

Generating Image-based Features for Improving Classification Accuracy of High Resolution Images

H. Ashoori, A. Alimohammadi, M. J. Valadan Zoej, B. Mojarradi

Geodesy and Geomatic Faculty, K.N.Toosi University of Technology, Tehran, Iran -
(hamed_ahoori, Alomoh_abb@yahoo.com, valadanzouj@kntu.ac.ir, mojaradi@albors.kntu.ac.ir)

KEY WORDS: Classification, Feature Generation, Texture

ABSTRACT:

High resolution images are rich sources of spatial information and details as well as noises. Because of the spectral similarities between different objects, pixel-based classifications of these images do not provide very accurate results. Incorporation of spatial data can lead to significant improvements in the accuracy of classifications. Several methods have been proposed for extraction of this information and generation of useful features for use in the classification process. In this study, different techniques for quantification of texture data including statistical, geostatistical, Fourier-based methods and gray level co-occurrence matrix have been investigated. New features have been generated by using spectral bands and the first PC resulting from the PCA transformation. The results have shown that the improvement of classification accuracy is significant and varies for different classes and features.

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 the conventional classification methods seems to be ineffective. Spatial information, which is a reach source of useful information especially in high resolution images, can be used to improve the classification accuracy. Texture quantization is an effective approach for utilization of the spatial information. There is no clear definition for image texture, but we can describe how the image texture looks e.g. fine, coarse, smooth or irregular, homogeneous and so forth. [1]

Many authors have introduced methods to quantify spatial relations between pixels and have used them as an input data in the classification. There are a wide range of texture quantization methods that can be classified in three main groups, statistical, structural and spectral based methods [2]. Statistical methods produce statistical measures of gray 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 this paper, five groups of features based on the first order statistics, Gray level co-occurrence matrix, geostatistics, Fourier transform and wavelet transform have been generated. The first three can be classified as the statistical and the last two as the spectral methods. Then different classifications resulting from the combination of different texture and spectral features have been evaluated.

2. GENERATED FEATURES

2.1 First Order Statistical Features

If (I) is the random variable representing the gray levels in the region of interest, the first order histogram $P(I)$ is defined as [1]:

$$P(I) = \frac{\text{number of pixels with gray level } I}{\text{Total number of pixels}}$$

Now different features can be generated by using the following equations:

2.1.1 Moment

$$m_i = E[I^i] = \sum_{I=0}^{N_g-1} I^i P(I) \quad i = 1, 2, 3, \dots \quad (1)$$

Where N_g = number of gray levels.

$m_1 = E[I]$ is the simple mean of pixels. Also 2nd, 3rd and other moments can be used.

2.1.2 Central Moments

$$\mu_i = E[(I - E[I])^i] = \sum_{I=0}^{N_g-1} (I - m_1)^i P(I) \quad (2)$$

2.1.3 Absolute Moments

$$\hat{\mu}_i = E[Abs(I - E[I])^i] \quad (3)$$

2.1.4 Entropy

$$H = -E[\log_2 P(I)] = -\sum_{I=0}^{N_g-1} P(I) \log_2 P(I) \quad (4)$$

2.1.5 Median: Median is the middle value in a set of numbers arranged in increasing order. Because the kernel size always covers odd number of pixels, median can be extracted simply by choosing the mid member of an array which contains gray levels of pixels that covered by the mask and then it is sorted.

2.1.6 Mode: Mode is the most frequent value of a random variable. So in an image mode is the most frequent pixel gray level.

2.1.7 Distance Weighted Mean: If the distance from center pixel is considered as the weight for computing the mean then near pixels have more contribution in the results.

$$Mean_w = \frac{\sum_{i=1}^{N_r} \sum_{j=1}^{N_c} \frac{1}{d_{i,j}} I(i, j)}{\sum_{i=1}^{N_r} \sum_{j=1}^{N_c} \frac{1}{d_{i,j}}} \quad (5)$$

2.2 Gray level Co-Occurrence Based Features: Haralick et.al [3] proposed this method to extract texture information from digital images. First Gray level co-occurrence matrix (GLCM) is produced and then several texture measures are computed from it. GLCM is a matrix that contains the number of each gray level pairs that are located at distance d and direction θ from each other. This matrix could be defined for different distances, angles and as well as for different lags.

$$GLCM_{d_1, d_2} = \frac{1}{R} \begin{bmatrix} \eta(0,0) & \eta(0,1) & \dots & \eta(0,N_{g-1}) \\ \eta(1,0) & \eta(1,1) & \dots & \dots \\ \dots & \dots & \dots & \eta(i,j) \\ \eta(N_{g-1},0) & \dots & \dots & \eta(N_{g-1},N_{g-1}) \end{bmatrix} \quad (6)$$

$\eta(i,j)$ #Pixel Pairs in lag (d_1, d_2) through

N_g : Number of Gray Levels

R : Total Number of Possible Pairs

In this research, following features have been generated from the GLCM matrix :

2.2.1 Mean

$$\mu_i = \sum_{i=0}^{N_{g-1}} \sum_{j=0}^{N_{g-1}} i \times P(i, j) \quad \mu_j = \sum_{i=0}^{N_{g-1}} \sum_{j=0}^{N_{g-1}} j \times P(i, j) \quad (7,8)$$

where $P(i,j)=GLCM(i,j)$

2.2.2 Variance

$$\sigma_i^2 = \sum_{i=0}^{N_{g-1}} \sum_{j=0}^{N_{g-1}} (i - \mu_i)^2 \times P(i, j) \quad (9)$$

$$\sigma_j^2 = \sum_{i=0}^{N_{g-1}} \sum_{j=0}^{N_{g-1}} (j - \mu_j)^2 \times P(i, j) \quad (10)$$

Mean and variance of GLCM are not the same as for the image because the frequency of occurrence of different pairs is modeled here.

2.2.3 Homogeneity (Inverse Differences Moment)

$$IDF = \sum_{i=0}^{N_{g-1}} \sum_{j=0}^{N_{g-1}} \frac{P(i, j)}{1 + (i - j)^2} \quad (10)$$

It assigns higher weight to the main diagonal of GLCM so it outputs higher value for images that have larger homogeneous areas.

2.2.4 Contrast

$$CON = \sum_{i=0}^{N_{g-1}} \sum_{j=0}^{N_{g-1}} (i - j)^2 P(i, j) \quad (11)$$

The more the distance from the main diagonal of GLCM the higher the weight that is assigned to the $P(i,j)$, so when the difference between neighboring pairs becomes large, the contrast increases.

2.2.5 Dissimilarity

$$CON = \sum_{i=0}^{N_{g-1}} \sum_{j=0}^{N_{g-1}} |i - j| P(i, j) \quad (12)$$

It works like contrast but gives lower weight to the difference of each gray level pairs.

2.2.6 Entropy

$$Entropy = - \sum_{i=0}^{N_{g-1}} \sum_{j=0}^{N_{g-1}} P(i, j) \ln(P(i, j)) \quad (13)$$

It outputs higher value for a homogeneous distribution of $P(i,j)$, and lower otherwise.

2.2.7 Angular Second Moment

$$ASM = \sum_{i=0}^{N_{g-1}} \sum_{j=0}^{N_{g-1}} (P(i, j))^2 \quad (14)$$

It is a measure of image smoothness. It outputs higher values when $P(i,j)$ is concentrated in a few places in the GLCM and lower if the $P(i,j)$ are close in value.

2.2.8 Correlation

$$Correlation = \sum_{i=0}^{N_{g-1}} \sum_{j=0}^{N_{g-1}} \frac{(i - \mu_i)(j - \mu_j) P(i, j)}{\sigma_i \sigma_j} \quad (15)$$

It measures linear dependency of gray levels on those of neighboring pixels.

2.3 Geostatistical Features

Geostatistics is the statistical methods developed for and applied to geographical data. These statistical methods are required because geographical data do not usually conform to the requirements of standard statistical procedures, due to spatial

autocorrelation and other problems associated with spatial data [4].

Semivariogram that represents half of the expectation of the quadratic increments of pixel pair values at the specified distance can quantify both spatial and random correlation between the adjacent pixels. [5] It is defined as:

$$\gamma(h) = \frac{1}{2n(h)} \sum_{i=1}^{N_r-h_1} \sum_{j=1}^{N_c-h_2} [DN(i, j) - DN(i + h_1, j + h_2)]^2 \quad (16)$$

That is the classical expression of variogram (h) here represents a vectorial lag between pixels. In this study direct variogram, madogram, cross variogram and pseudo-cross variogram have been used. The first two operate separately for each image bands and the second two operate for pairs of image bands.

2.3.1 Direct Variogram

In this approach the following equation is used to estimate:

$$\gamma(h) = \frac{1}{2n(h)} \sum_{i=1}^{N_r-h_1} \sum_{j=1}^{N_c-h_2} (DN(i, j) - DN(i + h_1, j + h_2))^2 \quad (17)$$

$n(h)$ is the number of pairs that are in mask filter.

2.3.2 Madogram: This is similar to direct variogram except squaring differences, but uses the absolute value of differences.

$$\gamma(h) = \frac{1}{2n(h)} \sum_{i=1}^{N_r-h_1} \sum_{j=1}^{N_c-h_2} |DN(i, j) - DN(i + h_1, j + h_2)| \quad (18)$$

2.3.3 Cross Variogram: Two image bands are used to quantify the joint spatial variability between bands.

$$\gamma_{m,n}(h) = \frac{1}{2n(h)} \sum_{i=1}^{N_r-h_1} \sum_{j=1}^{N_c-h_2} \{[DN_m(i, j) - DN_m(i + h_1, j + h_2)]^2 + [DN_n(i, j) - DN_n(i + h_1, j + h_2)]^2\} \quad (19)$$

2.3.4 Pseudo-cross Variogram: It is similar to direct variogram, but uses pairs which are from two different bands (m,n).

$$\gamma_{m,n}(h) = \frac{1}{2n(h)} \sum_{i=1}^{N_r-h_1} \sum_{j=1}^{N_c-h_2} [DN_m(i, j) - DN_n(i + h_1, j + h_2)]^2 \quad (20)$$

2.4 Fourier Based Features

Fourier transformation, transforms a signal from space/time domain to frequency domain. The amplitude and phase coefficients are two outputs of a Fourier transformation. So different texture patterns could be identified by their Fourier coefficients but because in this research one value for each pixel is required, raw Fourier coefficients couldn't be used. Several features can be generated using sum of the Fourier amplitude under different masks [6]. These are comprised ringing, sectorial, horizontal and vertical which are shown in figure 1.

$$F(u, v) = \frac{1}{N} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) e^{-j \frac{2\pi}{N} (ux+vy)} \quad (21)$$

$$\text{Fourier Amplitude: } A(u, v) = |F(u, v)|^2$$

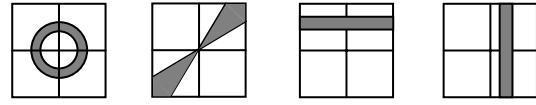


Figure 1. Different mask which can be used to generate features from Fourier coefficients

3) Wavelet Based Features

Mallat (1989) developed the multiresolution analysis theory using an orthonormal wavelet basis. The multiresolution wavelet transform decomposes a signal into low frequency approximation and its high frequency detail information at a coarser spatial resolution. [7]

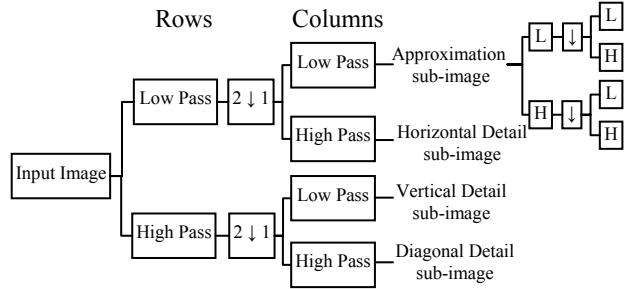


Figure 2. Decomposition procedure using multiresolution analysis

New features could be generated using wavelet transform outputs. We generated four defined features using approximation output of first and second level of transformed image. To obtain second level we use first level approximation to do the same decomposition. Those were defined by different authors [7] and they are:

$$LOG = \sum_{i=1}^k \sum_{j=1}^k \log(P(i, j)^2) \quad (22)$$

$$SHAN = -\sum_{i=1}^k \sum_{j=1}^k P(i, j) * \log(P(i, j)) \quad (23)$$

$$ASM = \sum_{i=k}^k \sum_{j=k}^k P(i, j)^2 \quad (24)$$

$$ENT = -\sum_{i=1}^k \sum_{j=1}^k Q(i, j) * \log|Q(i, j)|$$

$$Q(i, j) = \frac{|P(i, j)|}{\sqrt{\sum_{i,j} |P(i, j)|}} \quad (25)$$

$P(i,j)$ is the (i,j) th pixel of approximation band of wavelet transformed image in specific level.

We used first and second level of wavelet transformation.

3. CASE STUDY

For evaluation of the above methods, we selected nine 128*128 size subsets from the IKONOS pan-sharpened ortho image of Pishva and its suburb area (Fig 3). Then we generated a test image of 384*384 pixels using those nine subsets. Nine land cover classes (each subset contains a specific land cover) were

defined in test image (Table 1), and two images were produced to be used as training and testing sets. Some of these classes are spectrally similar and therefore they are confused in the pixel-based classification algorithms but they display different textures.

No.	Class Name	Sample
1	Urban Area	
2	Road	
3	Cultivated 1	
4	Cultivated 2	
5	Cultivated 3	
6	Cultivated 4	
7	Tree	
8	Row Tree	
9	Dense Trees	

Table 1. Classes and their image samples



Figure 3. Selected Subset

The features as mentioned in the table (2) were generated by using different kernel sizes (3 by 3 to 25 by 25) for each image bands and also for the first PC. So there were four features for each formulation, one for each of the four original image bands. In geostatistical method cross variogram and pseudo cross variogram were produced by using the six image band pairs. For the case of first PC, cross features of geostatistical methods could not be generated because there was only one band.

Fourier based features were generated in two ways: first by using low frequencies (from center of mask with half mask size symmetrically around center for all four masks) and second by using high pass masks (the remaining mask area), but sectorial one being one type.

For GLCM and geostatistical methods, one length lags in four main directions were used and numerous features were generated. ((1,0), (0,1), (1,1) and (-1,1)).

For evaluation of different features the images were classified using the maximum likelihood classification method and different inputs as follow:

In the first stage each generated feature from the original bands together with the image bands were used. In the second stage, usefulness of a combination of spectral bands and features of one group (belonging to a similar feature group) were tested.

Statistical (First Order)	GLCM based	Geo- statistical	Fourier based	Wavelet Based
Mean (first moment)	Contrast	Direct Variogram	Horizontal	LOG Energy
Distance weighted mean	Dissimilarity	Madogram	Vertical	Shano's Index
2 nd moment	ASM	Cross Variogram	Ringing	Angular second moment
3 rd moment	Entropy	Pseudo Cross Variogram	Sectorial	Entropy
4 th moment	Homogeneity			
First central moment		Mean i		
Variance (2 nd central moment)		Mean j		
Skewness (3 rd central moment)		Variance i		
Kurtosis (4 th central moment)		Variance j		
First absolute moment		Correlation		
Second absolute moment				
Entropy				
Median				
Mode				

Table 2. List of generated features in five groups

Features generated by the use of the first PC were combined with the spectral bands and were used for classification of the images.

Independent test samples were used for evaluation of the accuracy of different classifications. Producer's accuracy, overall accuracy, variance and mean of the producer and normalized overall accuracy were considered.

4. RESULTS AND CONCLUSIONS

Results have shown that combination of textural and spectral features can improve the classification accuracy noticeably. Therefore texture features are capable of capturing and using more information in the classification process. An important advantage is that, these features are generated from the image itself and no external information and data are required. Depending on the goal of classification, e.g. extracting a specific class, increasing the overall or mean accuracy or producing an output with the acceptable minimum accuracy, appropriate features can be selected and used. It may also be argued that the ratio of mean to variance of accuracies stated as:

$$\frac{\text{Overall or Mean Accuracy}}{\text{The Variance of the Accoracies}}$$

is a better criteria for evaluation of different results than the overall accuracy. Because the overall accuracy works in favor of dominant classes and in many cases accuracy of all classes may be as important as the accuracy of dominant class.

Results of this research have shown that use of the first PC as the input for feature generation lead to results comparable with those of the four spectral bands. So, use of the first PC for generation of new features is preferable. Because its production is not so time consuming and the number of features and classification time are significantly reduced.

When the spectral bands are used, combination of all features of each group doesn't lead to better results than using each feature separately with the spectral band. This observation may be attributed to the high correlation between features and the half phenomena. When all features are used for example in the first order statistical method there are 14×4 features plus four spectral bands. But when all features of PC1 in each group are used, because of the reduced number of features (e.g. 14 for the first order statistical method), better results are obtained.

In the case of the first order features, best results are obtained by using 'Median', 'Weighted Mean' and 'Moments' respectively. But 'Central moments' and 'Absolute moments' are not as useful as others. Use of the 'Median' have lead to best results in terms of the mean accuracy so it may be regarded as the best feature in this group for enhancement of the accuracy of all classes rather than a certain class. As it is expected use of all features from the original bands together has led to poor results. As compared to others when first PC is used, 'Entropy' and using all generated features have shown the best performance.

About GLCM based features the results have shown that 'Contrast' and 'ASM' are the best ones and 'Homogeneity', 'Correlation', 'Entropy', 'Mean i' and 'Mean j', 'Variance i', 'Variance j' and 'Dissimilarity' are in next positions. But use of all features together has lead to deterioration of the results because of the high dimensionality of the feature space. To solve this problem we used the average of generated feature in four directions so the number of features decreased to $\frac{1}{4}$ and all directions significantly contributed in the generated feature space. This modification led to better results. In the case of PC1, these results are a little different and all features together also work well.

In the case of geostatistical features, experimental results demonstrated that 'Pseudo Cross Variogram', 'Madogram', 'Cross Variogram' and 'Direct Variogram' were the bests respectively.

In Fourier based features there is no explicit difference between different features and using all features together have led to good results.

In wavelet method, performance of features could be ranked as 'Angular Second Moment', 'Shanon's Index', 'Log Energy' and 'Entropy'. Second level of wavelet generally has led to better results.

When all features are compared, it could be concluded that the geostatistical and GLCM feature led to better overall accuracies. Wavelet, first order and Fourier based features are in next positions.

If results of features are compared in terms of their mean accuracy, GLCM based and geostatistical features are in higher levels and first order and Fourier methods are in next positions and wavelet occupies the last position.

As long as the mean accuracy/accuracies variance is considered, GLCM, geostatistics and first order statistics methods, have led to better results, which is an indicator of the generality of their usefulness for improvement of the classification.

Results of spectral classification and classifications of combination of textural features and spectral band which led to best men accuracy and best normalized mean accuracy have shown in Table 3.

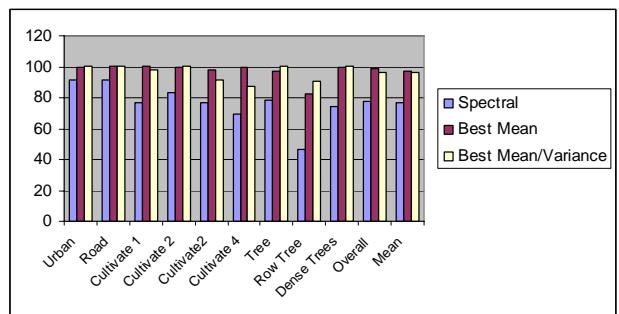


Table 3. Obtained results of spectral classification and best mean and mean/variance of accuracies obtained from combining new features with spectral bands

References

- Sergios Theodoridis, 1999. *Pattern Recognition*, Academic Press.
- Kenneth R. Castleman, 1996. *Digital Image Processing*, Prentice-Hall.
- R.M. Haralick, K. Shanmugam and I. Dinstein, 1973. Textural features for image classification, *IEEE Transactions on Systems, Man and Cybernetics*, 3(6), pp 610-621.
- <http://www.geo.ed.ac.uk>
- M. Chica-Olmo and F. Abarca-Hernandez, 2000. Computing geostatistical image texture for remotely sensed data classification, *Computers & Geosciences* 26, pp. 373-383.
- Pratt, 2001. *Digital Image Processing*, John Wiley & Sons.
- Soe Win Myint, Nina S.-N. Lam, and John M.Tyler, Wavelets for Spatial Feature Discrimination: Comparisons with Fractal, Spatial Autocorrelation, and Spatial Co-Occurrence Approaches, *Photogrammetric Engineering & Remote Sensing*, 70(7), pp. 803-812.