Input Textural Feature Selection By Mutual Information For Multispectral Image Classification

Mounir Ait kerroum, Ahmed Hammouch and Driss Aboutajdine

Abstract—Texture information plays increasingly an important role in remotely sensed imagery classification and many pattern recognition applications. However, the selection of relevant textural features to improve this classification accuracy is not a straightforward task. This work investigates the effectiveness of two Mutual Information Feature Selector (MIFS) algorithms to select salient textural features that contain highly discriminatory information for multispectral imagery classification. The input candidate features are extracted from a SPOT High Resolution Visible (HRV) image using Wavelet Transform (WT) at levels \((l = 1.2)\). The experimental results show that the selected textural features according to MIFS algorithms make the largest contribution to improve the classification accuracy than classical approaches such as Principal Components Analysis (PCA) and Linear Discriminant Analysis (LDA).

Keywords—Feature Selection, Texture, Mutual Information, Wavelet Transform, SVM classification, SPOT Imagery.

I. INTRODUCTION

ONE of the well-known problems in remotely sensed imagery classification and many pattern recognition applications is the important number of textural features extracted form multispectral bands. This obstacle makes the processes of learning and recognition more difficult. Therefore, it is necessary to reduce the dimensionality of input feature space. The most commonly used techniques for dimensionality reduction are the PCA [8] [10] and the LDA [14] [16]. However, their components are not necessarily the best for such classification [3] [13] [14].

Recently, several works of research [3] [4] [5] [6] [7] [9] [15] have successfully adopted Shannon’s mutual information [1] [2] in many feature selection problems by maximizing scheme between input candidate features and the output class labels. This concept was formalized by Battiti [3] to solve its FRnk problem (i.e., Feature Reduction from \(n \) to \( k \)) as follows:

Given an initial set \( F \) with \( n \) features and the output class \( C \), find the subset \( S \subset F \) of \( (k < n) \) features that is "maximally informative" about the output class \( C \), i.e., that maximizes the mutual information \( I(S; C) \) between input candidate features in \( S \) and the output class labels \( C \).

Because of the difficulty in direct computing of mutual information between multiple variables (i.e., \( I(S; C) \)) [3] [6] [15], a greedy scheme is proposed by Battiti in its MIFS algorithm [3] to solve the FRnk problem. In [4], Kwak and Choi improved MIFS [3] under assumptions that the information of all features is distributed uniformly in their MIFS-U algorithm. These two algorithms (i.e., [3] and [4]) select the next candidate feature as the one that maximizes the information about the output class and simultaneously minimizes a quantity proportional to the redundant information with already selected ones in \( S \).

These algorithms were sufficiently studied that we can now apply them on real textures. The objective is to evaluate their effectiveness in selecting the most informative textural features that improve the classification accuracy.

The performances of the retained algorithms and those of PCA and LDA, will be evaluated by reference to the ground truth using SVM classifier [19] [20]. We note that the applicability of the maximum mutual information criterion, with similar scheme to textural feature selection, was successfully tested on various textures from Brodatz Album [9] and has been used for multispectral image classification [18].

The SPOT image that we are working on refers to a forest area in the region of Rabat, Morocco. It was acquired in December 1999 with 20m of space resolution. From this image two test zones are extracted with respective size of 140x90 and 90x140 pixels, occupied essentially by five forest classes: a natural forest of eucalyptus coppice, a eucalyptus plantation, cork-oak forest, pine forest, and acacia forest. The sixth class corresponds to the sea. With regard to this, in pattern recognition and in image processing, it is often preferable to do some linear transformation to get better representation of color space [12]. In our case, we propose to use the color of each pixel in image as a 3-dimensional vector \(X=(r, g, y)\) where 

\[
r = \frac{R}{R + G + B}, \quad g = \frac{G}{R + G + B}, \quad y = \frac{B}{R + G + B}
\]

is a lightness variable. The color space \((r, g, y)\) is often used in color modeling and in pattern recognition. The variables \(R, G\) and \(B\) correspond respectively to the three channels: red, green and near infrared of our SPOT image.

The rest of this paper is organized as follows. In section (II) we give a brief view on the WT and its application to extract some textural features from our SPOT image. In section (III), we will deal with the SVM classifier in brief. The complete results and their comments will be presented in section (IV). Finally, a conclusion and some perspectives follow in in section (V).

II. TEXTURAL FEATURE EXTRACTION BY WAVELET TRANSFORM

The wavelet transform is a powerful mathematical tool to analyze our SPOT image by representing its textures at different scales and to extract some local features of variable sizes, variable frequencies, and at variable locations [10] [22]. At one-dimensional discrete signal, the standard wavelet transform consists in decomposing the input signal using two filters,
a low-pass filter and a high-pass filter, into two subbands (approximation A and details D). To decompose this signal, the process of filtering and downsampling is applied iteratively to the low-frequency subbands. At each level of this process, the high-frequency subband is preserved. When the decomposition reaches its highest level, both the low and high-frequency subbands are retained. Each level of transformation is regarded as a decomposition of the input signal into two mutually orthogonal subspaces.

The extension of the wavelet transform to two dimensional signals (image) is straightforward using the four 2-d filters, obtained by tensor product of the two one-dimensional filters, in horizontal and vertical directions. The four subbands obtained by these 2-d filters, after decomposition at level \( l \), are the images of Approximation(\( A_l \)) and details (i.e., Horizontal(\( H_l \)), Vertical(\( V_l \)) and Diagonal(\( D_l \))). At the outputs of these filters, the total number of samples will obviously be doubled. The decimation (downsampling) by 2 is needed to keep the number of samples the same as before [22].

To characterize the textures, we calculate on a sliding window of size \( 7 \times 7 \) pixels, from each subband resulting from the decomposition, the five following parameters:

\[
\begin{align*}
\text{ENT} &= \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} x(m,n)^2 \log(x(m,n)^2) \\
\text{MEAN} &= \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} |x(m,n)| \\
\text{MNT}_k &= \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (x(m,n) - \mu)^k; \quad k = 3, 4 \\
\text{ENE} &= \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} x(m,n)^2 
\end{align*}
\]

Where \( x(m,n) \) is an element of wavelet coefficient matrix. \( M \) and \( N \) represent horizontal and vertical size of \( x \) matrix. For more details on the choices of the highest decomposition level at two, the biorthogonal wavelet, and the windows size of the image bloc as \( 7 \times 7 \) pixels, see the previous work [10]. The five textural features are the entropy (ENT), the texture mean (MEAN), the third moment (MNT4), and the energy (ENE). The total number \( |F| \) of features extracted by WT at level 1 is \( |F| = 60 \) (i.e., 4 subbands \( \times \) 5 textural features \( \times \) 3 channels) and \( |F| = 105 \) at second level (i.e., 7 subbands \( \times \) 5 textural features \( \times \) 3 channels).

### III. SUPPORT VECTOR MACHINE CLASSIFIER

The Support Vector Machine (SVM) is a modern classifier which uses kernels to give optimal decision boundary to separate classes in higher dimensional feature spaces. It was successfully evaluated on several pattern recognition problems [6] [17]. The SVM algorithm was originally introduced by vapnik [19] [20] in his work on structural risk minimization. For example, in two classes problem (positive and negative sets of samples), the basic form of linear SVM classifier tries to find an optimal hyperplane that separates the set of positive samples from the set of negative ones. In this work, we use the LIBSVM package [21] that supports multi-class problem to classify our different textures.

### IV. EXPERIMENTAL RESULTS

#### A. Classification Strategy

The experiments presented here concern the classification of six classes of real textures listed in the two forest zones. Five forest classes: The eucalyptus coppice, the eucalyptus, the cork-oak, the pine and the acacia. The sixth class corresponds to the sea. The three other classes (road, urban zones and ground) will not be classified because of their small number of pixels and their distribution on small pieces (see ground truth in figure 1 above). The training cannot be done on these classes.

The data base is divided into two parts: the training base and the test base. The training base is built on the following way: we extract manually, from the two studied forest zones, six various textures corresponding to our six classes. Then, we calculate, on a sliding window on these zones, the textural features using WT at levels \((l=1,2)\). Each level corresponds to a training base of 600 observations equally partitioned between the six classes. The test bases correspond only to the pixels located in homogeneous zones, i.e., far away from the frontier zones between pieces. Table II gives the samples per class in each test base.

The selected textural features \( |S| \), according to the algorithms MIFS, MIFS-U, and the retained components of PCA and LDA (marked by (*) in table I) are used as input data for SVM classifier to show their performances. We note here that we retain the components of PCA and LDA that correspond to a variance proportion of 99.90% and 99.99% respectively. To implement the MIFS and MIFS-U algorithms, we estimate mutual information using the histogram approach with 18 bins. To control the redundant information, we use three values of \( \beta \) (0.35, 0.5 and 0.65).

#### B. Comparative Performances

Table I shows the comparative performances of all retained algorithms in terms of global accuracy classification and dimensionality reduction. The highest performances are obtained according to the MIFS-U algorithm with \( \beta = 0.65 \). Only three textural features are selected from the input candidate ones \((|F| = 105)\) extracted by WT at level 2. The lowest performances are obtained by PCA and LDA methods except for one particular case (MIFS, \( \beta = 0.65 \)) where the selected features are too small \((|S| = 2)\).

The algorithms MIFS and MIFS-U outperform the classical approaches PCA and LDA for both textural feature extraction methods \((WT, l=1,2)\). Table I also shows that the more the \( \beta \) value gets closer to 1, the more the selected textural features decreases and vice-versa. Compared to MIFS-U, the MIFS algorithm of Battiti uses fewer textural features.

Figure 1 shows the classification results of the two studied forest zones produced by SVM classifier. This latter uses the three informative textural features selected by MIFS-U algorithm from the input ones extracted by WT at level 2. The confusion matrix of this classification is given by table III.
TABLE I
COMPARATIVE PERFORMANCES IN TERMS OF GLOBAL ACCURACY CLASSIFICATION (%) AND DIMENSIONALITY REDUCTION

<table>
<thead>
<tr>
<th>Texture extraction method</th>
<th>WT (l = 1)</th>
<th>WT (l = 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algorithms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>β = 0.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIFS</td>
<td>59.66</td>
<td>58.13</td>
</tr>
<tr>
<td>MIFS-U</td>
<td>62.82</td>
<td>72.78</td>
</tr>
<tr>
<td>β = 0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIFS</td>
<td>68.22</td>
<td>67.59</td>
</tr>
<tr>
<td>MIFS-U</td>
<td>67.61</td>
<td>68.39</td>
</tr>
<tr>
<td>β = 0.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIFS</td>
<td>71.84</td>
<td>71.40</td>
</tr>
<tr>
<td>MIFS-U</td>
<td>70.31</td>
<td>68.39</td>
</tr>
<tr>
<td>PCA</td>
<td>56.77</td>
<td>57.79</td>
</tr>
<tr>
<td>LDA</td>
<td>61.26</td>
<td>58.57</td>
</tr>
</tbody>
</table>

TABLE II
SAMPLE SIZES PER CLASS IN THE TEST BASE

<table>
<thead>
<tr>
<th>Textures</th>
<th>Eucalyptus coppice</th>
<th>Eucalyptus</th>
<th>Oak-cork</th>
<th>pin</th>
<th>Acacia</th>
<th>Sea</th>
</tr>
</thead>
<tbody>
<tr>
<td>samples</td>
<td>79</td>
<td>2914</td>
<td>2679</td>
<td>587</td>
<td>91</td>
<td>201</td>
</tr>
</tbody>
</table>

TABLE III
CONFUSION MATRIX OF OUR SPOT IMAGE CLASSIFICATION USING THE SELECTED TEXTURAL FEATURES BY MIFS-U(β = 0.65) ALGORITHM FROM INPUT ONES EXTRACTED BY WT AT LEVEL 2

<table>
<thead>
<tr>
<th>Eucalyptus coppice</th>
<th>Eucalyptus</th>
<th>Oak-cork</th>
<th>pin</th>
<th>Acacia</th>
<th>Sea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eucalyptus coppice</td>
<td>53.16</td>
<td>2.53</td>
<td>64.3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Eucalyptus</td>
<td>23.68</td>
<td>75.88</td>
<td>0.45</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Oak-cork</td>
<td>17.36</td>
<td>6.16</td>
<td>76.19</td>
<td>0</td>
<td>0.30</td>
</tr>
<tr>
<td>pin</td>
<td>3.41</td>
<td>3.92</td>
<td>16.70</td>
<td>33.73</td>
<td>42.25</td>
</tr>
<tr>
<td>Acacia</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>17.58</td>
<td>82.42</td>
</tr>
<tr>
<td>Sea</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

V. CONCLUSION

So far, we have evaluated the effectiveness of two algorithms (MIFS and MIFS-U) in selecting the most informative textural features for multispectral image classification. These algorithms were evaluated on real textures using WT at levels (l=1,2) and the SVM classifier. Taken into account the deforestation of the two studied forest zones, the performances of the MIFS-U algorithm, compared to the classical approaches PCA and LDA and the MIFS algorithm [3], offer promising results for multispectral imagery classification. These performances are promising to integrate this algorithm in our future work for an automatized forest cartography system based on the satellite imagery classification. This work also proves that Shannon’s mutual information is a powerful tool in the context of dimensionality reduction for multispectral imagery classification.
REFERENCES