International Journal of Applied Earth Observation
and Geoinformation 9 (2007) 32–40

INTERNATIONAL JOURNAL OF
APPLIED EARTH OBSERVATION
AND GEOINFORMATION

www.elsevier.com/locate/jag

Road detection from high-resolution satellite images using artificial neural networks

M. Mokhtarzade *, M.J. Valadan Zoej [1]

K.N. Toosi University of Technology, Geodesy and Geomatics Faculty, No. 1364, Mirdamad Cross, Valiasr St., Tehran 19967-15433, Iran

Received 3 September 2004; accepted 2 March 2006

Abstract

This article treats the possibility of using artificial neural networks for road detection from high-resolution satellite images on a part of RGB Ikonos and Quick-Bird images from Kish Island and Bushehr Harbor, respectively. Attempts are also made to verify the impacts of different input parameters on network's ability to find out optimum input vector for the 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 found that when the input parameters are made up of spectral information and distances of pixels to road mean vector in a 3×3 window, the network's ability in both road and background detection can be improved in comparison with simple networks that simply use spectral information of a single pixel in their input vector.

Keywords: Classification; Feature extraction; Network design; Back-propagation

1. Introduction

Arial and satellite images are promising data sources for map generation and update of available maps to support activities and missions of government agencies and consumers. Full exploitation of these data sources depends on automatic techniques for object extraction from satellite and aerial imageries.

The availability of high-resolution imagery from next-generation remote sensors has motivated and increased a sense of urgency for new automatic feature extraction. The motivation stems from the realization that the sheer volume of collected images will quickly overwhelm the capacity of image community to carry out timely and costly manual feature extraction tasks.

Roads, as one of the most important man-made objects, are subjects of great concern to be extracted semi-automatically. Many researches have been conducted for this purpose. Geometrically constrained template matching (Gruen et al., 1995; Vosselman and Knecht, 1995), active contours or snakes (Neuenschwander et al., 1995; Trinder and Li, 1995; Gruen and Li, 1997), fuzzy set and morphological operators (Mohammadzadeh et al., 2004) constitute some of the semi-automatic methods for road extraction.

Tupin et al. (1998) and Jeon et al. (2002) implemented the road network extraction from SAR images, semantic networks and genetic algorithms, respectively. Fiset and Cvayas (1997) and Zhang and Baltsavias (2000), respectively, verified knowledge-based image analysis for updating road networks and for automatic road extraction. One can find an excellent survey paper

* Corresponding author. Fax: +98 311 6243610.
E-mail addresses: m_mokhtarzade@yahoo.com (M. Mokhtarzade), ValadanZouj@kntu.ac.ir (M.J.V. Zoej).
[1] Fax: +98 21 8786213.

0303-2434/$ – see front matter
doi:10.1016/j.jag.2006.05.001

by Mayer et al. (1998) on road detection in digital images.

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.

In the present research, road detection is performed on high-resolution pan-sharpened RGB Ikonos and Quick-Bird satellite images, using artifi-cial neural network algorithms. In these kinds of images, roads could be regarded as elongated hom-ogeneous regions contrasting the background with a distinct spectral behavior. Based on this model, a variety of different network structures were implemented. The impact of different input parameters on network's functionality was also tested. Finally, the results were compared with maximum-likelihood classifier.

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 neighbors (the output of other neurodes), to compute an output and to send that output to its neighbors (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 that receive, process and present the final results, respectively (Fig. 1).

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 fulfill a specific task; it 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 Eq. (1):

$$E = \frac{1}{2}\sum_{j=1}^{L}(d_j - o_j^M)^2 \qquad (1)$$

In this equation d_j and o_j represent the 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 the previous values as illustrated below:

$$\begin{cases} \Delta w_{i,j} = -\eta \dfrac{\partial E}{\partial w_{i,j}} \\ \Delta w_{i,j}(t+1) = \Delta w_{i,j} + \alpha\, \Delta w_{i,j}(t) \end{cases} \qquad (2)$$

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

Recalling asks 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, also being called generalizing.

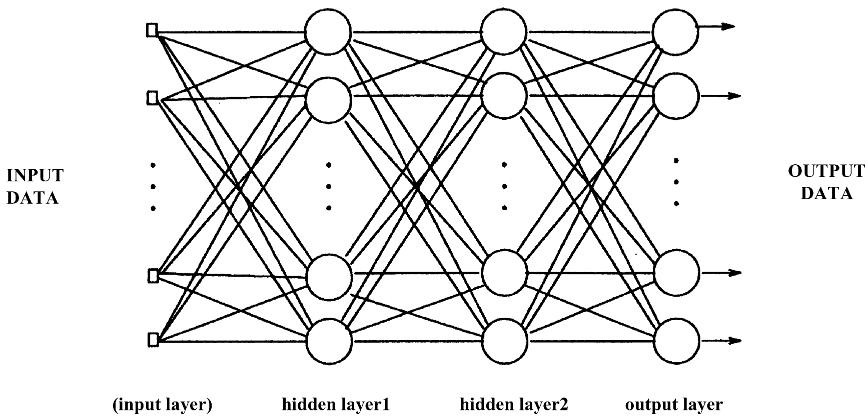

Fig. 1. A typical neural network.

One of the most important advantages of neural networks as compared 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 of dealing with multi-source data, because they do not require explicit modeling of the data from different sources. There is, therefore, no need to treat them independently as in the case of many statistical methods (Benediktsson et al., 1990). They are also clear of the problem in statistical multi-source analysis of specifying how much influence each data source should have in classification (Benediktsson et al., 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 being 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 enable the network to learn more complicated problems, but there will be an associated rise in training time (Foody et al., 1995).

The most important factor responsible for 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 Quick-Bird and Ikonos. For this 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 are emphasized.

The learning rate and momentum parameters have a major influence on the success of learning process and 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, comprising defining hidden layer size, input parameters selection, η and α 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 and enhanced with linear function. Fig. 2 shows the original image and its manually produced reference map, which is used in accuracy assessment.

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, which is self-programmed in Delphi, is implemented. The output layer consists of one neurode that expresses the 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 of the same size as input image is obtained, being 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 Khazenie, 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. Addition of the multiplication of correspondent values in whole binary image and output matrix produces a value that can be considered as road detection correctness coefficient (RCC) when divided to the road pixel number. When the binary image is inversed, background detection correctness coefficient (BCC) can be obtained in a similar method as well. The third parameter is root mean square error (RMSE),

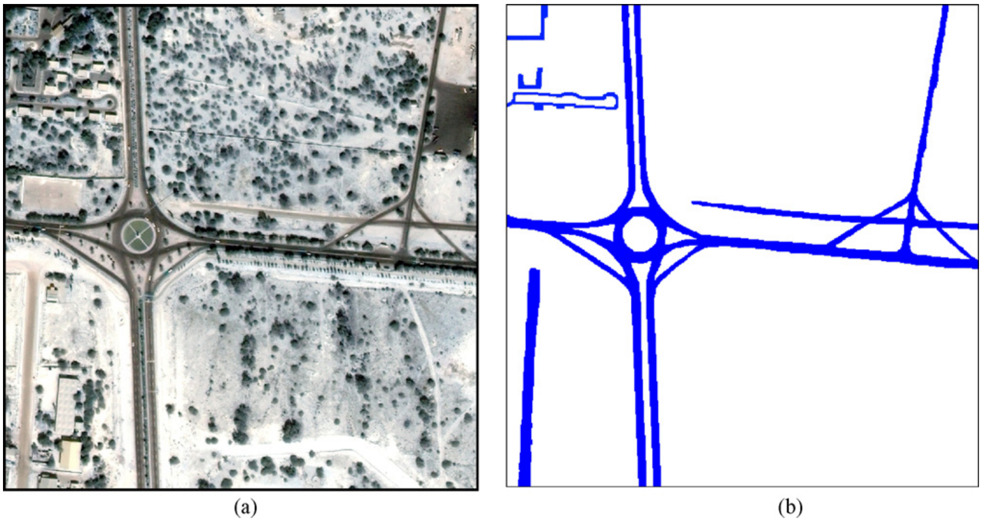

Fig. 2. (a) RGB Ikonos image from Kish Island in Iran and (b) manually produced reference map.

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 and kappa coefficient can be calculated the same way as conventional classification methods. Overall accuracy 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; 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 found 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 continuation, 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 is simply entered to the network as its input parameters after normalizing the RGB values between 0 and 1.

Thus, three neurodes are designed in input layer in charge of receiving spectral values for each pixel in the entire image. Fig. 3 shows the network structure; the results are presented in Table 1.

The network, with five hidden neurodes, is very unstable in the sense that functionality of networks is highly dependent on initial weight assignment and that the results vary in multiple implementations.

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

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

Table 1
Three spectral values as input parameters

Hidden neurodes	Best iteration	RCC	BCC	RMSE	Kappa coefficient	Overall accuracy
5	15000	77.66	88.96	0.2459	67.92	93.81
10	5000	73.29	88.85	0.2240	68.34	94.66
15	5000	73.60	89.97	0.2204	69.46	94.75
20	10000	74.79	90.85	0.2235	69.30	94.47

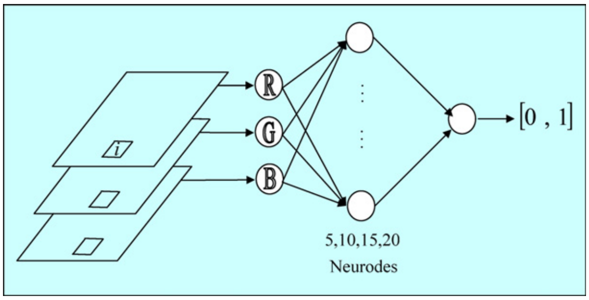

Fig. 3. Network structure when spectral values form input parameters.

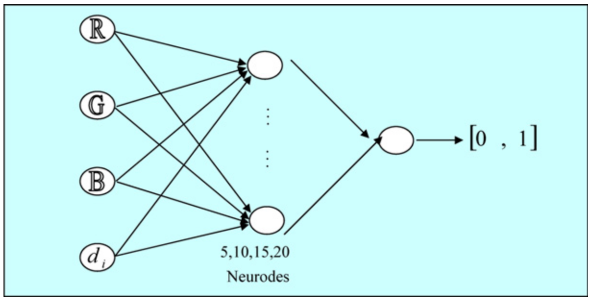

Fig. 4. Network structure when distance parameter is added to input parameters.

for each pixel as below:

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

$$d_i = \frac{1}{441.673}\left[(R_\text{m} - R_i)^2 + (G_\text{m} - G_i)^2 + (B_\text{m} - B_i)^2\right]^{1/2} \qquad (3)$$

While $[R_\text{m} \quad G_\text{m} \quad B_\text{m}]_i^\text{T}$ is the mean of road pixels in training set, 441.673 is the maximum imaginable distance in the spectral space.

This parameter is included in the input vector to inject more spectral information to the network in order to increase discrimination ability of the network between road and background pixels and to reduce the iteration times in learning stage.

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

While background comprises different objects with different spectral behaviors, distance parameter, accentuating spectral differences, has improved network ability in background detection and 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.

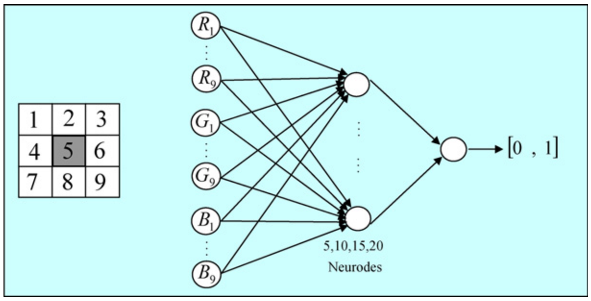

Fig. 5. Network structure when neighbor pixels form input parameters.

4.3. Participating neighbor pixels in input parameters

In this section, normalized spectral information in a 3×3 window around each pixel is extracted as nine red, nine green and nine blue values to form input parameters in that order. Accordingly, input layer involves 27 neurodes.

The motivation comes from emphasizing the homogeneity property of roads in high-resolution satellite images. Since homogeneity is a characteristic that can be recognized with respect to neighbor pixels, spectral information of all pixels in the designed window participates in input vector generation. Fig. 5 shows the network structure, while the results are presented in Table 3.

Table 2
Normalized distance as forth input parameter

Hidden neurodes	Best iteration	RCC	BCC	RMSE	Kappa coefficient	Overall accuracy
5	1000	71.43	92.60	0.2155	69.45	94.30
10	1000	71.50	92.31	0.2157	69.78	94.41
15	2000	72.46	92.94	0.2162	70.06	94.45
20	2000	72.52	93.26	0.2149	69.63	94.32

Table 3
Participating neighbor pixels in input parameters

Hidden neurodes	Best iteration	RCC	BCC	RMSE	Kappa coefficient	Overall accuracy
5	3000	80.56	86.78	0.2813	61.53	91.71
10	3000	80.99	86.51	0.2742	63.20	92.20
15	3000	80.42	87.57	0.2668	65.66	92.94
20	4000	80.80	88.31	0.2509	68.17	93.59

Table 4
Spatial information and normalized distance as input parameters

Hidden neurodes	Best iteration	RCC	BCC	RMSE	Kappa coefficient	Overall accuracy
5	1000	75.60	93.85	0.2011	71.31	95.11
10	1000	75.20	94.60	0.2002	71.60	95.15
15	1500	75.51	95.57	0.2014	71.86	95.15
20	1500	76.60	95.24	0.2008	72.18	95.17

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

As roads are presented like homogeneous areas in high-resolution satellite images, participating neighbor pixels in input parameters enables the network to extinguish road pixels more efficiently, causing the RCC parameter to improve.

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

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

4.4. Spatial information and normalized distance as input parameters

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

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

Finally, as a comparison with statistical methods, the results from maximum-likelihood classification method and the best network from Sections 4.1 and 4.4 are shown together in Fig. 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 Harbor in Iran was chosen as input image to evaluate network's behavior on these kinds of images.

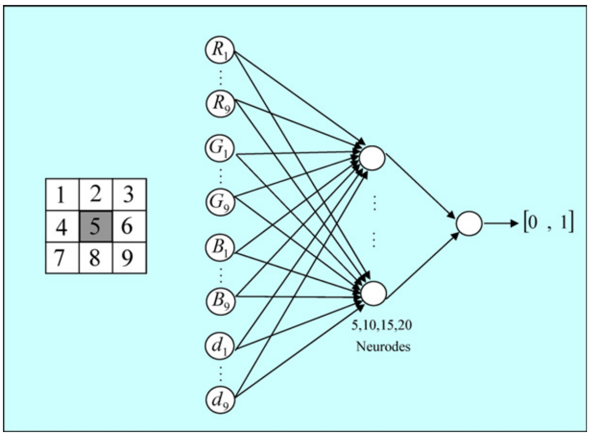

Fig. 6. Network structure when neighbor pixels with their normalized distances form input parameters.

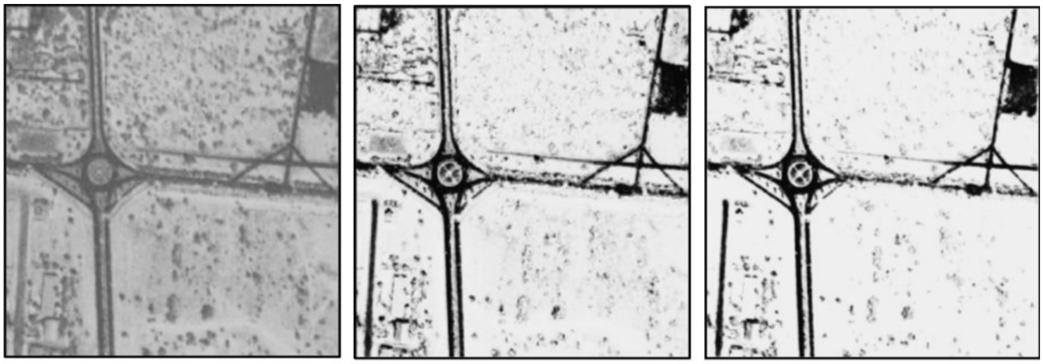

Fig. 7. Left image: maximum-likelihood result, middle image: simple BNN with 10 neurodes in hidden layer from Section 4.1, right image: improved BNN with 15 neurodes in hidden layer from Section 4.4.

Fig. 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 is implemented and therefore 36 neurodes are designed in input layer. The set is made up of spectral values and normalized distances of all pixels in the surrounding window.

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

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

Fig. 9a is the result of maximum-likelihood method, which uses the same training set and which has been shown for comparison between neural networks and

Fig. 8. (a) RGB Quick-Bird image from Bushehr Harbor in Iran and (b) manually produced reference map.

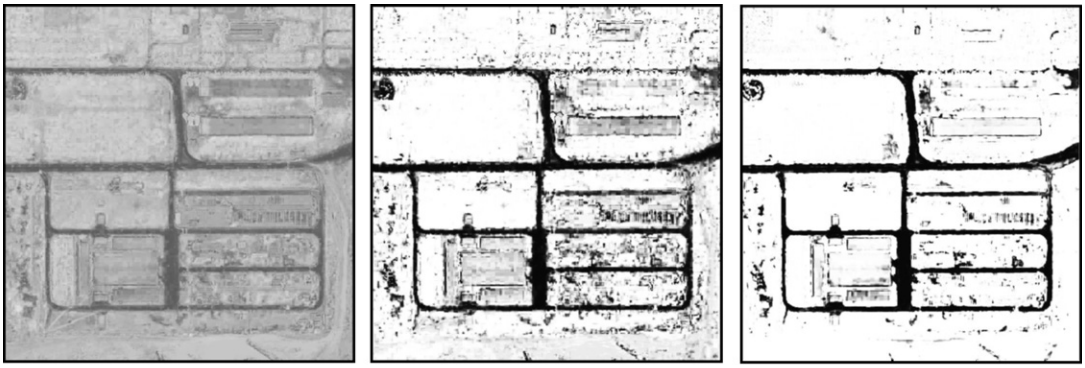

Fig. 9. (a) Maximum-likelihood results; (b) (3|6|1) neural network's results; (c) (36|10|1) network's results.

statistical methods. Fig. 9b is the result of optimum simple neural network that uses only the spectral information of a single pixel in its input layer. Fig. 9c is the answer to optimum improved network whose input vector consists of spectral values and normalized distances of all pixels in a 3×3 window.

6. Summary and discussion

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

As roads are homogeneous areas in high-resolution images, employing neighbor 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, also reducing the requested hidden layer size and iteration time.

It was found that there is no need to design more than 10 neurodes in hidden layer as it does not improve results noticeably. In fact, it 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.

Using edge information and texture parameters in input vector for network's training and recalling could be a good suggestion for further study in order to improve network's ability in road detection.

References

Benediktsson, J.A., Swain, P.H., Erosy, O.K., 1990. Neural network approaches versus statistical methods in classification of multi-source remote sensing data. IEEE Trans. Geosci. Remote Sensing 28, 540–551.

Civco, D.L., Waug, Y., 1994. Classification of multi-spectral, multi-temporal multi-source spatial data using artificial neural networks. In: Proceeding of the ASPRS 1994 Annual Convection, Reno, NV, USA, pp. 123–133.

Fiset, R., Cvayas, F., 1997. Automatic comparison of topographic map with remotely sensed images in a map updating perspective: the road network case. Int. J. Remote Sensing 18 (4), 991–1006.

Foody, G.M., McCulloch, M.B., Yates, W.B., 1995. Classification of remotely sensed data an artificial neural network: issues related to training data characteristics. Photogrammetr. Eng. Remote Sensing 61 (4), 391–401.

Gruen, A., Agouris, P., Li, H., 1995. Linear feature extraction with dynamic programming and globally enforced least squares matching. In: Gruen, A., Kuebler, O., Aqouris, P. (Eds.), Automatic Extraction of Man-Made Objects from Aerial and Space Images. Brikhauser, Basel, pp. 83–94.

Gruen, A., Li, H., 1997. Semi-automatic linear feature extraction by dynamic programming and LSB-snakes. Photogrammetr. Eng. Remote Sensing 63 (8), 985–995.

Heerman, P.D., Khazenie, N., 1992. Classification of multi-spectral remote sensing data using a back-propagation neural network. IEEE Trans. Geosci. Remote Sensing 30 (1), 81–88.

Jeon, B.K., Jang, J.H., Hong, K.S., 2002. Road detection in space born SAR images using a genetic algorithm. IEEE Trans. Geosci. Remote Sensing 40 (1), 22–29.

Mayer, H., Baumgartner, A., Steger, C., 1998. Road extraction from aerial imagery. In: CV Online: On-line Compendium of Computer Vision.

Mohammadzadeh, A., Tavakoli, A., Valadan Zoej, M.J., 2004. Automatic linear feature extraction of Iranian roads from high resolution multi-spectral satellite imagery. In: XXth ISPRS Congress, Istanbul, Turkey, pp. 764–768.

Neuenschwander, W., Fua, P., Szekely, G., Kubler, O., 1995. From Ziplock snakes to Velcro surfaces. In: Gruen, A., Kuebler, O., Aqouris, P. (Eds.), Automatic Extraction of Man-Made Objects from Aerial and Space Images. Brikhauser, Basel, pp. 105–114.

Paola, J.D., Schowengerdt, R.A., 1997. The effect of neural network structure on a multi-spectral land-use/land-cover classification. Photogrammetr. Eng. Remote Sensing 63 (5), 535–544.

Richard, J.A., 1993. Remote Sensing Digital Image Analysis: Introduction, second ed. Springer, New York, ISBN 0-387-5480-8.

Trinder, J., Li, H., 1995. Semi-automatic feature extraction by snakes. In: Gruen, A., Kuebler, O., Aqouris, P. (Eds.), Automatic Extraction of Man-Made Objects from Aerial and Space Images. Brikhauser, Basel, pp. 105–114.

Tupin, F., Maitre, H., Mangin, J.F., Nicolas, J.M., Pechersky, E., 1998. Detection of linear features in SAR images: applications to road network extraction. IEEE Trans. Geosci. Remote Sensing 36 (2), 434–453.

Vosselman, G., Knecht, J., 1995. Road tracing by profile matching and Kalman filtering. In: Gruen, A., Kuebler, O., Aqouris, P. (Eds.), Automatic Extraction of Man-Made Objects from Aerial and Space Images. Brikhauser, Basel, pp. 165–274.

Yang, G.Y., 1995. Geological mapping from multi-source data using neural networks. M.Sc. Thesis. University of Calgary, Canada.

Zhang, C., Baltsavias, E., 2000. Knowledge-based image analysis for 3D edge extraction and road reconstruction. Int. Arch. Photogrammetry Remote Sensing 33 (Part B3), 1008–1015.