Feature extraction and classification are two important steps in the process of strip steel surface defect recognition. Traditional methods of defect feature extraction are not of scale and rotation invariance. Moreover, traditional methods of defect classification have a conflict between efficiency and accuracy. In order to solve the above two problems, a novel recognition method is proposed in this paper. On one hand, the novel defect feature extraction scheme is realized by building sampling benchmark scale (SBS) information for training dataset and using gradient magnitude and gradient orientation co-occurrence matrix (GMGOCM), gray level and gradient orientation co-occurrence matrix (GLGOCM), and moment invariant features. On the other hand, $K$-nearest neighbor and $R$-nearest neighbor algorithms are used to prune training dataset, and amplification factors of pruned samples are used to improve least squares twin support vector machine (LSTWSVM) classifier in efficiency and accuracy. The experimental results show that the novel recognition method can not only realize defect feature extraction with scale and rotation invariance but also realize defect classification with high efficiency and accuracy.

KEY WORDS: strip steel; surface defect; feature extraction; classification; GMGOCM; GLGOCM; LSTWSVM.

1. Introduction

In recent years, with increasing demand for high quality steel products and intensified market competition, surface defect recognition is becoming a very important technology to detect and control the strip steel quality for iron and steel enterprises. In general, there are four steps in the process of strip steel surface defect recognition, including defect preprocessing, defect segmentation, defect feature extraction and defect classification. Defect feature extraction and defect classification are the crucial steps and the study of them is becoming more and more widely and deeply.1–5) In reference1) extended haar rectangle features are extracted and weak classifier adaptive enhancement classification method is used. In reference3) hough transform is used to classify defects with well defined geometric shapes, and principal component analysis and self-organizing maps are used to classify defects with complex shapes. In reference3) a new defect detection algorithm is provided with multifractal features and neural work classifier. Reference4) used local binary pattern algorithm to extract defect features, and employed decision tree and support vector machine as classification method. Reference5) proposed a scheme based on mixed features and combining classifiers, in which mixed features include geometric feature, gray feature, texture feature, projection feature, and transform feature, and combining classifiers include neural network and support vector machine.

References1–5) can effectively extract features and realize classification for the strip steel surface defect. However, samples with the same type of defect may be different in scale and rotation in real problems, which requires the algorithm of defect feature extraction to be of scale and rotation invariance. Although references1–5) have partly adopted some algorithms with scale and rotation invariance, such as moment invariant6) algorithm, they did not fully consider scale and rotation invariance in the process of defect feature extraction. So it is necessary to find some novel defect feature extraction algorithms with scale and rotation invariance. On the other hand, traditional support vector machine (SVM)7) is considered an effective method for the strip steel surface defect classification. Similarly, there is a conflict between accuracy and efficiency in traditional SVM. So more effective versions of SVM are produced, such as generalized eigenvalue proximal support vector machine,8) twin support vector machine (TWSVM)9) and least squares twin support vector machine (LSTWSVM).10) However, these classifiers still need to be improved to reach high accuracy and efficiency.

For the strip steel surface defect recognition, the novel method proposed in this paper can not only realize scale and rotation invariance but also realize defect classification with high efficiency and accuracy.
rotation invariance in defect feature extraction but also obtain high accuracy and efficiency in defect classification. In the process of defect feature extraction, firstly, a defect image is resampled by using sampling benchmark scale (SBS) information of training dataset. Then edge area features of the defect image are extracted based on gradient magnitude and gradient orientation co-occurrence matrix (GMGOCM), internal area features of the defect image are extracted based on gray level and gradient orientation co-occurrence matrix (GLGOCM), and moment invariant features of the defect image are extracted. Experiments proved that GMGOCM and GLGOCM can realize scale and rotation invariance in defect feature extraction. In the process of defect classification, firstly, training dataset is pruned by combining K-nearest neighbor and R-nearest neighbor algorithms. Then amplification factors of pruned samples are used to improve LSTWSVM. Experiments also proved that this algorithm can realize high accuracy and efficiency in defect classification.

This paper is structured as follows. Section 2 introduces strip steel surface defect recognition system. In section 3 methods of feature extraction with scale and rotation invariance are proposed. Section 4 focuses on the classification algorithm with high efficiency and accuracy. Some testing experiments and results are described in section 5. Some conclusions are drawn in section 6.

2. Surface Defect Recognition System

In general, the defect recognition system for strip steel surface includes two parts: surface image acquisition and defect image processing. The main function of surface image acquisition is to collect real-time surface images and transmit those images to the second part. Surface image acquisition system is always made up of computer vision detector,11) which is shown in Fig. 1. This system includes CCD sensors, light sources, acquisition boards, data transport interfaces, and parallel computers. The subsystem of CCD sensors, which includes many parallel high-speed linear CCD cameras, is used to scan the upper and lower surfaces of strip steel and obtain surface images. The subsystem of light sources is used to provide light for CCD cameras. The subsystem of acquisition boards, which always adopts digital signal processing system, is used to read, store, and transmit those images to parallel computers. It is also used to realize some parts of preprocessing for defect images, such as rapid defect area detection. According to defect area detection, defect images are reserved and other images are rejected. Data transfer interfaces, which adopts gigabit Ethernet, are used to transmit defect images. The subsystem of parallel computers is used to receive, store, process and synthesize defect images, and generate complete defect information.

Defect image processing system is used to realize defect preprocessing, segmentation, feature extraction, and classification. Defect preprocessing can realize rapid defect area detection, image filtering, image enhancing, etc. Rapid defect area detection can be finished on acquisition boards. And image filtering can filter the impulse noise, Gaussian noise, etc. Image enhancing can realize background correction and gray stretch for defect images. Defect segmentation can realize accurate defect positioning, and separate defect from background, and extract defect area. Defect feature extraction can obtain many valid features, such as geometric feature, gray feature, texture feature, and transform feature. Defect classification can classify many types of defects by using multi-class classifier. Defect feature extraction with scale and rotation invariance and defect classification with high efficiency and accuracy are discussed in this paper.

3. Feature Extraction Scheme

There may be differences in scale, rotation, translation, brightness and viewing angle for the same type of defect images. In the process of pattern recognition, these differences will affect the result of feature extraction and reduce the following classification accuracy. On a real strip steel production line of enterprises, there are surface defect acquisition instruments with consistent environment and fixed location that can avoid obvious differences in brightness and viewing angle. Image segmentation method can extract edge area and internal area of the defect image, which can avoid differences in translation. However, the differences in scale and rotation can not be avoided. So feature extraction methods with scale and rotation invariance should be adopted in the strip steel surface defect recognition. Moment invariant features are of good scale and rotation invariance6) and can describe geometrical features of a defect image well.
Though moment invariant features are useful, they are not sufficient. So some novel feature extraction methods with scale and rotation invariance are proposed in this section.

3.1. SBS of Defect Images

There are many forms with different scale and rotation for a defect sample in training dataset. Defect images with different rotation can be obtained by simulation. If the simulated angle is $\theta$, then this type of defect has $\frac{360}{\theta}$ different images. It can be seen that $\theta$ determines the accuracy and efficiency of simulation. In order to keep the balance between them, $\theta$ is selected as 22.5°. 16 images with different angles for a type of defect sample are obtained.

![Fig. 2](image)

**Fig. 2.** 16 images with different rotation angles: (a) rotation angle is 0°, 22.5°, 45° and 67.5°; (b) rotation angle is 90°, 112.5°, 135° and 157.5°; (c) rotation angle is 180°, 202.5°, 225° and 247.5°; (d) rotation angle is 270°, 292.5°, 315° and 337.5°.

Firstly, rotate all image samples in training dataset with 4 angles in Fig. 2(a). And get 16 scales for every image sample. Secondly, calculate the height-width ratio $R_{hw}$ according to every scale. Then, analyze the sizes of these samples with the same $R_{hw}$, and determine a rational SBS information $S_{hw}$. Finally, build an SBS table to show the relationship between $R_{hw}$ and $S_{hw}$.

![Fig. 3](image)

**Fig. 3.** The scheme of extracting SBS image.

![Fig. 4](image)

**Fig. 4.** A crack defect image, and its edge and internal areas.

Though different samples with the same type of defect may be different in size, they can be concerned with the same $R_{hw}$. If a type of sample exists in the training dataset, SBS table can reflect benchmark scale information for that type of testing sample. Moreover, SBS table can reflect all scales of training samples and rotated samples with different angles. It can be concerned that benchmark scale can not be affected by different rotation. So, SBS table shows the reference scale for a testing sample.

For a defect image, firstly, calculate $R_{hw}$ and determine $S_{hw}$ from the SBS table. Then, resample the image with interpolation to make it larger or smaller. Finally, SBS image of the image with benchmark scale can be obtained. To mention that, there are three types of interpolation to resample: bilinear interpolation, nearest neighbor interpolation and bicubic interpolation. The first one is used in this paper. Feature extraction is implemented on SBS image in this paper. The scheme of extracting SBS image is shown in Fig. 3, which can reduce the effects which the difference in scale has on defect feature extraction.

3.2. GMGOCM Features Extraction

For a defect image, its edge and internal areas can be determined by defect preprocessing and segmentation. Figure 4 shows the edge and internal areas of a defect image. Because the gray changes in edge area are drastic, they are usually described with gradient. GMGOCM is proposed based on gradient vector information and gray-level co-occurrence matrix. GMGOCM can completely express the statistical characteristics of gray changes, orientations and mixed of them for all pixels in an edge area.

Define the edge area of a defect image to be $L$, the number of pixels to be $N_L$ and the $i$th pixel to be $p_i$ with gray level $f_i = f(x_i, y_i)$, where $(x_i, y_i)$ is the coordinate of the $i$th pixel ($i = 1, 2, \ldots, N_L$). Suppose $W_{p_i}$ is a set with elements of all edge area pixels in the $3 \times 3$ window with center point $p_i$. The number of all pixels in $W_{p_i}$ is $n_i$ ($n_i < 9$). The gray levels for pixels in $W_{p_i}$ and its nearest neighbors are $f(x_i + \Delta x_1, y_i + \Delta y_1)$ ($\Delta x_1, \Delta y_1 = -2, -1, 0, 1, 2$). Gradient magnitude $(G_Y)_{p_i}$ and gradient orientation $(G_O)_{p_i}$ are obtained by statistic analysis for all pixels in $W_{p_i}$. Firstly, Gradient ampli-
tude and gradient orientation for every pixel in $W^L_{ij}$ can be calculated just like the following:

$$W^L_{ij} = \{k,l | k,l \in [-1,0,1] \text{ and } f(x_i+k,y_j+l) \in W^L_{ij} \}$$

$$dx_i(k,l) = (f(x_i+k+1,y_j+l) - f(x_i+k-1,y_j+l))/2$$

$$dy_j(k,l) = (f(x_i+k+1,y_j+l+1) - f(x_i+k,y_j+l-1))/2$$

$$g_0(k,l) = \text{atan}[dy_j(k,l)/dx_i(k,l)]$$

$$G_{Mi}(k,l) = [dx_i(k,l)^2 + dx_i(k,l)^2]^{0.5} \quad k,l \in W^L_{ij}$$

$$G_{Mi}(k,l) = \begin{cases} g_0(k,l) + \pi, & dx_i(k,l) < 0 \\ g_0(k,l), & \text{other} \end{cases}$$

(1)

where $atan$ is arctangent function. $(G_0)^y_i$ is defined as average of all gradient amplitudes in $W^L_{ij}$, which is shown in the following:

$$(G_0)^y_i = \frac{1}{n_i^L} \sum_{k,l \in W^L_{ij}} G_{Mi}(k,l) \quad i=1,2,\ldots,N^L \quad (2)$$

Gradient orientation of $P^L_{ij}$ is turned to 0° in order to keep rotation invariance. Then the other pixels in $W^L_{ij}$ are correspondingly turned. $(G_0)^y_i$ is defined as average of all gradient orientations in $W^L_{ij}$ that have been turned. Its formula is just like the following:

$$(G_0)^y_i = \frac{1}{n_i^L} \sum_{k,l \in W^L_{ij}} \text{MOD}(G_{Mi}(k,l) - G_{Mi}(0,0) + \pi, \pi) \quad i=1,2,\ldots,N^L \quad (3)$$

where $\text{MOD}$ represents modulus after division.

In order to improve the calculation efficiency, $(G_0)^y_i$ and $(G_0)^x_i$ are normalized into domain $[1,2,\ldots,32]$. Suppose the largest gradient magnitude and gradient orientation are $(G_0)^{y_{\text{max}}}_{\text{max}}$ and $(G_0)^{x_{\text{max}}}_{\text{max}}$ respectively. $\text{INT}$ means the value is rounded towards the nearest integer, then the normalized $(G_0)^{y_{\text{32}}}_{32}$ and $(G_0)^{x_{\text{32}}}_{32}$ are shown in the following:

$$(G_0)^{y_{\text{32}}}_{32} = 1 + \text{INT}\left( (G_0)^y_i \times \frac{31}{(G_0)^{y_{\text{max}}}_{\text{max}}} \right) \quad i=1,2,\ldots,N^L \quad (4)$$

$$(G_0)^{x_{\text{32}}}_{32} = 1 + \text{INT}\left( (G_0)^x_i \times \frac{31}{(G_0)^{x_{\text{max}}}_{\text{max}}} \right) \quad i=1,2,\ldots,N^L \quad (5)$$

For the edge area of a defect image, define a $32 \times 32$ matrix $H^L$. The element $H^L(k,l)$ is the number of the pixels with gradient amplitude $k = (G_0)^{y_{\text{32}}}_{32}$ and gradient orientation $l = (G_0)^{x_{\text{32}}}_{32}$. The probability of $(k,l)$ in $H^L$ is defined as $p(k,l)$, and it can be calculated just like the following:

$$p^L(k,l) = H^L(k,l) / \sum_{k=1}^{32} \sum_{l=1}^{32} H^L(k,l) \quad (6)$$

According to $H^L$ and the algorithm in references,(17,18) calculate the corresponding GMGOCM features, such as mean of gradient amplitudes $\mu^L_{0^1}$, mean of gradient orientations $\mu^L_{0^2}$, variance of gradient amplitudes $\delta^L_{0^1}$, variance of gradient orientations $(\delta^L_{0^2})^2$, entropy of gradient amplitudes $E_{0^1}^L$, entropy of gradient orientations $E_{0^2}^L$, inertia moment $I_{0^1}^L$, and correlation coefficient $I_{0^2}^L$.

Their calculating formulas are just like the following:

$$\mu^L_{0^1} = \frac{\sum_{k=1}^{32} (k \times \sum_{l=1}^{32} p^L(k,l))}{\sum_{k=1}^{32} \sum_{l=1}^{32} p^L(k,l)} \quad (8)$$

$$\mu^L_{0^2} = \frac{(\sum_{k=1}^{32} (k - \mu^L_{0^1})^2 \times \sum_{l=1}^{32} p^L(k,l))}{\sum_{k=1}^{32} \sum_{l=1}^{32} p^L(k,l)} \quad (9)$$

$$\mu^L_{0^2} = \frac{(\sum_{k=1}^{32} (l - \mu^L_{0^2})^2 \times \sum_{l=1}^{32} p^L(k,l))}{\sum_{k=1}^{32} \sum_{l=1}^{32} p^L(k,l)} \quad (10)$$

$$E_{0^1}^L = -\frac{\sum_{k=1}^{32} p^L(k,l) \log \left( \sum_{l=1}^{32} p^L(k,l) \right)}{\sum_{k=1}^{32} \sum_{l=1}^{32} p^L(k,l)} \quad (11)$$

$$E_{0^2}^L = -\frac{\sum_{l=1}^{32} p^L(k,l) \log \left( \sum_{l=1}^{32} p^L(k,l) \right)}{\sum_{k=1}^{32} \sum_{l=1}^{32} p^L(k,l)} \quad (12)$$

$$C_{0^2}^L = \frac{\sum_{k=1}^{32} \sum_{l=1}^{32} (l - \mu^L_{0^2}) (k - \mu^L_{0^1}) \times \sum_{l=1}^{32} p^L(k,l)}{\sqrt{E_{0^1}^L} \times \sqrt{E_{0^2}^L}} \quad (15)$$

### 3.3. GLGOCM Features Extraction

The internal area of a defect image is of gray and texture distribution features. GLGOCM is proposed based on gray level and gradient orientation information, which can reflect gray distribution, orientation distribution and texture distribution in the internal area.

Define the internal area of a defect image to be $\Omega$, the number of pixels to be $N^\Omega$, and the $j^{th}$ pixel to be $P^\Omega_j$ with gray level $f_j = f(x_j,y_j)$, where $(x_j,y_j)$ is the coordinate of the $j^{th}$ pixel $(j=1,2,\ldots,N^\Omega)$. Suppose $W^\Omega_j$ is a $3 \times 3$ window with center point $P^\Omega_j$. The gray levels for pixels in $W^\Omega_j$ and its nearest neighbors are $f(x_j+\Delta x_j,y_j+\Delta y_j)$ ($\Delta x_j,\Delta y_j=-1,0,1,2$). In order to reduce the effect of rotation, gradient orientation of $P^\Omega_j$ is rotated to 0°. Then the other pixels in $W^\Omega_j$ are correspondingly turned. $(G_0)^{y_j}_{2}$ is defined as average of all gradient orientations in $W^\Omega_j$ that have been turned. It is calculated just like the following:

$$dx_i(k,l) = f_j(x_i+k+1,y_j+l) - f_j(x_i+k,y_j+l)$$

$$dy_j(k,l) = f_j(x_i+k,y_j+l+1) - f_j(x_i+k,y_j+l)$$

$$g_0(k,l) = \text{atan}[dy_j(k,l)/dx_i(k,l)]$$

$$G_{0^L}(k,l) = \begin{cases} g_0(k,l) + \pi, & dx_i(k,l) < 0 \\ g_0(k,l), & \text{other} \end{cases} \quad k,l = -1,0,1 \ldots \quad (16)$$

$$(G_0)^{y_j}_{2} = \frac{1}{9} \sum_{k=-1}^{0} \sum_{l=-1}^{0} \text{MOD}(G_{0^L}(k,l) - G_{0^L}(0,0) + \pi, \pi) \quad j=1,2,\ldots,N^\Omega \quad (17)$$

Then $f_j$ and $(G_0)^{y_j}_{2}$ are normalized into domain $[1,2,\ldots,32]$. Suppose the largest gray level and gradient orientation are $\max$ and $(G_0)^{y_{\text{max}}}^{\Omega}$ respectively. The normalized $(f_j)^{\Omega}$ and $(G_0)^{y_{\text{32}}}^{\Omega}$ are shown in the following:
\[
(f_j^\Omega)^{32}_1 = 1 + \text{INT} \left( f_j \times \frac{31}{f_{\text{max}}} \right) \\
\text{if } j = 1, 2, \ldots, N^\Omega \quad (18)
\]

\[
(G_0^\Omega)^{32}_2 = 1 + \text{INT} \left( \frac{G_0^\Omega}{f_{\text{max}}} \right) \\
\text{if } j = 1, 2, \ldots, N^\Omega \quad (19)
\]

For the internal area of the defect image, define a 32 \times 32 matrix \( H^\Omega \). The element \( H^\Omega(k,l) \) is the number of the pixels with gray level \( k = (f_j^\Omega)^{32}_1 \) and gradient orientation \( l = (G_0^\Omega)^{32}_2 \). The probability of \((k,l)\) in \( H^\Omega \) is defined as \( p(k,l) \), and it can be calculated just like the following:

\[
p^\Omega(k,l) = H^\Omega(k,l) / \sum_{l=1}^{32} H^\Omega(k,l) \quad (20)
\]

According to \( H^\Omega \) and the algorithm in references, \cite{17,18} calculate the corresponding GLGOCM features, such as mean of gray levels \( \mu^\Omega \), mean of gradient orientations \( \mu^\Omega \), variance of gray levels \( (\delta^\Omega)^2 \), variance of gradient orientations \( (\delta^\Omega)^2 \), entropy of gray levels \( E^\Omega \), entropy of gradient orientations \( E^\Omega \), mixed entropy \( E^\Omega \), inertia moment \( I^\Omega \), and correlation coefficient \( C^\Omega \). Their calculating formulas are just like the following:

\[
\mu^\Omega = \sum_{k=1}^{32} (k \times \sum_{l=1}^{32} p^\Omega(k,l)) \quad (21)
\]

\[
\mu^\Omega = \sum_{k=1}^{32} (l \times \sum_{l=1}^{32} p^\Omega(k,l)) \quad (22)
\]

\[
(\delta^\Omega)^2 = \sum_{k=1}^{32} (k - \mu^\Omega)^2 \times \sum_{l=1}^{32} p^\Omega(k,l) \quad (23)
\]

\[
(\delta^\Omega)^2 = \sum_{k=1}^{32} (l - \mu^\Omega)^2 \times \sum_{l=1}^{32} p^\Omega(k,l) \quad (24)
\]

\[
E^\Omega = -\sum_{k=1}^{32} \sum_{l=1}^{32} p^\Omega(k,l) \times \log(\sum_{l=1}^{32} p^\Omega(k,l)) \quad (25)
\]

\[
E^\Omega = -\sum_{l=1}^{32} \sum_{l=1}^{32} p^\Omega(k,l) \times \log(\sum_{l=1}^{32} p^\Omega(k,l)) \quad (26)
\]

\[
E^\Omega = \sum_{k=1}^{32} \sum_{l=1}^{32} (p^\Omega(k,l) \times \log(p^\Omega(k,l))) \quad (27)
\]

\[
I^\Omega = \sum_{k=1}^{32} \sum_{l=1}^{32} (k - l)^2 \times p^\Omega(k,l) \quad (28)
\]

\[
C^\Omega = \sum_{k=1}^{32} \sum_{l=1}^{32} ((k - \mu^\Omega)(l - \mu^\Omega) \times p^\Omega(k,l)) / (\delta^\Omega)^2 \quad (29)
\]

4. Classification Algorithm Based on Improved LSTWSVM

4.1. \( \rho \)-LSTWSVM

LSTWSVM\cite{19} is a least squares version of SVM, which combines the principles of TWSVM\cite{20} and proximal support vector machine (PSVM).\cite{21} Just as TWSVM, LSTWSVM finds two optimal separating hyper-planes. Similarly to PSVM, LSTWSVM solves two linear equations instead of quadratic programming problems (QPPs). So LSTWSVM is fit for cross plane and large scale datasets. Improved LSTWSVM based on amplification factor \( \rho \) (\( \rho \)-LSTWSVM) is proposed in this paper. It can realize high efficiency and accuracy classification for strip steel surface defect. Suppose the training dataset \( X = [X_1, X_2, \ldots, X_i] \) denotes sample matrix in class +1, and \( Y = [Y_1, Y_2, \ldots, Y_i] \) denotes sample matrix in class -1. \( \rho \)-LSTWSVM can be described as the following two primal QPPs:

\[
\begin{align*}
\text{min} & \quad \frac{1}{2} \sum_{i=1}^{n} \xi_i^2 + C \sum_{i=1}^{n} \zeta_i^2 \\
\text{s.t.} & \quad \rho(\phi(A,C)x_i + e_1b_i) = \xi_i \quad (30) \\
\rho(\phi(B,C)x_i + e_2b_i + e_1) = \zeta_i \quad (31)
\end{align*}
\]

where, \( C_1 \) and \( C_2 \) are penalty factors, \( \rho \) is any arbitrary kernel, \( e_1 \) and \( e_2 \) are vectors of ones, \( \| \cdot \| \) is L2 norm, \( A = [A_1, A_2, \ldots, A_1] \) is the pruned sample matrix in class +1, \( B = [B_1, B_2, \ldots, B_2] \) is the pruned sample matrix in class -1, \( C = [A^T, B^T] \).

\( \rho \)-LSTWSVM has two major improvements: one is that it pruned the original training datasets \( X \) and \( Y \) into \( A \) and \( B \) respectively, the other is that it added two diagonal matrices \( \rho_1 \) and \( \rho_2 \). Pruning training dataset can improve the efficiency of \( \rho \)-LSTWSVM, and adding amplification factors \( \rho_1 \) and \( \rho_2 \) can enhance the weight of pruned samples, which can ensure the accuracy of \( \rho \)-LSTWSVM.

On substituting the constraints into objective function of QPP (30) and setting the gradient of objective function with respect to \( w_1 \) and \( b_1 \) to zeros, the following result is obtained:

\[
\begin{align*}
\left[ \begin{array}{c}
\phi(A,C)x_1^+ \rho_1 \phi(A,C)x_1^+ \\
e_1 \rho_1 \phi(A,C)x_1^+
\end{array} \right] & = \left[ \begin{array}{c}
\phi(A,C)x_1^+ \rho_1 \phi(A,C)x_1^+ \\
e_1 \rho_1 \phi(A,C)x_1^+
\end{array} \right] b_1 \\
C_1 & \left[ \begin{array}{c}
\phi(B,C)x_1^- \rho_2 \phi(B,C)x_1^- \\
e_2 \rho_2 \phi(B,C)x_1^-
\end{array} \right] & = \left[ \begin{array}{c}
\phi(B,C)x_1^- \rho_2 \phi(B,C)x_1^- \\
e_2 \rho_2 \phi(B,C)x_1^-
\end{array} \right] b_1 \\
& = -C_1 & \left[ \begin{array}{c}
\phi(B,C)x_1^- \rho_2 \phi(B,C)x_1^- \\
e_2 \rho_2 \phi(B,C)x_1^-
\end{array} \right] b_1 \\
\left[ \begin{array}{c}
w_1 \\
b_1
\end{array} \right] & = -\frac{1}{C_1} \left[ \begin{array}{c}
G^T G + H^T H
\end{array} \right]^{-1} \left[ \begin{array}{c}
G^T G + H^T H
\end{array} \right] b_1
\end{align*}
\]

where \( G = [\rho_1 \phi(A,C) \rho_1 e_1] \) and \( H = [\rho_1 \phi(B,C) \rho_2 e_2] \).

Similarly, the result of QPP (31) is:

\[
\begin{align*}
\left[ \begin{array}{c}
w_2 \\
b_2
\end{array} \right] & = \frac{1}{C_2} \left[ \begin{array}{c}
G^T G + H^T H
\end{array} \right]^{-1} \left[ \begin{array}{c}
G^T G + H^T H
\end{array} \right] b_2 \\
& = 0
\end{align*}
\]

4.2. Pruning Method and Amplification Factor

\( \rho \)-LSTWSVM is implemented based on two processes of pruning training dataset and adding amplification factors for pruned samples. K-nearest neighbor\cite{22} and R- nearest neighbor methods are used to realize these processes. K-nearest
neighbor method can capture critical samples in training dataset, while \( R \)-nearest neighbor method can prune non-critical samples. Adding amplification factors for pruned samples can ensure classification accuracy. Their steps are:

Step 1: For sample matrix \( X \), calculate \( d_j = ||X_j - X_i|| \) \((i,j = 1,2,\cdots,I)\).

Step 2: Determine matrix element \( k_{ij} \) of \( K \)-nearest neighbors for every sample in \( X \) according to \( d_j \) and optimal \( K \).

\[
k_{ij} = \begin{cases} 
d_j & \text{if } X_j \text{ is } K \text{-nearest neighbor of } X_i \\
0 & \text{otherwise}
\end{cases} \quad i,j=1,2,\cdots,I \ldots (35)
\]

Step 3: calculate \( K \)-nearest neighbor frequency \( t_j \) and average distance \( d_j \) of \( X \). Where, \( \text{ONE} \) means setting one for non-zero element.

\[
t_j = \sum_{i=1}^{l} \text{ONE}(k_{ij}), \quad d_j = \frac{1}{t_j} \sum_{i=1}^{l} k_{ij} \quad i,j=1,2,\cdots,I \ldots (36)
\]

Step 4: Sort all \( d_j \) and all samples of \( X \) according to \( t_j \) and obtain the sorted \( d_j^i \) \((j = 1,2,\cdots,I)\) and \( X^i = [X_1^i, X_2^i, \cdots, X_I^i] \).

Step 5: obtain \( R \)-nearest neighbors of \( X_i \) with center point \( X_i^j \) and radius \( R = d_j^i \) and determine the number of samples \( m_i \) in that region.

Step 6: Prune these \( m_i \) samples from \( X \) and determine \( A_i = X_i^j \) and \((\rho_1)_{i1} = m_i \).

Step 7: Repeat step 5 and 6 to determine \( A_i \) and \((\rho_1)_{i1} \) until \( X \) is pruned to be empty matrix. Where, \( i = 1,2,\cdots,n1 \).

Pruned sample matrix \( A \) and amplification factor \( \rho_1 \) for training dataset \( X \) can be obtained by using the above method. Similarly, sample matrix \( B \) and amplification factor \( \rho_2 \) for training dataset \( Y \) can also be obtained.

5. Experiments

5.1. Building of Experimental Dataset

There are more than thirty types of defects during the process of strip steel production. And six typical types of defects are studied in the paper, such as crack, bruise, coil, damage, hole and scarring, which are shown in Fig. 5. If heating temperature is not consistent or the stress is too large after deformation, crack defect will appear on both sides with the shape of scale or water ripple. If the machine is not smooth, then bruise defect will appear because of galling. It looks like black under high temperature, while white under normal temperature. Coil defect appears because of edge damage and curling. Showing as undulating waves, damage defect appears because of uneven deformation. Hole defect appears because of the tearing of the billet steel, which is not continuous and perforates through upper and lower surface. Scarring defect appears because of unclean surface of billet steel, which is like irregular pit.

In order to test the efficiency of the novel method proposed in this paper, typical samples of strip steel surface defect are obtained from national large steel plants. 1 000 samples including six types of defects are selected. These selected samples are all complete images including defect and background areas. During the process of recognition, these images should be firstly disposed by preprocessing and segmentation, which have been done according to References 13–16. Then defect images with edge and internal areas are obtained. Moreover, in order to obtain images with different scales and rotations, these 1 000 defect images are rotated and resized in five times and 5 000 images are generated. In this way, the experimental dataset with 5 000 samples are built. The number of all samples for each type of defect is shown in Table 1.

On the other hand, in order to satisfy the demand of scale and rotation invariance in defect feature extraction, these defect images are required to transform into SBS images by searching SBS table. So SBS table of experimental dataset should be firstly obtained by implementing the scheme mentioned in section 3.1.

5.2. Experimental Results and Analysis

In order to verify the performance of strip steel surface recognition method proposed in this paper, some experiments are done to test the novel feature extraction and improved classifier. All experiments are performed by using Microsoft visual C++ 6.0 on a PC with an Intel P4 processor (2.8 GHz) and 2 GB RAM.

Firstly, verification experiments are done for feature extraction algorithms proposed in this paper. An image is randomly selected from 1 000 defect images. 4 images with different scales and rotations for this selected image can be obtained from the experimental dataset. 5 SBS images can be obtained by implementing the scheme shown in Fig. 3. Then GMGOCM features and GLGOCM features can be extracted from 5 SBS images, which are shown in Tables 2 and 3. It can be seen that GMGOCM and GLGOCM feature values are nearly the same for 5 defect images with different scales and rotations. All these experimental results verify that feature extraction methods proposed in this paper have scale and rotation invariance.

Secondly, testing experiments are done to manifest the efficiency of pruning samples and adding amplification factor, \( \rho \)-LSTWSVM and binary tree21) are combined to realize

![Fig. 5.](https://example.com/fig5.jpg) Images of six types of defects.

<table>
<thead>
<tr>
<th>Defect type</th>
<th>Crack</th>
<th>Bruise</th>
<th>Coil</th>
<th>Damage</th>
<th>Hole</th>
<th>Scarring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Samples</td>
<td>1,000</td>
<td>750</td>
<td>800</td>
<td>750</td>
<td>850</td>
<td>850</td>
</tr>
</tbody>
</table>
multi-class classification for strip steel surface defect. Radial basis function (RBF) is employed as ϕ of ρ-LSTWSVM. Kernel radius of RBF is obtained from $[2^{10}, 2^4]$, and parameters $C_1$ and $C_2$ are selected from $[2^{20}, 2^4]$. In the processes of testing experiments, 25-dimensional feature vector is extracted for every sample in the experimental dataset. These 25 features are composed of 18 GMGOCM and GLGOCM features and 7 moment invariant features. So each experimental sample is a 25-dimensional feature vector. All these samples are normalized into $[0, 1]$. 90 percent of these samples are training samples, while the others are testing samples. The classification results with different parameter $K$ are shown in Table 4. And the comparative results of using $ρ$ (u-$ρ$) and not using $ρ$ (n-$ρ$) are also shown in Table 4. It can be seen that the larger the parameter $K$ is, the shorter the training time and testing time are. When $ρ$ is not used, the accuracy will rapidly reduce as $K$ becomes larger and larger. When $ρ$ is used and $K \leq 6$, the classification accuracy is nearly the same. However, the classification accuracy will rapidly reduce when $K > 6$. So, in order to keep the balance between the time and accuracy, parameter $K$ is proposed as 6.

Finally, comparative experiments are done in order to manifest the superiority of the novel recognition method proposed in this paper. On one hand, the differences are shown between the novel scheme with GMGOCM, GLGOCM, and moment invariant features extraction (GGFE) and traditional features extraction (TFE) scheme in reference.17 On the other hand, comparable classification results are shown among SVM, LSTWSVM and ρ-LSTWSVM. In TFE scheme, methods of gray-level co-occurrence matrix features extraction, moment invariant features extraction and gray features extraction are employed. In the method of gray-level co-occurrence matrix features extraction, six features with horizontal nearest neighbors as well as the same six features with vertical nearest neighbors are extracted. These six features are angular second moment, contrast, correlation, sum entropy, sum variance and difference variance.17) In the method of moment invariant features extraction, seven features with size and rotation invariance are extracted.6) In the method of gray features extraction, six features including mean, variance, skewness coefficient, kurtosis coefficient, energy, and entropy are extracted. According to TFE scheme, typical 25 features are obtained. Similarly, 25 features are extracted according to GGFE scheme. Parameters for three classifiers are all chosen according to the methods in the above experiments. Parameter $K$ in ρ-LSTWSVM is set as 6. The final experimental results are shown in Table 5. GGFE and TFE have little difference in extracting time, which proves that GGFE scheme have stable efficiency. As far as training time and testing time are concerned, SVM classifier takes the longest time, and $ρ$-

Table 2. GMGOCM feature values for 5 defect images with different scales and rotations.

<table>
<thead>
<tr>
<th>θ</th>
<th>size</th>
<th>$μ_1$</th>
<th>$μ_2$</th>
<th>$(δ_1)^2$</th>
<th>$(δ_2)^2$</th>
<th>$E_1$</th>
<th>$E_2$</th>
<th>$E_3$</th>
<th>$I_{MO}$</th>
<th>$C_{MO}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0°</td>
<td>26×67</td>
<td>19.6800</td>
<td>15.5600</td>
<td>25.8943</td>
<td>83.9131</td>
<td>2.9702</td>
<td>2.5960</td>
<td>4.4809</td>
<td>120.9676</td>
<td>0.0279</td>
</tr>
<tr>
<td>0°</td>
<td>13×34</td>
<td>19.5764</td>
<td>15.4987</td>
<td>25.7362</td>
<td>83.9787</td>
<td>2.9358</td>
<td>2.5896</td>
<td>4.4637</td>
<td>120.7842</td>
<td>0.0213</td>
</tr>
<tr>
<td>0°</td>
<td>52×134</td>
<td>19.6362</td>
<td>15.5253</td>
<td>26.0012</td>
<td>84.1235</td>
<td>2.9654</td>
<td>2.5535</td>
<td>4.3671</td>
<td>120.9796</td>
<td>0.0259</td>
</tr>
<tr>
<td>68°</td>
<td>60×42</td>
<td>19.3804</td>
<td>15.6994</td>
<td>26.2197</td>
<td>84.4188</td>
<td>2.9672</td>
<td>2.5780</td>
<td>4.5396</td>
<td>120.7497</td>
<td>0.0271</td>
</tr>
<tr>
<td>113°</td>
<td>67×26</td>
<td>19.6375</td>
<td>15.3125</td>
<td>25.9311</td>
<td>85.0523</td>
<td>2.9329</td>
<td>2.5684</td>
<td>4.5189</td>
<td>121.1375</td>
<td>0.0280</td>
</tr>
</tbody>
</table>

Table 3. GLGOCM feature values for 5 defect images with different scales and rotations.

<table>
<thead>
<tr>
<th>θ</th>
<th>size</th>
<th>$μ_1$</th>
<th>$μ_2$</th>
<th>$(δ_1)^2$</th>
<th>$(δ_2)^2$</th>
<th>$E_1$</th>
<th>$E_2$</th>
<th>$E_3$</th>
<th>$I_{MO}$</th>
<th>$C_{MO}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0°</td>
<td>26×67</td>
<td>14.0509</td>
<td>18.2176</td>
<td>30.3695</td>
<td>34.6517</td>
<td>2.9283</td>
<td>3.0653</td>
<td>5.0675</td>
<td>71.6352</td>
<td>0.0269</td>
</tr>
<tr>
<td>0°</td>
<td>13×34</td>
<td>14.3526</td>
<td>18.1123</td>
<td>30.7548</td>
<td>34.5681</td>
<td>2.9017</td>
<td>3.1015</td>
<td>5.0745</td>
<td>72.0356</td>
<td>0.0218</td>
</tr>
<tr>
<td>0°</td>
<td>52×134</td>
<td>14.5017</td>
<td>18.5034</td>
<td>30.5249</td>
<td>34.9712</td>
<td>2.9316</td>
<td>3.0882</td>
<td>5.1139</td>
<td>71.9165</td>
<td>0.0224</td>
</tr>
<tr>
<td>68°</td>
<td>60×42</td>
<td>14.4041</td>
<td>18.2176</td>
<td>30.5249</td>
<td>34.6517</td>
<td>2.9997</td>
<td>3.1015</td>
<td>5.1267</td>
<td>72.1143</td>
<td>0.0224</td>
</tr>
<tr>
<td>113°</td>
<td>67×26</td>
<td>14.4685</td>
<td>18.3874</td>
<td>30.4202</td>
<td>33.9878</td>
<td>2.9905</td>
<td>3.0473</td>
<td>5.0015</td>
<td>71.9009</td>
<td>0.0208</td>
</tr>
</tbody>
</table>

Table 4. Testing results with different $K$ by using $ρ$ and not using $ρ$.

<table>
<thead>
<tr>
<th>$K$</th>
<th>Training time (s)</th>
<th>Testing time (s)</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>u-$ρ$/n-$ρ$</td>
<td>u-$ρ$/n-$ρ$</td>
<td>u-$ρ$/n-$ρ$</td>
</tr>
<tr>
<td>1</td>
<td>14.157/14.156</td>
<td>0.9670/0.9670</td>
<td>96.90/96.90</td>
</tr>
<tr>
<td>2</td>
<td>9.9651/9.8142</td>
<td>0.7495/0.7494</td>
<td>95.80/95.80</td>
</tr>
<tr>
<td>3</td>
<td>3.8902/3.7805</td>
<td>0.3508/0.3509</td>
<td>95.40/95.40</td>
</tr>
<tr>
<td>4</td>
<td>3.1116/3.0221</td>
<td>0.2601/0.2601</td>
<td>95.40/95.40</td>
</tr>
<tr>
<td>5</td>
<td>2.4565/2.3773</td>
<td>0.2232/0.2231</td>
<td>91.80/91.80</td>
</tr>
<tr>
<td>6</td>
<td>1.6132/1.5542</td>
<td>0.1648/0.1647</td>
<td>82.60/82.60</td>
</tr>
<tr>
<td>7</td>
<td>1.0842/1.0402</td>
<td>0.1046/0.1044</td>
<td>75.20/75.20</td>
</tr>
</tbody>
</table>

Table 5. Testing results of three feature extraction methods and three classifiers.

<table>
<thead>
<tr>
<th>Recognition method</th>
<th>Training time (s)</th>
<th>Testing time (s)</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFE + SVM</td>
<td>195.16</td>
<td>1.1733</td>
<td>91.20</td>
</tr>
<tr>
<td>TFE + LSTWSVM</td>
<td>4.8731</td>
<td>14.156</td>
<td>0.5165</td>
</tr>
<tr>
<td>TFE + ρ-LSTWSVM</td>
<td>3.1116</td>
<td>0.2601</td>
<td>94.00</td>
</tr>
<tr>
<td>GGFE + SVM</td>
<td>195.16</td>
<td>1.1733</td>
<td>93.20</td>
</tr>
<tr>
<td>GGFE + LSTWSVM</td>
<td>4.8753</td>
<td>14.156</td>
<td>0.5179</td>
</tr>
<tr>
<td>GGFE + ρ-LSTWSVM</td>
<td>3.1116</td>
<td>0.2601</td>
<td>95.40</td>
</tr>
</tbody>
</table>
LSTWSVM classifier is the fastest, which proves pruning samples used in $\rho$-LSTWSVM classifier can improve the efficiency. The accuracy of GGFE scheme is higher than that of TFE scheme when both of them are combined with the same classifier, which proves GGFE can improve the classification accuracy. However, for the same feature extraction scheme, the accuracy of $\rho$-LSTWSVM classifier is higher than that of SVM classifier, and is almost the same with that of LSTWSVM classifier, which proves $\rho$-LSTWSVM classifier can ensure the accuracy. In conclusion, the recognition algorithm with GGFE and $\rho$-LSTWSVM is of high efficiency and accuracy.

6. Conclusions

The novel algorithm proposed in this paper has been used in the strip steel surface defect recognition. The novel feature extraction scheme is to obtain SBS image for every sample by searching SBS table. And GMGOCM features of edge area, GLGOCM features of internal area, and moment invariant features of defect area are extracted based on SBS image. The novel classification algorithm improves the efficiency by pruning samples with $K$-nearest neighbor and $R$-nearest neighbor. And the improved LSTWSVM with amplification factor $\rho$ is used to ensure the classification accuracy. The simulation experiments prove that the feature extraction scheme is of scale and rotation invariance, which can improve the recognition accuracy. Moreover, they also prove pruning samples and adding amplification factors can improve the recognition accuracy and ensure the recognition efficiency. The final comparative experiments have clearly shown the merits of the novel recognition method proposed in this paper.

Acknowledgment

The authors wish to thank the reviewers and associate editor. Their comments and advice promote the paper more solid. The authors also give thanks to sponsor and support from National Natural Science Foundation of China (Grant No. 61050006).

REFERENCES

19) G. Fung and O. L. Mangasarian: Proc. of the 7th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, ACM, NY, (2001), 77.