

# ROAD DETECTION FROM HIGH RESOLUTION SATELLITE IMAGES USING ARTIFICIAL NEURAL NETWORKS

M.J Valadan Zoej<sup>a</sup>, M. Mokhtarzade<sup>b</sup>

<sup>a, b</sup> Faculty of Geodesy and Geomatics Eng., K.N.Toosi University of Technology, Tehran, Iran

<sup>a</sup> ValadanZouj@kntu.ac.ir, <sup>b</sup> m\_mokhtarzade@yahoo.com

Commission WG III/4

**KEY WORDS:** Remote Sensing, Extraction, Neural, Networks, Learning, High resolution, IKONOS, Quickbird

## ABSTRACT:

In this article, the possibility of using artificial neural networks for road detection from high resolution satellite images is tested on a part of RGB Ikonos and Quick-Bird images from Kish Island and Bushehr Harbour respectively. Then, the effects of different input parameters on network's ability are verified to find out optimum input vector for this problem. A variety of network structures with different iteration times are used to determine the best network structure and termination condition in training stage.

It was discovered when the input parameters are made up of spectral information and distances of pixels to road mean vector in a 3\*3 window, network's ability in both road and background detection can be improved in comparison with simple networks that just use spectral information of a single pixel in their input vector.

## 1. INTRODUCTION

Roads are one of the most important manmade objects in map production from satellite images that must be extracted accurately. Nowadays when satellite images have improved highly in both spatial and spectral resolutions and they are available in short time intervals, automatic road extraction is one of the challenges in remote sensing and photogrammetry.

Road detection can be considered as the first step in road extraction. It is defined as the process of assigning a value to each pixel that can be used as a criterion to extinguish road and non-road pixels.

Artificial Neural Networks (ANN) are computational systems, inspired from biological neural networks, in which a set of input parameters is related to an output set by a transformation encoded in the network weights (Yang,1995). In this respect they can be used as a function to receive input data from satellite images as their input parameters and calculate the road detection criterion mentioned above.

In this article the possibility of using ANNs as road detectors is tested on an RGB Ikonos and Quick-Bird images and the effects of input parameters on network's functionality are verified.

## 2. ARTIFICIAL NEURAL NETWORKS (ANNs)

Neural Networks are made up of simple processing units called nodes or neurodes. The main task associated with a neurode is to receive input from its neighbours (the output of other neurodes), compute an output and send that output to its neighbours (Yang, 1995).

Neurodes are usually organized into layers with full or random connections between successive layers. There are three types of layers: input, hidden and output layers in charge of receiving, processing and presenting the final results respectively (Figure 1).

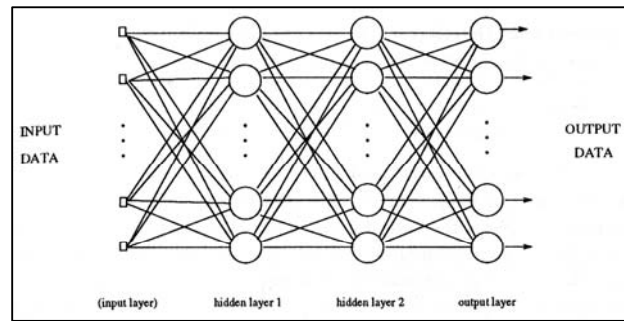


Figure (1): A typical neural network

There are two main stages in the operation of an ANN classifier: learning and recalling.

Learning (training) is the process of adapting or modifying the connection weights so that the network can fulfil a specific task and is usually done in an iterative way. This process is mainly carried out using a training set which comprises some known input-output samples. This kind of training is called training with a teacher or supervised learning.

Back-propagation is the most common learning algorithm that was discovered by Rumelhart and Parker independently in the early 1980s. It is an iterative gradient algorithm designed to minimize the Error function. The error function is shown in equation 1.

$$E = \frac{1}{2} \sum_{j=1}^L (d_j - o_j^M)^2 \quad (\text{Eq.1})$$

Where  $d_j$  and  $o_j$  represent desired output and current response of the neurode "j" in the output layer respectively and "L" is the number of neurodes in the output layer. In an

iterative method, corrections to weight parameters are computed and added to previous values as below:

$$\begin{cases} \Delta w_{i,j} = -\eta \frac{\partial E}{\partial w_{i,j}} \\ \Delta w_{i,j}(t+1) = \Delta w_{i,j} + \alpha \Delta w_{i,j}(t) \end{cases} \quad (\text{Eq. 2})$$

Where  $w_{i,j}$  is weight parameter between neurode i and j,  $\eta$  is a positive constant that controls the amount of adjustment and is called learning rate,  $\alpha$  is a momentum factor that can take on values between 0 and 1 and "t" denotes the iteration number. The parameter  $\alpha$  can be called smoothing or stabilizing factor since it smoothes the rapid changes between the weights (Yang, 1995).

Recalling refers how the network can operate based on what it has learned in the training stage. It is actually using the trained network for interpolation and extrapolation and is called generalizing either.

One of the most important advantages of neural networks with respect to conventional statistical methods is that they are distribution-free operators because the learning and recalling depend on the linear combination of data pattern instead of the statistical parameters of the input data (Civco and Waug, 1994). Neural networks are also highly capable to deal with multi-source data because they do not require explicit modelling of the data from different sources and therefore there is no need to treat them independently as in many statistical methods (Benediktsson and Swain, 1990). They also avoid the problem in statistical multi-source analysis of specifying how much influence each data source should have in classification (Benediktsson and Swain, 1990).

### 3. NETWORK DESIGN

Road detection from satellite images can be considered as a classification process in which pixels are divided into road and background classes. Recent researches have shown ANNs to be capable of pattern recognition and classification of image data. For using neural networks in road detection, input layer is consisted of neurodes the same number as input parameters and output layer is made up of just one neurode that shows the belief of network whether the input parameters can represent a road pixel or not. Usually one hidden layer is sufficient, although the number of neurodes in the hidden layer is often not readily determined (Richard, 1993). More neurodes in hidden layer enables the network to learn more complicated problems, but there would be an associated increase in training time (Foody, 1995).

The most important factor in employing ANNs is to decide what type of information should be extracted from input image to be fed through the network as its input parameters. The discrimination ability of the network is highly affected by chosen input parameters. Roads are presented as linear features in low resolution images while they are displayed as homogeneous areas in high resolution satellite images like QuickBird and Ikonos. For that reason, input parameters for road detection from low resolution images should be able to distinguish shape patterns while for high resolution images homogeneity and spectral characteristics seem more important to be emphasized.

The learning rate and momentum parameters have a major influence on the success of learning process and they should be defined by user in advance. This assignment is problem dependent.

Another factor that affects network ability is the number of iterations done in training stage. If the network is trained more than what is needed, training samples will be memorized by the network that reduces the discrimination ability of the network. Therefore termination conditions should be assigned accurately to avoid over-training problem.

Accordingly network design, consisted of defining hidden layer size, input parameters selection,  $\eta$  and  $\alpha$  assignment and termination conditions, is a crucial stage that must be performed before applying neural networks.

### 4. METHODOLOGIES

As a case study a part of an RGB Ikonos image with the size of 550\*550 pixels from Kish Island in Iran is chosen which is enhanced with linear function. Figure 2 shows the original image and its manually produced reference map which is used in accuracy assessment.

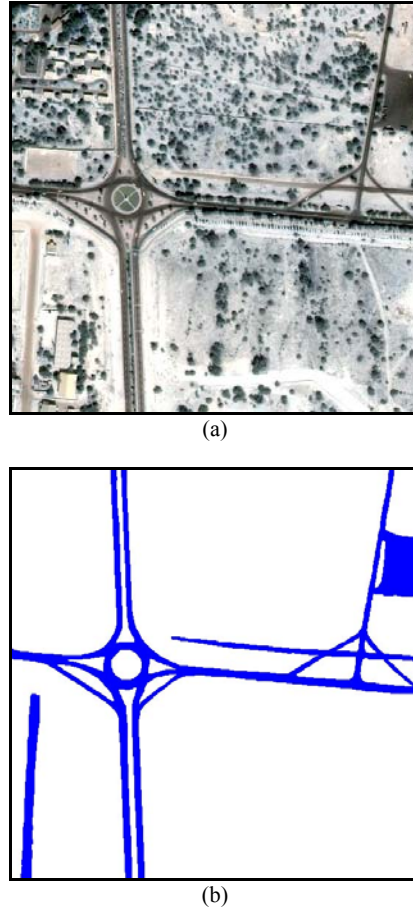


Figure (2) a) RGB Ikonos Image from Kish Island in Iran. b) Manually produced reference map.

500 road and 500 background pixels are chosen as training set to be used in learning stage. It is recommended to have representative pixels of all present objects in the training set.

A Back-propagation Neural Network (BNN) with one hidden layer is implemented which is self-programmed in Delphi. The output layer consists of one neurode that expresses network's response by a number between 0 and 1 as background and road pixel respectively. When the trained network is performed on entire pixels, a 2D matrix the same size as input image is obtained that is called output matrix.

An adaptive strategy is used to avoid trail and error learning rate and momentum assignment. In this method both parameters are adjusted downwards as half after some training intervals if the overall training error has increased and upward 1.2 times if the overall error has decreased (Heerman and Khazeinie, 1992). Therefore, the initial learning rate and momentum are not crucial to the success of training stage. Also training speed is increased because the learning rate is adjusted to the highest value that does not cause instability (Paola and Schowengerdt, 1997).

For accuracy assessment, a binary image is extracted from image-truth, assigning 1 to road and 0 to background pixels. Adding the multiplication of correspondent values in whole binary image and output matrix produce a value that when is divided to the road pixel number can be considered as road detection correctness coefficient (RCC). When the binary image is inverted, in a similar method, background detection correctness coefficient (BCC) can be obtained too. The third parameter is root mean square error (RMSE), assuming the difference between network's responses and expected results (1 for road and 0 for background pixels) as error values. After putting a threshold on output matrix, overall accuracy can be calculated the same way as conventional classification methods. This parameter is actually the percentage of correctly classified pixels to all available pixels in entire image.

Four input parameter types are designed to be fed through the network and for each one different hidden layer sizes are tested. Each network is trained with several iteration numbers to prevent over-training problem and find out the best termination condition. It was discovered that RMSE can be the most reliable parameter for this reason as it begins to deteriorate when the network is about to get over-trained. In continue each input vector type is evaluated in a distinctive section.

#### 4.1 Spectral values as input parameters

In this section spectral information for each pixel, after normalizing the RGB values between 0 and 1, is simply entered to the network as its input parameters.

Thus, three neurodes are designed in input layer in charge of receiving spectral values for each pixel in entire image. Figure 2 shows the network structure and obtained results are presented in Table 1.

The network with 5 hidden neurodes is very unstable which means networks functionality is highly dependent on initial weight assignment and results vary in multiple implementations.

Larger hidden layer size enables the network to modulate more complicated problems, but in this case classification problem dose not seem so complex since the results are quite the same for different networks.

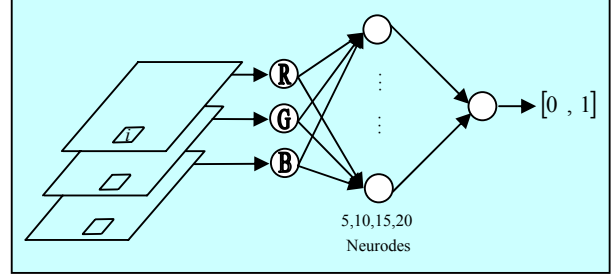


Figure (3): Network Structure when spectral values form input parameters.

Table 1: Three spectral values as input parameters

Hidden Neurods	Best Iteration	RCC	BCC	RMSE	Overall Acc.
5	15000	77.69	88.98	0.2457	93.83
10	5000	73.31	88.87	0.2238	94.66
15	5000	73.62	90.00	0.2202	94.78
20	10000	74.82	90.87	0.2234	94.49

#### 4.2 Including input parameters with normalized distance

In this section the normalized distance of each pixel to the road mean point in the spectral space is added to the input parameters. Distance parameter is calculated for each pixel as below:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix}_i, \begin{bmatrix} R \\ G \\ B \end{bmatrix}_{mean}$$

$$d_i = \frac{1}{441.673} \left[ (R_m - R_i)^2 + (G_m - G_i)^2 + (B_m - B_i)^2 \right]^{1/2} \quad (\text{Eq. 3})$$

While,  $\begin{bmatrix} R_m & G_m & B_m \end{bmatrix}_i^T$  is the mean of road pixels in training set.

Consequently, four neurodes are designed in the input layer to receive input parameters. Figure 4 shows the network structure and Table 2 presents obtained results.

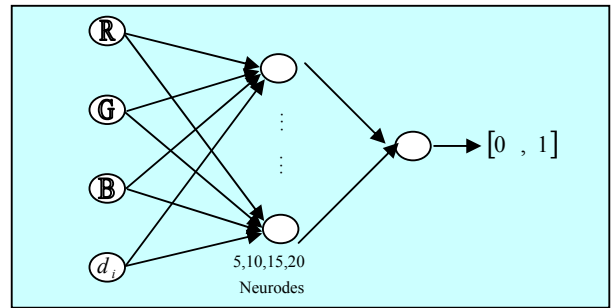


Figure (4): Network Structure when distance parameter is added to input parameters

Table2: Normalized distance as forth input parameter

Hidden Neurods	Best Iteration	RCC	BCC	RMSE	Overall Acc.
5	1000	71.45	92.63	0.2153	94.32
10	1000	71.52	92.33	0.2155	94.44
15	2000	72.49	92.97	0.2160	94.47
20	2000	72.55	93.28	0.2146	94.34

While background is consisted of different objects with different spectral behaviours, distance parameter, accentuating on spectral differences, has improved network ability in background detection and has brought about decrease in RMSE value. The decrease in RCC can be interpreted as numerical problems since  $d_i$  values for road pixels are very small in amount near zero.

### 4.3 Participating neighbours pixels in input parameters

In this section, normalized spectral information in a 3\*3 window around each pixel is extracted as 9 red, 9 green and 9 blue values to form input parameters in that order. Accordingly, input layer involves 27 neurodes. Figure 5 shows the network structure and obtained results are presented in Table 3.

The comparison between Tables 3 and 1 shows that RCC has increased highly, while BCC has decreased in value which means network's ability in road detection has improved while background detection ability has been deteriorated. Also spike noises were observed that caused increase in RMSE and decrease in overall accuracy.

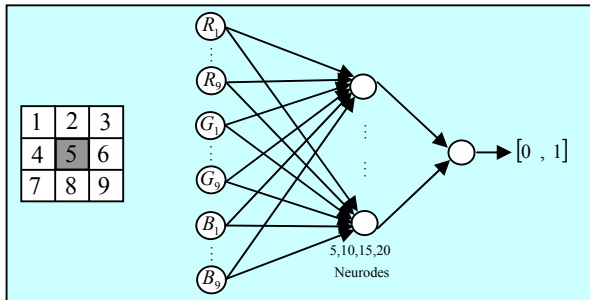


Figure (5): Network structure when neighbour pixels form input parameters

Table 3: Participating neighbour pixels in input parameters

Hidden Neurods	Best Iteration	RCC	BCC	RMSE	Overall Acc.
5	3000	80.58	86.80	0.2811	91.73
10	3000	81.01	86.54	0.2739	92.22
15	3000	80.45	87.59	0.2666	92.96
20	4000	80.83	88.33	0.2507	93.61

As roads are presented like homogeneous areas in high resolution satellite images, participating neighbour pixels in input parameters enables the network to extinguish road pixels more efficiently and it caused RCC parameter to be improved.

By increasing the hidden layer size and iteration time, results got improved but network's training stage took more time in return.

While designed input parameters in 4.2 and 4.3 have opposite results as improving background and road detection ability in comparison with simple network in 4.1, it seems the combination of both input parameters can make the network more powerful in both sides. This combination is examined in 4.4.

### 4.4 Spatial information and normalized distance as input parameters

In this section the normalized distances of all 9 pixels in the mentioned window to the mean vector of road pixels are added to previous stage's input parameters. Thus the input layer is consisted of 9 red, 9 green, 9 blue and finally 9 normalized distances to the road mean vector which means 36 neurodes are designed in this layer. Figure 6 shows network's structure and Table 4 presents obtained results.

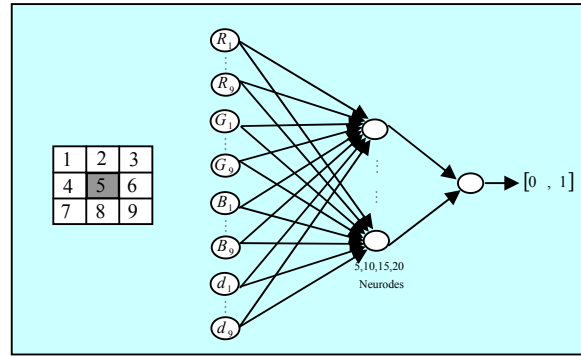


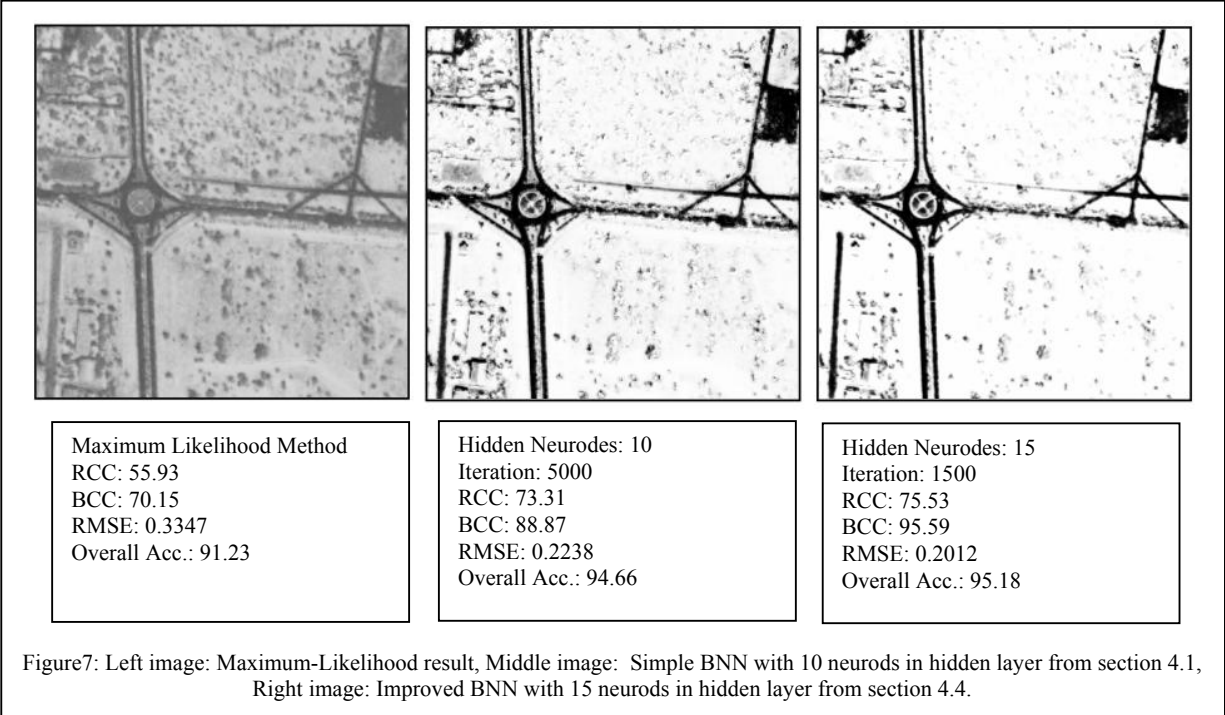
Figure (6): Network structure when neighbor pixels with their normalized distances form input parameters

Table (4): Spatial information and normalized distance as input parameters

Hidden Neurods	Best Iteration	RCC	BCC	RMSE	Overall Acc.
5	1000	75.62	93.88	0.2008	95.13
10	1000	75.22	94.62	0.1999	95.18
15	1500	75.53	95.59	0.2012	95.18
20	1500	76.63	95.26	0.2006	95.19

The comparison between Table 4 and Tables 1, 2, 3 shows that designed input parameters in this section could improve network's ability in both road and background detection. Although input layer enlargement can make training stage more time consuming, but this problem is compensated to some extent by decrease in requested hidden layer size and iteration time.

Finally, as a comparison with statistical methods, obtained results from Maximum-Likelihood classification method and best network from 4.1 and 4.4 are shown together in Figure 7 with their accuracy assessment parameters



**5. NETWORK'S FUNCTIONALITY ON QUICK-BIRD IMAGES**

In this section a part of an RGB Quick-bird image from Bushehr harbour in Iran is chosen as input image to evaluate network's behaviour on this kind of images.

Figure 8 shows the original image and its manually produced reference map which is used in accuracy assessment.

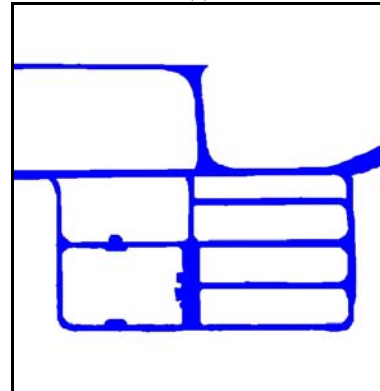
Two input parameter types are implemented. In the first case only spectral values are used in input vector formation and therefore three neurodes are designed in input layer. In the second case the suggested input parameter set, which is made up of spectral values and normalized distances of all pixels in surrounding window, is implemented and therefore 36 neurodes are designed in input layer.

A variety of networks with different neurode numbers in hidden layer are used and each network was trained with multiple iteration times to discover best iteration time when the network is not over-trained.

The optimum network structures and iteration times are selected considering computed accuracy assessment parameters. Obtained results and their accuracy assessment parameters are shown in Figure 9.



(a)



(b)

Figure (8) a) RGB Quick-Bird Image from Bushehr harbour in Iran. b) Manually produced reference map

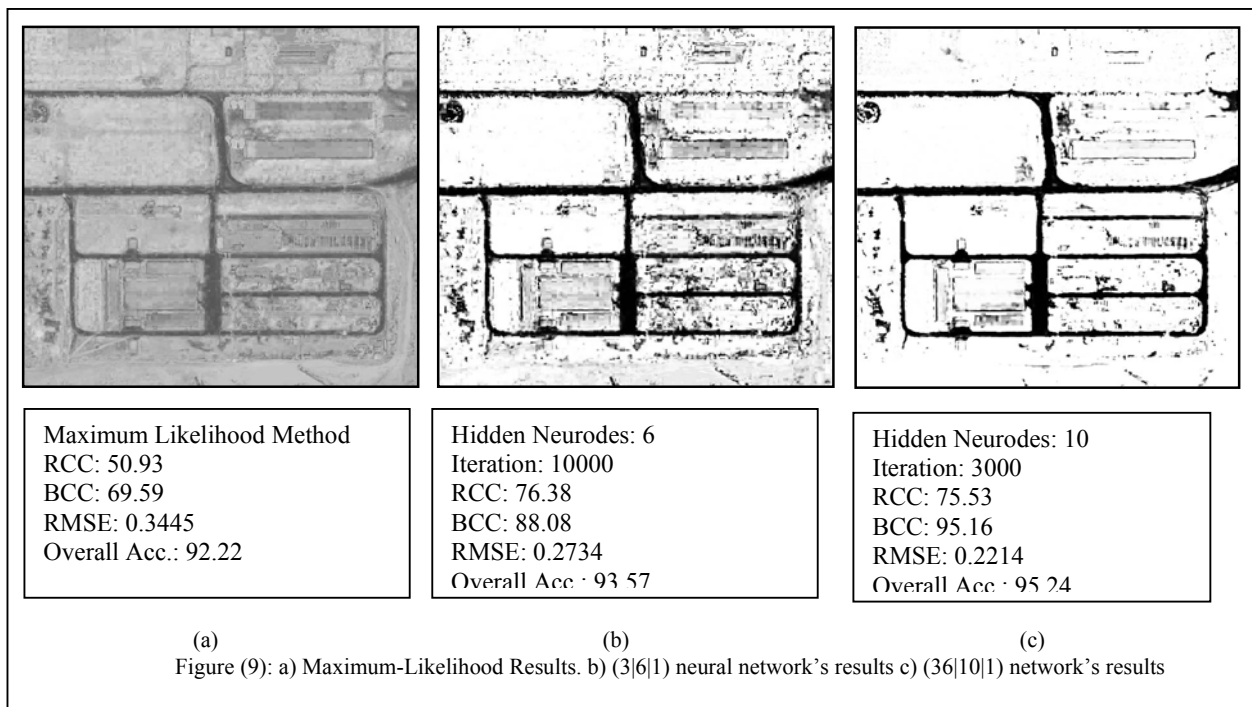


Figure 9.a is the result of Maximum Likelihood method which uses the same training set and has been shown for comparison between neural networks and statistical methods. Figure 9.b is the result of optimum simple neural network that just uses spectral information of a single pixel in its input layer. Figure 9.c is the answer of optimum improved network that its input vector is consisted of spectral values and normalized distances of all pixels in a 3\*3 window.

##### 5. Summary and discussion

In this article the impact of input parameters on neural network's ability for road detection from high resolution satellite images was tested on multi-spectral Ikonos and QuickBird images. A back-propagation neural network was implemented with different hidden layer sizes and it was trained with different iteration times to prevent over-training problem.

As roads are homogeneous areas in high resolution images, employing neighbour pixels in input parameters can improve road detection ability of the network, while using the distance of each pixel to the road mean vector can develop network's ability in background recognition.

The combination of both mentioned input parameters made the network powerful in both road and background detection and also reduced the requested hidden layer size and iteration time.

It was discovered that there is no need to design more than 10 neurodes in hidden layer as it does not improve results noticeably and just makes the training and recalling stages more time consuming.

RMSE proved to be the most reliable parameter to be used as termination condition since it begins to deteriorate when the network is about to get over-trained.

##### References

- Benediktsson, J.A., Swain, P.H., and Erosy, O.K. (1990). "Neural Network Approaches Versus Statistical Methods in Classification of Multisource remote Sensing data", IEEE Transaction on Geosciences and Remote Sensing, Vol. 28: pp. 540-551.
- Civco, D.L., and Waug, Y. (1994). "Classification of Multispectral, Multitemporal, Multisource Spatial Data Using Artificial Neural Networks". In proceeding of the ASPRS 1994 Annual Convection, Reno, NV, USA. pp. 123-133.
- Foody, G.M., McCulloch, M.B., and Yates W.B. (1995). "Classification of Remotely Sensed Data an Artificial Neural Network: Issues Related to Training Data Characteristics", Photogrammetric Engineering and Remote Sensing, Vol. 61, No. 4, pp. 391-401.
- Heerman, P.D., and Khazenie, N. (1992). "Classification of Multi-spectral Remote Sensing data Using a Back-Propagation Neural Network", IEEE Transactions on Geoscience and remote sensing, Vol. 30, No. 1, pp. 81-88.
- Paola, J.D., and Schowengerdt, R.A. (1997). "The Effect of Neural Network Structure on a Multi-Spectral Land-Use/Land-Cover Classification", Photogrammetric Engineering and Remote Sensing, Vol. 63, No. 5, pp. 535-544.
- Richard, J.A. (1993). "Remote Sensing Digital Image Analysis: Introduction". Second Edition, Springer, ISBN 0-387-5480-8, New York.
- Yang, G.Y. (1995). "Geological Mapping from Multi-Source Data Using Neural Networks", MSc Thesis, University of Calgary, Canada.