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

B. Mojaradi [a], H. Ashoori [b*], M. Mirghasempour [c]

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

Commission VII, WG VII/4

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

ABSTRACT:

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:

$$J = \text{trace}\left((S_w)^{-1} * S_b \right) \tag{1}$$

$$S_w = \sum_{i=1}^{L} P_i E\left\{ (X - m_i)(X - m_i)^T \middle| \omega_i \right\} = \sum_{i=1}^{L} P_i \sum\nolimits_i \tag{2}$$

$$S_b = \sum_{i=1}^{L} P_i \, (m_i - m_0)(m_i - m_0)^T = \sum_{i=1}^{L-1} \sum_{j=i+1}^{L} P_i P_j (m_i - m_j)(m_i - m_j)^T \tag{3}$$

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 $((x_1, y_1), (x_n, y_n)$, where xi є R^d and y_i є {+1, -1}) and a nonlinear mapping, φ(·), usually to a higher dimensional space, $\Re^d \xrightarrow{\Phi(\bullet)} \Re^H$ (H > d). 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:

$$\min_{w,\xi_i,b}\left\{\frac{1}{2}\|w\|^2 + C\sum_i \xi_i\right\} \tag{4}$$

Subject to the following constraints:

$$\begin{aligned} y_i(\Phi^T(x_i)w + b) &\geq 1 - \xi_i & \forall i = 1,...,n \\ \xi_i &\geq 0 & \forall i = 1,...,n \end{aligned} \tag{5}$$

Where w and b define a linear regressor in the feature space, nonlinear in the input space unless φ $(x_i) = x_i$. In addition, ξ_i and C are, respectively, a positive slack variable and the penalization applied to errors (Figur.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.

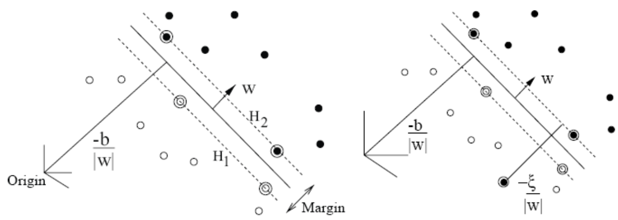

Figure 1. Left: The Optimal Decision Hyperplane in a linearly separable problem. Right: Linear decision hyperplanes in nonlinearly separable data can be handled by including Slack variables i. Figures adapted from (Sch¨olkopf, B., Smola, A., 2001)

An SVM is trained to construct a hyperplane $\varphi^T(x_i)w+b = 0$ for which the margin of separation is maximized. Using the method of Lagrange multipliers, this hyperplane can be represented as:

$$\sum_i \alpha_i y_i \phi(x_i).\phi(x) = 0 \tag{6}$$

Where the auxiliary variables α_i are Lagrange multipliers. Its solution reduces to: Maximize:

$$L_d \equiv \sum_i \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_i y_i y_j \phi(x_i).\phi(x_j) \tag{7}$$

Subjects to the constraints:

$$0 \leq \alpha_i \leq C, \qquad \sum_{i,j} \alpha_i y_i = 0$$

In order to solve problems with k classes we must reformulate the problem. Given a classifier (w_j , b_j), j є {0, ..., 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 × 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.

Class Name	No. of Pixels
Row of Trees	184
Dense Tree type 1	220
Grass	247
Highway	158
Uncultivated area	452
Building	185
Dense Tree type 2	623
Bush	479

Table 1. Selected Samples and their size

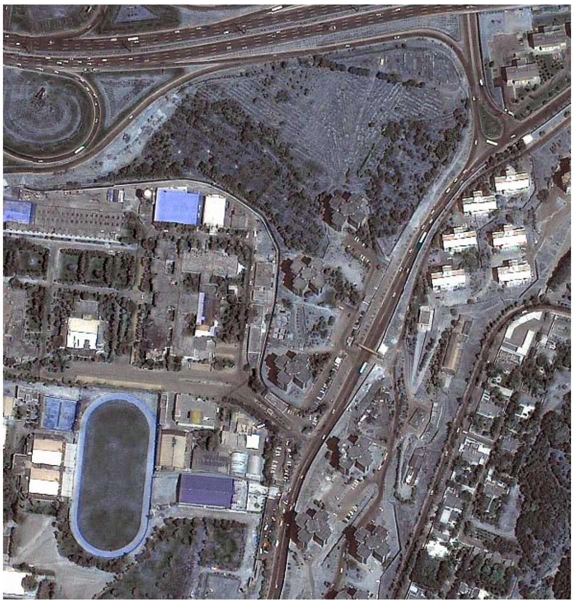

Figure 2. A subset of used image

In first order statistical group, textural features based on the Mean, Mode, Median, Central Moment 1, Central Moment2 (Variance), Central Moment3 (Skewness) and Central Moment4 (Kurtosis) were generated.Moreovr, in Gray level Co-Occurrence based features Row Variance, Column Variance and Correlation have been produced using four main distance vectors ((1,1),(1,0),(0,1),(1,-1)). In Geostatistical group, Madogram and Direct variogram features were also generated by applying the same distance vectors. Four main frequency masks including Ring, Slice, Horizontal and Vertical form were used to generate Fourier based features, both high-pass and low pass forms were also consider generating features. And the last group of features were wavelet based features which contain Log Energy, Shannon's Index, Angular Second Moment, Entropy, Log Energy features, in this method first and second steps of wavelet transform were used to generate mentioned features from.

In all groups different kernel sizes ranging from 3×3 to 27×27 were used to generate textural features. These kernels are applied over four spectral bands to generate features individually. As a result for each image band 612 features are derived that yielded 2448 textural features for four spectral bands. Hence, dimensionality of feature space increases to 2452 with 4 spectral bands.

3.1 Experiment 1 - Textural features advantage

The goal of this experiment is assessment 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 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.

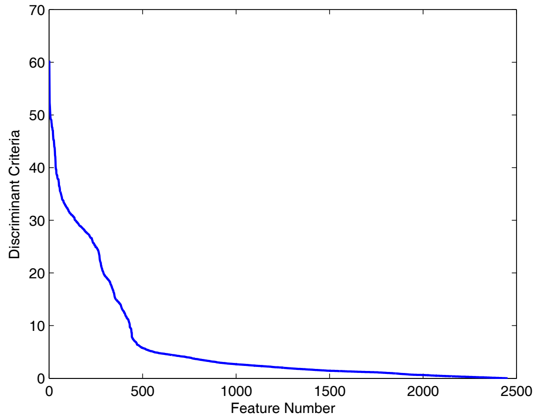

Figure 3. Discriminate potential of generated features

Investigation of ranked features show that wavelet based features had 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 compare to textural features of blue,green and red bands the textural features which derived from near infra-red band were 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 four 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 texture 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 texture 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.

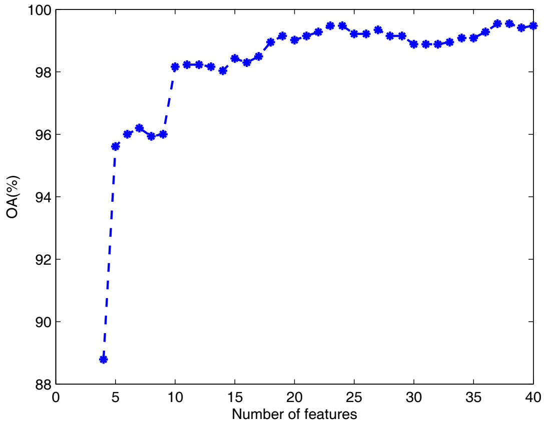

Figure 4. Overall accuracy for different number of combining features using 40% as training and QDC classifier

3.2 Experiment 2 – Training sample size effect

The goal of this experiment is assessment 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 experiment1 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 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.

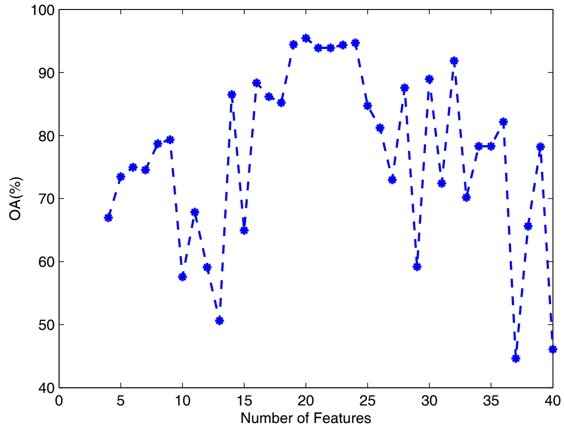

Figure 5. Overall accuracy for different number of combining features with 15 pixel training size and SVM classifier

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.

Class Name	QDC 37 Feature, 40 percent training	QDC 4 bands, 40 percent training	SVM 22 features, training sample size15 pixel	QDC 4 bands, training sample size15 pixel	SVM 4 bands, training sample size15 pixel
Row Trees	100	37.27	78.70	46.15	41.42
Dense Tree type 1	99.24	76.52	93.17	57.07	42.93
Grass	100	100	92.24	100	100
Highway	97.87	100	88.81	95.10	95.80
Uncultivated area	99.26	97.78	99.54	83.30	96. 8
Building	100	96.40	94.12	98.24	88.24
Dense Tree type 2	100	94.64	99.34	85.53	54.11
Bush	100	89.55	97.63	89.44	42.46
OA	99.67	89.52	95.4283	83.44	66.96
AA	99.55	86.52	92.9441	81.85	70.22

Table 2. Classification accuracy of test data

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. So it

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:

Kenneth R, Castleman.,1996. Digital Image Processing. Prentice-Hall

Vapnik, V.N.,1998, *Statistical Learning Theory*. John Wiley & Sons, New York

Fletcher, R.,1987, *Practical Methods of Optimization*. 2nd Edition, John Wiley & Sons, Inc.

References from Other Literature:

Ashoori, H., Alimohammadi, A., Valadan Zoej, M. J., Mojarradi, B.,2006. Generating Image-based Features for Improving Classification Accuracy of High Resolution Images, May, ISPRS Mid-term Symposium, Netherlands.

Sch¨olkopf, B., Smola, A.,2001, Learning with Kernels – Support Vector Machines, Regularization, Optimization and Beyond. MIT Press Series) partially available from http://www.learning-with-kernels.org

Lin, Y., Lee, Y., Wahba, G.,2000, Support Vector Machines for classification in nonstandard situations. Department of Statistics TR 1016, University of Wisconsin-Madison