Markov Model and Data Mining Approach for PCB Component Defect

Vaddin Prathiba¹, Dr. M Nagendra²

¹ Research Scholar, Department of Computer Science, Rayalaseema University, AP, India
² Professor, Department of Computer Science & Technology, Sri Krishnadevaraya University, AP, India

Corresponding Author: Vaddin Prathiba

Abstract: Currently the various electronic components which are manufactured use surface mount technology (SMT). This technology is a higher end assembly technique which produces printed circuit boards with very tiny electronic components. Due to the device area issues, the demand of PCBs is increasing. In order to cater this issue, high-volume production is demanded. Key challenge with PCB manufacturers is to maintain the quality of PCB with zero defects and assured quality. But due to the changing technologies in PCB fabrication, component placements and soldering of surface mount technology, defects are increasing in terms of the number and the type of defect. Various approaches have been proposed in this field of defect detection, based on the image processing techniques and data mining technique. To overcome these issues of PCB defects, we propose a new approach for the defect detection in the printed circuit boards for surface mount device inspection. According to the proposed approach defective points are extracted by detecting highlighted areas and then invalid areas are recognized and removed. To achieve this we utilize color features for the three channels, whereas invalid areas are recognized and removed by comparing the features of target and reference objects. Color feature extraction is carried out by applying proposed color distribution model. After extracting the features, classification technique is applied to classify the given the database into two classes: (a) defect detected and (b) no defect. In order to perform classification, Markov model is utilized which considers image features (i.e. color features). Proposed approach is implemented using MATLAB tool by combining image processing and data mining approach. Experimental study shows the efficiency of the proposed approach in terms of defect detection.

Keywords: PCB Defect Detection, Image Processing, Feature Distribution, data mining

Date of Submission: 26-07-2017
Date of acceptance: 05-08-2017

1. Introduction

The Printed circuit board inspection and testing is a crucial task for industries during manufacturing of boards. Various uncertainties occurs during manufacturing which causes error in PCB such as orientation, joint defect etc. Therefore an inspection system is required to perform inspection on the boards to asses manufacturing process resulting in better quality of manufacturing. PCBs can be inspected on various levels. During manufacturing, bare printed circuit boards are used widely. These boards are used for placing the components on it [1]. Various approaches have been developed by researchers for PCB inspection system. These approaches are based on computer vision system and data mining approach. According to computer vision system, methods are classified into two categories as (i) referential image approach and (ii) non-referential approach.

In reference image based approach, input test image is compared with the existing base image which is not having any defect. In [1] image subtraction method is used, verification of dimension method is discussed in [2] and matching the template is described in [3]. During image subtraction approach, test image is subtracted from the original perfect image. In dimension verification method, test image dimensions are verified by comparing with original test image and finally in template matching method, image features are extracted and matched with the original image.

As discussed before, another approach is based on data mining. Actually [4] presented a datamining approach based on decision-tree classification approach. This method is applied on semiconductor datasets. In order to perform the classification, various parameters are considered which effect soldier bumping process such as physical parameter or chemical parameters. But according to this process, controllable parameters only can be achieved, it doesn’t prove the effectiveness on classification of component defects.

Similarly, another approach was introduced by [5] for wafer classification using datamining approach. Wafer failure data is classified into 4 classes such as systematic failure, systematic attraction, random and others. In order to classify, spatial features are extracted and classification performance is carried out by applying neural network classifier.

DOI: 10.9790/0661-1904053847 www.iosrjournals.org 38 | Page
Further [6] discussed data mining approach for similarity measurements. In this article, feature selection and feature reduction are discussed for similarity measurement.

Manual inspection of PCBs becomes time consuming and lesser reliable compared to automated scheme. By taking this into account, a new approach for PCB defect detection is proposed in [7]. This scheme is mainly based on computer vision scheme which performs image enhancement, image filtering and thresholding process. This scheme uses template matching approach by keeping one reference image as perfect image. There are various researches going on for PCB defect detection and classification. In existing approaches such as computer vision techniques which includes erosion, connected component opening, closing and dilation

Pixels at the edge of image are removed when erosion scheme is applied which causes information loss for visual perception. In dilation process, gaps in binary image are filled with foreground pixels which creates difficulty to analyze the defects. Similarly, connected components opening and closing causes extra defect during visualization. Due to these pre-processing stages of computer vision technique, PCB defect detection still remains a challenge for the researches. Another approach as we discussed is data mining, which is based on the collection of data and learning of their pattern. In this field also various approaches have been proposed based on different classifiers which included ID3, decision tree, Naïve Bayes, decision tree etc. The main key issue with these algorithms is efficient approach for feature selection and feature reduction. Lack of feature selection and reduction causes more computation time and complexity. Due to these challenges, a new scheme is required for PCB defect detection and inspection. In order to overcome these issues, we propose a hybrid scheme for PCB defect detection by combining computer vision approach and data mining approach. Here for computer vision approach we use color distribution feature with multilevel thresholding and for data mining we use Gaussian mixture modeling approach. According to proposed approach, input image is passed through the computer vision schemes and achieved classification results. At same time, same input image and its parameters are passed through data mining approach and classification is achieved. Novelty of the proposed approach is that if data is miss-classified during computer vision, then based on data mining approach correct classification is achieved. Rest of the manuscript is organized as follows: section 2 provides a brief literature survey or related work, in section 3 a hybrid approach is proposed and discussed, section 4 provides results and discussion and finally in section 5 the conclusion is discussed.

II. Literature Survey

The In this section we discuss about most recent works carried out by researchers in this field of PCB defect detection and classification. B. Kaur et. al. [8] introduced PCB defect detection system by applying image processing technique. This approach utilizes reference based image method while using image subtraction approach. In this method, authors show results of defect detection i.e. missing holes , line break and wrong hole.

Feature extraction also plays important role in image processing schemes. In order to employ this method for automated PCB defect detection Z. Ibrahim et.al. [9] Used wavelet based feature extraction technique. Wavelets are used here to compute the difference of image. Based on the difference, defects are localized by using proposed defect localization scheme. Another scheme for defect detection using image processing technique is presented by C. Ma et al. in [3]. In this paper, a robot is constructed for vision imaging system. Iterative approach is used for segmentation based on the variance of the clusters. T. J. Mateo et al. [10] used template matching approach based on image processing and along with this a classification approach is also merged with image processing scheme. Classification is carried out using high intensity feature of the image. In [11],F. Xie et al. used genetic programming approach for defect detection and classification. Wen-Yen et al. [12] classified defects into two sub-categories which are functional and cosmetic defects. Functional defects are crucial to handle and cause into PCB failure whereas cosmetic defects are related to visualization of PCB image. Mainly 14 types of known defects are presented in the literature for single layer PCB which are tabulated in table 1. These defects are detected using image processing schemes such as A. TeohOnget.a. [13] used real time approach for PCB defect detection using computer vision system. These methodologies which are related to automate PCB defect detection having critical issue related to reliability of detection results. In order to overcome these schemes based on data mining also proposed by researchers. A. Kusiak et al. [14] discussed about data mining schemes for fault detection in PCBs using data mining algorithms. According to data mining approach, a pattern of dataset is required which helps to match the pattern of test image with the dataset image. These patterns are extracted based on the features of images. With this consideration Feng Zhang et al. [15] proposed a data mining scheme for PCB defect classification. In this method a latent variable model is constructed from the dataset image and incorporated with logistic regression model which provides interdependencies between the components of the PCBs. Based on this a probabilistic model is formulated which uses maximum-likelihood principle component analysis for classification. In [16], sensor based approach is used for PCB defect detection. In this method 2D-high precision sensor is utilized which provides high precision measurement. A scalable architecture is used for motion controlling of robot along with x and y direction and a user interface is also developed using Java platform. Similarly, in [17] sensor based approach is utilized which uses 2D sensor. In this approach PIC microcontroller is used to control the motion.
TABLE 1. Known defects for single layer PCB

<table>
<thead>
<tr>
<th>Defect Number</th>
<th>Name of Defect</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hole missing</td>
</tr>
<tr>
<td>2</td>
<td>Over etching</td>
</tr>
<tr>
<td>3</td>
<td>Conductor missing</td>
</tr>
<tr>
<td>4</td>
<td>Under etching</td>
</tr>
<tr>
<td>5</td>
<td>Breakout</td>
</tr>
<tr>
<td>6</td>
<td>Open-circuit</td>
</tr>
<tr>
<td>7</td>
<td>Unnecessary Short</td>
</tr>
<tr>
<td>8</td>
<td>Shorting</td>
</tr>
<tr>
<td>9</td>
<td>Spur</td>
</tr>
<tr>
<td>10</td>
<td>Mouse-bite</td>
</tr>
<tr>
<td>11</td>
<td>Wrong hole size</td>
</tr>
<tr>
<td>12</td>
<td>Under Etching</td>
</tr>
<tr>
<td>13</td>
<td>Conductor closeness</td>
</tr>
<tr>
<td>14</td>
<td>Pin Hole</td>
</tr>
</tbody>
</table>

2. PROPOSED MODEL

This section describes about proposed methodology for automated PCB defect detection system. This work comprises automated scheme of object detection and recognition for object extraction such as diode, resistors, transistors, IC from PCB image. PCB chip contains joints in solder, protective coating and markings. According to proposed approach gray level intervals are determined by applying multilevel thresholding scheme. In order to remove invalid areas present in PCB image, color distributions are computed and compared with reference image based on the coordinates of chromaticity.

2.1. Image capturing

To employ proposed algorithm, images are captured with help of PCB inspection instruments. Images are captured when LEDs are ON which helps to achieve the solder joint regions. Captured images are stored in the host computer and a database is created for varied PCB images. These images are resized and stored in the database with size of $512 \times 512$ and three channels of color channel combinations i.e. R, G and B. Figure 1 shows image capturing procedure. As discussed before, PCB contains marking coatings etc. are displayed in figure 1(a) and figure 1 (b) shows the image capturing during LED illumination.

![Figure 1. Image capturing procedure](image)

Proposed scheme contains two section for defect detection: (i) computer vision and (ii) data mining. According to computer vision scheme images are captured and stored. By taking this image in account we perform extraction of PCB components and localization. In next section we discuss the process of extraction of PCB components.

2.2. PCB Component Extraction

Various schemes have been proposed during last decade for object detection but accuracy of detection still remains challenging issue. In this work, we propose a new algorithm based on intensity level variation thresholding. This approach performs three steps on an input image: (i) image pre-processing (ii) color space conversion and (iii) histogram construction. According to first stage (pre-processing) input image is sharpened and brightness is adjusted with the help of equation 1.

$$I_E = [(l + (b_c + 1) \times 512) - 512] \times [(C_c + 1) \times 1] + 512$$

Here $l$ denotes original input image, enhanced image is represented by $I_E$. coefficient of brightness is denoted as $b_c$ which varies from $[-0 \text{ to } 1]$ same as coefficients of contrast which is denoted by $C_c$. This
Markov model and data mining approach for PCB Component Defect

In second stage, image is converted to grayscale which gives the color information of image by comparing lower level and higher level color details and in third stage histogram is constructed which provides intensity variation of image.

2.3. Color feature extraction

Color features are very significant components during visual perception system. In this scheme, colors of input image are considered for feature extraction with the help of Gaussian mixture modeling approach.

In order to achieve this objective, a framework is formulated which forms a chromaticity model in 𝒰′𝑽′ space and this model is used in mixture modeling, to compute the distribution of colors. With the help of this color distribution model, features are extracted and matched with the reference object. Matched object with reference image are extracted for further processing.

2.3.1. Color distribution model

In this section color distribution modeling is presented which utilizes detected objects and project them in LUV space. Components of LUV space are projected in chromaticity space which can be defined as

For 𝒰′ space

\[ \mathcal{U}' = \frac{\mathcal{U}}{\xi} + \mathcal{U}_n \]

Similarly for 𝒫′ space

\[ \mathcal{P}' = \frac{\mathcal{P}}{\xi} + \mathcal{P}_n \]

Where \( \mathcal{U}_n \) and \( \mathcal{P}_n \) are standard constants of illuminant region.

Highlighted pixel in the detection are employed by applying the rule mentioned below

\[ H_{\text{pixel}} = \arctan \left( \frac{c_{\text{max}} - \mathcal{P}' - \mathcal{P}_\ell}{\mathcal{U}' - \mathcal{U}_\ell} \right) \times \frac{180}{\pi} \]

(4)

\( c_{\text{max}} \) is maximum coordinate value. According to this rule, each channel of image is represented in a horizontal line between color feature vectors. For \( \mathcal{R} \) channel angle variation is noted as (89° to 304°), for \( \mathcal{G} \) channel (89° to 192°) and for blue channel (190° to 304°).

In next stage, expectation minimization approach is applied to model the feature which provides probability density function in the form of mixture model.

Let \( f_1, f_2, \ldots, f_n \) is a feature vector for input image on which Gaussian model is applied which results in \( d–\text{dimensional} \) euclidean space which is denoted as equation (5).

\[ \phi(f | \mathcal{M}_k, \mathcal{C}_k, \mathcal{W}_k) = \sum_{k=1}^{m} \mathcal{W}_k p_k(f), \mathcal{W}_k \geq 0 \]

(5)

This model contains \( m \) number of mixtures with \( \mathcal{M} \) mean, \( \mathcal{C} \) denotes covariance matrix and weight of pixel is denoted by \( \mathcal{W}_k \). Corresponding to three channels mixture models are denoted as \( \{\mathcal{M}_{k,r}, \mathcal{C}_{k,r}, \mathcal{W}_{k,r}\} \), \( \{\mathcal{M}_{k,g}, \mathcal{C}_{k,g}, \mathcal{W}_{k,g}\} \) and \( \{\mathcal{M}_{k,b}, \mathcal{C}_{k,b}, \mathcal{W}_{k,b}\} \) which corresponds to each group of pixels to corresponding channel.

2.3.2. Object recognition

In order to extract the same type of component, three consideration are performed : (1) same shares are present in reference and detected component (2) pixels values of subset and detected image belongs from one set and (3) outside pixels values from the subset. The relative difference on mean distribution probability is evaluated according to the formula, relative difference on distribution is computed to assist for selection of abovementioned considerations. According to consideration 1 and 2, mixture model’s difference \( M_{\text{dist}}(\alpha, i) > 0 \) is

\[ \text{prob}_1(\alpha, i) = \frac{\text{Prob}_{\mathcal{M}_1}(\alpha, i) - \text{Prob}_{\mathcal{M}_2}(\alpha, i)}{\text{prob}_{\text{max}}(\text{Prob}_{\mathcal{M}_1}(\alpha, i), \text{Prob}_{\mathcal{M}_2}(\alpha, i))} \]

(6)
2. 3.3. Pre-processing steps for PCB defect detection

Figure 2 shows pre-processing steps and defect detection steps. According to this stage, a database is created which contains all reference images, an input test image is given for processing which is passed through image thresholding, filtering, denoising and detection classification.

\[
\begin{align*}
    b_1(p_1, p_2) &= g(p_1, p_2) - m(p_1, p_2) \\
    b_2(p_1, p_2) &= g(p_1, p_2) - m(p_1, p_2)
\end{align*}
\]

where \(b_1\) represent positive image and negative image is denoted by \(b_2\). Next stage performs image difference modeling which provides difference between two images i.e. reference image and dataset image. In this process, pixel by pixel comparison is performed by applying XOR logical operator. Image subtraction and this XOR operation are similar in nature. Positive and negative difference of pixels are combined together which provides defective region.

Later, image addition is performed which combines two different images by applying OR operation.

\[
b_3(p_1, p_2) = g(p_1, p_2) + m(p_1, p_2)
\]

where \(b_3\) is the combined image.

A hole can be characterized as a foundation locale that is encompassed by an associated foreground of closer pixels. To fill the holes, we have performed dilation, intersection evaluation and complementation on the input image. Object comparator counter uses total number of counts between two images. This operation provides change detection in the original image.

3. MARKOV MODEL FOR DATA MINING PCB CLASSIFICATION

This section describes about the Markov model classification process. In order to formulate the Markov Model we use finite automata based probabilistic transition approach. This approach classifies the stages of the given dataset by using deterministic emission function. State transitions probability is time dependent which observes the automation process.

In this work we use this approach to classify the PCB defect detection prediction in the dataset using data mining approach. Steps of this approach are given below:

1. Input data sequence is given as \(\mathcal{X} = \{X_1, X_2, ..., X_n\}\) which consists \(S\) states and a markov model \(\mathcal{M}\)
2. Let Markov model \(\mathcal{M}\) with \(S\) states, the transitions probability (\(P\)) of this between \(S \times S\) dimension matrix \(\mathcal{D}\) with the elements \(D_{ij}\), \(P_{ij}\) is the probability of transition from state \(i\) to \(j\)
3. By using this probability transition matrix, one or more sequence also can be used to perform the training and the training set is given as

\[
\alpha^* = \arg \max_{\alpha} p(\mathcal{X}|\mathcal{M}, \alpha)
\]

Where \(\alpha\) represents the training set

In the first stage if the input sequence is similar to the states of the build Markov Model, then the observation probability can be given as
Markov model and data mining approach for PCB Component Defect

\[ p(X) = p(F|X_S)p(X_1|F) \prod_{s=2}^{S} p(x_s|x_{s-1}) \]  

(7)

In other case the training set is estimated based on the maximum probability. This is achieved based on the maximum likelihood criterion, which is given below

\[ P(x_s = l|x_{s-1} = k) = \frac{n_{kl}}{n_k} \]  

(8)

\[ n_{kl} \] is the representation of time for \( S \) to follow the input given sequence as training set, number of visited states are given as \( n_k \), finally the training set vector can be achieved as

\[ P(x_s = 1|x_{s-1} = k) \]  

(9)

### III. Results And Discussion

In this section we describe achieved results for PCB defect detection using PCB image dataset with annotations of component. Each image is rotated and stored in database along with dimension parameters for each PCB. These images are depicted in figure 3.

![PCB Image Processing Steps](image)

(a) Input image  
(b) Binarized image  
(c) Rotated image  
(d) Rotated Binarized image  
(e) Gray scale image  
(f) Thresholding image  
(g) Filtered image

**Figure 3. PCB image processing steps**

In figure 3, we show pre-processing steps applied on PCB images, these steps include image binarization, gray scale conversion, image thresholding and image filtering. Complete process in is implemented with the help of MATLAB 2013b tool. For noise consideration, white Gaussian noise parameters are applied. Further, this noisy image is filtered using median filtering approach. Finally, this image is considered for feature extraction analysis for varied test case consideration. In order to compute, to measure the performance of the system, we use various statistical parameters which includes: (i) True Positive Rate (ii) False Positive Rate (iii) Precision (iv) Recall and (v) ROC area. Performance analysis of the proposed approach is mentioned in the given section by considering the Markov model using various classification approaches.
True positive rate can be computed as
\[ TPR = \frac{TP}{TP + FN} \] (10)

\( TP \) denotes the true positive values, \( FN \) is the representation of false negative values.

False positive rate computation is carried out using below given equation
\[ FPR = \frac{FP}{FP + TN} \] (11)

Precision is computed using
\[ precision = \frac{TP}{TP + FP} \] (12)

False score is defined as
\[ f\text{score} = \frac{2TP}{2TP + FP + FN} \] (13)

Kappa measurement is given as
\[ Kappa = \frac{observed \ Class - Expected \ Class}{1 - Expected \ Class} \] (14)

### TABLE 2. Classification performance results

<table>
<thead>
<tr>
<th>Classifier</th>
<th>TP</th>
<th>FP</th>
<th>PR</th>
<th>Recall</th>
<th>ROC</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Markov Model</td>
<td>4.3</td>
<td>7.5</td>
<td>4.2</td>
<td>4.39</td>
<td>4.5</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>3.2</td>
<td>8</td>
<td>5.1</td>
<td>3.5</td>
<td>4.3</td>
<td>F</td>
</tr>
</tbody>
</table>

In order to show robust performance, a comparative study is presented by considering multiple classifiers and performance is measured aforementioned measurement parameters. For this analysis, complete database is divided into two cases where in first case 50% data is used for training purposed and remaining 50% data is considered for testing. Figure 4 shows obtained performance for test case 1 where conventional classifiers such as neural network, KNN (K-Nearest Neighborhood), SVM (Support Vector Machine) are compared with proposed classification model. Study shows that proposed approach gives better accuracy, FP, TP, precision and recall performance when compared with state-of-art classification model. Table 3 shows a comparative analysis for test case 1 where various parameters are compared by considering different classification schemes such as neural network, KNN and SVM.

### TABLE 3. Comparative analysis for test case 1

<table>
<thead>
<tr>
<th>Classifier</th>
<th>TP</th>
<th>FP</th>
<th>PR</th>
<th>Recall</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neural Network</td>
<td>8.22</td>
<td>7.93</td>
<td>0.01</td>
<td>2.13</td>
<td>81.22</td>
</tr>
<tr>
<td>KNN</td>
<td>0.5</td>
<td>4.55</td>
<td>1.38</td>
<td>0.24</td>
<td>83.57</td>
</tr>
<tr>
<td>SVM</td>
<td>2.61</td>
<td>2.37</td>
<td>3.44</td>
<td>4.21</td>
<td>87.81</td>
</tr>
<tr>
<td>Proposed</td>
<td>5.84</td>
<td>0.22</td>
<td>6.27</td>
<td>7.31</td>
<td>91.29</td>
</tr>
</tbody>
</table>

![Comparative Analysis case 1](image)

**Figure 4.** Comparative analysis for test case 1.
TABLE 4. Comparative analysis for test case 2

<table>
<thead>
<tr>
<th>Classifier</th>
<th>TP</th>
<th>FP</th>
<th>PR</th>
<th>Recall</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neural Network</td>
<td>1.2</td>
<td>0.34</td>
<td>95.21</td>
<td>90.1</td>
<td>90.52</td>
</tr>
<tr>
<td>KNN</td>
<td>5.27</td>
<td>5.88</td>
<td>96.27</td>
<td>91.61</td>
<td>91.28</td>
</tr>
<tr>
<td>SVM</td>
<td>7.63</td>
<td>8.97</td>
<td>97.49</td>
<td>96.16</td>
<td>97.04</td>
</tr>
<tr>
<td>Proposed</td>
<td>7.84</td>
<td>4.22</td>
<td>97.27</td>
<td>97.31</td>
<td>98.29</td>
</tr>
</tbody>
</table>

Similarly, table 4 depicts performance comparison for test case 2.

![Comparative Analysis Case 2](image)

**Figure 5.** Comparative analysis for test case 2.

Similarly, we present another comparative study by considering various classifiers. According to this case study, all available images are used for both training and testing process. This analysis shows that proposed approach obtains better accuracy when compared with test case 1 and other state-of-art models for classification techniques. Recent study shows that wavelet features provide better classification performance hence, here a classification study is presented by considering wavelet features and proposed color distribution modeling. In this analysis, both test cases are considered for comparative analysis. Figure 6 shows a comparative analysis for test case-1 where 50% data is taken for training and testing model.

![Feature based Analysis](image)

**Figure 6.** Comparative analysis for test case 1.

In this study, wavelet based feature extraction model [18] is combined with proposed classification approach. Experimental study shows that proposed combination of feature extraction and classification provides significant performance in terms of classification accuracy. Similarly, case 2 analysis is also presented in figure 7.
IV. Conclusion

In this work we present a robust approach for PCB defect detection and classification by extraction of the PCB component. This approach is a combination of computer vision image processing and data mining technique. According to proposed approach, image dataset is created by taking reference image and their dimensions are considered for data mining features. Initially image is passed through image processing stage which provides the defect detection of component. In order to validate and improve the performance of defect detection and classification, Markov Model classification approach is applied where dimensions are considered as feature and matched with the input image’s dimension for prediction of the defect. Results evaluation shows that proposed approach achieves accuracy of 94 % in terms of classification. Markov model based technique shows significant performance for defect detection and classification. Due to promising nature of this technique, in future this approach can be used for various other objectives such as medical image classification, hyper spectral image classification or content based image retrieval systems.

References


