</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decomposition method</td>
<td>[\text{Haynen}<em>{T</em>{\text{phase}}} ], [\text{Barnes II}<em>{T</em>{\text{phase}}} ], [\text{Holm II}<em>{T</em>{\text{phase}}} ], [\text{Cloude}<em>{T</em>{\text{phase}}} ], [H/A/\alpha]<em>{T</em>{\text{phase}}} ], [\text{Freeman I}<em>{s,d,v} ], [\text{Freeman II}</em>{v,G} ], [\text{Yamaguchi VI}<em>{s,d,v,h} ], [\text{Krogager}</em>{s,d,h} ], [\text{Yamaguchi III}<em>{s,d,v} ], [\text{VanZyl}</em>{s,d,v} ], [\text{Pauli}_{T} ]</td>
<td>63</td>
</tr>
<tr>
<td>Original polarimetric descriptors</td>
<td>[{H/A/\alpha, H(1-A), H(1-H)A, HA, (1-H)(1-A), \beta, \gamma, \delta} ]</td>
<td>10</td>
</tr>
<tr>
<td>Discriminators</td>
<td>{\lambda_{i}, \psi_{i}, T_{i}, \alpha_{i}, \phi_{i}}_{i=1,3}</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>({S_{\text{inv}}^{\ast}S_{\text{inv}}^{T}} / \left{S_{\text{inv}}^{T}S_{\text{inv}}\right}) (x, y, x', y' = (H,V))</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>(\sum_{A}S_{A,11}S_{A,22}^{2})</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>(\text{DR} = {S_{\text{inv}} - S_{\text{inv}}}, {S_{\text{inv}} - S_{\text{inv}}}, {S_{\text{inv}} - S_{\text{inv}}}, {S_{\text{inv}} - S_{\text{inv}}}, {S_{\text{inv}} - S_{\text{inv}}}, {S_{\text{inv}} - S_{\text{inv}}})</td>
<td>6</td>
</tr>
<tr>
<td>Total number of descriptors</td>
<td>127</td>
<td></td>
</tr>
</tbody>
</table>

Indeed, as the incoherent decomposition method uses averaging of the scattering signal and the coherent decomposition method uses the full image resolution, we think that these two methods can complete each other to describe the behavior of different urban surfaces in the case of
very high spatial resolution. These features have been computed using POLSARpro [17] and Matlab codes.

2.3. Mutual Information:

MI is a measurement of variable dependency in the case of nonlinear correlation, without assuming their distribution. Generally, MI describes quantitatively the relationship between two features or between a feature and a class. The notation for the description of MI based on Shannon’s information theory [2] is taken from [3]. The entropy measurements, $H$, of a random variable, $X$, with the probability density function (pdf), $p_x$, and joint entropy of $X$ and $Y$ with joint pdf, $p_{x,y}(x, y)$, are defined by:

$$H(X) = -\sum_i p_i(x_i) \log p_i(x_i), \quad H(X, Y) = -\sum_i \sum_j p_{i,j}(x_i, y_j) \log p_{i,j}(x_i, y_j)$$

(1)

MI is a measurement of mutual dependency of two variables based on the entropy and is defined by:

$$MI(X; Y) = H(X) + H(Y) - H(X, Y) = \sum_i \sum_j p_{i,j}(x_i, y_j) \log \left( \frac{p_{i,j}(x_i, y_j)}{p_i(x_i) p_j(y_j)} \right)$$

(2)

The main goal of the feature subset selection is to achieve the maximum relevance between a selected feature and a class, noted $D(S_m, C)$, and the minimum redundancy between selected features, noted $R(S_m)$. According to Eq. (2), the Relevance, Redundancy and mRMR evaluation function are defined by:

$$D(S_m, C) = \frac{1}{|S_m|} \sum_{x_i \in S_m} MI(X_i; C), \quad |S_m| = m$$

$$R(S_m) = \frac{1}{|S_m|} \sum_{x_i \in S_m} MI(X_i; X_j)$$

$$\Rightarrow \hat{S}_m = \arg \max_{S_m} [U(S_m) = D(S_m) - R(S_m)]$$

(3)

Where $X_i$ is the feature variable and $C$ is the class variable. $S_m = \{x_1, ..., x_m\}$ is a feature set variable, and $C = \{c_1, c_2, ..., c_k\}$ is a class label set with $k$ classes. $\Omega_{S_m}$ is the domain of variable $S_m$.

2.4. Genetic Algorithm:

GA is one of widely common iterative searching algorithm that aims at optimizing an evaluation function. It is inspired from natural evolution derived from the Darwinian theory of survival of the fittest individuals and was introduced by John Holland in the early 1970s. A Genetic feature subset selection process consists in four basic steps: a population generation (individuals), an evaluation of each individual according to an evaluation function and a selection process, termination (it may be based on iteration number, feature number, threshold, or population convergence), and a result validation to verify the validity of the selected solution [1]. Our proposed SSM method based on GA follows this classical scheme. One individual is a subset of a given number of features. Here, the possible maximal number of features is 127. The evaluation function is mRMR computed from MI theory. Besides, the GA implementation has the following particularities:

Parameter initialization: five parameters are initialized: the population size or number of individuals in one generation, the maximal number of generations (iteration), the pressure factor;
which defines the strength of selection method and influence the explored domain in the search space, the size of the feature subset, and the evaluation threshold.

**Evaluation**: the evaluation function is based on mRMR [3], Eq. (3). At each stage, the algorithm sorts the individuals according to their evaluation values.

**Selection**: the best individuals are selected for cross-over if their evaluation measurement is above the evaluation threshold.

**Crossover**: individuals are randomly introduced to new generation through crossover probability operator. This one is defined by an asymmetric distribution [3]:

\[
f(r,a,i) = \text{round}\left[(i-1)\frac{e^{ar} - 1}{e^a - 1}\right], \quad i \in [1, N]
\]

where, \( r \in [0,1] \) is a random variable with uniform distribution, \( N \) is the number of individuals, \( i \) is the individual index, and \( a \) is a positive constant which is known as pressure factor.

The algorithm is implemented in Matlab and adapted from the source codes provided by the authors in [3].

### 3. Results

#### 3.1. GA parameter initialization:

To get the optimal number of generation (iteration), we consider the relationship between the iteration number and the maximal evaluation function value obtained for different sizes of feature subset (3, 4, 5, …, up to 20). Figure 1 (a) shows that after several iterations for any set size, the evaluation function value does not vary anymore. To select the number of descriptors (that is the size of the researched subset), the classification accuracies (overall accuracy) for different set sizes (3, 4, 5, …, up to 20) are estimated. Figure 1 (b) shows that the optimal feature number could be five, leading to an overall accuracy of around 83.9% and iteration number of 16.

![Figure 1: (a) Evaluation function value of the optimal subset consisting in 3 to 20 descriptors, obtained by the SSM from 127 descriptors, vs iteration number. (b) Overall accuracy from SVM classification vs number of descriptors per subset.](image)

#### 3.2. GA-MI selection result:

The result of the selection for subset consisting in 3 to 6 features is shown in Table 2. According to the previous remarks, the optimal subset may refer to the subset with 5 features. In addition, the © EARSeL and University of Warsaw, 2014, ISBN 978-83-63245-65-8, DOI: 10.12760/03-2014-26, Zagajewski B., Kyckö M., Reuter R. (eds.)
Figure 2 shows the MI values between each class and each descriptor retained in the optimal subset.

**Table 2.** List of the selected descriptors by the GA-MI SSM for subset consisting in 3, 4, 5 or 6 features through 127 features.

<table>
<thead>
<tr>
<th>#</th>
<th>Descriptor</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>[Barnes I]-T_{12}-phase, Polarization-Fraction (PF) , Span</td>
</tr>
<tr>
<td>4</td>
<td>Shannon Entropy-I (SE), Lueneburg Anisotropy (LA), T_{33,I}, [Cloude]-T_{12}-phase</td>
</tr>
<tr>
<td>5</td>
<td>Shannon Entropy-I (SE), H(1-A), T_{33,I}, T_{12}-phase, [H/A/α]-T_{11-I}</td>
</tr>
<tr>
<td>6</td>
<td>Span, T_{33,I}, A_{12}, T_{11-I}, [Holm I]-T_{23}-phase, [Barnes I]-T_{12}-phase</td>
</tr>
</tbody>
</table>

**Figure 2:** Mutual information values between class and descriptor from the optimal selected subset of 5 features selected by GA-MI SSM through 127 descriptors.

**Figure 3:** Overall Accuracy (%) of SVM classification over different extraction descriptors methods. The Overall accuracy (%) of the SVM classification is presented in Figure 3. The results indicate that the optimal subset with 5 features has the highest overall accuracy (83.9%) compared to the used of classical target decomposition methods. The SVM classification result from the optimal subset is shown in Figure 4 (b).
Figure 4: (a) RAMSES SAR image represented in Pauli decomposition RGB color. (b) SVM classification result using the optimal 5-descriptor subset selected by GA-MI SSM from 127 descriptors.

4. Conclusions

Experimental results on a polarimetric data of the GA-MI SSM indicate that the proposed method seems to lead to a good feature selection as it shows good classification performance. Thus, such radar image seems interesting for urban surface mapping. For our data and on this area, the result accuracy is close to the one obtained with the Yamaguchi polarimetric decomposition into 3 components.

The proposed GA-MI SSM has several advantages. Firstly, the use of polarimetric radar images for mapping can appear discouraging for a non-expert because many polarimetric decompositions and attributes exist and one can have difficulties to decide which one could be the more appropriate. Our SSM brings one solution to this.

Secondly, our method is easy to implement and is fast. Thirdly, MI has capability to handle non-linear correlation between feature and class as well as anti-correlation of a pair of features. This criterion is independent from the data values and the shape of distributions. Finally, this method can be applied to other kinds of land cover and sensors as a future work.

References


