Arabian Journal for Science and Engineering (2019) 44:2925–2932
https://doi.org/10.1007/s13369-018-3329-5

RESEARCH ARTICLE - COMPUTER ENGINEERING AND COMPUTER SCIENCE

Surface Defects Classification of Hot-Rolled Steel Strips Using Multi-directional Shearlet Features

Mohammed Waleed Ashour[1] · Fatimah Khalid[1] · Alfian Abdul Halin[1] · Lili Nurliyana Abdullah[1] · Samy Hassan Darwish[2]

Received: 1 August 2017 / Accepted: 17 May 2018 / Published online: 5 June 2018

Abstract
In this paper, a method combining the use of discrete shearlet transform (DST) and the gray-level co-occurrence matrix (GLCM) is presented to classify surface defects of hot-rolled steel strips into the six classes of rolled-in scale, patches, crazing, pitted surface, inclusion and scratches. Feature extraction involves the extraction of multi-directional shearlet features from each input image followed by GLCM calculations from all extracted sub-bands, from which a set of statistical features is extracted. The resultant high-dimensional feature vectors are then reduced using principal component analysis. A supervised support vector machine classifier is finally trained to classify the surface defects. The proposed feature set is compared against the Gabor, wavelets and the original GLCM in order to evaluate and validate its robustness. Experiments were conducted on a database of hot-rolled steel strips consisting of 1800 grayscale images whose defects exhibit high inter-class similarity as well as high intra-class appearance variations. Results indicate that the proposed DST–GLCM method is superior to other methods and achieves classification rates of 96.00%.

Keywords Steel surface classification · Manufacturing defects detection · Discrete shearlet transform · Hot-rolled steel strips · Gray-level co-occurrence matrix · Principal component analysis · Support vector machines

1 Introduction

Steel is a strong and robust material that is used to produce high-quality strips [1]. The quality of the flat steel surface is a crucial parameter determining the quality of the final product [2]. Hot-rolled steel is a common type of steel that is roll-pressed at temperatures over 1700°F, which is well above the recrystallization temperature for most steels [3]. This makes it more pliable and easier to work with. During the manufacturing process, a variety of surface defects for hot-rolled steel products are reported to be very high [4–6]. Many types of defects exhibit large inter-group similarity and intra-group diversity, meaning that it becomes very challenging to differentiate between defect types [7]. Coupled with the lack of defect-type standardization, the characterization and classification can vary from mill to mill and operator to operator. Further manifestations of certain defects can also change due to variations in the production process.

During production, manual inspection of strips is time-consuming and error-prone due to human factors (e.g., fatigue and inconsistency [8]). Thus, automated methods are desirable to guarantee the consistency of inspection with a reasonable degree of confidence. Specifically, image processing and computer vision techniques can be used where steel strip images can be acquired using specific sensing hardware and analyzed using specialized computer algorithms [2]. Specifically, computer vision methods can extract discriminative visual features from a material's surface image and those features are fed into a decision-making engine (e.g., supervised learning classifier) to determine the actual surface defect [2,9,10].

Generally, feature extraction methods can be extracted from the frequency and spatial domains [11]. Spatial-domain features deal with pixel value changes with respect to the scene [9]. One example is the gray-level co-occurrence matrix (GLCM), which is shown to be an effective feature for machined surfaces classification [12]. The GLCM is a texture descriptor that calculates how often specific pixel

✉ Mohammed Waleed Ashour
eng.m.ashour@gmail.com

[1] University Putra Malaysia, Serdang, Selangor, Malaysia

[2] Pharos University in Alexandria, Alexandria, Egypt

value pairs with specific spatial relationships occur within an image. However, since GLCM and its variants are merely spatial descriptors, they cannot capture the high- or low-frequency components of the multi-directional defects on a surface [13,14]. Frequency-domain features are therefore useful in such situations by modifying the spectral transform of an image. The discrete shearlet transform (DST), Gabor filters (GF) [15] and wavelet transforms (WT) [14] are popular techniques that have previously been used in early surface defects classification work. DST uses multiple directions at various decomposition scales, whereas shearlet is ideal in two-dimensional smooth functions approximation with discontinuities along $C2$-curves that produce almost the optimal properties of approximation [16]. Since steel strip surface defects vary widely (e.g., various types, shapes and orientations) [2,7], we postulate that shearlet transform might be highly suitable [10]. The proposed feature method in this paper is termed DST–GLCM for the task of surface defects classification of hot-rolled steel strips. The rest of this article is organized as follows: Sect. 2 provides a background of the features extraction methods. Section 3 discusses the features selection and reduction method of the extracted data using principal component analysis (PCA). Section 4 introduces the dataset and experimental setup. Section 5 presents the results and discussion. Finally, this work is concluded in Sect. 6.

2 Features Extraction

Features extraction and encoding is one of the most crucial stages in the design of a good classification system [17]. It mainly deals with the extraction of image properties resulting from low-level image processing operations. It is crucial that the extracted features are highly descriptive and discriminative of the class they are trying to represent [18]. In this section, we discuss the features used in this work, both the frequency and spatial domains.

2.1 Discrete Shearlet Transform

Shearlet transform was designed to be applied as a framework of affine systems to extract geometrical features from multi-dimensional signals. Over the past years, Guo and Labate [19] studied the construction of irregular shearlet systems and provided a framework for the construction of a variety of discrete directional multi-scale systems with the ability to detect orientations inherited from continuous shearlet transform. In 2008, they proposed a Fourier integral operator with respect to a Parseval frame of shearlets and the matrix representation of the method was proved to be sparse and well organized [20]. In 2010, W. Lim proposed the extended DST, adding extra basis elements to shearlet systems to obtain orthonormal basis for each shear matrix [16]. DST is a compactly supported shearlet generated by separable functions constructed using multi-resolution analysis [21]. DST can be configured to have various shapes, bandwidths, center frequencies and orientations. The main purpose of choosing DST in our work is because it provides a more flexible theoretical tool for approximating two-dimensional smooth functions with discontinuities and extract singularities of an image properly [22].

The two-dimensional DST of an image $h[x, y]$ of size $N_x \times N_y$ is a space frequency representation that can be expressed as two nested one-dimensional DST [23], i.e.,

$$S[n_x, n_y, k_x, k_y] = \sum_{l_x=0}^{N_x-1} A_y[n_y, l_x, k_x, k_y] e^{-\frac{2\pi^2 l_x^2}{k_x^2}} e^{\frac{i2\pi l_x n_x}{N_x}}, \quad (1)$$

where

$$A_y[n_y, k_x, k_y] = \sum_{l_y=0}^{N_y-1} H[k_x, l_x + k_y] e^{\frac{2\pi^2 l_y^2}{k_y^2}} e^{\frac{i2\pi l_y n_y}{N_y}}. \quad (2)$$

$H[k_x, k_y]$ represents the DFT of $h[x, y]$. $S[n_x, n_y, k_x, k_y]$ contains the 2D frequency index $[k_x, k_y]$ at the pixel $[n_x, n_y]$. Equation 1 is valid when k_x and k_y are positive. If both are positive and $k_y = 0$, the 2D DST is defined as:

$$S[n_x, n_y, k_x, 0] = \sum_{l_x=0}^{N_x-1} A_y[l_x, k_x] e^{-\frac{2\pi^2 l_x^2}{k_x^2}} e^{\frac{i2\pi l_x n_x}{N_x}}, \quad (3)$$

where

$$A_y[K_x] = \frac{1}{N_x N_y} \sum_{i_x=0}^{N_x-1} \sum_{i_y=0}^{N_y-1} h[i_x, i_y] e^{-\frac{i2\pi k_x i_x}{N_x}}; \quad (4)$$

if $K_x = 0$ and $K_y = 0$ is positive, the 2D DST can be obtained by interchanging the subscripts x and y in (3) and (4). If $K_x = 0$ and $K_y = 0$, the 2D ST is computed as

$$S[n_x, n_y, 0, 0] = \frac{1}{N_x N_y} \sum_{i_x=0}^{N_x-1} \sum_{i_y=0}^{N_y-1} h[i_x, i_y]. \quad (5)$$

Shearlet transform can decompose an image into $\boldsymbol{2 \times 2^{nj} + 1}$-directional sub-bands in the horizontal cone and the vertical cone at any scale, respectively. The $\boldsymbol{j}$ is denoted as the decomposition level, and n_j refers to the direction parameter (*ndir*). The schematic diagram of applying DST–GLCM method is illustrated in Fig. 1. The aim of this method is firstly to extract all the significant information from the textured surface image by means of the DST and, secondly, to calculate the GLCM for all the extracted sub-bands. This is

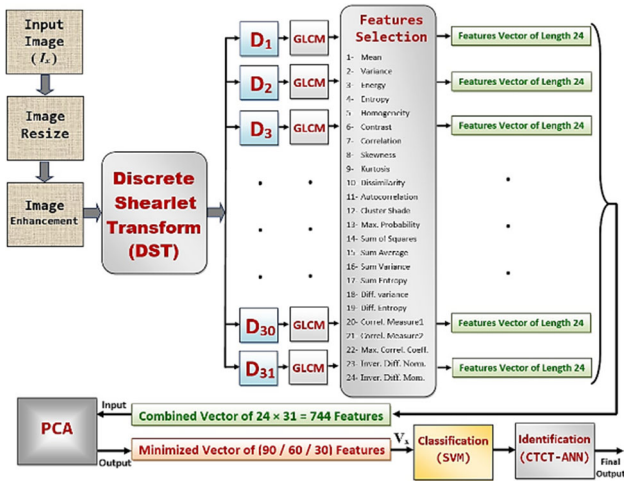

Fig. 1 Schematic diagram of DST–GLCM

performed in order to detect the information nonvisible by the GLCM-based approaches while still capitalizing on the effectiveness of the traditional co-occurrence matrix.

In this work, we choose the parameters of five decomposition levels, with [2, 4, 8, 16] shearing directions and filter of sizes 32 × 32, 32 × 32, 16 × 16 and 8 × 8 for the decomposition levels. Note that these parameters returned the best classification rates in our experiments. The output of this algorithm is the shearlet coefficients we acquire from each image. So, with the aforementioned filter structure and for every input image ($\boldsymbol{I}_x$) of size 200 × 200, there are $1 + 2 + 4 + 8 + 16 = 31$ filters and each one of them has $200 \times 200 = 40{,}000$ coefficients. For each one of the 31 generated sub-band images from the five-level decomposition procedure, GLCM is computed and then the set of 24 statistical features are computed from all the resultant GLCMs. This method results in a feature vector of size 744 features.

2.2 Gabor Filters

Generally, Gabor filters can be configured to have various shapes, bandwidths, center frequencies and orientations by the adjustment of suitable parameters. By varying these parameters, i.e., the tunable orientation and radial frequency bandwidths, a filter can be made to pass any elliptical region of spatial frequencies. It optimally achieves joint resolution in space and spatial frequency domains. In this work, the two-dimensional GF is used to extract the high-frequency components of every input image in twelve orientations, specifically 0°, 15°, 30°, 45°, 60°, 75°, 90°, 105°, 120°, 135°, 150° and 165°. This is so that all possible directions are taken into account, hence offering the best texture description for the surface image. The parameters' values of the used GF were as follows: wavelength = 10, phase offset = 0, aspect ratio = 0.5, and bandwidth = 1. These parameters were chosen after performing a number of trials in order to select the ideally suited values for extracting the most significant features. After extracting Gabor coefficients from the input image, GLCM is computed for each one of the twelve generated sub-bands and then a set of 24 statistical features are computed from all the resultant GLCMs. This method results in a feature vector of size 288 features.

2.3 Biorthogonal Wavelet

Biorthogonal wavelets, which are symmetric, prevent the contents of an image from shifting between sub-bands while allowing extension at the boundaries. The two-dimensional wavelet transforms (2D-WT) usually decompose images into four coefficients at level j. The approximation, and the details in three orientations (horizontal, vertical and diagonal). In this work, the two-dimensional biorthogonal wavelet transform is employed to decompose input image into four multiple levels. Every decomposing level produces one approximation and three details in (horizontal, vertical and diagonal). The aim is to extract the frequency components that represent all the surface defects of an image. In this process, the biorthogonal wavelet was applied on the input image, to be decomposed into 4 sub-images, where each one has $(m \times n)/2$ size: $L_{10}L_{10}$, $L_{10}H_{10}$, $H_{10}L_{10}$ and $H_{10}H_{10}$. In the $L_{10}L_{10}$ sub-image, there are "low frequencies" in both directions: "horizontal and vertical." In the $L_{10}H_{10}$ sub-image, there are "low frequencies in horizontal direction and high ones in the vertical direction." In the $H_{10}L_{10}$ sub-image, there are "high frequencies in horizontal direction and low ones in the vertical direction," and in the $H_{10}H_{10}$ sub-image, there are high frequencies in both directions. The biorthogonal wavelet was then applied again on $L_{10}L_{10}$ image with size $(m \times n)/2$ to get four new sub-images, each one with $(m \times n)/4$ size: $L_{11}L_{11}$, $L_{11}H_{11}$, $H_{11}L_{11}$ and $H_{11}H_{11}$. The same process continues for two more times to have sub-images with $(m \times n)/8$ and $(m \times n)/16$, respectively. For each one of the twelve generated sub-band images from the four-level decomposition procedure, GLCM is computed and then the set of 24 statistical features are computed from all the resultant GLCMs. This method results in a feature vector of size 288 features.

2.4 Gray-Level Co-occurrence Matrix

The gray-level co-occurrence matrix was introduced by Haralick et al. [24]. It is generally referred to as the matrix whose entries are transitions between all pairs of two gray levels. The gray-level transitions are calculated based on two parameters, displacement and angular rotation, giving four gray-level co-occurrence matrices at 0°, 45°, 90°, 135° orientations. In this paper the GLCM is computed for the input image in the four directions; then, the same set of 24 statistical

features is calculated from the generated four GLCMs. All the calculated statistical features from the all the matrices are then combined in one high-dimensional feature vector. The dimensionality of the resultant feature vector space is then reduced using PCA. This method results in a feature vector of size 96 features.

2.5 Statistical Features Computation

The proposed statistical features set aims to minimize the number of features extracted from every image. Several tests were performed on various types of features to come up with the optimum set consisting of the features in Eqs. (6–29). The statistical features set is computed from every GLCM that is calculated from an extracted sub-band image based on the earlier-mentioned methods DST, GF, WT and the original GLCM. The statistical features set comprises the following features:

$$\text{Mean: } \mu = \sum_{i=0}^{G-1} ip\,(i) \tag{6}$$

$$\text{Variance: } \sigma^2 = \sum_{i=0}^{G-1} (i-\mu)^2\, p\,(i) \tag{7}$$

$$\text{Energy: } \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} [p(i,j)]^2 \tag{8}$$

$$\text{Entropy: } \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} P(i,j)\log_2[p(i,j)] \tag{9}$$

$$\text{Homogeneity: } \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} \frac{1}{1-(i-j)^2} P\,(i,j) \tag{10}$$

$$\text{Contrast: } \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} (i-j)^2\, P\,(i,j) \tag{11}$$

$$\text{Correlation: } \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} \frac{ijp(i,j)-\mu_x\mu_y}{\sigma_x\sigma_y} \tag{12}$$

$$\text{Skewness: } \mu_3 = \sigma^{-3} \sum_{i=0}^{G-1} (i-\mu)^3\, p\,(i) \tag{13}$$

$$\text{Kurtosis: } \mu_3 = \sigma^{-4} \sum_{i=0}^{G-1} (i-\mu)^4\, p\,(i) \tag{14}$$

$$\text{Dissimilarity: } \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} |i-j|.p(i,j) \tag{15}$$

$$\text{Autocorrelation: } \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} (ij)p(i,j) \tag{16}$$

$$\text{Cluster shade : } \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} (i+j-\mu_x-\mu_y)^3 p(i,j) \tag{17}$$

$$\text{Maximum probability: } \max(i,j)p(i,j) \tag{18}$$

$$\text{Sum of squares : } \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} (i-\mu)^2\, p(i,j) \tag{19}$$

$$\text{Sum average : } \sum_{i=0}^{2G-2} ip_{x+y}(i) \tag{20}$$

$$\text{Sum variance } \sum_{i=0}^{2G-2} (i-\text{SumAverage})^2 p_{x+y}(i) \tag{21}$$

$$\text{Sum entropy: } \sum_{i=0}^{2G-2} p_{x+y}(i)\log(p_{x+y}(i)) \tag{22}$$

$$\text{Difference variance : } \sum_{i=0}^{G-1} (i-\mu_{x-y})^2 p_{x-y}(i) \tag{23}$$

$$\text{Difference entropy: } -\sum_{i=0}^{G-1} p_{x-y}(i)\log(p_{x-y}(i)) \tag{24}$$

Information measures of correlation 1:

$$\frac{\text{Entropy} - HXY1}{\max\{HX, HY\}} \tag{25}$$

Information measures of correlation 2:

$$\sqrt{1-\exp[-2(HXY2-\text{Entropy})]} \tag{26}$$

Maximum correlation coefficient: (second largest eigenvalue of $Q)^{0.5}$, where

$$Q(i,j) = \sum_{k=1}^{G} \frac{P(i,k)p(j,k)}{p_x(i)p_y(k)} \tag{27}$$

Inverse difference normalized:

$$\sum_i\sum_j \frac{p(i,j)}{1+|i-j|} \tag{28}$$

Inverse difference moment normalized:

$$\sum_i\sum_j \frac{p(i,j)}{1+(i-j)^2}, \tag{29}$$

where

$$P_x = \sum_{j=1}^{G} p(i,j), \quad and \quad P_x = \sum_{i=1}^{G} p(i,j)$$

$$P_{x+y}(k) = \sum_{i=1}^{G} \sum_{j=1 i+j=k}^{G} p(i,j), \quad \text{for } k = 2, 3, \ldots 2G$$

$$P_{x-y}(k) = \sum_{i=1}^{G} \sum_{j=1 |i-j|=k}^{G} p(i,j), \quad \text{for } k = 0, 1, \ldots G$$

$$HX = \sum_{i=1}^{G} p_x(i) \log(p_x(i))$$

$$HY = \sum_{j=1}^{G} p_y(j) \log(p_y(j)).$$

The function $p(i, j)$ stands for (i, j)th entry or value in a matrix of an image, where (G) is the total number of intensity levels in the image, while μ_x, μ_y and σ_x, σ_y denote the mean and standard deviations of the row and column sums of the matrix, respectively.

3 Features Selection and Reduction

In general, features selection and reduction is applied to the high-dimensional feature vectors in order to remove useless features and redundant information. Large number of features that represent a single image may increase the complexity of the classifier which in turn may slow down the training process and affect badly the final classification accuracy [25]. In this work, the features reduction is performed by applying PCA to minimize the size of the combined input vectors at the feature extraction level. PCA is an effective statistical analysis technique for producing a lower-dimensional representation of larger dataset [26]. The main idea of PCA is based on a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called the principal components [27]. Usually, the number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it be orthogonal to (i.e., uncorrelated with) the preceding components. In this stage, the aim is to reduce the dimensional size of feature vector obtained from features extraction methods to the lowest possible size which allows better discrimination between different classes and produces the highest classification accuracy rates. The obtained number of features using the four methods DST, GF, WT and GLCM was 744, 288, 288 and 96, respectively. Figure 2 shows the detailed process of features extraction and reduction for all the four features methods. The PCA is applied several times on the extracted feature vectors to generate a reduced number of features than the original for every image. After conducting a number of trials, on reducing the dimensional size of the feature vectors to 30, 60, 90 and 120, the dimensions of size 90 features seem to be superior in classification than the others.

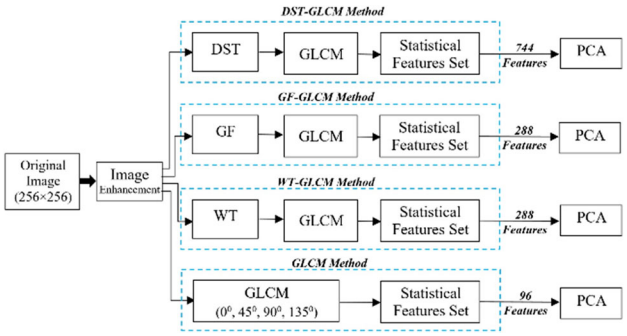

Fig. 2 Features extraction and reduction for DST, GF, WT and GLCM methods

4 Experimental Work

The dataset used in this research is the Northeastern University (NEU) surface defect database [7], where six types of typical surface defects of the hot-rolled steel strip are collected as: rolled-in scale (RS), patches (Pa), crazing (Cr), pitted surface (PS), inclusion (In) and scratches (Sc). The database includes 1800 grayscale images divided into six different types of typical surface defects each of 300 samples. The original resolution of each image is 200 × 200 pixels. The main challenges of this database are (i) high intra-class differences, as shown in Fig. 3a, and (ii) low inter-class difference as shown in Fig. 3b.

The proposed framework presented in Fig. 4 utilizes the concept of multi-level image analysis. Firstly, each input image undergoes an enhancement process using histogram equalization, and then the feature extraction process is performed followed by calculating the GLCM in zero degrees for all the extracted coefficients. The set of 24 statistical features is then computed for the generated GLCMs. This aims to uncover patterns not visible through the GLCM-based approaches while still capitalizing on the effectiveness of the traditional co-occurrence matrix. Next, all the extracted features from the input image are combined in one high-dimensional feature vector. This vector is reduced to the size of 90 features using PCA. The steps of the proposed image analysis algorithm are summarized as follows:

1. DST- based method for (I_x).
   - Extract DST coefficients at five different scales and orientations.
   - For every sub-band image 1:31

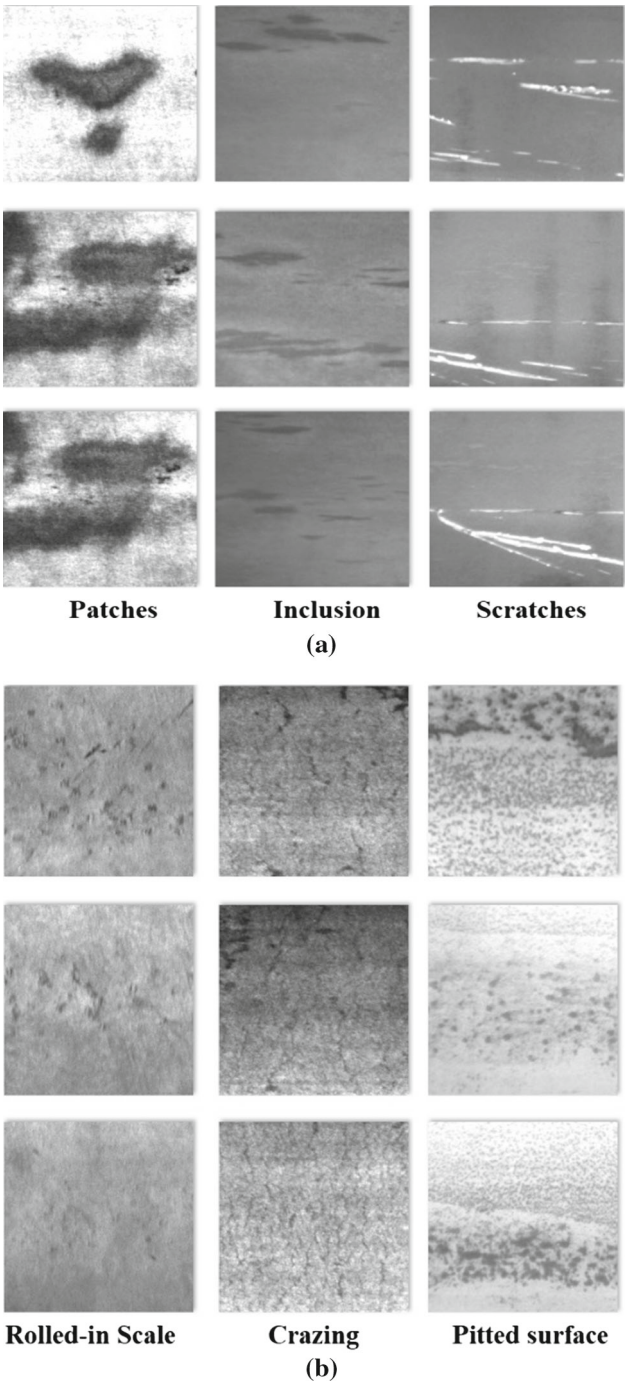

Fig. 3 **a** Patches, inclusion and scratches defects. **b** Rolled-in scale, crazing and pitted surface defects

- Calculate the GLCM at (0°).
- Extract the set of 24 statistical features for the generated GLCMs.

2. GF-based method for (I_x).

- Extract multiple Gabor features by using 2D-GF in twelve different orientations (0°, 15°, 30°, 45°, 60°, 75°, 90°, 105°, 120°, 135°, 150°, 165°).

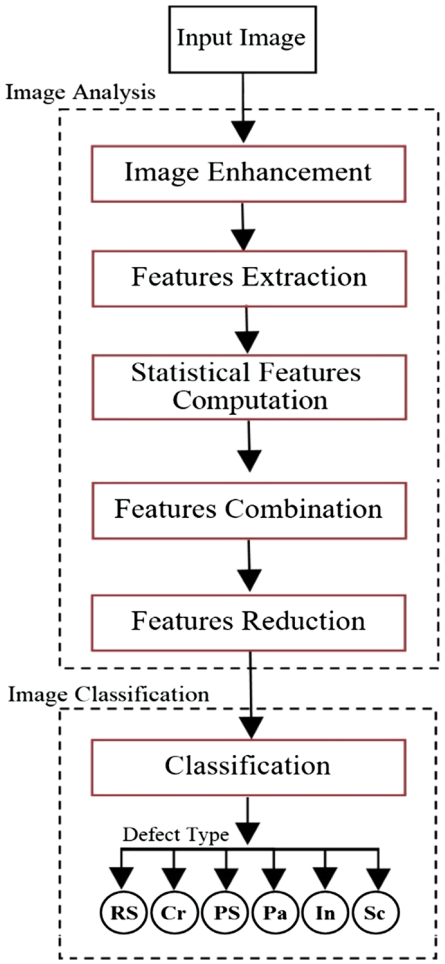

Fig. 4 Proposed workflow

- For every sub-band image 1:12
- Calculate the GLCM at (0°).
- Extract the set of 24 statistical features for the generated GLCMs.

3. WT-based method for (I_x).

- Extract biorthogonal wavelet coefficients for four decomposition levels and three orientations: horizontal (0°), diagonal (135°) and vertical (90°).
- For every sub-band image 1:12
- Calculate the GLCMs at (0°).
- Calculate the set of 24 statistical features for the generated GLCMs.

4. GLCM-based method for (I_x).

- Extract GLCM coefficients at the four orientations (0°, 45°, 90°, 135°).
- For every sub-band image 1:4
- Extract the set of 24 statistical features

Table 1 Overall classification of different methods with evaluation set (50–50%)

Feature extraction method	Feature dimension		Samples classified	SVM classification rate (%)
	Original	PCA-based		
GLCM	96	90	731/900	81.22
WT–GLCM	288	90	779/900	86.55
GF–GLCM	288	90	809/900	89.88
DST–GLCM	744	90	847/900	94.11

Table 2 Overall classification of different methods with evaluation set (75–25%)

Feature extraction method	Feature dimension		Samples classified	SVM classification rate (%)
	Original	PCA-based		
GLCM	96	90	379/450	84.22
WT–GLCM	288	90	403/450	89.55
GF–GLCM	288	90	418/450	92.88
DST–GLCM	744	90	432/450	96.00

The SVM setup as presented in our previous work [15] was used for the classification in the final stage. The feature vectors of every method are fed individually into our SVM classifier with RBF for final classification of the defects. The evaluation of classification stage is accomplished with two data partition sets for training and testing as (50–50%) and (75–25%). Random sets of images per class are chosen for training and the remaining for testing. The reason of using multiple partition sets in classification is to achieve the best discrimination using the lowest number of training samples, and thus reduce classification complexity and increase the time of defect detection. The next section presents the results and discussion of this research.

5 Results and Discussion

The classification results of the proposed framework are presented in Tables 1 and 2 showing the two training–testing splits of 50–50 and 75–25%, respectively. The overall classification results are presented for the four different features extraction methods DST–GLCM, GF–GLCM, WT–GLCM and the original GLCM. The results can clearly show that the highest classification accuracy rate was achieved using the evaluation set (75–25%). The DST–GLCM method achieved the highest accuracy rates among all other methods in both classification sets. As presented in Table 2, the highest classification accuracy rate was 96.00% and it was achieved by the DST–GLCM with the (75–25%) set. The second highest accuracy was achieved using Gabor filters, the third was using wavelets, and then finally the lowest rate was achieved using

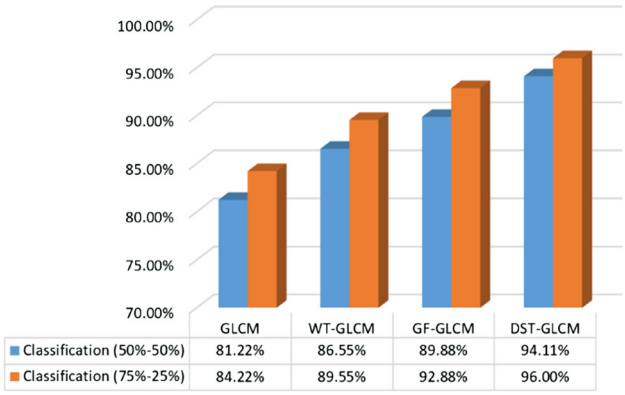

Fig. 5 Overall results of different methods

	Pa	In	Sc	RS	Cr	PS	Precision
Pa	74	0	0	1	0	0	98.67 % 1.33 %
In	0	74	1	0	0	0	98.67 % 1.33 %
Sc	0	2	72	0	0	1	96.00% 4.00 %
RS	1	0	0	71	1	2	94.66% 5.34%
Cr	0	0	0	2	71	2	94.66% 5.34%
PS	0	1	0	2	2	70	93.33% 6.67%
Recall	98.67% 1.33 %	96.00% 4.00 %	98.67% 1.33 %	93.33% 6.67%	96.00% 4.00 %	93.33% 6.67%	96.00% 4.00%

Fig. 6 Confusion matrix of DST–GLCM method

the original GLCM. The summary of the overall results is presented in Fig. 5. The confusion matrix of the DST–GLCM method which is presented in Fig. 6 shows the classification results for the six classes. The (Pa), (In) and (Sc) classes have achieved higher classification rate than the others. This was due to the large differences in appearance that exist in the intra-class samples. The inter-class defect which has similar aspects in appearance achieved lower accuracy rate.

6 Conclusions and Future Work

In this paper a high discriminative feature extraction method called DST–GLCM was presented for accurate defects classification of steel strip surfaces. A dataset of 1800 hot-rolled steel strips was used for classification. Our technique has shown an improvement in classification accuracy compared to other methods. These improvements were achieved by taking advantages of the combined features extracted using DST–GLCM along with the set of 24 statistical features. The reduced feature set using PCA managed to maintain its discriminative prowess, which further helped in speeding up the whole process. Overall accuracy rate of 96.00% was reported using the proposed method, while accuracy rates of 92.88,

89.55 and 84.22% were achieved, respectively, using GF–GLCM, WT–GLCM and original GLCM methods.

References

1. Davim, P.J. (ed.): Surface Integrity in Machining. Springer, New York (2010)
2. Neogi, N.; Mohanta, D.K.; Dutta, P.K.: Review of vision-based steel surface inspection systems. EURASIP J. Image Video Process. **1**, 50 (2014)
3. Metal Supermarkets: Difference between hot and cold rolled steel (2014). https://www.metalsupermarkets.com/difference-between-hot-rolled-steel-and-cold-rolled-steel/. Accessed 23 Jan 2017.
4. Xu, K.; Liu, S.; Ai, Y.: Application of shearlet transform to classification of surface defects for metals. Image Vis. Comput. **35**, 23–30 (2015)
5. Sharifzadeh, M.; Alirezaee, S.; Amirfattahi, R.; Sadri, S.: Detection of steel defect using the image processing algorithms. In: Multitopic Conference, 2008. INMIC 2008. IEEE International, pp. 125–127. IEEE (2008)
6. Luiz, A.M.; Flávio, L.P.; Paulo, E.A.: Automatic detection of surface defects on rolled steel using computer vision and artificial neural networks. In: IECON 2010—36th Annual Conference on IEEE Industrial Electronics Society, pp. 1081–1086. IEEE (2010)
7. Song, K.; Hu, S.; Yan, Y.: Automatic recognition of surface defects on hot-rolled steel strip using scattering convolution network. J. Comput. Inf. Syst **10**(7), 3049–3055 (2014)
8. Smith, C.J.; Adendorff, K.: Advantages and limitations of an automated visual inspection system. S. Afr. J. Ind. Eng. **5**(1) (2012)
9. Puig, D.; Garcia, M.A.: Pixel-based texture classification by integration of multiple feature extraction methods evaluated over multisized windows. Int. J. Pattern Recognit. Artif. Intell. **21**(07), 1159–1170 (2007)
10. Kutyniok, G.; Labate, D.: Shearlets: Multiscale Analysis for Multivariate Data. Birkhäuser, New York (2012)
11. Li, H.; Wang, X.; Tang, J.; Zhao, C.: Combining global and local matching of multiple features for precise item image retrieval. Multimed. Syst. **19**(1), 37–49 (2012)
12. Ashour, M.W.; Hussin, M.F.; Mahar, K.M.: Supervised texture classification using several features extraction techniques based on ANN and SVM. In: IEEE/ACS International Conference on Computer Systems and Applications, pp. 567–574 (2008)
13. Zheng, D.; Zhao, Y.; Wang, J.: August. Features extraction using a Gabor filter family. In: Proceedings of the Sixth Lasted International conference, Signal and Image Processing, Hawaii (2004)
14. Guo, K.; Kutyniok, G.; Labate, D.: Sparse multidimensional representations using anisotropic dilation and shear operators. In: Chen, G., Lai, M.J. (eds.) Wavelets and Splines (Athens, GA, 2005), pp. 189–201. Nashboro Press, Nashville (2006)
15. Ashour, M.W.; Halin, A.A.; Khalid, F.; Abdullah, L.N.; Darwish, S.H.: Texture-based classification of workpiece surface images using the support vector machine. Int. J. Softw. Eng. Appl. **9**(10), 147–160 (2015)
16. Lim, W.Q.: The discrete shearlet transform: a new directional transform and compactly supported shearlet frames. IEEE Trans. Image Process. **19**(5), 1166–1180 (2010)
17. Lee, M.; Hur, S.; Park, Y.; April. An obstacle classification method using multi-feature comparison based on 2D LIDAR database. In: 12th International Conference on Information Technology—New Generations (ITNG), pp. 674–679 (2015)
18. Ping Tian, D.: A review on image feature extraction and representation techniques. Int. J. Multimed. Ubiquitous Eng. **8**(4), 385–396 (2013)
19. Guo, K.; Kutyniok, G.; Labate, D.: Sparse multidimensional representations using anisotropic dilation and shear operators(2006)
20. Easley, G.; Labate, D.; Lim, W.Q.: Sparse directional image representations using the discrete shearlet transform. Appl. Comput. Harmon. Anal. **25**(1), 25–46 (2008)
21. Kutyniok, G.; Labate, D. ed.: Shearlets: Multiscale Analysis for Multivariate Data. Springer Science & Business Media (2012).
22. Vivek, C.; Audithan, S.: Texture classification by shearlet band signatures. Asian J. Sci. Res. **7**(1), 94 (2014)
23. Pradhan, P.M.; Cheng, C.H.; Mitchell, J.R.: S-transform based approach for texture analysis of medical images. In: 2014 International Conference on High Performance Computing and Applications (ICHPCA), pp. 1–4. IEEE (2014)
24. Haralick, R.M.: Statistical and structural approaches to texture. Proc. IEEE **67**(5), 786–804 (1979)
25. Hassaballah, M.; Abdelmgeid, A.A.; Alshazly, H.A.: Image features detection, description and matching. In: Image Feature Detectors and Descriptors, pp. 11–45. Springer, New York (2016)
26. Sachin, D.: Dimensionality reduction and classification through PCA and LDA. Int. J. Comput. Appl. **122**(17) (2015)
27. Kambhatla, N.; Leen, T.K.: Dimension reduction by local principal component analysis. Dimension **9**(7), 1493–1516 (2006)