Applying wavelets transform, rough set theory and support vector machine for copper clad laminate defects classification

Te-Sheng Li *

Department of Industrial Engineering and Management, Minghsin University of Science and Technology, Hsinchu, Taiwan

A R T I C L E  I N F O

Keywords:
Copper clad laminate
Discrete wavelet transform
Inverse discrete wavelet transform
Feature extraction
Rough set theory
Support vector machine

A B S T R A C T

In this paper, we present a multi-resolution approach for the inspection of local defects embedded in homogeneous copper clad laminate (CCL) surfaces. The proposed method does not just rely on the extraction of local textural features in a spatial basis. It is based mainly on reconstructed images using the wavelet transform and inverse wavelet transform on the smooth sub-image and detail sub-images by properly selecting the adequate wavelet bases as well as the number of decomposition levels. The restored image will remove regular, repetitive texture patterns and enhance only local anomalies. Based on these local anomalies, feature extraction methods can then be used to discriminate between the defective regions and homogeneous regions in the restored image. Rough set feature selection algorithms are employed to select the feature. Rough set theory can deal with vagueness and uncertainties in image analysis, and can efficiently reduce the dimensionality of the feature space. Real samples with four classes of defects have been classified using the novel multi-classifier, namely, support vector machine. Effects of different sampling approach, kernel functions, and parameter settings used for SVM classification are thoroughly evaluated and discussed. The experimental results were also compared with the error back-propagation neural network classifier to demonstrate the efficacy of the proposed method.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

Visual inspection plays a vital role of quality control in manufacturing systems. Manual inspection is subjective and highly dependent on the inspector's expertise. In this study, we use wavelet transform combining rough set feature selector and a novel classifier, support vector machine to detect and classify copper clad laminate (CCL) surface defects. Most CCL surface defects are tiny involving obvious faulty items such as pinholes, stains, scratches, strips and other ill-defined defects. These unanticipated defects are small in size, refractive in light and cannot be described using explicit measures, making automatic defect detection difficult.

Because CCL defects are similar to texture patterns, the inspection task in this study can be classified as texture detection and classification. In this environment, one must solve the problem of detecting small surface defects that break the local homogeneity of a texture pattern. Most defect detection methods for texture surfaces generally involve computing a set of textural features in a sliding window. The system searches for significant local deviations in the feature values. The most difficult task in this approach is feature extraction and feature selection which completely embody the texture information in the image. There is no arbitrary approach for selecting and judging the appropriate features to use. Therefore, selecting an adequate feature set for a new texture in the training process can be very time-consuming, requiring the help of expert knowledge. The more features extracted from the texture image, the more sophisticated the classifier such as Bayes (Gonzalez & Woods, 1992), maximum likelihood (Cohen, 1992), and neural networks (Van Hulle & Tollenaere, 1993) needed to discriminate the texture classes.

Numerous methods have been proposed to extract textural features either directly from the spatial domain or from the spectral domain. In the spatial domain, the more reliable and commonly used features are second-order statistics derived from spatial gray-level co-occurrence matrices (Haralick, Shanmugam, & Dinstein, 1973). They have been applied to wood inspection (Conners, McMillin, Lin, & Vasquez-Espinosa, 1983), carpet wear assessment (Siew & Hogdson, 1988), and roughness measurements for machined surfaces (Ramana & Ramamorthy, 1996).

In the past, Fourier transform and Gabor filter were two of the spectral domain approaches used to find the more adequate textural features that were less sensitive to noise and intensity variations. The Fourier transform is a global approach that characterizes only the spatial-frequency distribution, but it does not consider this information in the spatial domain. The Gabor filters are well regarded as a joint spectral representation for analyzing textured images with highly specific frequency and orientation characteristics. This technique extracts features by filtering the textured image with a set of Gabor filter banks characterized by the
frequency, the sinusoid orientation, and the Gaussian function scale. However, obtaining the optimal Gabor filter design is difficult because human intervention is required to select the appropriate filter parameters for the texture under study. If the center frequency of a selected Gabor filter does not match any of the important harmonics in the textured image, only noisy information is produced (Pichler, Teuner, & Hosticka, 1996).

Recently, multi-resolution decomposition schemes based on discrete wavelet transforms (DWT) have received considerable attention as alternatives for textural feature extraction. The multi-resolution wavelet representation allows an image to be decomposed into a hierarchy of localized subimages at different spatial frequencies (Chen & Lee, 1997). It divides the 2-D frequency spectrum of an image into a lowpass (smooth) subimage and a set of highpass (detail) subimages. The textural features are then extracted from the decomposed subimages in different frequency channels and different resolution levels. Chen and Kuo (1993) proposed a tree-structured wavelet transform for texture classification. They used an energy criterion to select the subimage for decomposition. A set of textural features is derived from the energy values of the dominant channels, and distance measures are then employed to discriminate between the texture classes. Lambert and Bock (1997) proposed a feature extraction approach for texture defect detection. The textural features were derived from the wavelet packet decomposition coefficients. Neural network and Bayes classifiers were used to evaluate the feature vector. The wavelet-based feature extraction methods were applied to industrial material inspection such as LSI wafers (Lee, Choi, Choi, & Cho, 1997; Maruo, Shibata, Yamaguchi, Ichikawa, & Ohmi, 1999), cold rolled strips (Sari-Sarraf & Goddard, 1998), and woven fabrics (Tsai & Hsieh, 1999).

Tsai and Hsieh (1999), Tsai and Hsiao (2001), and Tsai and Huang (2003) proposed a global approach based on a Fourier image reconstruction scheme for inspecting surface defects in textures. Their method does not depend on local texture features but detecting defects via the high-energy frequency components in the Fourier spectrum. In the restored image of a textured surface, the regular region with periodic lines in the original image will have an approximately uniform gray level, whereas the defective region will be distinctly preserved.

Rough set theory (RST), which was introduced by Pawlak in the early 1980s, (Pawlak, 1982) is a new mathematical tool that can be employed to handle uncertainty and vagueness. It focuses on the discovery of pattern in inconsistent data (Pawlak, 1996; Slowinski & Stefanowski, 1989) and can be used as the basis to perform formal reasoning under uncertainty, machine learning, and rule discovery (Yao, Wong, & Lin, 1997). The basic philosophy behind rough set theory is based on equivalence relations or indiscernibility in the classification of objects (Walczak & Massart, 1999). It does not need any a priori knowledge, such as probability in statistics, or the basic probability assignment in the Dempster–Shafer theory of evidence or membership grade in the fuzzy set theory. Using rough set theory as a data mining tool, it can explore the hidden patterns in the data set and obtain the decision rules. That is, rough set theory can be used for (a) reduction of feature spaces; (b) finding hidden patterns; and (c) generation of decision rules (Kusiak, 2001; Pawlak, 1982). In the past decades, the rough set theory has been successfully applied to many real-world problems in medicine, pharmacology, engineering, banking, financial and market analysis (Ahna, Cho, & Kim, 2000; Kusiak, 2001; the related methodology refers Refs. Pawlak (1982), Pawlak (1996)).

From Bayes classifiers to neural networks, there are many possible choices for classifier selection. Among these, support vector machines (SVM) would appear to be a good candidate because of their ability to generalize in high-dimensional spaces, such as spaces spanned by texture patterns. The appeal of SVM is based on its strong connection to the underlying statistical learning theory. That is, an SVM is an approximate implementation of the structural risk minimization (SRM) method (Vapnik, 1995; Vapnik, 1998). For several pattern classification applications, SVM have already been shown to provide better generalization performance than traditional techniques, such as neural networks (Schokopf et al., 1997).

The aim of this paper is to illustrate the potential of DWT, RST and SVM in CCL defects detection and classification. The proposed method incorporates the DWT, inverse DWT, and RST with the SVM classifier, using the selected features fed directly into the SVM classifier. As a result, the features obtained from the restored defects image are nonlinearly mapped into the SVM architecture. Since SVM was originally developed for two-class problems, its basic scheme is extended to multi-texture classification by adopting the one-against-all and one-against-one decomposition methods. This works by applying SVM to separate one class from all the other classes, or separate each pair among all classes. Therefore, feature detection and classification are both performed in accordance with a unique criterion referred to as the classification rate.

This paper is organized as follows. Section 2 describes the DWT and inverse DWT transform in image, the rough set theory for defects feature selection and classification approach used to determine the multi-class defects by SVM. Section 3 presents experimental results on four categories of CCL surface defects classification using SVM classifier. Section 4 discusses the effects of the proposed method. The paper is concluded in Section 5.

2. Methodology

2.1. Wavelet transform and reconstruction

Wavelets are functions generated from one single function by dilations and translations. The basic idea of the wavelet transform is to represent any arbitrary function as a superposition of wavelets. Any such superposition decomposes the given function into different scale levels where each level is further decomposed with a resolution adapted to that level (Mallat, 1989).

This process continues until some final scale is reached. The values or transformed coefficients in approximation and detail images (sub-band images) are the essential features, which are as useful for texture discrimination and segmentation. Since textures, either micro or macro, have non-uniform gray-level variations, they are statistically characterized by the values in the DWT transformed sub-band images or the features derived from these sub-band images or their combinations. In other words, the features derived from these approximation and detail sub-band images uniquely characterize a texture.

The 2-D wavelet analysis operation consists of filtering and downsampling horizontally using 1-D lowpass filter L and highpass H to each row in the image and produces the coefficient matrices $f_L(x, y)$ and $f_H(x, y)$. Vertically filtering and down-sampling follows, using the lowpass and highpass filters L and H to each column in $f(L)$ and $f(H)$, and produces four subimages $f(LL)$, $f(LH)$, $f(HL)$, and $f(HH)$ for one level decomposition. The 2-D pyramid algorithm can iterate on the smooth subimage $f(LL)$ to obtain four coefficient matrices in the next decomposition level.

For images, there exist an algorithm similar to the 1-D case for 2-D wavelets and scaling functions obtained from 1-D ones by tensorial product. This kind of 2-D DWT leads to a decomposition of approximation coefficients at level j in four components: the approximation at level $j + 1$, and the details in three orientations (horizontal, vertical, and diagonal). The following chart describes the basic decomposition steps for images.
The inverse DWT performs a single-level 2-D wavelet reconstruction with respect to either a particular wavelet or particular wavelet reconstruction filters that you specify. The following chart describes the basic decomposition steps for images (see Fig. 1).

### 2.2. Support vector machine classifier

SVMs (Chapelle, Haffner, & Vapnik, 1999; El-Naqa, Yongyi, Werneck, Galatsanos, & Nishikawa, 2002; Kim, Jung, Park, & Kim, 2002; Kim, Jung, & Kim, 2003; Liyang, Yongyi, Nishikawa, & Werneck, 2005a; Liyang, Yongyi, Nishikawa, & Yulei, 2005b; Song, Hu, & Xie, 2002; Vapnik, 1995) have recently been proposed as popular tools for learning from experimental data. The reason is that SVMs are much more effective than other conventional nonparametric classifiers (e.g., the neural networks, nearest neighbor, k-NN classifier) in terms of classification accuracy, computational time, stability to parameter setting. One of the neural network models that fulfill these requirements is Support vector machine (SVM). The SVM obtains high generalization performance without the need a priori knowledge even when the input dimension space is high. Moreover, this is a model that allows more accurate formal assessment of the generalization performance. This fits well within the framework of this proposal. Below we attempt a rigorous assessment of the generalization performance of the supervised network. We also provide a methodology for SVM model selection well fitted to the number of transferred patterns (i.e., model order selection). However, some notations and key concepts should be defined first.

#### 2.2.1. Optimal separating hyperplane

Let \( \{(\mathbf{x}_i, y_i), i = 1, \ldots, |N|\} \) be a training example set \( S \), each example \( \mathbf{x}_i \in \mathbb{R}^d \) belongs to a class labeled by \( y_i \in \{-1, 1\} \). The goal is to define a hyperplane which divides \( S \) such that all the points with the same label are on the same side of the hyperplane while maximizing the distance between the two classes. This means to find a pair \((\mathbf{w}, b)\) such that

\[
y_i (\mathbf{w} \cdot \mathbf{x}_i + b) > 0, \quad i = 1, \ldots, |N|,
\]

where \( \mathbf{w} \in \mathbb{R}^d \) and \( b \in \mathbb{R} \). The pair \((\mathbf{w}, b)\) defines a separating hyperplane

\[
\mathbf{w} \cdot \mathbf{x} + b = 0.
\]

If there exists a hyperplane satisfying, the set \( S \) is said to be linearly separable and we can change \( \mathbf{w} \) and \( b \) so that

\[
y_i (\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \quad i = 1, \ldots, |N|.
\]

Accordingly, we can know the minimal distance between the closest point and the hyperplane is \( 1/||\mathbf{w}|| \). Besides, among these separating hyperplanes, the OSH is a hyperplane, which the distance to the closest point is maximal. Hence, in order to find the OSH, we must minimize \( ||\mathbf{w}||^2 \) under constraint Eq. (3).

Since \( ||\mathbf{w}||^2 \) is convex, we can minimize it under constraint Eq. (3) by means of the classical method of Lagrange multipliers. If we denote with \( \mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_N) \) the \( N \) nonnegative Lagrange multipliers associated with constraint Eq. (3), the problem of finding OSH is equivalent to the maximization of the function

\[
W(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j,
\]

where \( \alpha_i \geq 0 \) and under constraint \( \sum_{i=1}^{N} \alpha_i y_i = 0 \).

Once the vector \( \mathbf{\alpha} = (\alpha_1, \ldots, \alpha_N) \) solution has been found, the OSH \((\mathbf{w}, b)\) has the following expansion:

\[
\mathbf{w} = \sum_{i=1}^{N} \alpha_i y_i \mathbf{x}_i
\]

while \( b \) can be determined from \( \mathbf{\alpha} \) and from the Kühn–Tucker conditions (Vapnik, 1995)

\[
\alpha_i (y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1) = 0, \quad i = 1, 2, \ldots, |N|.
\]

By the way, the corresponding training examples \((\mathbf{x}_i, y_i)\) with nonzero coefficients \( \alpha_i \) are called support vectors. At last, the decision function of classifying a new data point \( \mathbf{x} \) can be written as:

\[
f(\mathbf{x}) = \text{sgn} \left( \sum_{i=1}^{N} \alpha_i y_i \mathbf{x}_i \cdot \mathbf{x} + b \right).
\]

#### 2.2.2. Non-linear support vector machines

The training example set that we want to classify is usually non-linear. To achieve better generalization performance, the input data can be mapped into a high-dimensional feature space first. Then the OSH is constructed in the feature space.

If \( \Phi(\mathbf{x}) \) denotes a mapping function that maps \( \mathbf{x} \) into a high-dimensional feature space as follows:

\[
W(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j).
\]

Now, let \( K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j) \), we can rewrite above equation as

---

Fig. 1. 2-D discrete wavelet reconstruction steps, where ↑ Upsample columns: insert zeros at odd-indexed columns. ↑ Upsample rows: insert zeros at odd-indexed rows. \( \square \) Convolve with filter \( X \) the rows of the entry. \( \Diamond \) Convolve with filter \( X \) the columns of the entry.
The features used in this study are mainly extracted from the geometric features obtained via the blob analysis detected from the restored images. There are 20 extracted features in this study, including area, perimeter, elongation, convex perimeter, compactness, number of holes, roughness, Euler number, invariant moments, and texture features. These features are used to classify defects in the restored images.

3.2. Wavelet transformation and inverse wavelet transformation

After obtaining the preprocessing defect images in previous experiments, we conduct the 2-D wavelet decomposition. In this experiment, we evaluate five orthogonal wavelet bases including Haar, Daublets D4 and D8, and Symmlets S8 and S20 and different resolution level. The appropriate wavelet bases and resolution level will effectively highlight the local anomalies in the homogeneous surface. If the number of multi-resolution levels is too small they cannot sufficiently separate defects from the repetitive texture pattern. However, a too large number of multi-resolution levels yields the fusion effect of the anomalies, and may result in false detection. Multi-resolution levels 3 and 4 are most appropriate to enhance defects in the restored image in these experiments (Tsai & Hsiao, 2001). The experiment evaluates the three selective detail subimage combinations, namely, the horizontal and diagonal detail subimage, vertical and diagonal detail subimage and vertical and horizontal detail subimage, respectively, with a repetitive horizontal line pattern defect in Fig. 4. The selective subimage effects on CCL defect image reconstruction are similar. However, the reconstructing horizontal and diagonal detail subimages can effectively enhance the defect regions in the restored image. It also efficiently separates defects from the background in the corresponding binary images. Therefore, the features extractions are preferred to be selected from the defects in horizontal and diagonal detail subimage. Fig. 5 shows the four reconstructed images from horizontal and diagonal detail subimages remove all non-defect areas in the original images and preserve only the defects of the CCL surface. The defects in the restored image can be efficiently separated from the background using a simple binary thresholding technique such as the one proposed by (Otsu, 1979).

3.3. Feature extraction and selection

The features used in this study are mainly extracted from the geometric features obtained via the blob analysis detected from the restored images. There are 20 extracted features in this study, including area, perimeter, elongation, convex perimeter, compactness, number of holes, roughness, Euler number, invariant moments, and texture features. Based on these features, the comparison with the error back-propagation neural network classifier is also discussed in this section.

2.3. Multi-class support vector machines

Two main approaches have been suggested for applying SVMs to multi-class classification. One is the one-against-all strategy to classify between each class and all the remaining; the other is the one-against-one strategy to classify between each pair. In each, the underlying basis has been to reduce the multi-class problem to a set of binary problems, enabling the basic SVM approach to be used. In the one-against-all approach, a set of binary classifiers, each trained to separate on class from the rest, is undertaken, and the input vector allocated to the class for which the largest decision value was determined (Hsu & Lin, 2002). The ith SVM is trained from the training samples where some examples contained in the ith class have “+1” labels, and other examples contained in the other classes have “-1” labels. Specifically, with this approach, where D is the number of classes

\[
W_j(x) = \sum_{i=1}^{D} a_{ij} K(x_i, x) + b_j,
\]

where \( K \) is a kernel function and must satisfy Mercer’s theorem (Vapnik, 1995). Finally, the decision function becomes:

\[
f(x) = \text{sgn}\left( \sum_{i=1}^{D} a_{ij} K(x_i, x) + b_j \right)
\]

Typical kernel functions are the following:

1. Linear: \( K(x, z) = x \cdot z \)
2. Polynomial: \( K(x, z) = (\gamma \cdot x \cdot z + \text{coef})^d \)
3. Gaussian Radial Basis: \( K(x, z) = \exp(-\gamma \cdot |x - z|^2) \)
4. Sigmoidal: \( K(x, z) = \tanh(\gamma \cdot x \cdot z + \text{coef}) \)

where \( \gamma \) and coef are constants and \( d \) is a degree.

The current inspection activity of CCL surface depends mainly upon the inspector’s visual inspection and feeling with a glove. This inspection activity could be very subjective and highly dependent on the experience of human inspectors. Among the 250 collected samples, there are four major classes of defects on the CCL surface that include convex dots, pinholes, flashes, and oxidation. Fig. 2 shows the four different defects categories and the number of the defects that belong to each class.

The major preprocessing steps in this study include smooth filtering, sobel filtering, and close morphological operations. The main objective of preprocessing is to perform the image processing operation on the source image to enhance and smooth images, accentuate or extract image edges, or remove “noise” from an image. They allow images to be separated into their low and high spatial-frequency components. The preprocessing results are shown in Fig 3, respectively.

3.1. Defects classification and preprocessing

The features used in this study are mainly extracted from the geometric features obtained via the blob analysis detected from the restored images. There are 20 extracted features in this study, including area, perimeter, elongation, convex perimeter, compactness, number of holes, roughness, Euler number, invariant moments, and texture features. Based on these features, the comparison with the error back-propagation neural network classifier is also discussed in this section.

2.3. Multi-class support vector machines

Two main approaches have been suggested for applying SVMs to multi-class classification. One is the one-against-all strategy to classify between each class and all the remaining; the other is the one-against-one strategy to classify between each pair. In each, the underlying basis has been to reduce the multi-class problem to a set of binary problems, enabling the basic SVM approach to be used. In the one-against-all approach, a set of binary classifiers, each trained to separate on class from the rest, is undertaken, and the input vector allocated to the class for which the largest decision value was determined (Hsu & Lin, 2002). The ith SVM is trained from the training samples where some examples contained in the ith class have “+1” labels, and other examples contained in the other classes have “-1” labels. Specifically, with this approach, where \( D \) is the number of classes

\[
W_j(x) = \sum_{i=1}^{D} a_{ij} K(x_i, x) + b_j,
\]

where \( a_{ij} \) are constants and \( D \) is a degree.

The current inspection activity of CCL surface depends mainly upon the inspector’s visual inspection and feeling with a glove. This inspection activity could be very subjective and highly dependent on the experience of human inspectors. Among the 250 collected samples, there are four major classes of defects on the CCL surface that include convex dots, pinholes, flashes, and oxidation. Fig. 2 shows the four different defects categories and the number of the defects that belong to each class.

The major preprocessing steps in this study include smooth filtering, sobel filtering, and close morphological operations. The main objective of preprocessing is to perform the image processing operation on the source image to enhance and smooth images, accentuate or extract image edges, or remove “noise” from an image. They allow images to be separated into their low and high spatial-frequency components. The preprocessing results are shown in Fig 3, respectively.

3.1. Defects classification and preprocessing

The current inspection activity of CCL surface depends mainly upon the inspector’s visual inspection and feeling with a glove. This inspection activity could be very subjective and highly dependent on the experience of human inspectors. Among the 250 collected samples, there are four major classes of defects on the CCL surface that include convex dots, pinholes, flashes, and oxidation. Fig. 2 shows the four different defects categories and the number of the defects that belong to each class.

The major preprocessing steps in this study include smooth filtering, sobel filtering, and close morphological operations. The main objective of preprocessing is to perform the image processing operation on the source image to enhance and smooth images, accentuate or extract image edges, or remove “noise” from an image. They allow images to be separated into their low and high spatial-frequency components. The preprocessing results are shown in Fig 3, respectively.
20 features as a full set of attributes, we will concentrate in the following section on rough set theory to feature selection.

Rough set approach to feature selection can be based on minimal description length principle (Rissanen, 1978) and tuning methods of parameters of the approximation spaces to obtain high quality classification based on selected features. In this study, the application of rough set theory for feature selection is in a closed loop. This method is based on searching first for short reducts or reduct approximations. We have implemented the feature selection using software ROSETTA (see http://www.idi.ntnu.no/~aleks/rosetta/rosetta.html). The ROSETTA software supports six feature selection algorithms, namely, Genetic algorithm, Johnson’s algorithm, Dynamic reducts (RSES), Exhaustive calculation (RSES), Johnson’s algorithm (RSES) and Genetic algorithm (RSES). Each method has the selection criteria that may be based on the genetic algorithm with fitness function, a simple greedy algorithm, or calculating the most occurring reducts across the decision table. Accordingly, the most occurring reducts subsets obtained by means of the six-feature selection algorithm are listed in Table 1. The reduct subsets consist of 18, 14 and 8 feature composition. These reducts composition will be put into the SVM classifier for CCL surface defect detection and classification.

3.4. Classification results using SVM

After obtaining the features extracted from the defect subimages, the support vector machine is used to classify four main defect categories on the CCL surface. The SVMs are constructed using LIB-SVM version 2.6 (Chang & Lin, 2001) software to validate the classification capability of the proposed feature extraction and feature selection approach, i.e., inverse wavelet reconstruction and rough set approach. To evaluate the performance of the proposed approach, the holdout and tenfold cross-validation random sampling methods have been adopted in the SVM classification. Four existing kernel functions for SVM classifier are also considered, namely, RBF, Linear, Polynomial and Sigmoid function with the original features set and reduced 18, 14, and 8 feature set, respectively. Thus, the total 250 collected samples with four classes were divided into 2/3 for training data set (i.e. 167 samples) and 1/3 for testing data set (i.e. 83 samples). In our evaluation, we used accuracy rate of

Fig. 2. Original defect images: (a) pine hole (119 samples), (b) convex dot (8 samples), (c) flash (53 samples), and (d) oxidation (50 samples).

Fig. 3. Preprocessing results: (a) pine hole, (b) convex dot, (c) flash, and (d) oxidation.

Fig. 4. Restored images from selective subimages: (a) the original image, (b) horizontal + diagonal subimage, (c) vertical + diagonal subimage, and (d) horizontal + vertical subimage.

Fig. 5. 2-D inverse DWT reconstruction images: (a) pinhole, (b) convex dot, (c) flash, and (d) oxidation.
classification as a figure of merit. Accuracy rate of classification is defined as the total number of correctly classified samples divided by the total number of samples classified. In Table 2, we summarized the results for the holdout method using four kernel functions with original and reduced feature sets. Although the generalization ability of the SVM is relatively robust to variations in the parameter settings, the method used to define the parameter $c$ and cost parameter $C$ in the kernel function RBF will insure that high accuracy is obtained. Accordingly, the parameters are set in the range of $c = [0.001, 200]$, $C = [0.1, 50,000]$ using grid combinations for testing the classification accuracy. The other parameters with different kernels are set as follows: $C_0$ is set from 0.1 to 100, and kernel order $p = 3, 4$ for Polynomial. These results indicate that the performance of the SVM classifier is not very sensitive to the value of the model parameter setting. The software can automatically search the best parameter setting for the accuracy rate of classification. The training time of the RBF and Linear kernel is around 20 s faster than those of the Polynomial and Sigmoid (implemented on a Pentium IV 1.8 GHz PC). Note that, as expected, the accuracy of the original feature set is higher than those of the reduced feature set. The best result of the classification is the Gaussian RBF kernel function and the accuracy rate of classification is over 80%, whereas the other kernel functions are almost near 80% accuracy. In addition, the 18 reduced feature set is obtained the better result than that of the 14 and 8 feature sets, respectively.

The SVM classifier as the same parameter setting as the holdout method is then further trained using the tenfold Cross-Validation method with the four kernel functions. In Table 3, we summarize the results of the Cross-Validation method that the classification accuracies are almost the same with the four kernel functions, accuracy between 68% and 81%. The best result is obtained using the 14 reduced features via Linear kernel, and it takes the longest training time for Sigmoid kernel. The accuracy of classification with the reduced feature set 14 and 18 are not significantly different from those of the original feature set.

Consequently, the performance of the holdout method is better than the Cross-Validation method as the classification accuracy is 74% - 85% higher than that of 68% - 81%. The efficiency also shows that the holdout method is higher than that of Cross-Validation method.

In order to compare the performance of the proposed approach, we also implement the error back-propagation neural network (BPNN) classifier which is the most commonly used feed forward neural networks with the same classification problem. As described in the previous section, the parameters of SVM are identical among the four kernel functions. The data set have been divided into 2/3 (167 data) for training, and 1/3 (83 data) for testing with four feature sets either for SVM or for back-propagation neural classifier. The single hidden layer (5 hidden neurons), learning rate 0.01, and momentum 0.8 are used to classify the defects on the CCL.

### Table 1

Results of feature selection

<table>
<thead>
<tr>
<th>No. of features</th>
<th>18 Features</th>
<th>14 Features</th>
<th>8 Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Area</td>
<td>1 Area</td>
<td>1 Perimeter</td>
<td></td>
</tr>
<tr>
<td>2 Perimeter</td>
<td>2 Perimeter</td>
<td>2 Perimeter</td>
<td></td>
</tr>
<tr>
<td>3 Feret minimum diameter</td>
<td>3 Feret minimum diameter</td>
<td>3 Feret mean diameter</td>
<td></td>
</tr>
<tr>
<td>4 Feret minimum angle</td>
<td>4 Feret maximum diameter</td>
<td>4 Convex perimeter</td>
<td></td>
</tr>
<tr>
<td>5 Feret maximum diameter</td>
<td>5 Feret mean diameter</td>
<td>5 Length</td>
<td></td>
</tr>
<tr>
<td>6 Feret maximum angle</td>
<td>6 Convex perimeter</td>
<td>6 Elongation</td>
<td></td>
</tr>
<tr>
<td>7 Convex perimeter</td>
<td>7 Compactness</td>
<td>7 Number of runs</td>
<td></td>
</tr>
<tr>
<td>8 Compactness</td>
<td>8 Feret elongation</td>
<td>8 Secondary axis (Bin.)</td>
<td></td>
</tr>
<tr>
<td>9 Number of holes</td>
<td>9 Roughness</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>10 Feret elongation</td>
<td>10 Length</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>11 Roughness</td>
<td>11 Breadth</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>12 Euler number</td>
<td>12 Elongation</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>13 Length</td>
<td>13 Number of runs</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>14 Breadth</td>
<td>14 Secondary axis (Bin.)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>15 Elongation</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>16 Number of Runs</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>17 Minimum pixel</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>18 Principal axis (Bin.)</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

### Table 2

Performance of holdout method (Accuracy%)

<table>
<thead>
<tr>
<th>No. of features</th>
<th>RBF</th>
<th>Linear</th>
<th>Polynomial</th>
<th>Sigmoid</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Train</td>
<td>Test</td>
<td>Train</td>
<td>Test</td>
</tr>
<tr>
<td>20</td>
<td>85.03</td>
<td>84.34</td>
<td>79.64</td>
<td>83.13</td>
</tr>
<tr>
<td></td>
<td>88.02</td>
<td>84.34</td>
<td>90.42</td>
<td>81.93</td>
</tr>
<tr>
<td></td>
<td>88.62</td>
<td>85.54</td>
<td>91.02</td>
<td>80.72</td>
</tr>
<tr>
<td>Average</td>
<td>87.22</td>
<td>84.74</td>
<td>87.03</td>
<td>81.93</td>
</tr>
<tr>
<td>18</td>
<td>88.62</td>
<td>84.34</td>
<td>70.66</td>
<td>72.29</td>
</tr>
<tr>
<td></td>
<td>86.83</td>
<td>81.93</td>
<td>81.43</td>
<td>81.93</td>
</tr>
<tr>
<td></td>
<td>86.83</td>
<td>84.34</td>
<td>88.62</td>
<td>83.13</td>
</tr>
<tr>
<td>14</td>
<td>89.82</td>
<td>79.52</td>
<td>88.02</td>
<td>77.11</td>
</tr>
<tr>
<td></td>
<td>92.81</td>
<td>83.13</td>
<td>89.22</td>
<td>79.52</td>
</tr>
<tr>
<td></td>
<td>90.42</td>
<td>81.93</td>
<td>88.62</td>
<td>77.11</td>
</tr>
<tr>
<td>8</td>
<td>86.83</td>
<td>78.31</td>
<td>83.83</td>
<td>75.90</td>
</tr>
<tr>
<td></td>
<td>87.43</td>
<td>79.52</td>
<td>81.44</td>
<td>78.31</td>
</tr>
<tr>
<td></td>
<td>91.02</td>
<td>81.93</td>
<td>82.04</td>
<td>77.11</td>
</tr>
<tr>
<td>Average</td>
<td>88.96</td>
<td>81.66</td>
<td>83.76</td>
<td>78.05</td>
</tr>
</tbody>
</table>
surface based on four different feature sets. The classification accuracies of four SVM classifier and BPN are shown in the Table 4. It is obvious that the more the feature of the classifier, the higher the accuracy of the classification. Moreover, the simpler the network structure, the lower the training time (18-5-1 equal to 30 s vs. 8-5-1 equal to 45 s). The best test result of the SVM is the classifier with RBF kernel, which is 82.43% better than those of the other three SVM classifiers as well as BPN classifier that average 78.81%, 79.21%, 78.83% and 81.93%, respectively. In addition, the classification accuracies of the reduced feature set via rough set feature selection between SVM classifiers and BPN classifier are above or near 80% accuracy. It is noted that the SVM classifier with RBF kernel function via rough set theory can reduce and select the significant feature in the classification because most of the classification accuracies are higher than that of BPN classifier. In SVM classifier, we can control the tradeoff between the complexity of decision rule and training error rate by searching the combination of C and gamma.

4. Discussion

Based on the experimental results and experience in CCL defect classification problems, we would like to highlight that:

1. The high classification accuracies of the multi-class SVM and BPN classifiers give insight into the features selected for defining the CCL defect signals. The conclusions drawn in the applications demonstrated that the inverse wavelet and geometry feature which well represent the CCL defects signals and by the usage of these selected feature via rough set a good distinction between classes can be obtained.

2. During the SVM training, most of the computational effort is spent on finding the optimal parameter with different kernel functions in order to obtain high classification accuracy. The SVMs map the features to higher dimensional space and then use an optimal hyperplane in the mapped space. This implies that though the original features carry adequate information for good classification, mapping to a higher dimensional feature space could potentially provide better discriminatory clues that are not present in the original feature space. One would not know the suitability of a kernel function and performance of the SVM until one has tried and searched with different parameter.

3. The proper selection of multi-resolution level and subimage combination can appropriately enhance defects in the restored image. Experimental results on a variety of CCL defects images have confirmed that three or four multi-resolution levels are generally sufficient for CCL defect detection application. On the other hand, the proper selection of decomposed subimage combinations for reconstruction can effectively enhance the defects with high directionality. The selective subimage effects on CCL defect image reconstruction are similar. However, the reconstructing horizontal and diagonal detail subimages can reveal the defect regions in the restored image. It also efficiently separates defects from the background in the corresponding binary images. Therefore, the features extractions are preferred to be selected from the defects in horizontal and diagonal detail subimage.

5. Conclusion

In this paper, we have conducted the classification problem of detecting CCL defects embedded in a homogeneous texture using
the inverse wavelet transform, rough set selector and SVM classifier. The proposed method not only works on local textural features in a spatial domain, but also based on an image restoration scheme using the multi-resolution wavelet transforms with the advantage of computational savings. With proper selection of a smooth sub-images and detail subimages in different multi-resolution levels for image restoration, the global repetitive texture pattern can be effectively removed and local anomalies can be enhanced in the restored image. Thus, the rough set selector is therefore used to conduct the feature selection in the original feature sets of the restored image. The high generalization performance of SVM classifier without the need of prior knowledge can map the lower feature space into higher feature space increasing the accuracy of classification. The experimental results have revealed that SVM with RBF kernel function obtained higher classification accuracy compared with the other kernel functions. The classification accuracies of the reduced feature set via rough set feature selection are above or near an 80% accuracy rate. It is noted that the SVM classifier with different kernel function via rough set theory can reduce and select the significant feature in the classification. Moreover, the reduced feature set well represented the capability of SVM, and kept with the suitable classification accuracy. The proposed method implemented in this study can promote the accuracy of the inspection and avoid human intervention for defect inspection in homogeneous CCL surface defects.

References


Lambert, G., & Bock, F. (1997). Wavelet method for texture defect detection. In IEEE international conference on image processing (pp. 201–204), Santa Barbara, CA.


