Research Journal of Applied Sciences, Engineering and Technology 4(11): 1471-1477, 2012
ISSN: 2040-7467

Submitted: November 25, 2011 Accepted: January 04, 2012 Published: June 01, 2012

A Novel Approach for Text Extraction from Complex Natural Scene

[1,2]Tongcheng Huang and [2]Liping Yin
[1]Institute of Laser and Information Technology,
[2]Department of Information Engineering, Shaoyang University, Shaoyang 422004, China

Abstract: In this study, hybrid Particle Swarm Optimization (PSO)-based Wavelet Neural Network (WNN) for natural scene text extraction is presented. In our approach, wavelet transformation is done on the different current and the wavelet coefficients are obtained. The wavelet coefficients are given as inputs to the wavelet neural network trained by Particle Swarm Optimization (PSO-WNN). The final network output of real text regions is different from those non-text regions. The experimental results demonstrate the effectiveness of the proposed method for complex natural scene.

Key words: Natural scene text extraction, Particle Swarm Optimization (PSO), Wavelet Neural Network (WNN), wavelet theory

INTRODUCTION

In a society driven by visual information and with the drastic expansion of low-priced cameras, vision techniques are more and more considered and text recognition is nowadays a fast changing field, which is included in a large spectrum, named text understanding. Recently, handheld scanners such as pen-scanners appeared to acquire small parts of text on a fairly planar surface such as that of a business card. Issues having an impact on image processing are limited to sensor noise, skewed documents and inherent degradations to the document itself. Based on this classical acquisition method, Optical Character Recognition (OCR) systems have been designed for many years to reach a high level of recognition with constrained documents, meaning those falling into traditional layout, with relatively clean backgrounds such as regular letters, forms, faxes, checks and so on and with a sufficient resolution (at least 300 dots per inch (dpi)). With the recent explosion of handheld imaging devices (HIDs), i.e. digital cameras, standalone or embedded in cellular phones or Personal Digital Assistants (PDAs), research on document image analysis entered a new era where breakthroughs are required: traditional document analysis systems fail against this new and promising acquisition mode and main differences and reasons of failures will be detailed in this section. Small, light, and handy, these devices enable the removal of all constraints and all objects, such as natural scenes in different situations in streets, at home or in planes may be now acquired! Moreover, recent studies (Kim *et al*., 2005) announced a decline in scanner sales while projecting that sales of HIDs will keep increasing over the next 10 years.

Challenge of natural scene text extraction: First of all, in order to understand challenges of this field, new imaging conditions and newly considered scenes need to be detailed. The new imaging conditions deal with:

- **Raw sensor image and sensor noise:** in low-priced HIDs, pixels of a raw sensor are interpolated to produce real colours, which can induce degradations. Demosaicing techniques, viewed more as complex interpolation techniques, are sometimes required. Moreover, sensor noise of an HID is usually higher than that of a scanner.
- **Viewing angle:** scene text and HIDs are not necessarily parallel creating perspective to correct.
- **Blur:** during acquisition, some motion blur can appear or be created by a moving object. All other kinds of blur, such as wrong focus, may also degrade even more image quality.
- **Lighting:** in real images, real (uneven) lighting, shadowing, reflections onto objects, respectively inter-reflections between objects may make colours vary drastically and decrease analysis performance.
- **Resolution and aliasing:** from webcam to professional cameras, resolution range is large and images with low resolution must also be taken into account. Resolution may be below 50 dpi which causes commercial OCR to fail. It may lead to aliasing creating fringed artefacts in the image. The newly considered scenes represent targets such as:
- **Outdoor/non-paper objects:** different materials cause different surface reflections leading to various degradations and creating inter-reflections between objects.

Corresponding Author: Tongcheng Huang, Institute of Laser and Information Technology, Shaoyang University, Shaoyang 422004, China

Fig. 1: Samples of natural scene images

- **Scene text:** backgrounds are not necessarily clean and white and more complex ones make text exraction from background difficult. Moreover scene text such as that seen in advertisements may include artistic fonts.
- **Non-planar objects:** text embedded in bottles or cans suffer from deformation.
- **Unknown layout:** there is no a priori information on structure of text to detect it efficiently.
- **Objects in distance:** distance between text and HIDs can vary, and character sizes may vary in a wide range, leading to a wide range of character sizes in a same scene.

The main challenge is to design a system as versatile as possible to handle all variability in daily life, meaning variable targets with unknown layout, scene text, several character fonts and sizes and variability in imaging conditions with uneven lighting, shadowing and aliasing. Our proposed solutions for each text understanding step must be context independent, meaning independent of scenes, colours, lighting and all various conditions. Hence we focus on methods which work reliably across the broadest possible range of natural scene images, such as displayed in Fig. 1.

Numerous application: As HIDs become more and more powerful, on-the-fly image processing becomes possible, opening up a new range of applications. Nevertheless, today's HIDs are easily connected to various networks and supplementary computing resources. Starting from sign recognition for foreigners for the 2008 Olympic Games in Beijing, automatic license plate recognition to driver assisted systems with text projection on windshields, various situations could be handled. Interesting applications such as mobile phones operating as fax machines even led to strict sanctions in Japanese bookstores!

Visually impaired people are directly affected by such research (Thillou *et al*., 2005). With an HID and sufficient resources, scene in daily life may be analyzed to give them access to text and, coupled with a text-to-speech algorithm, make them "read" book covers, banknotes, labels on office doors, medicine labels and so on. For the blind community, such devices are really expected. Another promising application is the one of visual landmark-based robot navigation.

Several kinds of robot navigation may be listed such as dead-reckoning, map-based navigation, positioning sensor-based navigation or landmark-based navigation, which can be divided into natural and artificial landmarks. Natural landmarks may be designed on purpose for indoor robot navigation, such as room numbers (Mata *et al*., 2001), but may also be part of real life such as natural scenes. Even if conditions of navigation are still constrained, natural landmark-based one is very promising and satisfying results already appeared. Hence either nameplates, information signs or any text embedded in images contain large quantities of useful semantic information. Text understanding may be useful in high level robot navigation, such as path planning or goal driven navigation. Applications are very numerous and currently only limited by imagination. Scene text is an important feature to understand for all these applications.

Overview of this paper: For the last decade, the wavelet neural network(WNN) method was noticed by many researchers (Yao *et al*., 1996; Eric and Griff, 2002). The wavelet theory (Daubechies, 1990; Mallat, 1989) provides a multiresolution approximation for discriminate functions. The WNN can thus exhibit better performance in function learning than the conventional feed-forward neural networks. Researchers have successfully applied WNNs in function approximation Huang (2008), motor drive control (Wai and Chang, 2003; Wai *et al*., 2003), robotics (Yoo *et al*., 2006) and Video OCR (Jing *et al*., 2002; Huang, 2008). Using neural networks to achieve learning (Ham and Kostanic, 2001; Widrow and Lehr, 1990) usually involves two steps, i.e., designing a network structure and deriving an algorithm for the learning process. The structure of the neural network governs the nonlinearity of the modeled function. The learning algorithm determines the rules for optimizing the weight values of the network within the training period. A typical WNN structure offers a fixed set of weights after the learning process. This single set of weights is used to capture the characteristics of all input data. However, a fixed set of weights may not be enough to learn the data set if the data are separately distributed in a vast domain and/or the number of network parameters is too small. It is prone to cause the curse of dimensionality with the factors taken into consideration increasing, which becomes the bottleneck for the improvement of its applications (Zhang *et al*., 2005; Lu *et al*., 2003).In this study, a new method for optimizing the structure of wavelet networks was developed by adopting a Particle

Swarm Optimization Algorithm (PSO) to optimize the time-frequency phase spot. Based on the operational data, the application of this method in complex natural scene text extraction indicates that it can overcome the defects of WNN with the advantage of higher responding speed and simpler network structure. Therefore it is more effective in complex natural scene text extraction.

In this study, we presented a hybrid Particle Swarm Optimization (PSO)-based Wavelet Neural Network (WNN) for natural scene text extraction. Firstly, we propose the PSO algorithm to train the wavelet neural networks. Meanwhile, the algorithm is applied in complex natural scene text extraction. Then, the background of the PSO is introduced. After that, the WNN is given and the PSO-WNN structure is put forward. At last, the experimental simulation results in complex natural scene text extraction is presented which show the effectiveness of the proposed method.

PSO algorithm: Particle Swarm Optimization (PSO) method stemmed from the research on the prey behavior of a swarm of bird (Kennedy and Ebergart, 1995) in 1995 and has been compared to genetic algorithms for efficiently finding optimal or near-optimal solutions in large search spaces. In PSO (Kennedy, 2000; Jui-Fang *et al.*, 2005), a point in the problem space is called a particle, which is initialized with a random position and search velocity. As in evolutionary computation paradigms, the concept of fitness is employed and candidate solutions to the problem are termed particles or sometimes individuals, each of which adjusts its flying based on the flying experiences of both itself and its companions. It keeps track of its coordinates in hyperspace which are associated with its previous best fitness solution, and also of its counterpart corresponding to the overall best value acquired thus far by any other particle in the population. Vectors are taken as presentation of particles since most optimization problems are convenient for such variable presentations.

In standard PSO algorithm, the PSO formulae define each particle as a potential solution to a problem in the M-dimensional space, and the position and the velocity of the ith particle can be represented as $\vec{x}_i = (x_{i1}, x_{i2}, ..., x_{iM})$ and $\vec{V}_i = (v_{i1}, v_{i2}, ..., v_{iM})$ respectively. Each particle also maintains the previous best position P_i = (p_{i1} , p_{i2} ,..., p_{iM}).The global best particle, which represents the fittest particle found in the entire swarm is denoted by P_{best1} = (p_{best1} , p_{best2} ,..., p_{bestM}). The new velocity of each particle is calculated according to the following formula (1):

$$v_{iM(best+1)} = \omega_i v_{iM(best)} + c_1\gamma_1(p_{iM} - x_{iM}) + c_2\gamma_2(P_{bestM} - x_{iM}) \quad (1)$$

where c_1 and c_2 are constants and are known as acceleration coefficients, ω_i is called the inertia factor ; γ_1 and γ_2 are two sparely generated uniformly distributed random numbers in the range of [0,1]; $v_{iM(best)}$ denotes current velocity of the Mth dimensional in the ith particle at iteration generation best; best denotes current iteration number. At each iteration (generation) the position of each particle is updated according to the following formula (2):

$$x_{iM} = x_{iM} + v_{iM} \quad (2)$$

The inertia weight ω is introduced in to improve PSO performance. Suitable selection of inertia weight ω provides a balance between global and local exploration and exploitation. The inertia weight ω is set according to the following formula (3):

$$\omega = \omega_{\max} - \frac{\omega_{\max} - \omega_{\min}}{MaxIter} Iter \quad (3)$$

where ω_{max} and ω_{min} are the initial weight and the final weight; MaxIter is the maximum iteration number, Iter is the current iteration number.

Wavelet neural network: Wavelet neural network (WNN) is a new network that is based on wavelet theory in which a discrete wavelet function is used as the node activation function. From an image application point of view, the proposed method, wavelet networks, is divided into a series of stages for further concerning, each of which is described in the follows. The WNN has a three-layer structure with n_{in} nodes in the input layer, n_h nodes in the hidden layer, and n_{out} nodes in the output layer, described as in the left part of Fig. 2.

The first part is choosing the mother wavelet. The Mexican hat wavelet (Min *et al.*, 2005) is selected as the basis in this study. It is defined by:

$$\psi(x) = (1-x^2)\, e^{-x_2/2} \quad (4)$$

The input to the kth neutron is:

$$S_k = \sum_{j=1}^{n_{in}} W_{j,k} \times x_j \quad (5)$$

where the x_j's, j=1,2, …,n_{in}, are the input variables, and $W_{j,k}$ denotes the weight of the link between the *i*th input and the kth hidden nodes. In order to control the magnitude and the position of the wavelet, the multiscaled wavelet function is used as the hidden node transfer function. The dilation parameter *a* of the first hidden node

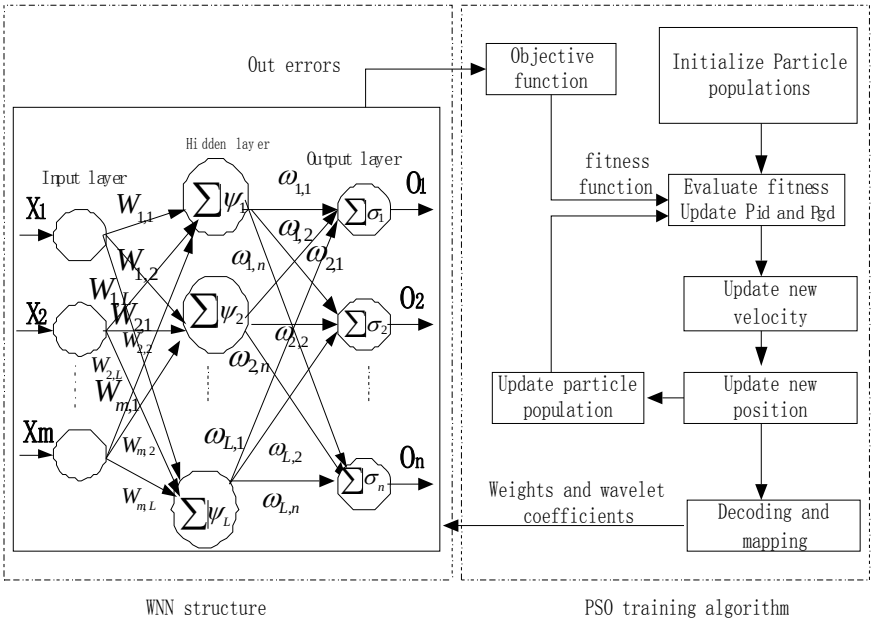


Fig. 2: Block diagram representation of PSO-WNN

(j = 1) is set as 1, i.e., $\psi_{1,b1}(x) = \psi(x - b1)$. For the second hidden node (j=2), the dilation parameter a is set as 2, i.e., $\psi_{2,b2}(x) = (1/\sqrt{2})\psi((x - b2)/2)$, where the output of the wavelet is scaled down by $1/\sqrt{2}$. Similarly, for the jth hidden node, the dilation parameter a is set as j. Hence, the output of the hidden layer of the WNN is given by:

$$\psi_{k,b_k} = \frac{1}{\sqrt{k}}\psi\left(\frac{S_k - b_k}{k}\right) \tag{6}$$

The output of the kth neutron is:

$$\psi_{k,b_k} = \frac{1}{\sqrt{k}}\left(1 - \left(\frac{S_k - b_k}{k}\right)^2\right) e^{-\frac{(\frac{s_k - b_k}{k})^2}{2}} \tag{7}$$

The output of the WNN is defined as:

$$o_I = \sum_{k=1}^{n_k} \psi_{k,b_k}(s(k))\omega_{Ik} = \sum_{k=1}^{n_k} \psi_{k,b_k}(\sum_{j=1}^{n_{in}} W_{j,k} \times x_j)\omega_{I,k} \tag{8}$$

Where $\omega_{l,k}$,k=1,2, , n_h and l=1,2, …,n_{out}, denotes the weight of the link between the kth hidden and lth output nodes.

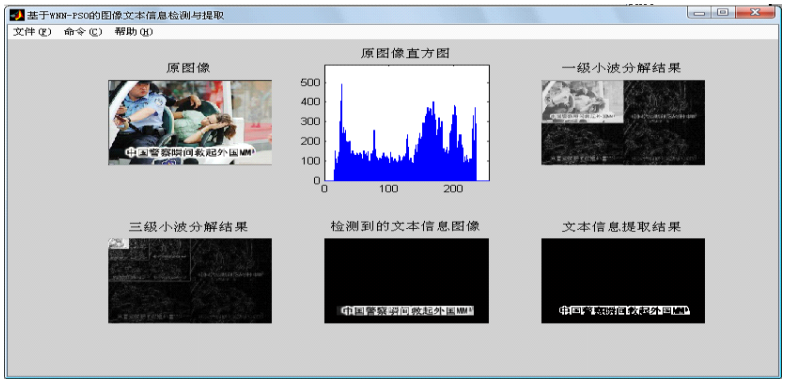
Not only weights but also both translating and scaling parameters are adjusted to minimize the error function during the training stage:

$$E = \frac{1}{2}\sum_{i=1}^{I}\sum_{k}^{K}\left(D_{ik} - O_{ik}\right)^2 \tag{9}$$

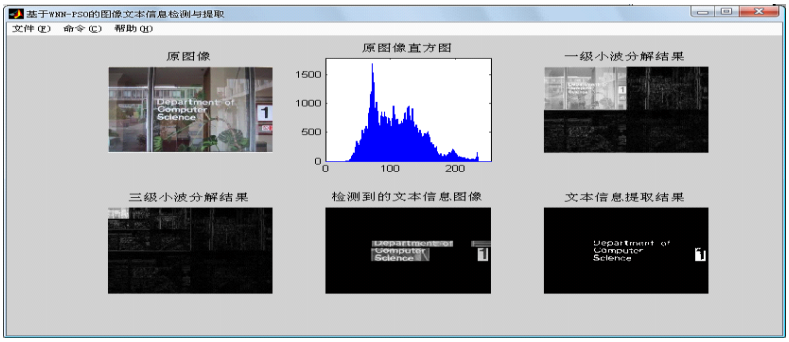
where I = 1,…,I is the number of training patterns, k = 1,…,K is the number of the objective, D_{ik} and O_{ik} represent the desired output value of $Node_{ik}$ and the actual net output value respectively.

Training with PSO: In the above mentioned WNN, the most commonly used study algorithm is the BP algorithm,BP algorithm essentially is the gradient degrade law, it is a partial search algorithm, namely searching partial optimum point in the partial scope along the most superior direction using the gradient as the inspiration information. However gradient degrade algorithm makes the network getting into partial minimum easily, thus causing the network training result to be unsatisfactory, the success probability of search is low, this makes BP network to have certain limitation in the application.

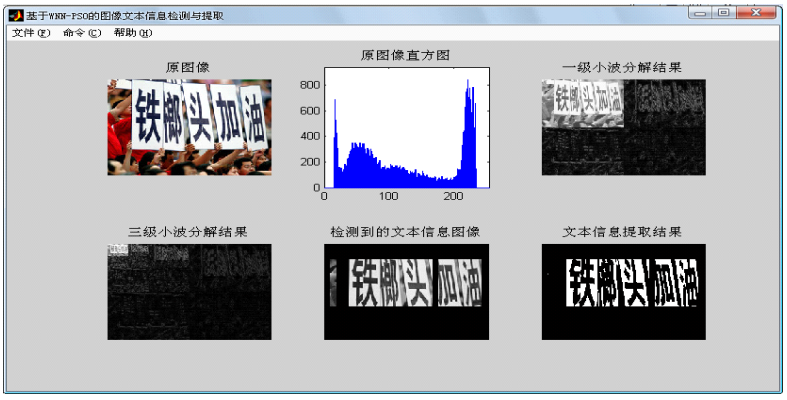
Based on the defects of the gradient degrade algorithm and the principle and advantage of PSO, the PSO algorithm is used to train WNN and the weights and bias values are updated. The structure of PSO-WNN is based on Fig. 1. Define the particle swarm's position vector X= (ω, a, b), in which ω is the connecting weight value of ever layer, a is the stretching parameter and b is translation parameter. The initial number of particle is 100. The position vector of each particle is:



(a)



(b)



(c)

Fig. 3: Some experimental results

present[i] = [ω_1, ..., ω_j, a_1, ..., a_j , b_1,..., b_j] ,i = 1, …, 100

There j is the nerve cell number of hidden layer. The fitness is defined as formula (10), ($\hat{y}$) is actual output after k times iteration, *y* is the ideal output:

$$J(k) = \sum_{m=1}^{M} e_m^k = \sum_{m=1}^{M} \left(y_m - \hat{y}_m^k\right)^2 , k = 1,2,\ldots,K \quad (10)$$

Where n is the sampling number of nonlinear function, $\hat{y}_m^k$ is the actual output of k^{th} iteration and of m^{th} input. The flowchart of the wavelet neural network based on PSO is as fellow:

- Define the initial value of the WNN parameters. Initial the position vector, velocity vector V, objective error *Egoal* = 1.0, and maximum iteration

number T_{max} of 100 particles, in which each particle producing at random. Establishing the minimum and maximum value position vector ω_j a_j b_j j = 1, 2, …, j^* of every particles, defining the minimum and maximum value of velocity vector V_m, m=1, 2, …, 100. Initial a_j as: producing aj with the number of j* in sector [0, 1] at random, then match aj in sector [0, 0.5] to [1, 5].

- Using formula (1), (2) to renew the position vector present and velocity V of every particle, and record the most superior position $pbest_i^k$ ($pbest_i^k$ is the k^{th} superior position vector of particle i) and record the most superior position vector $gbest^k$ overall, in addition record the fitness $Epbest_i^i$ and $Egbest^d$ according to $pbest_i^k$ and $dbest^k$.
- If $Egbest^d \leq Egoal$ or the maximum iteration number is T_{max}, the training is over. Else turn to 2).
- Substitute gbest (the overall most superior number) for the weight value of the wavelet network to calculate the output of WNN and gain the fitness curve.
- During simulation, the accelerating factors are $c_1 = c_2 = 1.4944$, the inertia factor ω is 0.7.

EXPERIMENTAL RESULTS

Experiments are performed on static images and video sequences. The frame size is 1024×768 in BMP or MPEG format. We convert the colored frames into gray-level before applying the proposed method. In Fig. 3a, b, c, the results of the proposed algorithm are illustrated step by step. The original images are decomposed into one average component sub-band and three detail component sub-bands. Those detail component sub-bands contain the key features of text regions. According to these features, the text regions are obtained using a neural network. The overall recognition results are satisfactory for use in complex natural scene text understanding. Performing the proposed method on natural scene video and combining its results with other video understanding techniques will improve the overall understanding of the natural scene.

CONCLUSION

In this study, a method for complex natural scene text extraction using wavelet neural networks and particle swarm optimization has been proposed. Due to the variable translation parameters in the network, the WNN becomes adaptive and is able to improve the learning ability of the neural network. Experimental results have been given to show the improved performance and solution stability of the WNN and the hybrid PSO. One limitation in this paper is that the choice of suitable parameters for the PSO and the neural network is quite difficult. Most parameters are determined by trial and error through experiments. Some possible future research directions can be identified. Multiobjective PSO could be studied, which are particularly good to handle complex natural scene text extraction and other multiobjective optimization problems.

ACKNOWLEDGMENT

This study is supported by Key Lab of Intelligent Information Processing, Chinese Academy of Sciences (IIP-2006-5), Hunan Provincial Natural Science Foundation (06JJ2031) and the Provincial Scientific Research fund (06A065) of Hunan Provincial Education Department.

REFERENCES

Daubechies, I., 1990. The wavelet transform, time-frequency locatization and signal analysis [J]. IEEE Trans. Inf. Theory, 36(5): 961-1005.

Eric, A.R. and L.B. Griff, 2002. Focused local learning with wavelet neural networks [J]. IEEE T. Neural. Networ., 13(2): 304-319.

Ham, F.M. and I. Kostanic, 2001. Principles of Neurocomputing for Science & Engineering [M]. McGraw-Hill, New York.

Huang, T.C., 2008. Research on VOCR and HOCR based on wavelet neural network theory [D]. Ph.D. Thesis, Shanghai University, Shanghai, China.

Jing, Z., C. Xilin, *et al*., 2002. A Robust Approach for Recognition of Text Embedded in Natural Scenes [C]. Proceedings of the 16th International Conference on Pattern Recognition, Quebec City, Canada, August 11-15, 3: 204-207.

Jui-Fang, C., C. Shu-Chuan, F.R. John and P. Jeng-Shyang, 2005. A parallel particle swarm optimization algorithm with communication strategies [J]. J. Inf. Sci. Engine., 21(4): 809-818.

Kennedy, J. and R.C. Ebergart, 1995. Particle swarm optimization [C].IEEE Proceedings of the 6th conference on neural networks. IEEE Neural Networks Council, pp: 1942-1948.

Kennedy, J., 2000. Stereotyping: Improving particle swarm performance with cluster analysis [C]. IEEE Proceedings of the International Conference on Evolutionary Computation, pp: 303-308.

Kim, J., S. Park and S. Kim, 2005. Text locating from natural scene images using image intensities [C]. Proceedings of Internation Conference Document Analysis and Recognition, pp: 655-659.

Lu, W.Z., H.Y. Fan and S.M. Lo, 2003. Application of evolutionary neural network method in predicting pollutant levels in downtown area of Hong Kong. Neurocomputing, 51(4): 387-400.

Mallat, S.G., 1989. A Theory for multiresolution signal decomposition:The wavelet representation. IEEE. Pattern Anal. 11(7): 674-693.

Mata, M., J.M. Armingol, A. Escalera and M.A. Salichs, 2001. A visual landmark recognition system for topologic navigation of mobile robots. Proceedings of Int. Conference on Robotics and Automation, pp: 1124-1129.

Min, H.Q., J.H. Zhu and X.J. Zheng, 2005. Obstacle avoidance with multiobjective optimization by PSO in dynamic environment [C]. Proc. Mach Learn. Cybern. Aug., 5: 2950-2956.

Thillou, C., S. Ferreira and B. Gosselin, 2005. An embedded application for degraded text recognition [J]. Eurasip Jour. Appl. Si. Pr. (Special Issue on Advances in Intelligent Vision Systems: Methods and applications,) 13: 2127-2135.

Wai, R.J. and J.M. Chang, 2003. Implementation of robust wavelet neural network sliding-mode control for induction servo motor Drive [J]. IEEE Trans. Ind. Electron., 50(6): 1317-1334.

Wai, R.J., R.Y. Duen, J.D. Lee and H.H. Chang, 2003. Wavelet neural network control for induction motor drive using sliding-mode design technique [J]. IEEE Trans. Ind. Electron. Aug., 50(4): 733-748.

Widrow, B. and M.A. Lehr, 1990. 30 years of adaptive neural networks: Perceptron, madaline and back propagation[C]. Proc. IEEE Sep., 78(9): 1415-1442.

Yao, S., C.J. Wei and Z.Y. He, 1996. Evolving wavelet neural networks for function approximation. Electron. Lett., 32(4).

Yoo, S.J., J.B. Park and Y.H. Choi, 2006. Adaptive dynamic surface control of flexible-joint robots using self-recurrent wavelet neural networks [J]. IEEE T. Syst. Man Cybern. B., 36(6): 1342-1355.

Zhang, J.M., S.H. Yuan and K.M. Xie, 2005. Application of artificial neural network in teaching quality assessment system [J]. Journal of Taiyuan University of Technology, Taiyuan City, China, pp: 36:37-39.