CORE
Metadata, citation and similar papers at core.ac.uk
Provided by TELKOMNIKA (Telecommunication Computing Electronics and Control)

TELKOMNIKA, Vol.13, No.2, June 2015, pp. 587 ~ 596
ISSN: 1693-6930, accredited A by DIKTI, Decree No: 58/DIKTI/Kep/2013
DOI: 10.12928/TELKOMNIKA.v13i2.1430

An Image Compression Method Based on Wavelet Transform and Neural Network

Suqing Zhang, Aiqiang Wang*
Information Engineer Department, Henan Vocational and Technical Institute,
Zhengzhou 450046, Henan, China
*Corresponding author, e-mail: 929795121@qq.com

Abstract

Image compression is to compress the redundancy between the pixels as much as possible by using the correlation between the neighborhood pixels so as to reduce the transmission bandwidth and the storage space. This paper applies the integration of wavelet analysis and artificial neural network in the image compression, discusses its performance in the image compression theoretically, analyzes the multi-resolution analysis thought, constructs a wavelet neural network model which is used in the improved image compression and gives the corresponding algorithm. Only the weight in the output layer of the wavelet neural network needs training while the weight of the input layer can be determined according to the relationship between the interval of the sampling points and the interval of the compactly-supported intervals. Once determined, training is unnecessary, in this way, it accelerates the training speed of the wavelet neural network and solves the problem that it is difficult to determine the nodes of the hidden layer in the traditional neural network. The computer simulation experiment shows that the algorithm of this paper has more excellent compression effect than the traditional neural network method.

Keywords: *Image Compression, Wavelet Analysis, Artificial Neural Network*

1. Introduction

Image compression is a technology which uses the minimum bit number to represent the image information with no or little distortion while ensuring the image quality. The process of image compression is to look for an appropriate encoding or transform method to reduce the data volume which can represent this image [1]. The starting point to compress the image data volume is to reduce the redundant data to represent the image. The image is stored in the machine in the form of data matrix, therefore, a series of transform is conducted on that data matrix to reduce the redundant part. Make effective coding on the processed data to reduce the coding space. When reading the image in the follow-up phase, get the original image after inverse transformation processing [2]. As a basic technology of image processing, image compression involves every link of image processing. At present, image compression plays an important role in the satellite imagery, space exploration, teleconference and medical imaging, since the image has a large data volume and higher requirements on the real time [3].

The research of image compression started from the pulse code modulation (PCM), which was proposed on the transmission of television image in 1948. The research in the 1950s and 1860s was limited to the intraframe coding of the image. Starting from the late 1960s, orthogonal transform and other methods had been brought forth and preliminary exploration had been made on the intraframe coding of the image (namely the moving image coding). The year of 1988 was a greatly significant year in the development of image compression coding when the video compression standard H.261 and the framework principle of the still image compression JPEG were basically determined and progress had been made on the fractal and neural network in the image compression coding [4]. With an increasing demand for application, the traditional compression methods have failed to meet the requirements of image processing in the compression efficiency and effect, therefore, high-quality and high-efficient smart image compression algorithm has become an emphasis and objective of international research. People have begun to break through the original coding theory and search for some new coding approaches to obtain a higher compression ratio and a better compression quality. There are mainly two research thoughts: one is to realize the existing compression algorithms with new technology with higher precision and the other is to look for brand-new image compression theory, algorithm and corresponding realization technology [5].

Received January 17, 2015; Revised March 29, 2015; Accepted April 20, 2015

This paper integrates wavelet theory and artificial neural network (ANN), replaces the excitation function in the neural network with wavelet function and applies the advantages of multi-resolution analysis into the neural network so as to obtain a more flexible network design and better network performance. It has the advantages such as large-scale parallel processing and distributed information storage as well as excellent adaptivity, self-organization, fault tolerance, learning function and associative memory function. This paper firstly introduces the basic principle of image compression. Then it elaborates and integrates wavelet analysis and ANN. In the wavelet neural network, training is only needed in the weight of the output layer while that of the input layer can be determined according to the relationship between the interval of the sampling points and the interval of the wavelet compactly-supported interval. Once determined, no training is necessary; thus, it greatly accelerates the training speed of wavelet neural network and solves the problem that it is difficult to determine the nodes in the hidden layer of the traditional neural network. The final part is the experimental simulation and analysis.

2. Image Compression Mechanism

The purpose of digital image compression is to reduce the necessary bit number to represent the image and represent the image more effectively so as to facilitate the image processing, storage and transmission. The compression of time domain can accelerate the transmission of various information source more quickly, more parallel operations can be opened in the existing main lines of communciation through the compression of frequency domain;the compression of energy domain can reduce the transmitter efficiency and the compression of space domain can compress the data storage space. In the image data, there are plenty of redundancies, including space redundancy, structural redundancy, knowledge redundancy, information entropy redundancy and visual redundancy, which makes it possible to transform a large digital image file into small digital image files so as to achieve the purpose of image compression through the reduction of redundant data.To reduce the image information redundancy by fully utilizing the visual characteristics of human eyes and the statistical characteristics of the image can ensure the image quality [6].

After processing an image with the process of Figure 1, the restored image is a lossy image with certain compression. In the entire processing, compression is generated from the quantization process and the coding process and the selection of quantization methods and the quantization effect directly affect the final image compression result.The commonly-used quantization methods include: scalar quantization, linear quantization, vector quantization and the mixed quantization coding method of different quantization methods adopted by the low-frequency and high-frequency sub-bands. The frequently-used coding methods include: Huffman coding, run-length coding, arithmetic coding and predictive coding suitable for still image [7].

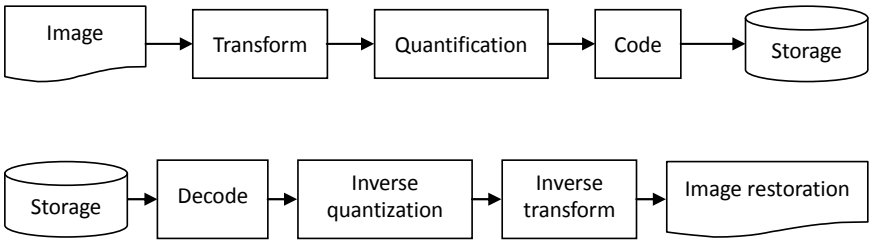

Figure 1. Image encoding and decoding process

3. Wavelet Neural Network

3.1. Wavelet Analysis

Wavelet, namely the wave in the small region, is a special waveform with limited length and an average value of 0. It has two features. One is small, in other words, it has compact support or approximate compact support in the time domain and the other is the alternatively positive and negative volatility, namely that the tributary component is 0.

3.1.1. Continuous wavelet transform (CWT)

Expand any function $f(t)$ in the space of $L^2(R)$ in the wavelet basis, call it as CWT of the function $f(t)$ and the transform formula is:

$$WT_f(a,b) =< f, \psi_{a,b} >= \frac{1}{\sqrt{|a|}} \int_R f(t) \cdot \overline{\psi\left(\frac{t-b}{a}\right)} dt \quad (1)$$

If the tolerance condition of the wavelet is satisfied, its inverse transformation is:

$$f(t) = \frac{1}{C_\psi} \int_{-\infty}^{+\infty} \frac{da}{a^2} \int_{-\infty}^{+\infty} WT_f(a,b) \cdot \psi\left(\frac{t-b}{a}\right) db \quad (2)$$

In this formula, $C_\psi = \int_R \frac{|\overline{\psi}(\overline{w})|^2}{|w|} dw \prec \infty$ is the tolerance condition of $\psi(t)$.

We can see it this way, Fourier analysis is to decompose the signals into the overlapping of a series of sine wave with different frequencies, likewise, and wavelet analysis is to decompose the signals into the overlapping of a series of wavelet functions. These wavelet functions are obtained from a mother wavelet function after the translation and scale. The wavelet analysis is better than Fourier analysis in that it has excellent localization nature in both time domain and frequency domain. Besides, since gradually-refined time-domain or frequency-domain sampling step-length is adopted in the high-frequency component, any detail of the objects can be focused [8].

3.1.2. Discrete wavelet transform

The image information is stored in the computer in the form of discrete signals, so it needs to discretize the continuous wavelet transform.

(i) Discretization of Scale and Translation

Discretize the scale factor a and the translation factor b of the continuous wavelet basis function $\psi_{a,b}(t)$ and get the discrete wavelet transform $WT_f(a,b)$ so as to reduce the redundancy of the wavelet transform coefficient. Discretize the scale factor a and the translation factor b in a power series, namely $a = a_0^m, b = b_0^m$ (m is an integer, $a_0 \neq 1$, but normally it is assumed that $a_0 \succ 1$) and get the following discrete wavelet function:

$$\psi_{m,n}(t) = \frac{1}{\sqrt{|a_0|}} \psi\left(\frac{t - n a_0^m b_0}{a_0^m}\right) = \frac{1}{\sqrt{|a_0|}} \psi\left(a_0^{-m} t - n b_0\right) \quad (3)$$

Its coefficient of correspondence is:

$$C_{m,n} =< f(t), \psi_{m,n} >= \int_{-\infty}^{+\infty} f(t) \overline{\psi_{m,n}(t)} dt \quad (4)$$

(ii) Binary Wavelet Transform

Binary wavelet transform is a special discrete wavelet transform. Assume $a_0 = 2$, $b_0 = 1$, and $\psi_{m,n} = 2^{-\frac{m}{2}} \psi\left(2^{-m} t - n\right)$.

The discrete wavelet transform is:

$$WT_f(m,n) =< m,n >= \int_{-\infty}^{+\infty} f(t) \psi_{m,n}(t) dt \quad (5)$$

The discrete binary wavelet transform is:

$$WT_f(m,n) =< m,n >= \int_{-\infty}^{+\infty} f(t) \psi_{m,n}(t) dt \quad (6)$$

3.1.3. Multi-resolution analysis

The concept of multi-resolution analysis is proposed when constructing orthogonal wavelet basis in Mallat, explaining the multi-resolution property of the wavelet from the concept of space and unify all the previous construction methods of orthogonal wavelet basis. The role of Mallat algorithm in the wavelet analysis is equal to the role of fast Fourier transform in the classic Fourier analysis.

Multi-resolution analysis can be vividly expressed as a group of nested multi-resolution sub-space [9]. Please see the Figure 2.

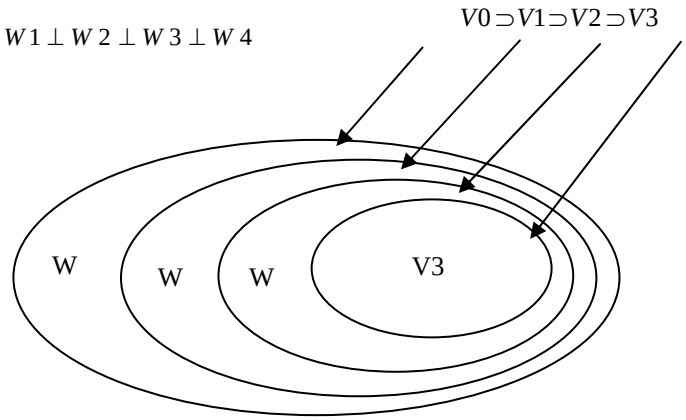

Figure 2. Nested multi-resolution sub-space

Assume that the frequency space of the original signal is V_0 and then it is decomposed into 2 sub-spaces: the low-frequency V_1 and the high-frequency W_1 after the first level of decomposition and V_1 is decomposed into the low-frequency V_2 and the high-frequency W_2 after the second level of decomposition.The decomposition process of such sub-space can be recorded as:

$$V_0 = V_1 \oplus W_1, V_1 = V_2 \oplus W_2, V_2 = V_3 \oplus W_3, \cdots, V_{N-1} = V_N \oplus W_N$$

Here, the symbol $\oplus$ refers to the orthogonal sum of two sub-spaces, V_f is the corresponding multi-resolution analysis sub-space to resolution 2^{-j}, the vector space W_j constituted by the dilation and translation of the corresponding wavelet function to the scaling function is the orthogonal complementary space of V_j, every W_j reflects the high-frequency sub-space of V_{j-1} space signal details, and V_j reflects the low-frequency sub-space of V_{j-1} space signal approximation. The following characteristics of sub-space can be obtained from the discrete wavelet frame:

$$V_0 = V_1 \oplus W_1 = V_2 \oplus W_2 \oplus W_1 = V_N \oplus W_N \oplus W_{N-1} \oplus \cdots W_2 \oplus W_1$$

This result demonstrates that limited sub-spaces can be approximate to the multi-resolution analysis sub-space V_0 with a resolution of 2^0=1.

3.2. Artificial Neural Network

Artificial neural network (ANN) is a complicated network system which is extensively interconnected by a large number of simple processing units similar to neuron. It is proposed on the basis of the research results of modern parallel neurology. It reflects some characteristics of human brain, however, it is not an actual description of neural network but its simplification,

abstraction and simulation. It presents the learning, summarization and classification features similar to human brain through the adjustments of interconnection strength. Therefore, the fundamental objective of neural network research is to explore the mechanism the human brain processes, stores and searches inforamtion so as to search the possibility to apply this principle to various signal processing [10]. The principle of artificial neural network is shown in Figure 3.

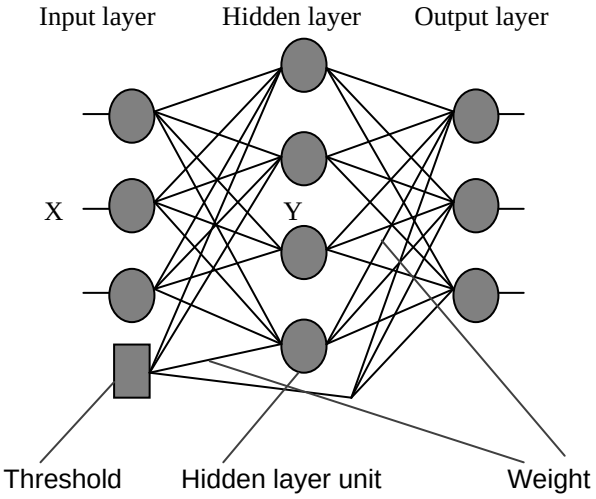

Figure 3. Principle of artificial neural network

ANN is a non-linear and self-adaptive information processing system interconnected by many processing units. It is raised based on the research results of modern neurology and it processes the information by simulating the way the brain neuralnetwork processes and memorizes information.

Artificial neural network has the following four basic characteristics:

(i) Non-linearity. Non-linear relationship is the general characteristic in the natural world. The brain wisdom is a non-linear phenomenon. Artificial neuron is either in the activation or suppression state, which is a kind of non-linear relationship mathematically. The network formed by the neurons with thresholds has better performance, which can enhance the fault tolerance and storage capacity.

(ii) Non-limitation. A neural network is usually extensively interconnected by many neurons. The overall behavior of a system not only depends on the characteristics of a signal neuron, but it may also determined by the interaction and interconnection of the main units. ANN simulates the non-limitation of the brain through the various interconnection of the units. Associative memory is a typical example of non-limitation.

(iii) Non-qualitation. ANN has self-adaptive, self-organization and self-learning capacities. Neural network can not only process the information with various changes, but the non-linear dynamic system changes continuously in the information processing. The iteration process is frequently adopted to describe the evolution process of the dynamic system.

(iv) Non-convexity. The evolution direction of a system depends on certain specific state function in a certain condition. For example, the extremum of energy function corresponds to the stable state of the system. Non-convexity means that such function has several extremums, therefore, the system has several stable equilibrium states, which will result in the diversity of system evolution [11].

3.3 Wavelet Neural Network Mechanism

As a newly-emerging mathematical modeling analysis method, wavelet neural network is a substitute for the feedforward neural network to approximate any function transform and its basic thought is to use waveron to replace neuron and build a connection between wavelet transform and neural network through consistent and approximate wavelet decomposition. It is

formed by integrating the latest-developed time-frequency localization with excellent wavelet transform and the self-learning function of the traditional artificial nerual network. The series to be obtained from translation and scale changes after wavelet decomposition have the property and classification characteristics of the common approximate function of the wavelet decomposition. Additionally, since it has introduced two new parameters, namely the scale factor and the translation factor, which makes it have more flexible and effective function approximation capability, stronger pattern recognition ability and fault tolerance ability. See the network structure of wavelet neural network and the excitation functions adopted in various layers (Figure 4). The excitation function can be named Sigmoid function [12].

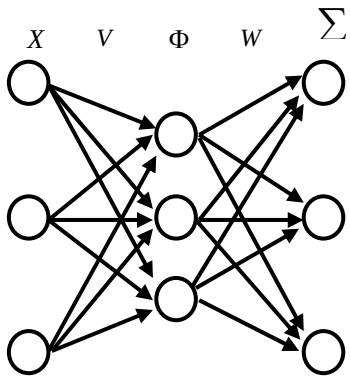

Figure 4. Multi-Input wavelet network

The network structure and the expression are basically the same with BP network, that is to say, it is formed by three layers: input layer, hidden layer and output layer. The difference is that the excitation function of the neuron in the hidden layer of BP network is Sigomoid function: $f(x)=1/(1+e^{-x})$ while the wavelet network uses the wavelet function $\psi(t)$,which can meet the admissibility condition as the excitation function. The specific valuation of $\psi(t)$ can be chosen according to the actual requirements. The commonly-seen excitation functions in the output layer include: Sigomoid function and linear Purline function.

4. Establishment of The Image Compression Algorithm Based on Wavelet Neural Network

This paper initializes the parameters of neural network with Morlet wavelet and the other types of wavelet networks are the same in the parameter setting steps except different time-frequency parameter. The expression of Morlet wavelet basis function is:

$$\psi(x)=\cos(1.75x)e^{-\frac{x^2}{2}} \tag{7}$$

Assume that the number of neurons in the hidden layer of three-layered neural network is M , the number of nodes in the input layer is L , the number of neurons in the output layer is N , w_{ji} is the connective weight from the jth neuron in the hidden layer to ith the neuron in the input layer, w_{kj} is the connective weight from the kth neuron in the output layer to the jth neuron in the hidden layer and $\psi(\frac{x-b_j}{a_j})$ is the excitation of the net output of the jth neuron in the hidden layer. Firstly, initialize w_{ji} according to the following steps:

(1) Firstly, take the random number uniformly distributed in the range of [-1,1] as the initial setting value of w_{ji} ;

(2) Then normalize w_{ji} by row;

$$w_{ji} = \frac{w_{ji}}{\sqrt{\sum_{i=1}^{L} w_{ji}^2}} (j = 1, 2, \cdots, M) \tag{8}$$

(3) Then multiple by a corresponding factor to the number of nodes L in the input layer, the number of neurons in the hidden layer M and the transfer function:

$$w_{ji} = C \square M^{\frac{1}{L}} \square w_{ji} (j = 1, 2, \cdots, M) \tag{9}$$

In this formula, C is a constant related to the transfer function in the hidden layer. The valuation of C is very important to the network. After several learning practices, the appropriate value for Morlet neural network is between1.9-2.1.

(4) Finally, associate with the training samples. Assume that the maximum value and minimum value of the input sample of the ith neuron in the input layer are $x_{\max}$ and $x_{\min}$ respectively, then:

$$w_{ji} = \frac{2w_{ji}}{x_{i\max} - x_{i\min}} (j = 1, 2, \cdots, M) \tag{10}$$

The w_{ji} obtained from the above steps is the initial weight from the input layer to the hidden layer. After initializing w_{ji}, the initial setting of the scale and translation parameter of wavelet is also very important to the network convergence. It usually can be divided into two circumstances:

(1)The number of nodes in the input layer is 1. Take the same value for the scale parameters a_i of every waveron and the translation parameter $b_j = (j-1)S / M (j = 1, 2, \cdots, M)$. In this formula, S is the number of training samples, and M is the number of neurons where the excitation function is located.

(2)The number of nodes in the input layer is >1. It can be known from the basic wavelet theory that if the time-domain center of the mother wavelet is t^* and the radius is $\Delta\psi$, then the concentrated time-domain area of wavelet translation is:

$$[b + at^* - a\Delta\psi, b + at^* + a\Delta\psi] \tag{11}$$

In order to make the wavelet scalability cover the entire range of the input vector, the initial setting of the scale and translation parameters should satisfy the following formulas:

$$\begin{cases} b + at^* - a\Delta\psi = \sum_{i=1}^{L} w_{ji} x_{i\min} \\ b + at^* + a\Delta\psi = \sum_{i=1}^{L} w_{ji} x_{i\max} \end{cases} \tag{12}$$

It can obtain from the above formula:

$$\begin{cases} a_j = \dfrac{\sum_{i=1}^{L} w_{ji} x_{i\max} - \sum_{i=1}^{L} w_{ji} x_{i\min}}{2\Delta\psi} \\ b_j = \dfrac{\sum_{i=1}^{L} w_{ji} x_{i\max} (\Delta\psi - t^*) + \sum_{i=1}^{L} w_{ji} x_{i\min} (\Delta\psi + t^*)}{2\Delta\psi} \end{cases} \tag{13}$$

The above formula requires the time-domain center and radius of the mother wavelet, which can be obtained through calculation.

The connective weight w_{kj} from the kth neuron in the output layer to the jth neuron in the hidden layer can be initialized through the following method:

(1)Firstly, take the random number uniformly distributed in the range of [-1,1] as the initial setting value of w_{kj} ;

(2)Then normalize w_{kj} :

$$w_{kj} = \frac{w_{kj}}{\sqrt{\sum_{i=1}^{M} w_{kj}^2}} (k = 1, 2, \cdots, N) \quad (14)$$

Image compression workflow of wavelet neural network is shown in Figure 5.

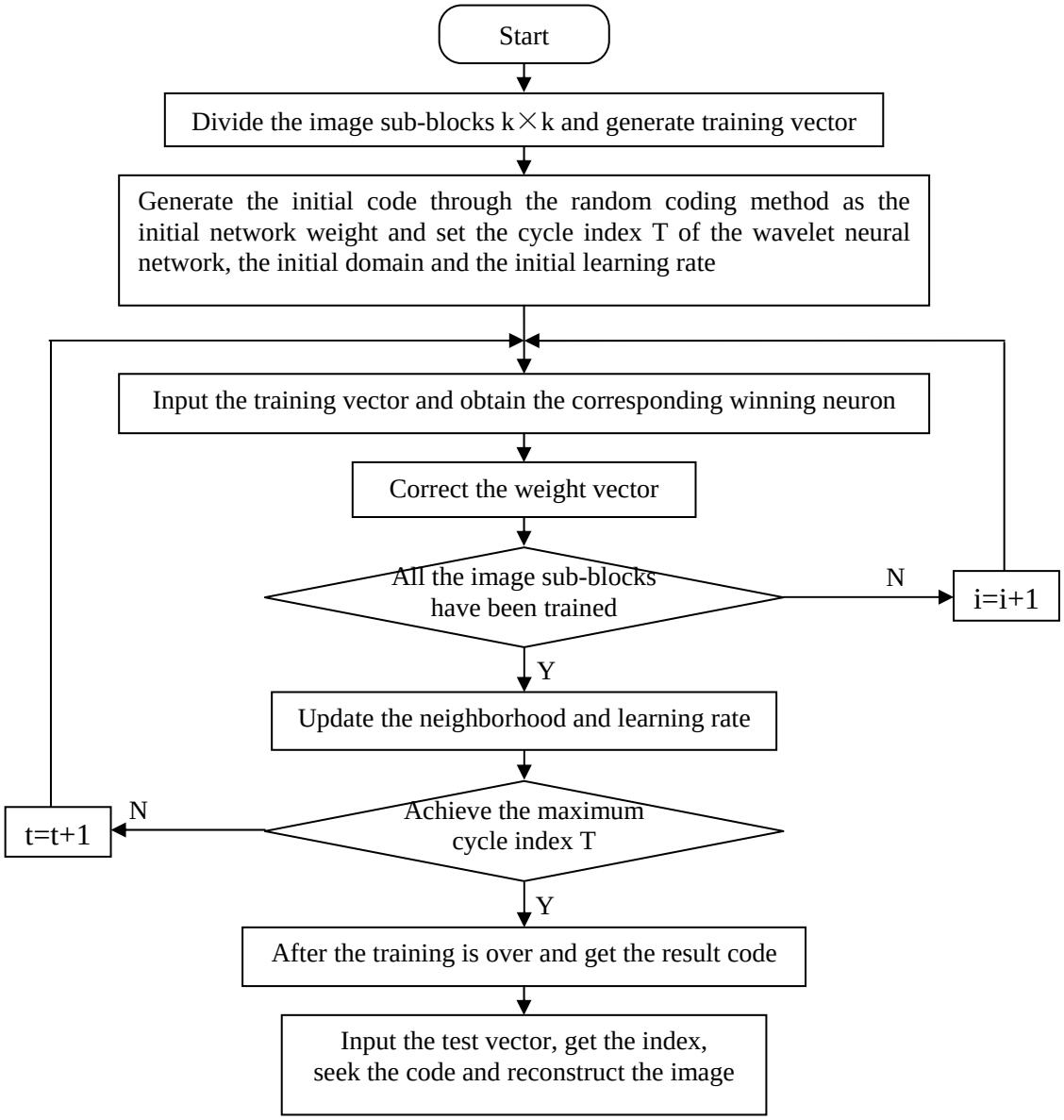

Figure 5. Image compression workflow of vector quantization of wavelet neural network

5. Experimental Simulation and Analysis

With “pout girl” as the original image, realize the iamge compression by using BP neural network and the method of this paper in MATLAB environment and the effects of the reconstructed images are indicated in Figure 6.

(a) Original image (b) BP neural network (c) Wavelet neural network

Figure 6. Simulation result of image compression with different algorithms

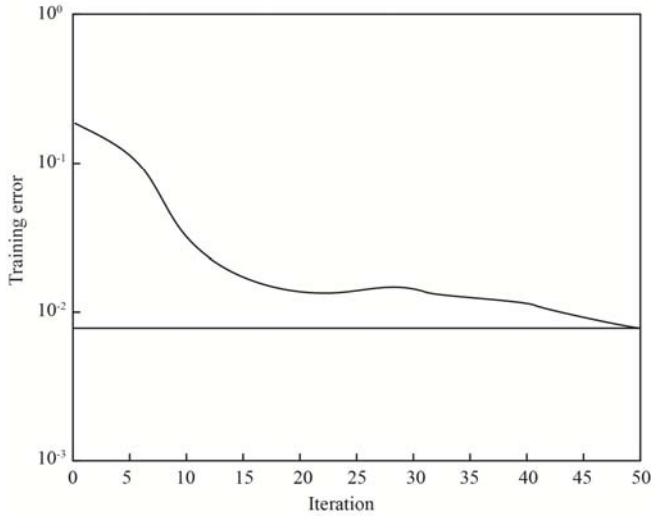

Figure 7. Wavelet neural network training error curve

It can be seen from the above figure that compared with the original image, the reconstructed image of Figure 6(b) has bad visual effect, obvious distortion and the image edge has big distortion and it is more blurred. Besides, the BP neural network training time increases and the higher compression ratio is obtained at the sacrifice of training time, which will direclty lead to the decrease of real time. However, only the weight of the output layer in the wavelet neural network needs to be trained while the weight of the input layer can be determined by the relationship between the interval of the sampling points and the interval of the wavelet compactly-supported interval. Once determined, the training speed of the wavelet neural network can be greatly accelerated. From Figure 6(c), it can be seen that the image adopting wavelet neural network is very clear, its details are more profound and it is very close to the original image.

6. Conclusion

With the increases of image pixels and the transmission rate, image compression technology has become one of the bottleneck technologies in the image processing. This paper has realized the application of wavelet neural network in the image compression and effectively enhanced the compression ability of image data. The comparison result with the traditional neural network training has demonstrated that the algorithm of this paper has had better compression efficiency and effects.

References

[1] Ehsan OS. An Algorithm for Real Time Blind Image Quality Comparison and Assessment. *International Journal of Electrical and Computer Engineering (IJECE)*. 2012; 2(1): 120-129.
[2] Wei F, Wenxing B. An Improved Technology of Remote Sensing Image Fusion Based Waveled Packet and Pulse Coupled Neural Net. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2012; 10(3): 551-556.
[3] Mario A. Rodríguez D, Hermilo SC. Refined Fixed Double Pass Binary Object Classification for Document Image Compression. *Digital Signal Processing*. 2014; 30(7): 114-130.
[4] Kartik S, Ratan KB, Amitabha C. Image Compression Based on Block Truncation Coding using Clifford Algebra. *Procedia Technology*. 2013; 10(3): 699-706.
[5] G Rosline N, S Maruthuperumal. Normalized Image Watermarking Scheme using Chaotic System. *International Journal of Information and Network Security (IJINS).* 2012; 1(4): 255-264.
[6] A Alfalou, C Brosseau, N Abdallah. Simultaneous Compression and Encryption of Color Video Images. *Optics Communications.* 2015; 338(1): 371-379.
[7] Roman S. New Simple and Efficient Color Space Transformations for Lossless Image Compression. *Journal of Visual Communication and Image Representation.* 2014; 25(5): 1056-1063.
[8] Hamid T, Aref M. Wavelet Neural Network Applied for Prognostication of Contact Pressure between Soil and Driving Wheel. *Information Processing in Agriculture*. 2014; 1(1): 51-56.
[9] Bhargav V, Biswarup D, Rudra P, etc. An improved Scheme for Identifying Fault Zone in A Series Compensated Transmission Line using Undecimated Wavelet Transform and Chebyshev Neural Network. *International Journal of Electrical Power & Energy Systems*. 2014; 63(12): 760-768.
[10] Yashar F, Narges P, Yuk FH, etc. Estimating Evapotranspiration from Temperature and Wind Speed Data using Artificial and Wavelet Neural Networks (WNNs). *Agricultural Water Management*. 2014; 140(7): 26-36.
[11] Khaled D, Tarek AT. Speaker Identification using Vowels Features through A Combined Method of Formants, Wavelets, and Neural Network Classifiers. *Applied Soft Computing*. 2015; 27(2): 231-239.
[12] Majid J, Abul K, AQ Ansari, etc. Generalized Neural Network and Wavelet Transform Based Approach for Fault Location Estimation of a Transmission Line. *Applied Soft Computing.* 2014; 19(6): 322-332.