Texture Fusion and Feature Selection Applied to SAR Imagery

Anne H. Schistad Solberg and Anil K. Jain

Abstract— The discrimination ability of four different methods for texture computation in ERS SAR imagery is examined and compared. Feature selection methodology and discriminant analysis are applied to find the optimal combination of texture features. By combining features derived from different texture models, the classification accuracy increased significantly.

I. INTRODUCTION

In the interpretation of synthetic aperture radar (SAR) images, texture [16] provides important information, in addition to image gray levels or the backscatter values alone. Several studies have shown that classification based on texture features can improve the accuracy of the interpretation [1], [2], [12], [17]. A large number of approaches for computation of image texture has been proposed in the literature. Some studies have compared the performance of different methods [2], [18] for SAR image analysis, but we are not aware of any study which has combined texture features derived from different models and then applied feature selection to find the optimal feature combination for a given classification task.

In this paper, we investigate the performance of texture features derived from the gray-level co-occurrence matrix [6], features based on local statistics, fractal features [11], and features derived from a lognormal random field model [4] for land-use classification of ERS-1 SAR images. The discrimination ability of the different methods for texture computation is examined and compared. Feature selection methodology and discriminant analysis are applied to find the optimal combination of texture features. All experiments are conducted on several SAR images to be able to provide generalizations of the results.

The remainder of the paper is organized as follows. In Section II the methods used for texture computation and evaluation are described. The experimental data and results are given in Section III. Finally, a discussion of the results and conclusions are provided in Section IV.

II. TEXTURE FEATURES AND CLASSIFICATION METHODOLOGY

A brief description of the selected texture features is given below. The classification methodology and the principles of feature transformation in terms of supervised discriminant analysis are then described. The input to the texture models will be an $I$-by-$I$ image $f(x, y)$, where $f(x, y)$ denotes the gray level of pixel $(x, y)$.

A. Features Derived from Gray-Level Co-Occurrence Matrices

The co-occurrence texture features are based on gray-level spatial dependencies [6]. A co-occurrence matrix, computed in a local window, contains the relative frequencies of all pairwise combinations of backscatter values at a certain distance $d$ and direction $\alpha$ within the window. From this matrix, a number of features can be computed.

With $N_g$ gray levels in the image, the dimension of the matrix will be $N_g \times N_g$. The $(i, j)$th element, $P_{ij}$, of this matrix is defined by

$$P_{ij} = \frac{p_{ij}^{d,\alpha}}{\sum_{i,j} p_{ij}^{d,\alpha}}$$

(1)

where $p_{ij}^{d,\alpha}$ is the frequency of occurrence of gray levels $i$ and $j$, separated by a distance $d$ and direction $\alpha = 0^\circ, 45^\circ, 90^\circ$ and $135^\circ$. The summation is over the total number of pixel pairs $L$, given $d$, in the window.

We compute the following texture features from the co-occurrence matrix: angular second moment (ASM), contrast (CONT), entropy (ENT), cluster shade (CLSH), inertia (INER), and inverse difference moment (IDM). These features are among the most commonly used co-occurrence features.

Prior to the computation of the co-occurrence matrix, the number of gray levels in the image needs to be reduced to a small number to get reliable estimates of the relative frequencies when texture is computed in a small window ($9 \times 9$ pixels). In this study, histogram equalization is first applied to the images to produce an image with gray levels which span the entire 8-bit range of pixel values, followed by a quantization into $N_g = 8$ gray levels.

B. Features Computed from Local Statistics

Based on local statistics, the following features are computed:

- power-to-mean ratio, $\text{PMR} = \sigma/\mu$, where $\sigma$ is the local standard deviation and $\mu$ is the local mean of the backscatter values;
- skewness, $\text{SKW} = \frac{E[(f(x,y) - \mu)^3]}{\sigma^3}$, where $E[]$ denotes the expectation operator;
- kurtosis, $\text{KUR} = \frac{E[(f(x,y) - \mu)^4]}{\sigma^4}$;
- contrast [15], $\text{CNT} = \frac{1}{W^2} \sum_{(i,j) \in \mathcal{N}} (f(x,y) - f(i,j))^2/\mu^2$, where $W$ is the number of pixels in a local neighborhood $\mathcal{N}$ around the pixel of interest $(x, y)$;
- homogeneity [15], $\text{HOM} = \frac{1}{W^2} \sum_{(i,j) \in \mathcal{N}} \frac{1}{1 + [(f(x,y) - f(i,j))/\mu]^2}$.

C. Fractal Features

Many natural textured surfaces can be described as a fractal surface. The key parameter describing a fractal surface is its fractal dimension. Several methods for computation of the fractal dimension of a surface exist. The most common methods are the variation method [3], the $\epsilon$-blanket method [13], and the box-counting method [11]. In previous comparative studies [8] of the performance of these three methods, the variation method was found to have the smallest variance, indicating the best stability. Based on this result, we will use the variation method for computation of the fractal dimension.

The variation method is based on the observation that the variation of a fractal surface within small boxes as a function of the box size produces good estimates of the fractal dimension [3]. The method computes the maximum variation, $V(\epsilon)$, of the image surface $f(x, y)$ in a square of side $\epsilon$. The fractal dimension $D$ is estimated as $3$ minus the slope of the least square fit of the data $\{\ln(\epsilon), \ln(V(\epsilon))\}$ as $\epsilon$ varies.

Several studies have shown that the fractal dimension alone does not capture all the textural properties (see, e.g., [11]). Another measure, called lacunarity, is often used in addition to the fractal dimension. The lacunarity value is small when texture is fine and it
is large when the texture is coarse. An estimate of lacunarity can be derived from the box counting algorithm [11].

D. Parameters of Lognormal Field Models

Following Frankot and Chellappa [4], we will model the SAR image using a multiplicative autoregressive random field (MAR). The parameters of the model will be used as texture descriptors.

Let the observed image \( f(x, y) \) be represented by a white-noise-driven multiplicative system, where \( g(x, y) = \ln f(x, y) \) follows a Gaussian autoregressive (AR) model

\[
g(x, y) = \sum_{r \in N} \theta_{r} g(x + r, y + s) - \mu_{g} + u(x, y).
\]

\( N \) is the neighborhood system, and \( \mu_{g} \) is the mean value of the logarithmic image \( g \). The noise process \( u(x, y) \) is an uncorrelated white noise with variance \( \sigma_{u}^{2} \).

We will use the least squares estimates for the parameters \( \theta_{r}, \sigma^{2}, \) and \( \mu_{g} \) [4]. The parameter estimates are computed in windows of size \( 9 \times 9 \) with \( N = \{(0, -1), (-1, -1), (-1, 0)\} \).

For land-use classification in Norway, the average size of typical regions is rather small. Thus, we need to limit the window size used in the texture computations. In this study, a window of size \( 9 \times 9 \) pixels will be used. Shork [14] compared the effect of varying the window sizes from \( 5 \times 5 \) pixels to \( 9 \times 9 \) pixels and found no significant differences in GLCM texture features. However, if the window size is increased to, e.g., \( 30 \times 30 \) pixels or larger, fractal features and/or frequency-based features might have better performance.

E. Classification Procedure

To test the performance of the texture features with respect to their discrimination ability, a classifier must be chosen. We can use either a nonparametric classifier, like the \( K \)-Nearest Neighbor classifier, which does not assume any particular statistical distribution of the texture features, or a parametric classifier based on a statistical model. To be able to perform a large number of classification experiments in a reasonable amount of time, we chose a simple quadratic classifier based on the multivariate normal distribution. Finding the best classifier for SAR texture feature vectors is a topic of future research and will not be considered in this paper. Assuming that class \( k \) has a probability distribution \( p_{k}(\mathcal{F}(x, y)) \), where \( \mathcal{F}(x, y) \) is the vector of texture features for pixel \( (x, y) \), the classification rule is defined by

\[
\text{Allocate pixel } (x, y) \text{ to class } k \text{ if } p_{k}(\mathcal{F}(x, y)) = \max_{l=1,\ldots,K} p_{l}(\mathcal{F}(x, y)),
\]

where

\[
p_{l}(\mathcal{F}(x, y)) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathcal{F}(x, y) - \mu_l)^T \Sigma_l^{-1}(\mathcal{F}(x, y) - \mu_l)}
\]

for class \( l \). The error rates reported for the classifiers are computed by the leave-one-out method using all the available ground-truth data.

F. Feature Transformations by Discriminant Analysis

Discriminant analysis is a well-known approach for dimensionality reduction and multivariate data projection [5]. The input features are projected into a space having fewer dimensions. The class labels are large when the within-class scatter is small. Thus, we need to limit the window size used in the texture computations. In this study, a window of size \( 9 \times 9 \) pixels will be used. Shork [14] compared the effect of varying the window sizes from \( 5 \times 5 \) pixels to \( 9 \times 9 \) pixels and found no significant differences in GLCM texture features. However, if the window size is increased to, e.g., \( 30 \times 30 \) pixels or larger, fractal features and/or frequency-based features might have better performance.

where the transform matrix \( \mathcal{H}_{k} \), with dimension \( (K-1) \times d \), is defined in terms of the within-class \( (S_{W}^{-1}) \) and the between-class \( (S_{B}^{-1}) \) scatter matrices. For the definition of the scatter matrices, see, e.g., [10]. This transformation has the important property that the patterns are projected from \( d \) dimensions to \( (K-1) \) dimensions while maintaining the scatter ratio to be constant. The rows of \( \mathcal{H}_{k} \) are the eigenvectors corresponding to the \( (K-1) \) nonzero eigenvalues of \( S_{W}^{-1} S_{B} \), and the new feature vector \( \mathcal{G}(x, y) \) has \( (K-1) \) components. To get a projection into a \( b \)-dimensional space (\( b \leq K-1 \)), the \( b \) eigenvectors corresponding to the \( b \) largest eigenvalues are chosen.

G. Feature Selection

To find the best subset of texture features from the set of \( d \) available features, Whitney’s algorithm for feature selection will be used [19]. The algorithm works in the following manner. First, the feature with the smallest probability of misclassification is selected. Then the feature which, in combination with the previously selected features, gives the smallest probability of misclassification, is selected, and so on. This feature selection process is suboptimal, because it does not perform an exhaustive evaluation of all the possible feature subsets. Whitney used a \( K \)-NN classifier with the leave-one-out method for error estimation. We have modified this algorithm to use a quadratic classifier and the leave-one-out method for error estimation.

III. EXPERIMENTAL RESULTS

A data set consisting of four ERS-1 SAR images from the fall of 1991 of Kjeller, Norway was used. The images were captured under different weather conditions, so the SAR signature varies greatly from one image to another. The images were captured at different dates, but from exactly the same path of the satellite track. Thus, they covered exactly the same area, and no co-registration or geometric correction was necessary. Fig. 1 shows a small part of the scene for all the four images. Large variations in SAR signature for the same ground-cover class were observed with different wind conditions (for water areas), temperature, and soil moisture content (for forest and agricultural areas).

A five-class classification problem was considered (\( K = 5 \)), consisting of the following common ground cover classes: water, urban areas, forests, and two classes of agricultural fields (plowed and unplowed). The data set was acquired in a project for monitoring soil erosion due to runoff from agricultural fields tilled in the fall. Between the acquisition of the SAR images, a large number of the agricultural fields has been tilled. Ground control samples were available from each acquisition date. The size of the training areas was typically 4000 pixels per class for each image. For agricultural areas, the training set varied from 400 pixels to 12000 pixels as the plowing season progressed.

A. Selection of Texture Features With Optimal Discrimination Ability

In order to determine the discrimination ability of different subsets of texture features, we addressed the following questions: a) Which texture feature extraction model performs the best when applied to different images? b) When fusing the features computed from different texture models, which feature combination is optimal?, and c) Can texture fusion improve the classification accuracy?

A comparison between the classification performance of the GLCM features, the local-statistics features, and features from the lognormal random field model is provided in Table I. On an average, the error rate was 24.1% for classification based on lognormal-field features, compared to 29.2% for GLCM features, and 38.6% for features from local statistics. Classification accuracy based on a speckle-reduced backscatter image, with no texture features, was 30.5% on an average.
Fig. 1. ERS-1 SAR subimages from October 17, October 23, November 22, and November 28, 1991 (from left to right).

TABLE I
A COMPARISON OF THE PERFORMANCE OF DIFFERENT METHODS FOR TEXTURE FEATURE EXTRACTION FOR SAR IMAGES

<table>
<thead>
<tr>
<th>Acquisition date</th>
<th>Mean only</th>
<th>GLCM features</th>
<th>Local statistics</th>
<th>Lognormal fields</th>
</tr>
</thead>
<tbody>
<tr>
<td>October 17</td>
<td>43.8</td>
<td>36.5</td>
<td>46.5</td>
<td>32.2</td>
</tr>
<tr>
<td>October 23</td>
<td>33.3</td>
<td>31.0</td>
<td>44.9</td>
<td>26.7</td>
</tr>
<tr>
<td>November 22</td>
<td>20.8</td>
<td>22.5</td>
<td>31.6</td>
<td>17.7</td>
</tr>
<tr>
<td>November 28</td>
<td>23.9</td>
<td>26.6</td>
<td>31.3</td>
<td>19.8</td>
</tr>
<tr>
<td>Average performance</td>
<td>30.5</td>
<td>29.2</td>
<td>38.6</td>
<td>24.1</td>
</tr>
</tbody>
</table>

In all the experiments, the features derived from the lognormal field model performed significantly better than the other texture features. The classification accuracy was always significantly improved by using the lognormal field parameters compared to using the speckle-filtered mean backscatter value alone. On the other hand, classification based on features from local statistics only always performed worse than classification based on only the mean backscatter value.

To see if the classification performance could be further improved by combining the features derived from the co-occurrence matrix, local statistics, fractal models, and lognormal field models, a feature selection experiment was performed. The six GLCM features (angular second moment, cluster shade, contrast, entropy, inertia, and inverse difference moment), the five features from local statistics (power-to-mean ratio, skewness, kurtosis, homogeneity, and contrast), the five parameters of the lognormal field model (\(\theta_1, \theta_2, \theta_3, \sigma^2, \mu_x\)), the two fractal features (lacunarity and fractal dimension), and the local mean value were merged into one large feature vector, containing a total of 19 features. A feature selection was performed on this set of 19 features, using Whitney’s method for feature selection [19]. As mentioned earlier, we replaced the K-NN classifier in the original Whitney’s algorithm by a quadratic classifier.

In our experiments the “best” subset with \(b\) features out of the total of 19 features was found for \(b = 1, \ldots, 19\). By the best subset of features we mean the feature subset which gives the lowest classification error as determined by Whitney’s modified method. Fig. 2 shows the classification error rate as a function of the number of texture features included in the classification for the original texture features and for features transformed based on discriminant analysis as described in Section II-F. The results, after applying a feature transform as defined by discriminant analysis, were not significantly different from classification based on the features selected by Whitney’s method using the same number of features. We also note that the lowest classification error was achieved with approximately ten features. Increasing the number of features beyond a certain point resulted in a higher error rate (this phenomenon is called the “curse of dimensionality.”) [9].

Table II explains the feature names, while Table III lists the ten best features using Whitney’s selection scheme for different ERS-1 SAR images. The features are listed in the order in which they were selected. Among the ten best features, we find reasonably consistent results for different ERS-1 SAR images, though with different individual rankings. Among the frequently selected features are the mean value, the lognormal parameters \(\theta_1, \theta_2, \theta_3,\) and \(\sigma^2\), the GLCM features Inertia and Cluster Shade, the statistical features power-to-mean ratio and kurtosis, and the fractal lacunarity. Fusion of the texture features also reduced the classification error rate from 24.1% using only the lognormal features to 17.1% on an average.

IV. DISCUSSION AND CONCLUSION

In this paper, we have investigated the performance of four different and commonly used methods for texture computation. The following methods for texture computations were included in this comparative study: a) gray-level co-occurrence matrices, b) local image statistics, c) lognormal field models, and d) fractal features.

Feature selection was applied to the pooled set of texture features to find the optimal subset of features. Combining features derived from different models significantly improved the classification accuracy by 2% compared to using features from the lognormal field model alone. A feature transformation defined by discriminant analysis produced similar classification accuracies as feature selection using Whitney’s modified method.

By applying feature selection to various SAR images, reasonably consistent classification results were found, although with different individual rankings of the selected features. The images used for this evaluation showed quite a large variation in backscatter values. Although the texture features were not invariant to natural SAR signature variations, the same features were selected in the best subset
Fig. 2. Classification error rate versus the number of features used. Original features: —; Discriminant analysis for a five-class problem gives four new features whose performance is marked as ‘+.’ The total number of features considered for each image varies because features with a high correlation were excluded to avoid singular covariance matrices in the classification.

TABLE III

<table>
<thead>
<tr>
<th>Image</th>
<th>Best features</th>
<th>Error rate (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>October 17</td>
<td>CLSH, SKEW, $\mu$, $\theta_1$, INER, LAC, $\theta_2$, PMR, SKEW, CONT</td>
<td>25.8</td>
</tr>
<tr>
<td>October 23</td>
<td>CLSH, SKEW, KURT, $\theta_1$, $\theta_2$, Mean, $\sigma^2$, LAC, $\theta_3$, $\sigma^2$, IDM</td>
<td>19.9</td>
</tr>
<tr>
<td>November 22</td>
<td>Mean, PMR, INER, $\theta_1$, $\theta_2$, CLSH, LAC, $\theta_3$, KURT, $\sigma^2$</td>
<td>14.4</td>
</tr>
<tr>
<td>November 28</td>
<td>Mean, $\theta_1$, $\theta_2$, $\theta_3$, $\sigma^2$, INER, CLSH, PMR, KURT, LAC</td>
<td>11.2</td>
</tr>
</tbody>
</table>

of features for each image. This robustness in feature selection is a very important property of an operational SAR image analysis system. To optimize the performance of applications involving land-use classification of SAR images, it might be wise to use images taken at certain times of the year and certain weather conditions.

Combining texture features computed using different models is an important topic for further research. Additional approaches to linear or nonlinear feature transformation based on, e.g., flexible discriminant analysis [7] or neural nets should be investigated.

Conclusion The most significant result of this study is that texture fusion and selection by combining texture features obtained by different models significantly improve the classification accuracy.

ACKNOWLEDGMENT

The authors would like to thank T. Taxt for his useful comments, and the Norwegian Defense Research Establishment, the State Pollution Control Authority (SFT), and the Norwegian Space Center for providing the ERS-1 SAR images.

REFERENCES

Evidence of Hot Spot Directional Signature from Airborne POLDER Measurements

F.-M. Bréon, V. Vanderbilt, M. Leroy, P. Bicheron, C. L. Walthall, and J. E. Kalshoven

Abstract—The POLDER instrument was flown during the BOREAS experiment over various sites and at various altitudes in the Canadian boreal forest and other nearby targets. The instrument design permits the acquisition of the directional signature of any surface cover. In particular, the high directional resolution of POLDER allows it to measure, with an unprecedented accuracy, the hot spot signature of natural targets. In this paper we present some typical examples of such highly anisotropic reflectance directional signatures. The ratio of the maximum reflectance (hot spot direction) to the minimum reflectance (broad area in the forward scattering hemisphere) varies with wavelength and canopy. It can be as large as six in the visible and three in the near IR.

I. INTRODUCTION

Directional effects on the reflectance from natural surfaces have received a great deal of interest in the past few years. This results in part from the fact that the use of wide field-of-view instruments for Earth monitoring yields large variations in the observation geometry. Because of these geometry variations, care must be exercised when comparing reflectance measurements acquired day after day. On the other hand, it is believed that directional reflectance measurements can be inverted to retrieve key parameters of the surface such as the leaf area index, the leaf orientation distribution, or the soil roughness. A key feature of the reflectance directional signature is the “hot spot” effect [1]–[5]. When the viewing direction coincides, i.e., is colinear, with the solar direction, all elements in the instantaneous field of view are illuminated by the direct solar beam—none are shadowed—which yields a local peak in reflectance. The backscattered reflectance gradient close to this particular direction is called the hot spot. Several models show that this gradient depends on the ratio of the horizontal and vertical scales of the canopy, or to the relative size and density of the leaves [6]–[10]. Thus, in principle, the hot spot reflectance gradient could be inverted for retrieving the relative size and spacing of elements making up the canopy [9], [11]. Moreover, the exact hot spot direction is of particular interest for measuring the leaf reflectance thanks to the lower dependence on the canopy structure [1].

However, the hot spot effect is difficult to measure in nature because of the radiometer’s own shadow [11]–[13]. In order to acquire meaningful measurements of the hot spot directional signature, it is necessary to have an instrument whose size is small compared to the distance to the target. Also, the instrument field of view must be larger than the typical scale of the target. On the other hand, because the hot spot directional extension is typically small, a high directional resolution is necessary. Thus, in practice, airborne measurements shall be preferred for meaningful hot spot measurements.

The POLDER instrument [14], which is described below, complies with the requirements expressed above. Thanks to its limited weight and size, it can be easily put on a small aircraft or helicopter.

Another problem in precise measurement of the hot spot signatures is the atmospheric effect on the reflectances. Atmospheric aerosols, as well as molecules, scatter solar light before it reaches the surface. The aerosol scattering phase function shows an intense peak in the forward direction. Most of the radiances scattered by aerosols reaches the surface with a change of the incoming direction of a few degrees. Thus, aerosol scattering yields a smoothing of the measured directional signature, which may greatly flatten the hot spot measurement. A very clear atmosphere is necessary to measure, even from low altitude, a clean hot spot signature. During the BOREAS experiment [15], [16], extremely clear atmospheric conditions were frequent with aerosol optical thickness at 500 nm on the order of 0.04. This optical thickness includes the remaining stratospheric aerosols from the Pinatubo explosion. Thus, the data which have been acquired are meaningful for hot spot analysis.

Other directional reflectance measurements have been acquired over forest canopies. However, this was either with a photographic method without a quantitative analysis [3], [17], or with an angular resolution which is not sufficient for hot spot studies [2], [18]–[24].

In this communication, we show that the POLDER instrument is very well adapted to the hot spot characterization, and we analyze a few such directional signatures. The measurements allow a quantification of the reflectance increase resulting from the hot spot effect.

II. THE POLDER INSTRUMENT

The POLDER instrument has been described elsewhere in the literature [14], [25]. It basically consists of a CCD camera with a very wide field of view lens. Thus, one two-dimensional (2-D) image corresponds to a large domain of viewing angles. The pixel size on the ground is proportional to the instrument altitude. It is 6 × 6 m² for an altitude of 1 km, which is equivalent to an angular resolution better...