Journal of University of Science and Technology Beijing
Volume 14, Number 5, October 2007, Page 437

Application of a new feature extraction and optimization method to surface defect recognition of cold rolled strips

Guifang Wu[1,2], *Ke Xu*[1], *and Jinwu Xu*[1]

1) National Engineering Research Center for Advanced Rolling Technology, University of Science and Technology Beijing, Beijing 100083, China
2) Electronic Information Engineering College, Henan University of Science and Technology, Luoyang 471003, China
(Received 2006-09-24)

Abstract: Considering that the surface defects of cold rolled strips are hard to be recognized by human eyes under high-speed circumstances, an automatic recognition technique was discussed. Spectrum images of defects can be got by fast Fourier transform (FFT) and sum of valid pixels (SVP), and its optimized center region, which concentrates nearly all energies, are extracted as an original feature set. Using genetic algorithm to optimize the feature set, an optimized feature set with 51 features can be achieved. Using the optimized feature set as an input vector of neural networks, the recognition effects of LVQ neural networks have been studied. Experiment results show that the new method can get a higher classification rate and can settle the automatic recognition problem of surface defects on cold rolled strips ideally.

Key words: cold rolled strip; surface defect; neural networks; fast Fourier transform (FFT); feature extraction and optimization; genetic algorithm; feature set

[*This work was financially supported by the National High Technology Research and Development Program of China (No.2003AA331080 and 2001AA339030) and the Talent Science Research Foundation of Henan University of Science & Technology (No.09001121).*]

1. Introduction

On account of the demand for steel production being enlarged in the domestic and international markets, requirements for surface quality has become greater and automatic surface defect detection and recognition of cold rolled strips [1] has attracted the attention of a lot more steel and iron enterprises. At the same time, production of cold rolled strips amounts to a great proportion of iron and steel production, and their surface qualities are the most concerning target of customers.

The most important stages in the surface inspection of cold rolled strips are surface defect detection and recognition. Surface defect recognition will be mainly discussed, and FFT (fast Fourier transform) is introduced to extract features, genetic algorithm is used to optimize the feature set, and neural networks are introduced to explore the defect recognition effect and applied to the onsite product line. The following will lay an emphasis on the analysis, and research the key problem of applying the new feature extraction and optimization method to surface defect recognition of cold rolled strips.

2. Spectrum analysis and feature extraction

When producing cold rolled strips, there are always chief defect types, such as, scratches, point breaks, feather roll imprints, white spots, roll imprints, edge foldings, rust, orange peels, emulsion marks, edge cracks, cracks, and so on. Some typical images of defects are shown in Fig. 1.

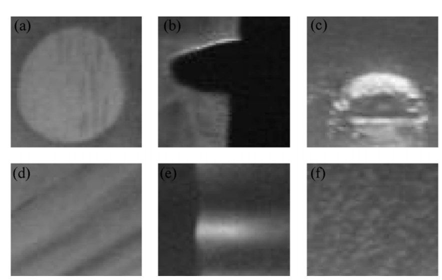

Fig. 1. Some of the typical defect images with a size of 64×64 pixels. There are a total of 6 samples of defect images listed: (a) white spots; (b) edge cracks; (c) point breaks; (d) feather roll imprints; (e) edge foldings; (f) original peels.

Corresponding author: Guifang Wu, **E-mail:** easyfancy@126.com

All the defect images have the size 64×64 pixels. They are then filtered by average to get average filtered images. By applying FFT [2] to the images, the spectrum images can be obtained. It is well known from the qualities of FFT that it can transform images from time-domain to frequency-domain, and can concentrate most of its energies on the center. It can not only reflect the gray features and geometrical features of defect images, but can also achieve fast convolution and object recognition. Therefore, regarding the gray values of spectrum images as feature information of defect images, one can recognize defects, especially the pixels near the central area of spectrum images. The typical spectrum images of FFT are shown in Fig. 2.

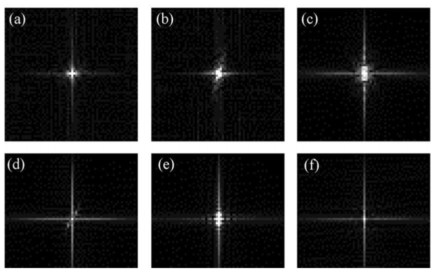

Fig. 2. Typical FFT images of Fig. 1 with a size of 64×64 pixels. There are a total of 6 samples of FFT images listed: (a) white spots; (b) edge cracks; (c) point breaks; (d) feather roll imprints; (e) edge foldings; (f) original peels.

According to the results of FFT, it can be seen that nearly all energies are concentrated on the center area and all the pixels are centrally symmetrical. At the same time, if all the pixels of FFT image are taken as the feature set, the feature set will be too large and unnecessary, and will affect recognition results contrarily. As a consequence, based on many experiments, the region shown as Fig. 3 is taken as the original feature set, where the region width a is 4 and the region height b is 4 too. Therefore, the chosen region totally includes 240 pixels, and it is regarded as the original feature set.

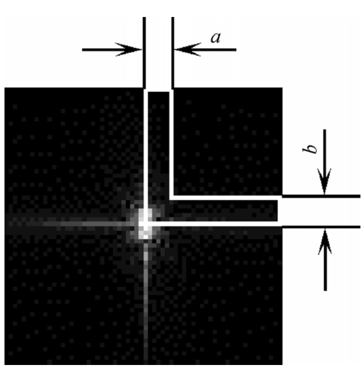

Fig. 3. Optimized original feature set region.

For complete analysis of FFT images, another feature named sum of valid pixels (SVP) is introduced. The definition of SVP is the count of all the pixels whose values are larger than a given threshold in the full FFT images. Therefore the region outside the selected region is taken into consideration because of the SVP extracted from full FFT images.

3. Genetic algorithm and feature optimization

After the feature extraction, an original feature set and SVP can be obtained, but the feature set includes 241 features, which is still very large for neural networks. Therefore, how to decrease the dimension of the feature set becomes necessary, and the process is called feature optimization. Feature optimization of surface defects is feature selection, which is to optionally regroup all features of defects to get a set of optimized character parameters, which can describe the defects more perfectly with fewer features. There are three traditional optimization algorithms, such as, the enumeration method, heuristic algorithm, and search algorithm. Genetic algorithm (GA) [3-5] is a random search algorithm for feature optimization.

3.1. Genetic algorithm

The GA [3-5] is based on natural selection and genetic mechanism, and its main thoughts stem from the Darwinian theory and Mendelism. It can use simple coding techniques to describe varied and intricate structures, and can process a simple genetic operation and natural selection to instruct learning and explore search orientation.

The simple genetic algorithm (SGA) can be described by an equation with 8 elements:

$$\mathrm{SGA}=(C,E,P_0,M,\varPhi,\varGamma,\varPsi,T) \tag{1}$$

where C is the coding method of individuals; E is the fitness function of individuals; P_0 is the initial population; M is the population size; $\varPhi$ is the selection operator; $\varGamma$ is the crossover operator; $\varPsi$ is the mutation operator; T is the termination condition of the GA.

The canonical genetic algorithm can be described as follows:

Step 1. Initialization of the populations' individuals and evaluation of the individuals' fitness.

Step 2. Selection of parents according to a preselected selection scheme (*e.g.*, roulette wheel, linear ranking, and truncation selection).

Step 3. Recombination of selected parents by exchanging parts of their genes.

Step 4. Mutation of some genes by a prespecified probability.

Step 5. Go to step 2.

3.2. Feature optimization

The process of using the GA to optimize the feature set is a process of setting the parameters of the GA to describe the problem best. The GA is used to optimize the region of the original feature set, which includes 240 features, to a smaller dimension.

(1) According to the region of the original feature set, the integer coding method is adopted to encode individuals, such numbers as 1, 2, 3, $\cdots$, 239, 240 are set to stand for every pixel of the region separately.

(2) It is known that the object function of mutual information entropy is a problem of the largest values. On the basis of experiments, and according to analysis of surface defects of steel strips, the mutual information entropy is taken as the fitness function of the GA, and it is shown as the following Eq. (2),

$$I(E,F)=\frac{1}{N}[I^{*}(E,F)-I(F)]=$$

$$\frac{1}{N}\left[\sum_{i=1}^{N}\sum_{j=1}^{N}I(w_i,x_j)-\frac{1}{N-1}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}I(x_i,x_j)\right] \tag{2}$$

Where

$$I(w_i,x_j)=\sum_{k=1}^{K}p(w_i,x_{j,k})\log_2\frac{p(w_i,x_{j,k})}{p(w_i)p(x_{j,k})} \tag{3}$$

$$I(x_i,x_j)=\sum_{h=1}^{K}\sum_{k=1}^{K}p(x_{i,h},x_{j,k})\log_2\frac{p(x_{i,h},x_{j,k})}{p(x_{i,h})p(x_{j,k})} \tag{4}$$

(3) Set individual size as 50 pixels and population size as 60 pixels based on experiments and initialize population by random selection from 240 features.

(4) Decide selection operator. According to the feature set of all the features, choose roulette as the selection operator.

(5) Because integer coding is used, and every data type of gene is an integer, the random assortment crossover and random mutation operator beyond a certain rate can be used.

(6) As there is no known optimum feature set in the beginning, the termination condition can be set as a value of maximum generation. When GA evolves to a certain generation, which is larger than the preassigned value of the maximum generation, the GA terminates.

Between the running of the GA and its termination, 60 optimized feature sets are achieved, and the most optimized feature set whose fitness value is greatest is selected as one part of the input vector of neural networks. The flowchart is shown in Fig. 4.

4. Neural networks and their application

Artificial neural networks [6-7] form the parallel distributed information processing system. It is a network comprising of many processing units, which are connected to each other by unidirectional signal paths. These processing units are called artificial neural units, and have the ability of local memory and local information processing. Every processing unit unidirectionally outputs to an expected connection and transmits the same signal, which is called the output signal of the processing unit. The output signal of the processing unit can be of any demanded mathematical type. Under complete locally restricted circumstances, information processing executed in every processing unit can have any definition, meaning information processing depends only on the memorized value of itself and the current value of the input stimulated signal of the processing unit. Neural networks is an abstraction and simplification of the human brain in microstructure and function, and reflects some fundamental characters, such as, parallel processing, learning, association, pattern classification, memory, and so on.

Obviously, according to the above characteristics of neural networks, applying neural networks to surface defect recognition of cold rolled strips is a good choice. The application effects of learning vector quantization (LVQ) neural networks are discussed in the following section.

4.1. LVQ neural network

Learning vector quantization [8] neural network consists of three parts, the input layer, hidden layer, and output layer. The hidden layer is also called competitive layer or Kohonen layer, and it is completely linked between the input layer and hidden layer, but just partly connected between the hidden layer and output layer, and every output neuron is connected with a different group of hidden neuron. The training process of the LVQ neural network is a process of continuous competition and weight updating process among all neuron units of the hidden layer. Fig. 5 is the topological structure of the LVQ neural network.

Given a sequence of input documents, an initial group of reference vectors $\boldsymbol{w}_k$ is selected. In each iteration, an x_i document is selected and the vectors $\boldsymbol{w}$ are updated, so that they adapt better to x_i. The LVQ algorithm works as follows.

For each class k, a weight vector $\boldsymbol{w}_k$ is associated. In each repetition, the algorithm selects an x_i input document and compares it with each weight vector $\boldsymbol{w}_k$, using the Euclidean distance $\|x_i - w_k\|$, so that the winner will be the weight vector $\boldsymbol{w}_c$ nearest to x_i, with c as the index of that weight vector:

$$\|x_i - w_c\| = \min_k \{\|x_i - w_c\|\} \tag{5}$$

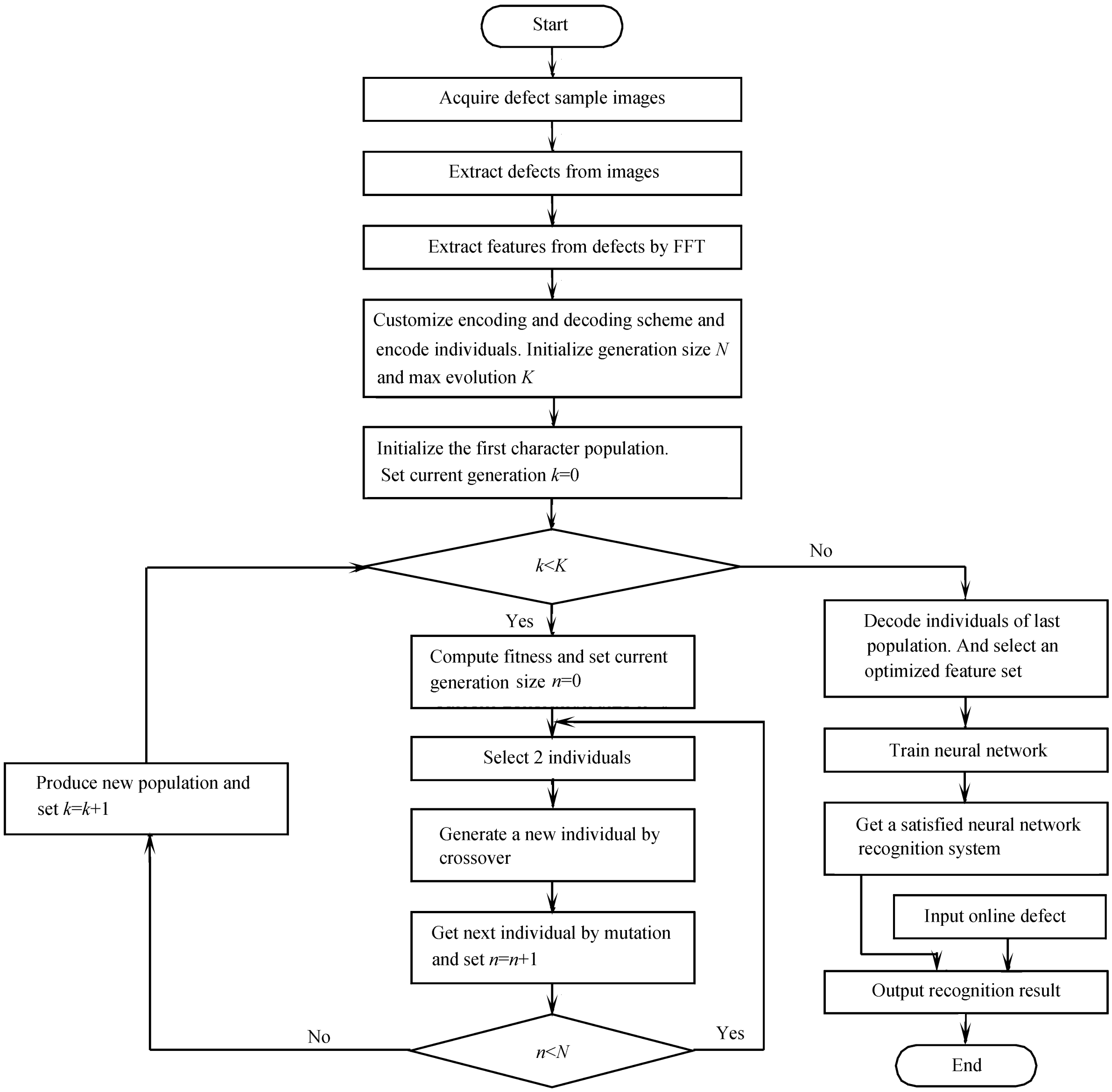

Fig. 4. Flowchart of the full recognition procedure, including spectrum analysis, feature extraction, feature optimization, training and test of neural networks.

The classes compete among themselves to find the most similar one to the input vector, so that the winner will be the one with less Euclidean distance with respect to the input document. Only the winner class will modify its weights using a reinforced learning algorithm, either positive or negative, depending on whether the classification is correct or not. Thus, if the winner class belongs to the same class as the input vector, it will increase the weights, moving slightly closer to the input vector (prize). On the contrary, if the winner class is different from the input vector class, it will decrease the weights, moving slightly away from the input vector (punishment). Let $x_i(t)$ be an input document at time t, and the weight vector $w_k(t)$ for the class k at time t. The following equation define the basic learning process for the LVQ algorithm:

$$w_c(t+1)=w_c(t)+s\alpha(t)[x_i(t)-w_c(t)] \tag{6}$$

where s=0, if $k\neq c$; s=1, if $x_i(t)$ and $w_c(t)$ belong to the same class; and s=−1, if they do not, and $\alpha(t)$ is the learning rate, $0<\alpha(t)<1$, a monotonically decreasing function of time. It is recommended that $\alpha(t)$ should initially be rather small, smaller than 0.1, and $\alpha(t)$ continues decreasing to give a threshold u, very close to 0. Usually, $\alpha(0)$ is always initialized in 0.005, decreasing linearly to u=0.001, according to the following equation where K is the number of classes,

$$\alpha(t+1)=\alpha(t)-\frac{\alpha(0)-u}{K} \tag{7}$$

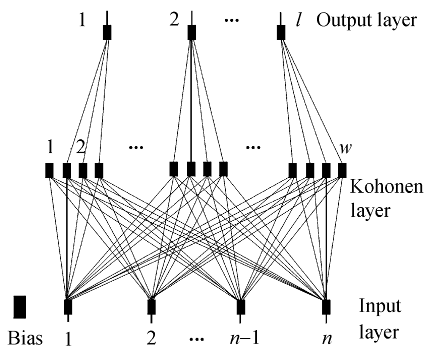

Fig. 5. Topological structure of LVQ neural networks.

4.2. Experimental results

In the surface defect recognition of cold rolled strips, set the input parameters of neural networks as the defect feature set optimized by the GA, and train the neural networks with plenty of defects sampled from a certain product line, so that the neural networks can adapt to the circumstance of the product line, and at last a contented neural networks system will be achieved. The detailed flowchart is shown in Fig. 4.

Taking out 241 defects from images acquired by a surface defect online inspection system of a certain product line, as training samples, and extracting all 51 features optimized by GA, Table 1 shows the training sample defect data distribution and training results, where SN is the sample number, AN is the accuracy number, and RR is the recognition rate. Meanwhile, AN1 and RR1 are tested by the original feature set, including 240 features, and AN2 and RR2 are tested under the optimized feature set including 51 features, with the help of the GA. Under different experimental conditions, the BP neural network and LVQ neural network are trained and tested repeatedly, and the test results are shown in Table 2.

Table 1. Training sample defect data distribution and training results

Item	Scratch	Point stick	Feather	White spot	Roll imprint	Edge folding	Rust	Orange peel	Emul-sion mark	Edge crack	Coil break	Total
SN	43	45	42	36	35	18	26	34	24	22	45	370
AN2 of BP	40	38	38	31	31	16	21	26	19	20	40	320
RR2 of BP / %	93.02	84.44	90.48	86.11	88.57	88.89	80.77	76.47	79.17	90.91	88.89	86.49
AN1 of LVQ	38	36	35	30	28	16	20	27	17	20	38	305
RR1 of LVQ / %	88.37	80.00	83.33	83.33	80.00	88.89	76.92	79.41	70.83	90.91	84.44	82.43
AN2 of LVQ	42	40	41	34	32	18	20	31	21	22	42	343
RR2 of LVQ / %	97.67	88.89	97.62	94.44	91.43	100.00	76.92	91.18	87.50	100.00	93.33	92.70

Table 2. Test sample defect data distribution and test results

Item	Scratch	Point stick	Feather	White spot	Roll imprint	Edge folding	Rust	Orange peel	Emul-sion mark	Edge crack	Coil break	Total
SN	52	46	44	44	38	19	20	32	22	26	54	397
AN2 of BP	47	38	39	33	30	16	15	21	16	22	45	322
RR2 of BP / %	90.38	82.61	88.64	75.00	78.95	84.21	75.00	65.63	72.73	84.62	83.33	81.11
AN1 of LVQ	45	34	36	31	26	14	14	24	15	21	47	307
RR1 of LVQ / %	86.54	73.91	81.82	70.45	68.42	73.68	70.00	75.00	68.18	80.77	87.04	77.33
AN2 of LVQ	48	37	41	37	32	17	15	25	15	24	48	339
RR2 of LVQ / %	92.31	80.43	93.18	84.09	84.21	89.47	75.00	78.13	68.18	92.31	88.89	85.39

Because the time complexity of the BP neural network is relative higher than that of the LVQ neural network, the BP neural network was not tested with 241 features, and only got AN2 and RR2. It can be seen from Table 1 and Table 2 that AN2 and RR2 are better than AN1 and RR1, and AN2 of LVQ and RR2 of LVQ are better than AN2 of BP and RR2 of BP. Therefore, it can be found that the recognition rate will be improved with the input feature optimized by the GA, and the BP neural network cannot work very well under multi-input-features and multi-defect-types. As a result, how to extract more accurate features of defects from cold rolled strips and how to search for the optimal feature set will influence the recognition accuracy of surface defects. At last, the two tables show that the results still have much room for improvement, especially for some defects, such as, emulsion marks, orange peels, rusts and so on. Therefore, it can be deduced that single neural networks still cannot get an optimal recognition accuracy under too many defect types and strong outside disturbances in the onsite product line. Consequently, more advanced neural networks to overcome the problem must be studied, or an integrated recognition method must be explored based on current recognition techniques to perform multi-level recognition to get a better recognition effect.

5. Conclusions

Researched feature extraction of defects by FFT detailedly, and applied the GA successfully to optimize the feature set. Using the optimized feature set as the input vector of the LVQ neural network and the BP neural network, to run surface defect recognition, a better result was obtained than that using the original features, which were not optimized. The method here not only improved recognition rate, but also saved recognition time.

From the results of the experiments, it must be pointed out that by using a single neural network it is difficult to obtain an optimal recognition effect at present, because of too many defect types and strong outside disturbances in the onsite product line. As a consequence, a more advanced neural network must be studied to overcome the problem, or an integrated recognition method must be explored to perform multi-level recognition based on current recognition technique to get a better recognition effect. At the same time, further efforts must be made to extract a more accurate feature set which plays an important role in surface defect recognition of cold rolled strips.

References

[1] K. Xu, J.W. Xu, and S.L. Lu, Surface inspection system for cold rolled strips based on image processing technique, *J. Univ. Sci. Technol. Beijing*, 6(1999), No.4, p.296.

[2] E. Chu and A. George, *Inside the FFT Black Box: Serial and Parallel Fast Fourier Transform Algorithms*, CRC Press, Boca Raton, 2000, p.32.

[3] X.J. Tang, *Genetic Algorithms with Application to Engineering Optimization* [Dissertation], University of Memphis, Tennessee, 2004, p.13.

[4] R.L. Haupt and S.E. Haupt, *Practical Genetic Algorithms*, Wiley Press, New Jersey, 2004, p.58.

[5] K. Man, K. Tang, and S. Kwong, *Genetic Algorithms: Concepts and Design*, Springer, New York, 1999, p.106.

[6] K.S. Narendra, Neural networks for control: theory and practice, *Proc. IEEE*, 84(1996), No.10, p.1385.

[7] C. Fyfe, *Artificial Neural Network* [Dissertation], University of Paisley, Paisley, 1996, p.22.

[8] T. Kohonen, Learning vector quantization, *Neural Networks*, 1(1988), No.6, p.303.