IMPROVING CLASSIFICATION ACCURACY VIA CONTEXTUAL FEATURE RANKING IN HIGH SPATIAL RESOLUTION SATELLITE IMAGERY

B. Mojaradi$, H. Ashoori**, M. Mirghasempour$

Department of Geodesy & Geomatics, Shahid Rajaee Teacher Training University, P.O.Box 16785/163, Tehran, Iran - $bmojaradi@gmail.com$, **Hamed_ashoori@yahoo.com$, *m.a.ghasempour@srttu.edu

Commission VII, WG VII/4

KEY WORDS: High Spatial Resolution, Classification, Texture Quantization, Feature Ranking, SVM

ABSTRACT: High Spatial Resolution, Classification, Texture Quantization, Feature Ranking, SVM

Texture quantization is a useful method for extraction spatial relevance between pixels which is used in humane brain for image interpreting. Beside the spectral bands textural features of high spatial resolution data can be used to improve classification accuracy. Depends on the land cover characteristics different textural features possibly are effective from large number of available textural features. So it is important to find proper features among available features for special case studies. In this paper efficient features are determined by ranking based on their ability for improving class separability. The quadratic discriminate classifier (QDC) and support vector machine (SVM) are used for data classification. Comparative tests on compact of texture features and training sample size to improve accuracy of QDC and SVM classifiers demonstrated that i) QDC is an efficient classifier with large number of training samples while due to using more texture features led to futile result in high dimensional feature space; ii) SVM generates accurate results in high dimensional feature spaces and can train with few training samples. Experimental results show 13% and 10% improvement in obtained average and overall accuracies respectively.

1. INTRODUCTION

Classification is the most common method of extracting information from remotely sensed data. In conventional classification methods only spectral data are used. High resolution images have more spatial information but do not have a high spectral resolution, so using conventional classification methods seems to be ineffective. To improve the classification accuracy, spatial information, which is a reach source of useful information and is the merit of this kind of images, could be used. Texture quantization is an effective approach for utilization of the spatial information.

Many authors have been introduced variety of methods to quantify spatial relations between pixels and used them as an input feature in the classification task. There are a wide range of texture quantization methods that are classified in three main groups, statistical, structural and spectral based methods (Kenneth R, Castleman, 1996). Statistical methods produce statistical measures of grey level variation; Structural methods assume that the texture pattern is composed of spatial arrangement of texture primitives, so their task is to locate the primitives and quantify their spatial arrangement; and Spectral features are generated using the spectrum obtained through image transformations such as Fourier transform.

In each method several parameters are set to produce different features like kernel size, distance vector, formulation. Hence, using different parameters result in large number of new features which could be generated. Each feature quantizes special characteristic of image texture, and depends on the variety of interested classes special method with special parameters should be generated and used to discriminate desire land covers. So huge number of texture features should generate and test to obtain the best combination. To get best result in an automatic procedure with short time consuming they should be ranked to find the best combination.

2. METHODS AND ALGORITHMS

In this paper to evaluate the potential of texture quantization method for classification of high resolution images, several feature generation methods were applied. And to test feature ranking for finding best features, feature ranking based on their ability to separate classes is used. Also for classification stage, QDC and SVM classifiers were used.

2.1 Generated features

In this research, First Order statistical, Gray level co-occurrence based, Geostatistical, Fourier based and wavelet based features were generated. For features formulation see (Ashoori, H, et al., 2006). This methods can be categorized in Statistical (first and second order), and spectral features. Features generated from wavelet could have multiresolution view to the image because they could be generated from different steps of wavelet transform. Generated features name are listed in experiment explanation.

2.2 Feature Ranking:

To evaluate discriminate potential of features, trace of between-class to within-class scattering matrix ratio is defined as follows:

\[
J = \text{trace}\left((S_w)^{-1}S_b\right)
\]

\[
S_w = \sum_{i=1}^{L} P_i E\left(\left(X - m_i\right)\left(X - m_i\right)^T\right) = \sum_{i=1}^{L} P_i \sum_{j=1}^{L} P_i
\]

\[
S_b = \sum_{i=1}^{L} P_i \sum_{j=1}^{L} P_i (m_i - m_j)(m_i - m_j)^T = \sum_{i=1}^{L} P_i \sum_{j=1}^{L} P_i (m_i - m_j)(m_i - m_j)^T
\]
Where $S_b$ means between-class scatter, $S_w$ refers to within-class scatter matrix, $P_i$ is to the prior probability of class $i$, $m_i$ is mean of class $i$, $L$ is number of classes.

2.3 Support Vector Machines

Support Vector Machines (SVM) have been recently proposed as a method for pattern classification and nonlinear regression. Their appeal lies in their strong connection to the underlying statistical learning theory where an SVM is an approximate implementation of the method of structural risk minimization (Vapnik, V.N., 1998). SVM has many attractive features. For instance, the solution of the quadratic programming (QP) problem (Fletcher, R., 1987) is globally optimized while with neural networks the gradient based training algorithms only guarantee finding local minima. In addition, SVM can handle large feature spaces (specially convenient when working with high dimensional data), can effectively avoid overfitting by controlling the margin and can automatically identify a small subset made up of informative points, namely support vectors (SV). Consequently, they have been used for particle identification, face recognition, text categorization, time series prediction, bioinformatics, texture classification, etc. Visit http://www.kernel-machines.org for publications and application resources. In the following, we summarize the “one-against-the-rest procedure” for classification purposes, in which, a classifier is obtained for each class. Given a labeled training data set $((x_1, y_1), (x_n, y_n), \ldots, (x_i, y_i))$, where $x_i \in \mathbb{R}^d$ and $y_i \in \{+1, -1\}$ and a nonlinear mapping, $\phi(\cdot)$, usually to a higher dimensional space, $\mathbb{R}^d \rightarrow \mathbb{R}^{H}$. In this paper instead of increasing dimensionality by linear or nonlinear mapping, generated textural features are used to increase dimensionality of space. The SVM method solves:

$$
\min_{w, b, \xi} \left\{ \frac{1}{2} \|w\|^2 + C \sum_i \xi_i \right\} 
$$

Subject to the following constraints:

$$
y_i (\Phi^T (x_i)w + b) \geq 1 - \xi_i \quad \forall i = 1, \ldots, n
$$

$$
\xi_i \geq 0 \quad \forall i = 1, \ldots, n
$$

Where $w$ and $b$ define a linear regressor in the feature space, nonlinear in the input space unless $\phi (x_i) = x_i$. In addition, $\xi_i$ and $C$ are, respectively, a positive slack variable and the penalization applied to errors (Figu.2). The parameter $C$ can be regarded as a regularization parameter which affects the generalization capabilities of the classifier and is selected by the user. A larger $C$ corresponds to assigning a higher penalty to the training errors.

In order to solve problems with $k$ classes we must reformulate the problem. Given a classifier $(w_j, b_j), j \in \{0, \ldots, k-1\}$ for each class, to assign a sample $x$ to a certain $k$ class we must calculate the output of the $k$ classifiers and select the one with the highest output. We then proceed as in the binary case. Full details on the solution can be found in (Lin, Y, et al., 2000).

3. EVALUATION AND EXPERIMENTS

In order to evaluate the proposed method, a sub-image of IKONOS data with the size of $1377 \times 1335$ pixels was used. This image was taken over the northeast Tehran city in June 2005. The data is a pansharped image that has 4 spectral bands, with a spatial resolution of 1 m. Eight classes including row trees, dense tree type 1, grass, highway, uncultivated area, building, dense Tree type 2 and bush are determined in the image which number of labelled samples per class is given in Table 1. A subset of image is shown in Figure 2.

Having selected samples for each class, the image was used to generate features; to prevent time consuming features were generated only for selected samples. Textural features were generated according methods pointed out in introduction.
### Table 1. Selected Samples and their size

<table>
<thead>
<tr>
<th>Class Name</th>
<th>No. of Pixels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row of Trees</td>
<td>184</td>
</tr>
<tr>
<td>Dense Tree type 1</td>
<td>220</td>
</tr>
<tr>
<td>Grass</td>
<td>247</td>
</tr>
<tr>
<td>Highway</td>
<td>158</td>
</tr>
<tr>
<td>Uncultivated area</td>
<td>452</td>
</tr>
<tr>
<td>Building</td>
<td>185</td>
</tr>
<tr>
<td>Dense Tree type 2</td>
<td>623</td>
</tr>
<tr>
<td>Bush</td>
<td>479</td>
</tr>
</tbody>
</table>

### 3.1 Experiment 1 - Textural features advantage

The goal of this experiment is to assess the effect of texture features on the classification accuracy. Since the large numbers of textural features have been generated, the features are ranked based on the trace of between-class to within-class scattering matrix ratio. The discriminate potential of each feature in terms of $J$ is shown in Figure 3.

![Figure 3. Discriminate potential of generated features](image)

Investigation of ranked features shows that wavelet-based features have high discrimination potential to classify selected classes. Then, first-order features, gray level co-occurrence, Fourier-based and geostatistical-based features were in next priority to contribute for classification respectively. As there are similar vegetation classes in land cover, it was observed that compared to textural features of blue, green, and red bands the textural features which are derived from near infra-red bands are more informative in terms of $J$.

To evaluate the effect of texture features, the image was classified with a fixed number of training data using SVM and Quadratic Discriminant Classifier (QDC) classifiers. The QDC was trained using 40 percent of training data with applying only 4 spectral bands. Moreover, in addition to spectral bands, textural features that got high separability ranking in feature ranking stage are used for image classification. This classification conducted by increasing feature space dimension thorough adding one by one textural feature to original 4 spectral bands. The class accuracies given by this scenario is shown in Figure 4 in terms of OA. As can be observed the best OA is yielded with 37 features (including 4 spectral bands and 33 textural features). The class accuracies, OA, and AA corresponding to using 4 spectral bands and 33 textural features for classification are shown in Table 2. The OA and average accuracies (AA) demonstrate that texture features are efficient and able to improve the classification accuracy. As can be seen for a fixed number of training data (40 percent) compared to using only 4 spectral features, the OA of QDC using 33 textural features yielded 10.15% improvement, this improvement is 13.03% for AA. As a result QDC that trained using 40 percent of training data obtains perfect accuracy with applying 37 features.
3.2 Experiment 2 – Training sample size effect

The goal of this experiment is to assess the effect of limited training sample size on the classification accuracy. For this purpose, the image was classified by SVM classifier using 15 pixels as training data for each class. This experiment was conducted as pointed out in experiment 1 using only spectral bands and considering different textural features. The classification results are shown in Figure 5, in terms of OA. As shown, the maximum accuracy is obtained in 22 features including 4 spectral bands and 18 textural features. The class accuracies, OA and AA corresponding to using 4 spectral bands and 22 features for classification are shown in Table 2. As can be observed, compare to QDC, the SVM shows better performance in dealing with limited sample size in high-dimensional space. In the other hand, the QDC needs more samples for training in high-dimensional space. It is worth mentioned the OA of SVM alter due to the limited sample size. In contrast, obtaining 95.43% OA using 15 pixels as training data for SVM is remarkable.

Moreover, image is classified by QDC with only 4 spectral bands and 15 pixels is used as training data. The classification results are shown in Table 2. It can be seen compared to using 15 pixels, 40 percent training data yielded 5.08% improvement on OA when only 4 spectral bands are applied for classification.

As can be observed, compare to QDC, the SVM shows better performance in dealing with limited sample size in high-dimensional space. In the other hand, the QDC needs more samples for training in high-dimensional space. It is worth mentioned the OA of SVM alter due to the limited sample size. In contrast, obtaining 95.43% OA using 15 pixels as training data for SVM is remarkable.

### Table 2. Classification accuracy of test data

<table>
<thead>
<tr>
<th>Class Name</th>
<th>QDC 37 Feature, 40 percent training</th>
<th>QDC 4 bands, 40 percent training</th>
<th>SVM 22 features, training sample size 15 pixel</th>
<th>QDC 4 bands, training sample size 15 pixel</th>
<th>SVM 4 bands, training sample size 15 pixel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row Trees</td>
<td>100</td>
<td>37.27</td>
<td>78.70</td>
<td>46.15</td>
<td>41.42</td>
</tr>
<tr>
<td>Dense Tree type 1</td>
<td>99.24</td>
<td>76.52</td>
<td>93.17</td>
<td>57.07</td>
<td>42.93</td>
</tr>
<tr>
<td>Grass</td>
<td>100</td>
<td>100</td>
<td>92.24</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Highway</td>
<td>97.87</td>
<td>100</td>
<td>88.81</td>
<td>95.10</td>
<td>95.80</td>
</tr>
<tr>
<td>Uncultivated area</td>
<td>99.26</td>
<td>97.78</td>
<td>99.54</td>
<td>83.30</td>
<td>96.8</td>
</tr>
<tr>
<td>Building</td>
<td>100</td>
<td>96.40</td>
<td>94.12</td>
<td>98.24</td>
<td>88.24</td>
</tr>
<tr>
<td>Dense Tree type 2</td>
<td>100</td>
<td>94.64</td>
<td>99.34</td>
<td>85.53</td>
<td>54.11</td>
</tr>
<tr>
<td>Bush</td>
<td>100</td>
<td>89.55</td>
<td>97.63</td>
<td>89.44</td>
<td>42.46</td>
</tr>
<tr>
<td>OA</td>
<td>99.67</td>
<td>89.52</td>
<td>95.4283</td>
<td>83.44</td>
<td>66.96</td>
</tr>
<tr>
<td>AA</td>
<td>99.55</td>
<td>86.52</td>
<td>92.9441</td>
<td>81.85</td>
<td>70.22</td>
</tr>
</tbody>
</table>

4. CONCLUSION

In this paper, an algorithm for classification of high spatial images using textural features is proposed. Having generated textural features, the proper textural features via feature ranking are chosen. The algorithm is investigated in two cases using textural features and only spectral bands with sufficient and limited training sample size. Comparing obtained accuracies the following results are yielded:—

- Results of experiment 1 demonstrate texture quantization and applying textural features beside spectral features improve classification accuracy significantly.
- Ranking generated features helps to find proper textural features for classification. As a result instead of using all features for classification task only efficient ones that increase discrimination between classes are used. Hence, this method suggests, before utilizing features, to investigate textural features on training and test area.
prevents to time consuming for generating a lot of features for whole image.

- Obviously, Maximum likelihood classification lead to deficient results in high dimensional feature spaces, or demands a large sample size of training data to perform good classification. In contrast, SVM classifier shows high performance in dealing with limited sample size in high dimensional space. Another merit of this method is that utilizing textural features can be substitute with kernels, as a function for increasing dimensionality, in SVM classifier.

REFERENCES

References from Books:


References from Other Literature:

