AN INTEGRATION OF WAVELET ANALYSIS AND NEURAL NETWORKS IN SYNTHETIC APERTURE RADAR IMAGE CLASSIFICATION*

Qiming Qin [a] , Robert R. Gillies [b] , Rongjian Lu [c] , Shan Chen [c]

[a] Institute of Remote Sensing and GIS, Peking University, China, 100871 qmqin@pku. edu.cn
[b] Department of Geography and Earth Resources, Department of Plants, Soils Biometeorology, Utah State University, Logan, UT 84322-4820 USA，voice: (435) 797 2664 email: rgillies@cnr.usu.edu
[c] Institute of Remote Sensing and GIS, Peking University, China, 100871

KEY WORDS: Image Classification, Wavelet Analysis, Neural Networks, Texture and Structural Characteristics, SAR

ABSTRACT:

In this paper, the concepts of wavelet analysis and neural networks are applied to the classification of shuttle imaging radar experiment C (SIR-C) synthetic aperture radar (SAR) data from a location in northwest China. Initially, the paper presents the visual elements of tone, texture and structural features on SAR imagery as important bases for image classification and target recognition. The wavelet analysis is used as a method to extract elements of texture and structural features; it involves deriving the energy of sub-image blocks through wavelet decomposition. A improved backpropagation neural network was applied to a multiresolution representation of six images comprising reflectance SAR data and those obtained by the wavelet transform. A simple scene was classified, yielding poplar trees and bushes. Where they were well differentiated the probability of yielding the correct classification was found to be 100%. Erroneous classification occurred in transition areas between cover types where the percentage of correct classification fell slightly. The results suggest that such an integrated approach to classification is applicable for SAR data that involves regular textures and structures with rather strong orientation of land features.

1. INTRODUCTION

Maximum the often quoted advantage of radar, i.e. being unaffected by cloud cover, is paramount to its use in certain regions of the world. Radar images are however characterized by distinctive properties that are in part present due to an image formation process that is quite different from that of conventional optical images, i.e. SPOT or Landsat TM imagery. In principle, many of the conventional algorithms that are applied in the analysis of multispectral data can be applied in the analysis of Synthetic Aperture Radar (SAR) data (Simard et al. 2000). There are no theoretical limitations to the number of features (or bands) used by any of these algorithms. On the other hand, due to the nature of SAR data, transformation tools are often required to provide collateral information to assist in the process of image classification.

Each homogeneous region of a SAR image contains certain characteristics that are important bases for target recognition and image classification. The significance of certain image interpretation elements is particularly useful to establish coherent information set that permits a robust classification of a SAR image. This study utilizes three elements for SAR image interpretation - namely tone, texture and pattern (i.e. combining elements of structure, orientation or direction).

Tone refers to the relative brightness of the pixel elements and represents a qualitative measure of microwave backscatter strength. Tone on SAR imagery mainly relies upon the backwards scattering character of the terrain object. For natural objects, rough surfaces such as mountains and agricultural fields produce powerful backscattered returns and in doing so, form a variety of textural and structural features. Smooth surfaces such as calm water and flat land surfaces act as specular reflectors so that most of the energy is reflected away from the imaging SAR. Differentiating between such specular reflectors will generally depend more upon recognizing pattern.

Texture from a SAR image is spatial information of the image tone variety repeated with a certain rule. That is to say it is the arrangement of tone and is manifested by an arrangement of variation in brightness. This variation in brightness has an important intrinsic property as a frequency of tonal change and it is this that is particularly useful in the discrimination of different areas of SAR illuminated areas.

Pattern presented on a SAR image is the composition formed regularly by some spatial characters of target objects in an area. Such spatial arrangements of objects on the ground may be systematic or random. They may exhibit structure in space, e.g., they may lie north – south and be parallel or form more intricate patterns. Nevertheless orientation and direction, whether part of a systematic pattern or not, can serve to varying degrees as a useful basis for the interpretation of SAR images.

The combination of such visual elements as a means to classification is by no means straightforward. For example, man-made target objects (i.e. urban) generally have regular geometrical character. However, the tones and textural features in the SAR imagery will vary along with antenna look direction. Bryan (Bryan 1979) found that different orientation angles formed between cultural targets and antenna look direction could produce dramatic differences in image gray tone: streets paralleling with radar track direction took on light lines tones, because the buildings alongside the streets play a major role in determining echo strength. On the other hand, the streets across track direction appeared as dark tone lines or had no presentation on the SAR image.

Wavelet transforms have been suggested as an analysis tool for the analysis of SAR images (Barbarossa and Parodi 1995). In such a framework the wavelet approach maps high frequency features. In this paper, a wavelet approach is proposed that detects and describes texture and pattern (hereafter referred to as structural features) in a simple SAR image; simple in the sense that the land cover classified only contains two types of vegetation. A neural network is then used to classify the image based upon six components as measured directly from the SAR and those derived by wavelet decomposition.

2. SAR DATA AND SITE DESCRIPTION

SAR data was obtained from the shuttle imaging radar experiment C (SIR-C). This SAR has multi-frequency and multi-polarization capabilities. Available frequencies were X (3.0 cm), C (5.8 cm) and, L (23.5 cm) and obtainable polarization combinations were HH, HV, VV and VH with 17 x 25m pixel resolution. The SAR image in Figure 1 is taken over the settlement of Pishan in Pei Shan county located in the southwest of Xin Jiang, northwest China. It is juxtaposed between the southern edge of the Takilimakan desert and the northern extent of the Kulun mountain range.
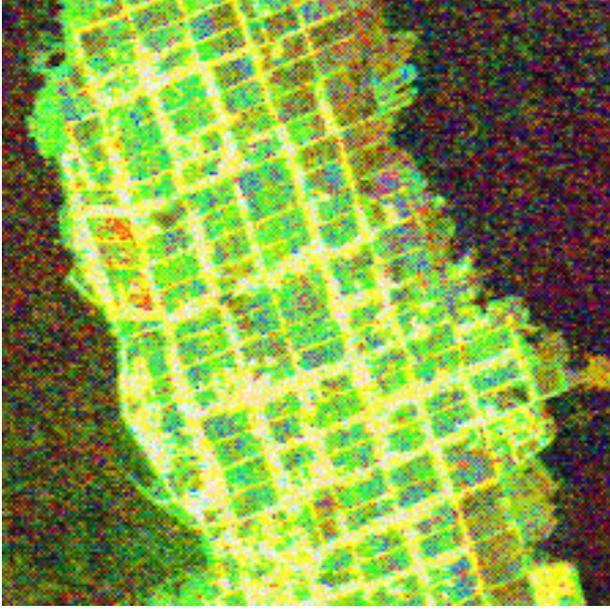

Figure 1.   A community of poplar trees ('white' lattice) interspersed by bushes (varying colours from cyan to turquoise) and surrounding background located in Pishan in Pei Shan county in the south-west of Xin Jiang, north-west China.

The original image was subset to a smaller one with dimensions of 224 by 224 pixels with three layers of polarization combinations L-HH, L-HV and C-HV. Figure 1 therefore shows the radar image as composed of three microwave wavelength data, the colours based on different frequencies and polarizations. Image display colour were assigned as follows: red is the L-band horizontally transmitted, horizontally received, L-HH(R); green is the L-band horizontally transmitted, vertically received, L-HV(G); and blue is the C-band horizontally transmitted and vertically received, C-HV(B). A visual interpretation of the image is indicative of a grid. Indeed

a grid of poplar trees with bushes in-between supplanted the original ground cover on the existing alluvial fans. This pattern and alignment was constructed so as to provide windbreaks for the settlements downwind. The ground coverage is therefore fairly simple and internally homogeneous.

3. FEATURE EXTRACTION AND CLASSIFICATION METHODOLOGIES

The 2D Wavelet Transform
The discussion begins with the definition of the orthonormal wavelet (Chui 1992): let $L^2(0, 2\pi)$ represent the set of all measurable functions defined from $(0, 2\pi)$ that satisfy $\int_0^{2\pi} |f(x)|^2 dx < \infty$. It is assumed the functions in $L^2(0, 2\pi)$ are expanded periodically into the real line IR = $(-\infty, +\infty)$, that is: $f(x) = f(x - 2\pi)$ are satisfied at each x. Let Ψ have unit length, then

$$\Psi_{j,k}(x) = 2^{j/2}\Psi(2^j x - k) \tag{1}$$

Equation 1 is referred to as the orthonormal wavelet. Where $\{\Psi_{j,k}\}$ is the canonical orthonormal basis of $L^2(IR)$, that is $\langle \Psi_{j,k}, \Psi_{l,m} \rangle = \delta_{j,l} \bullet \delta_{k,m}$ and $\forall f \in L^2(IR)$, so that,

$f(x) = \sum_{j,k}^{\infty} C_{j,k}\Psi_{j,k}(x)$ in which:

$$\lim_{M1,N1,M2,N2\to\infty} = \left\| f - \sum_{j=-M2}^{N2} \sum_{k=-M1}^{N1} C_{j,k}\Psi_{j,k} \right\|^2 = 0$$

The orthonormal wavelet whose rank is j has a degenerate matrix from rank 0 to rank j. This nature in the wavelet $\Psi(x)$ has the advantage of an improved ability in edge detection in digital images – in this case SAR. At the same time, $\Psi(x)$ can also detect the peak signal at multi-resolution.

The orthonormal wavelet whose rank is j has a degenerate matrix from rank 0 to rank j. This nature in the wavelet $\Psi(x)$ has the advantage of an improved ability in edge detection in digital images – in this case SAR. At the same time, $\Psi(x)$ can also detect the peak signal at multi-resolution.

There are a number of ways to accomplish wavelet decomposition of a 2-dimensional (2D) digital image. For the purposes of this research the Stéphanne Mallat pyramid algorithm (Mallat 1989) was adopted as follows: let H represent the H_n operator (i.e. high pass filter), G represent the G_n operator (i.e. low pass filter), subscripts r and c represent row and column respectively and j is defined as before, then:

$$C^{j+1} = H_c H_r C^j \tag{2}$$

$$d^{j+1,1} = G_c H_r C^j \tag{3}$$

$$d^{j+1,2} = H_c G_r C^j \qquad (4)$$

$$d^{j+1,3} = G_c G_r C^j \qquad (5)$$

In practice, in the process of texture and structural feature extraction in a digital image, the operator $G_c H_r$ in equation 3 smoothes the column vector and finds such differences as exist between the objects in the rows. Likewise, $d^{j+1,1}$ can detect the change in the edge of objects in the horizontal direction and, the operator in equation 4 can detect the change in the edge of objects in the vertical direction, while the operator in equation 5 can detect any change in the diagonal direction.

To extract texture and structural features of ground objects in an image with *S. Mallat*'s pyramid decomposing algorithm, a wavelet base is selected along with the number of levels (N) to be decomposed. In this particular case, Daubechies's orthonormal wavelet (Daubechies 1998) was adopted as the wavelet base. The reasons for this were as follows:

    (a)   Orthonormal, which has direct ratio to the size of the support set (2N).

    (b)   Continuous degenerate matrix.

    (c)   The smoothness and large degenerate matrix means that it is better at differentiating frequencies. At the same time, similar to a low-pass filter, it can keep the low frequency component of the original image without obvious blur effects.

From the point of view of filtering, it is preferable to maximize N, but in image decomposing, this will result in more boundary effects -- the higher the decomposing level, the greater the boundary effects on the image. Moreover, computation time is expanded as N^2. Through experimentation, N equal to three was selected where the results were better.

The application of a two dimensional Discrete Wavelet Transform (DWT) expands the image I into a sum of four components at N resolution levels. In essence, the wavelet transform operation is separable and consists of two one dimensional operations along the rows and the columns of I:

    i.   From the *first* row to the mth row in I, a 1D DWT is performed to generate $I \otimes H_r$ and $I \otimes G_r$.

    ii.   From the *first* column to the mth column in $I \otimes H_r$ and $I \otimes G_r$ a 1D DWT is performed to generate four components $\{I \otimes (G_r, H_c), I \otimes (H_r, G_c), I \otimes (G_r, G_c)\}$ and $I \otimes (H_r, H_c)$.

The original image is divided into four components after 2D DWT. These comprise one low frequency component $I(1)_0 = \{I \otimes (H_r, H_c)\}$ and three high frequency components:

$I(1)_1 = \{I \otimes (G_r, H_c), I \otimes (H_r, G_c), I \otimes (G_r, G_c)\}$.

The $I \otimes (G_r, H_c)$ component contains horizontal edge information, the $I \otimes (H_r, G_c)$ contains vertical edge information, and the $I \otimes (G_r, G_c)$ contains diagonal edge information. Steps 1 and 2 are repeated with the low frequency image $I(1)_0$, to produce four components at each new level. So, N level pyramid decomposing will result in *3N+1* components.

At each level, there is one low frequency component and three high frequency components. The three high frequency components contain textural and structural information in the horizontal, vertical and diagonal directions respectively at each level of decomposition. Recognition features were constructed from the components as follows: the original digital image was decomposed into one low frequency component (E1) and three high frequency components denoted as E2, E3 and E4. After decomposing the low frequency component at level 1 another set of three high frequency components were created at level 2, respectively denoted as E5, E6 and E7 along with one low frequency component (now E1). Finally, DWT was performed to E1 at level 2 generating four components at level 3 (E1, and three high frequency components respectively denoted as E8, E9 and E10). Making use of the 9 high frequency components, recognition features in the diagonal, vertical and horizontal were integrated as follows.

$$\lambda 1 = E4/(E2 + E3) \qquad (6)$$

$$\lambda 2 = E7/(E5 + E6) \qquad (7)$$

$$\lambda 3 = E10/(E8 + E9) \qquad (8)$$

where λ1, λ2 and λ3 represent, at each decomposing level, the ratio of the energy of the edge of ground objects in the diagonal direction to the sum of the energy of the edge of ground objects in the horizontal and vertical at that scale. These three recognition features are invariable throughout orientations of the image through 90, 180, and 270 degrees.

The **Improved** Backpropagation Neural Network Classifier

An Artificial Neural Network (ANN), also referred to as a Neural Network is 'an interconnected assembly of simple processing elements, units or nodes, whose functionality is loosely based on the animal neuron. The processing ability of the network is stored in the inter-unit connection strengths, or weights, obtained by a process of adaptation to, or learning from, a set of training patterns' (Gurney 1997). Neural Networks are often used for cluster analysis and image classification.

ANN models include Back Propagation, Counter Propagation, Hopfield Nets, Adaptive Resonance Theory (ART) nets, Kohonen Self-Organization Feature Maps (SOFM) etc. In this study, the Feedforward multi-layer network based Back Propagation model (BP) was adopted. The BP model is applicable to a wide class of problems (Paola and Schowengerdt 1995). In the BP model, the training algorithm to be developed is based on Back Propagation (Rumelhart et al. 1986) in which the signaling errors go backwards from output to input nodes through nets. In the training process each iteration is divided into two stages after the image data are input to the input layer. The outline of the BP algorithm consists of the following steps:

(a) Initiate BP Nets. The weight matrixes (W_{ji}, W_{kj}) and the threshold values (,) are initialized as randomized real numbers within the range -1.0 to +1.0 where i represents i^{th} node of input layer, j represents j^{th} node of hidden layer, and k indicates k^{th} node of output. W_{ji} is the weight matrix between input layer and hidden layer, W_{kj} is the weight matrix between hidden layer and output layer.

(b) Input the values of the training pixels (samples) and the target value for correct output.

(c) Calculate the value at each network neurons using equation 9:

$$net_j = \sum W_{ji} O_i \qquad (9)$$

where net_j is the input value of jth hidden neuron, and O_i is the input value of ith input neuron. W_{ji} indicates the weight between the ith neuron of the input layer and the jth neuron of the hidden layer. The output value of the hidden neuron is evaluated as:

$$O_j = f(net_j) \qquad (10)$$

where O_i is the output value of the jth hidden neuron and f is the Sigmoid activation function. The following function specifies f:

$$f(x) = \cfrac{1}{1 + exp\left[-\cfrac{(x + \mu)}{\mu_0}\right]} \qquad (11)$$

where μ is a threshold vector and μ_0 is used to adjust the shape of the Sigmoid activation function.

(d) Calculate the value of output neuron:

$$net_k = f(W_{kj} O_j) \qquad (12)$$

where net_k is the input value of the kth output neuron, and O_i is the input value of the jth hidden neuron. W_{kj} indicates the weight between the jth neuron of hidden layer and the kth neuron of output layer. The corresponding output value is:

$$O_k = f(net_k) \qquad (13)$$

where f is Sigmoid activation function as specified earlier.

(e) Calculate the output layer error and the hidden layer error.

$$d_k = O_k(1 - O_k)(O_k - t_k) \qquad (14)$$

$$e_j = O_j(1 - O_j)\sum_k W_{kj} d_k \qquad (15)$$

In the equation 14, d_k is the reference error of the kth neuron in the output layer, t_k is the target output, and in the equation 15 e_j is the reference error of jth neuron in the hidden layer.

(f) Calculate the output layer error using the error function E.

$$E = \frac{1}{2} \sum_{i=1}^{N} \sum_{k=1}^{P} \left(O_{pk} - t_{pk}\right)^2 \qquad (16)$$

where, O_{pk} is the observed output value and t_{pk} is the arget value. P is the number of output neurons and N is the number of samples.

(g) Adjust the connection weight matrix and thresholds.
The adjustment of weight matrix W_{kj} and threshold γ_{kj} between output layer and hidden layer follows according to:

$$W_{kj}(m+1) = W_{kj}(m) + \alpha O_j d_k + \eta \Delta W_{kj}(m) \qquad (17)$$

$$\gamma_{kj}(m+1) = \gamma_{kj}(m) + \alpha d_k \qquad (18)$$

where m is the number of iterations, α is the learning rate, ΔW_{kj} is a matrix representing the change in the matrix W_{kj} and, η is a momentum factor; it is used to allow the previous weight change to influence the weight change in the present model iteration.
The adjustment of the weight matrix W_{kj} and the threshold θ_{ji} between the hidden layer and the input layer:

$$W_{ji}(m+1) = W_{jij}(m) + \beta O_i e_j + \eta \Delta W_{jij}(m) \qquad (19)$$

$$\theta_{kj}(m+1) = \theta_{ji}(m) + \beta e_j \qquad (20)$$

where β is the learning rate, ΔW_{ji} is a matrix representing the change in matrix W_{ji}.

The learning process is implemented by a change of the connection weight matrix and thresholds, in which the error function E gives the greatest gradient descent direction to change the connective weight, the weights connected with different neurons are updated by equations (17) and (19). It results in gaining the best weight coefficient sets.

(h) Iteration (a) through (f) continues until the value of the error is less than a desired threshold or the iterations times exceed a specified time. This means 'training process' is complete. The learning of BP nets by least error function rule completes its non-linear mapping from input to output.

(i) Input the digital image to be classified to the BP network that has completed the learning process, you can then generate a complete image classification.

In practice, because BP algorithm adopts the simple gradient descent, the rate of convergence is very slow, and the local minima often occurs. Thus iterative process cannot converge to global optimum solution, and especially big trainings and lots of input parameters will remarkably hamper the learning effect. So Simulated Annealing algorithm is introduced to globally optimize in the networks learning.

Simulated Annealing algorithm(SA) was put forward by S. Kirkpatrick(1983). Its postulate was contrasting the solution of some kind of optimizing with the heat balance of energetics

statistics,and try to simulate the heat object annealing so that we can find the global optimum solution.

Suppose $S=\{S_1,S_2,\ldots,S_n\}$ is the set of all the possible combinations(or states).$C:S \to R$ is the non-minus objective function ,and so $C(S)\geqslant 0$ means the cost of the solution is S_i. It is clear that the optimizing combination can be described formly with finding $S^* \in S$, so that

$$C(S^*) = \min C(S_i), S_i \in S$$

.Simulated Annealing processes mostly as follows:

Procedure $SA(i_0,T_0)$ /* i_0 is initial state， T_0 is the initial value of control parameter， $C_i = C(S_i)$ */

   (1) $S := S_{i_0}; k = 0;$ /*S is current state*/

   (2) Repeat

   (3) Repeat

   (4) $S_j := Generate(S);$

   (5) If $C_j \leq C^*$ Then $S := S_j;$

   (6) Else If $Accept(j,S)$ Then $S_s := S_j;$

   (7) Until 'inner-loop stop criterion' /* "Inner-loop stop criterion" means the number of iterations of the SA in the temperature T */

   (8) $T_{k+1} = Update(T_k); k \leftarrow k+1;$ /* Tthe velocity of temperature's decline at a time with function update(T_k) */

   (9) Until 'final stop criterion' /* the finish of SA */

In Above algorithm, Generate(S) in the step(4) means generate the next state S_j at random from N. If $C_j \leq C^*$,then accept j as the new current state,otherwise only accept j as the new current state with some probability .All that is the function of Accept(j,S).

Usually function Accept processes as follows.

Procedure $Accept(j,S)$ /*in the step(6),only when $C_j > C^*$,call Accept */

   (1) If $\exp\left\{-(C_j - C^*)\big/_k T_k\right\} > random(0,1)$

   (2) Then Accept:=True
        Else Accept:=False

The aim of using Simulated Annealing algorithm is to get globally optimize. In the errors are reducing process, disturbance at random in some degree can get over the restriction of the local minima, and ensure the system keep away form disturbing when it converges to global optimum solution.This is just the problem which Simulated Annealing algorithm has settled.

For aim of globally optimize, the following function specifies the random disturbing:

$$W_{ij} = W_{ij} \times \left[1 + \frac{1}{loop\_time} Random(-1,+1)\right] \quad (21)$$

Where Loop_time is the number of iterations, Random(-1,+1) initialized as randomized real numbers within the range –1 to +1 .From the above state we can see the Simulated Annealing algorithm as the gradient descent with noises , and when the temperature which identify the noise intensity is 0(the number of iterations is infinity),it is just the gradient descent.

4. APPLICATION TO SAR IMAGE

Image classification belongs to the division of patterns in eigen space. If it is supposed that existing samples x1, x2, x3, x4, ... , xn, in an image belong to certain categories C1, C2, C3, C4, ...,Cm , (m<n), it is possible to select n samples to extract feature of each ground targets. The goal then of establishing supervised samples is to make use of multispectral features and couple them with those of texture and structural features and use them for training the BP nets. In preparation for the classification of the entire 224 by 224 scene, six input fields were used which comprised the training pixels – the three original channels (polarization combinations L-HH, L-HV and C-HV) and the three energy components (1, 2 and 3) as derived from the wavelet decomposition. These are referred to as the target samples. Thus, the BP nets have six nodes in the input layer and three nodes comprised the output layer based upon the desired classification (poplar, bushes and background). Broadly speaking, the number of nodes in the hidden layer is arbitrary although general guidelines exist e.g. (Lippmann, 1987). In general, the more nodes in the hidden layer, the better the result of image classification but it takes a longer time for the network to learn the necessary knowledge for the classification and often results in a reduction of the network's ability to generalize. The problem is therefore achieving a balance between accuracy and the time required for training. Through experimentation, four nodes were defined for the hidden layer since this was found to generate an optimal classification. In order to train the BP nets with the target samples, the input data was rescaled to comply within the limits of the Sigmoid activation function and set to 0.9 and 0.1 respectively. Table 1 shows the possible responses of the output layer processing element. Data for training the network was accomplished by defining a 10 10 pixel window (100 samples of input data of each type) from the first three channels (polarization combinations L-HH, L-HV and C-HV). For the second three channels, the decomposed elements (1, 2 and 3) the window must contain sufficient resolution to preserve the essential information: considering the characteristic of texture and structure in high-resolution SAR images and the requirement of three level DWT, a window size of 32 32 pixels was adopted.

Class	Target Output		
	O_1	O_2	O_3
Poplar Trees	0.9	0.1	0.1
Bushes	0.1	0.9	0.1
Background	0.1	0.1	0.9

Table 1. Response of the output layer processing element

In the training process, an iteration is divided into two stages after the data are input into the input layer. First, the vector of the hidden-layer neuron is computed by the Sigmoid activation function; then the vector of the output-layer neuron is computed by the Sigmoid activation function. Second, the error between the observed output and the desired output is calculated at the

output neurons. If the observed error exceeds the desired error, the output signals are fed forward to the hidden layer. In the same way, the signals in the hidden layer are fed forward to the input layer, in which it is also calculated to the hidden-layer error. Finally, the weights between the hidden layer and the output layer, and the weights between the input layer and hidden layer are adjusted. The iterative procedure involving the previous two stages continues until the output layer error is within the specified threshold.

Upon completion of the iterative training procedure the entire 224 by 224 SAR scene was classified. The result of the classification of the SAR image is shown in figure 2. In figure 2 the blue, turquoise and yellow colours are associated respectively to poplar, bushes and background. For the two-element vegetation classification into poplar and bushes in well differentiated regions the probability of yielding poplar and bushes classification was found to be 100%. When the scene includes transition regions, i.e. boundaries between trees and bushes, the resulting probability of correct classification is reduced marginally. The same is observed in the boundaries between the trees and the background. It is worthwhile stating at this point that cultivated areas like those exhibited in figure 1 often exhibit significantly different intensity (digital number) values; this is evident from a cursory inspection of figure 1. This will lead to confusion in decision boundaries and so, result in erroneous classification when conventional classification schemes are applied. The addition of wavelet decomposition fields reduces the confusion that might arise from consideration of the reflectance fields alone as areas of poplar and bushes exhibit a similar high energy content – i.e. 1, 2 and 3.
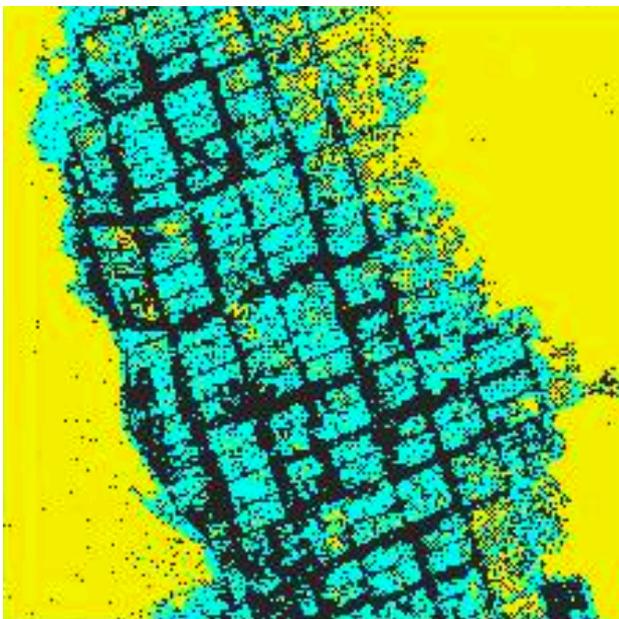

Figure 2. The result of classification of the SAR image.

5. CONCLUSIONS

This paper demonstrates the integration of multispectral SAR data coupled to representations of the image by wavelets. The resulting nonlinear nature of the data forms a complex set for classification. A modified neural network is therefore employed as a means to classification. The proposed approach has shown good results on a three-element classification into poplar, bushes and background in a region where the features are visibly well differentiated. This illustrates that SAR imagery can be classified effectively where there are distinct regions of each class and when additional information is provided on texture and structural features with rather strong orientation. An important consideration in the whole process of classification of SAR images, however, is that it involves a multiple-exploitation of techniques to interpret land surface characteristics. This particular combination of methods was consistent in this context but extension into further environments, both different and more complex, requires further research.

REFERENCES

1. BRYAN, M.L., 1979, The effect of radar azimuth angle on cultural data. *Photogrammetric Engineering and Remote Sensing*, 45, 1097-1107.

2. BARBAROSSA, S., and PARODI, L., 1995, SAR image classification by wavelets, *IEEE International Radar Conference*, 462-467.

3. CHUI, C.K., 1992, *An introduction to wavelets*, (Boston Academic Press).

4. DAUBECHIES, I., 1998, *Orthonormal bases of compactly supported wavelets*. Comm. Pure and Applied. Math, 41, 909-996.

5. GURNEY, K., 1997, *An introduction to neural networks*, (UCL Press Limited)

6. LIPPMANN, R.P., 1987, An introduction to computing with neural nets. *IEEE ASSP Magazine*, 4, 22.

7. MALLAT, S., 1989, A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 11, 674-693.

8. PAOLA, J.D. and SCHOWENGERDT, R.A., 1995, A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban and land use classification. *IEEE Transactions on Geoscience and Remote Sensing*, 33(4), 981-996.

9. RUMELHART, D.E., HINTON, G.E., and Williams R.J., 1986, Learning internal representations by error propagation. Parallel Distributed Processing: Explorations in the Microstruction of Cognition. D.E. Rumnelhart and J.L McClelland, (Eds.). Cambridge, MA, The MIT Press. I: 318-362.

10. SIMARD, M., SAATCHI, S. and DeGRANDI, G., 2000, Comparison of a decision tree and maximum likelihood classifiers: application to SAR image of tropical forest, *IEEE 2000 International Geoscience and Remote Sensing Symposium*, 5, 2129-30.

11. Kirkpatrick, S., C. D. Gelatt Jr., M. P. Vecchi, "Optimization by Simulated annealing", *Science*, 220, 4598, 671-680, 1983.