

## The Improved K-nearest Neighbor Solder Joints Defect Detection

Meiju Liu<sup>1, a</sup>, Lingyan Li<sup>1, b\*</sup> and Wenbo Guo<sup>1, c</sup>

<sup>1</sup> Department of Information and Control Engineering, Shenyang Jianzhu University,  
Shenyang 110168, Liaoning, China.

<sup>a</sup> 11404178112@qq.com, <sup>b</sup> 1159558936@qq.com, <sup>c</sup> 2217020304@qq.com

**Keywords:** AOI; Feature extraction; Improved K-nearest neighbor algorithm; Solder joint defect detection.

**Abstract.** Aiming at the problems such as defect misstatements, omissions are prone to happen when automatic optical inspection (AOI) system detects Printed Circuit Board (PCB) solder joints. The article puts forward a kind of method based on improved K-nearest neighbor to test and classify the quality of solder joints. Firstly, the original images collected by industrial camera should be pretreated, and solder joints should be positioned by using the method of template matching. Secondly, the features of solder joints should be extracted and selected usefully through the experiments. Finally, the improved K-nearest neighbor algorithm based on effective feature is used to test and classify solder joints. Experiments show that the improved K-nearest neighbor algorithm has higher accuracy and stronger adaptability than neural network algorithm used for classification. What's more, the cost of testing is also reduced effectively. So we can conclude that the improved K-nearest neighbor algorithm is useful for solder joints testing.

### Introduction

With the development of PCB assembly technique, the traditional manual visual inspection method is becoming more and more difficult. This method is affected by subjective factors easily, so it often happens to check wrongly or leak detection. At present, there are some common detection classification methods such as AOI system [1-4], K-nearest neighbor method [5-8], BP neural network algorithm [9,10], etc. The common detection algorithms used by AOI inspection system still exist error-prone, and they are always affected by external factors. The accuracy of detection can't meet the requirements, and the process of using is too complicated [11,12]. By comparison, K-nearest neighbor method possessed the full capability of simple method, unsupervised training ability, without prior model, etc. But it also has its own disadvantages, for example, easily affected by the surrounding noise and isolated points in the sample. And the value of K will influence the classification result at the same time.

Aiming at the disadvantages of K-nearest neighbor method, some methods are proposed to improve K-nearest neighbor method. First of all, the samples in the original space are mapped in a high-dimensional kernel space. The purpose is to prominent characteristic differences between the samples in different categories, and make the samples linear separable or approximately linear separable. Then K-nearest neighbor method is used to classify in the kernel space, so the classification error of nonlinear samples was reduced effectively. Experiments indicate that the improved K-nearest neighbor method obtained the better classification effect.

### Algorithm Description

**K-nearest Neighbor Method.** Suppose there are  $m$  categories, and each category has a sample  $(w_1, w_2, \dots, w_m)$ . For the classification of unknown sample  $x$ , firstly calculated the distance from the known category of samples  $x_i^k$ , and then to find the closest sample. So the category of sample  $x$  is subordinate to the class. The following is the discriminant function of  $w_i$  class:

$$d_i(x) = \min_k \|x - x_i^k\|, k = 1, 2, \dots, N_i \quad (1)$$

If  $d_m(x)$  is equal to  $\min d_i(x)$ , we can get the conclusion:  $x$  is included in  $w_m$ . This method is only defined according to the nearest sample. To avoid the accident of one sample, it should find  $K$  nearest neighbor samples. Then it is belonged to the kind of the most samples. This method is usually recorded as KNN.  $K$ -nearest neighbor method has perfect theoretical basic, and it is an important kind of classification methods. But the classification effect is poorer when the boundary of the sample is nonlinear.

**Improved K-nearest Neighbor Method.** To improve  $K$ -nearest neighbor algorithm, the unclassified samples in the original space should be mapped to a high-dimensional feature space  $\phi(x)$ . In this way, the differences between samples stand out. Then use  $K$ -nearest neighbor algorithm to classify the samples in high-dimensional space.

$$x \rightarrow \phi(x) \quad (2)$$

Unclassified samples mapped in the high-dimensional space are converted into:  $\phi(x_1), \phi(x_2), \dots, \phi(x_N)$ . And the distance formula between the samples is:

$$d_i(x) = \|\phi(x_i) - \phi(x_j)\|^2 \quad (3)$$

Next  $K$ -nearest neighbor algorithm is used to classify the kinds of the samples in high-dimensional space. The results show that the improved  $K$ -nearest neighbor algorithm can reduce the error, and get better classification results.

### Experiment Based on the Improved K-nearest Neighbor Method

The main types of defective solder joints have more solder, less solder, leakage welding and connection welding, etc, as shown in Fig. 1; Solder joint defects exist differences in many aspects such as type, location, shape. So the training samples need to be preprocessed and sample features should be obtained. The features are translated into a high-dimensional space and then the transformed features are trained to form the trainer.

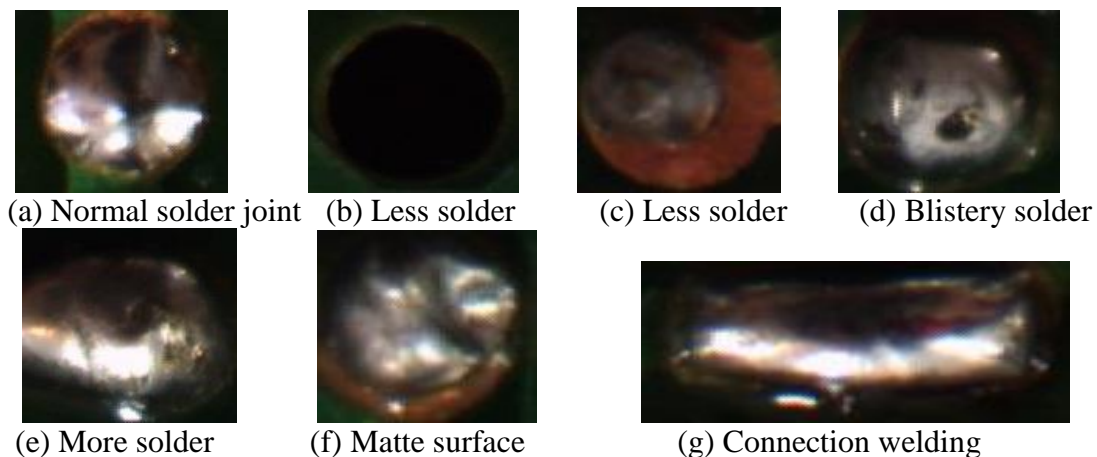


Figure 1. The main types of solder joints

This article is in view of the PCB solder joint detection. Because the background of PCB is fixed, template matching method may be used for the position of solder joints. Then extract the characteristics of the solder joints according to the feature and gray histograms. Finally, obtain the classifier through training the model based on the improved  $K$ -nearest neighbor method, the specific process of this experiment is shown as follows.

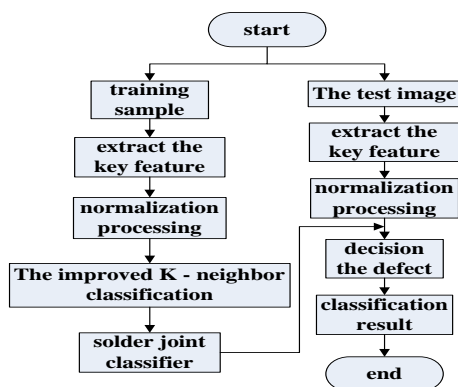


Figure 2. The process of this experiment

**The Location of Solder Joints Based on Template Matching.** The location of solder joints is fixed, and the center of the solder joints is closely related to the particular location of PCB. Aiming at the interest region, we can create a template to improve the matching speed. The basic process of solder joint location as follows:

Step1. Firstly, determine the region of interest (ROI). It only needs to make sure the coordinates of top-left and bottom-right points. Using the function of `gen_rectangle()` will generate a rectangle region. Then the function of `area_center()` is used to find the center of rectangle region. Finally, the function of `reduce_domian()` is used to obtain ROI from the image, as shown in the Fig. 3;

Step2. Establish a ROI template by using the function of `create_shape_model()`. It also needs the function of `inspect_shape_model()` to watch the template created. In addition, the function of `get_shape_model_contours()` which can obtain the contour outline of the template is used for the following matching, as shown in the Fig. 4;

Step3. Open another image to match the template after created. In another word, find the part of template in the new image. It can be finished by using the function of `find_shape_model()`. It still needs to be transformed to display after finding the template, as shown in the Fig. 5; the functions of `vector_angle_to_rigid()` and `affine_trans_contour_xld()` are used to calculate a rigid affine transformation. The role of the transformation is to transform the reference images into the current images.

Step4. Make sure the location of every solder joint according to the relationship between each solder joint center and the template center after matching, as shown in Fig. 6;



Figure3. The region of interest

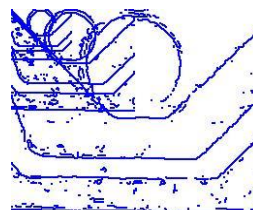


Figure 4. The outline of template

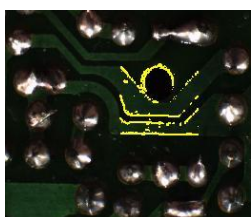


Figure 5. The template matching

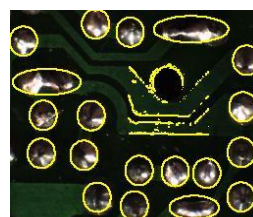


Figure6. The location of solder joints

**The Feature Extraction of Solder Joints.** Feature extraction is the precondition of defect detection, so you need to accurately extract the features of the solder joint. And then select effective

features as the detection basis through the experiment. The article selects the characteristics as follows:

1) The normalized area (A). The area of the solder joint is only related to its boundary, and here use the method of calculating the number of pixels in solder joint (including boundary) to normalize the area. Assuming that set 1 as the length of the square pixel, the formula of solder joint area:

$$A = \sum 1 \quad (4)$$

2) The circumference of the solder joint region (C). The region is showed by chain codes, such as the outline of solder joint  $x_0L_0L_1 \cdots L_n$ , the start point coordinate  $x_0$ , the Code word is shown as  $L_i$ . There is using the method of Euclidean metric, so the formula of C is shown as follows:

$$\Delta C = \begin{cases} 1 & L_i \in \{(0, 2, 4 \cdots)\} \\ \sqrt{2} & L_i \in \{(1, 3, 5 \cdots)\} \end{cases} \quad (5)$$

$$C = \sum_{i=0}^n \Delta C \quad (6)$$

3) The hydraulic radius of solder joint (R).

$$R = A / C \quad (7)$$

4) The sphericity of solder joint (S). The radius of inscribe circle is shown as  $r_i$ , and the radius of circumcircle circle is shown as  $r_c$ . So the formula of S is shown as follows:

$$S = r_i / r_c \quad (8)$$

5) The moment of solder joint ( $m_i$ ).

$$m_i = E[n] = \sum_{n=0}^{N-1} h^i p(n) \quad (9)$$

If  $i=1$ , it would represent the mean of the image.

6) The central moment of solder joint.

$$u_i = E[(h - E(h))^i] = \sum_{h=0}^{N-1} (h - m_1)^i p(h) \quad (10)$$

If  $i=2$ , it would represent the variance of gray value.

Feature selection is not the more, the better. It is necessary to select the typical features. Various features have large differences in such as number, unit and so on. So it needs to be normalized:

$$X_i = (Z_i - Z) / S \quad (11)$$

The normalized feature is shown as  $X_i$ , and  $Z_i$  represents the unnormalized feature.  $Z$  is the mean value of  $Z_i$ , and  $S$  is the variance of  $Z_i$ .

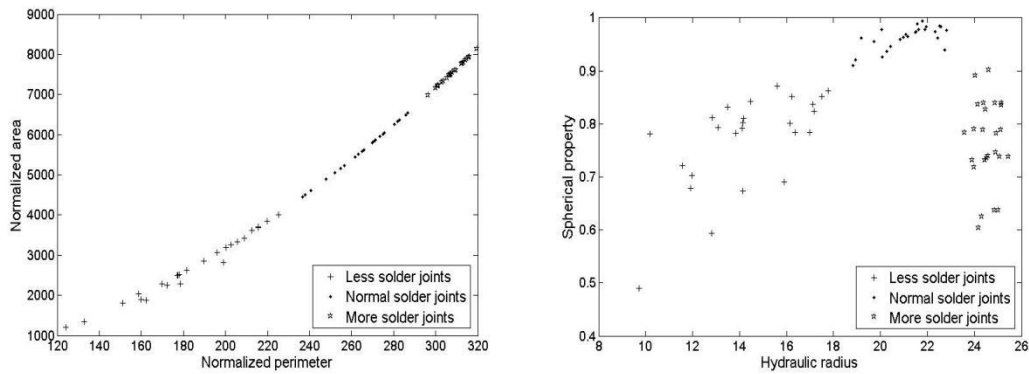
**The Experiments of Solder Joints Defect Detection.** The steps of the experiment are shown as follows:

Step1. Firstly, the representative samples are selected as the training samples. The features should be extracted and transformed into a high-dimensional feature space. And the distribution of the representative features should be described in the feature space, as shown as Fig. 6;

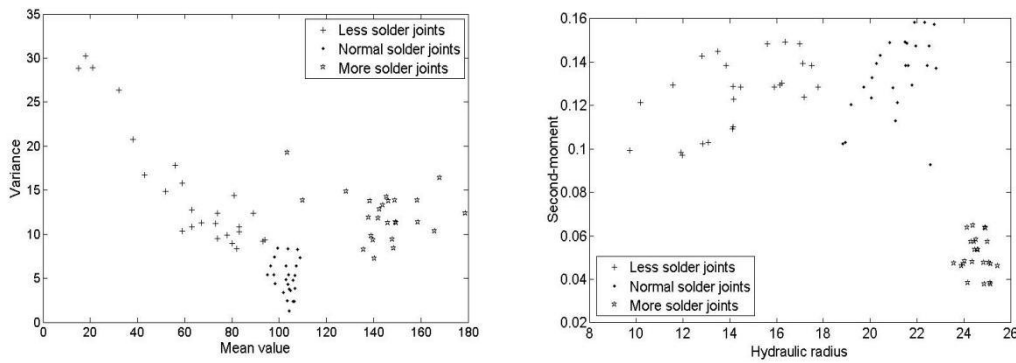
Step2. Then the function of `create_class_knn()` can be used to generate a classifier based on the improved K-nearest neighbor method. And the function of `add_samples_image_class_knn()` is used

to transform the training samples into the data that the classifier can recognize. The function of `train_class_knn()` is used to form a taxonomic tree.

Step3.Finally, the functions of `classify_image_class_knn()` and `region_to_mean()` are used to detect and classify the testing samples and sign the unqualified samples, as shown as Fig. 7;



(a) Normalized circumference and area (b) Hydraulic radius and sphericity property



(c) Mean value and variance (d) Hydraulic radius and second-moment

Figure 7. The distribution of samples in various feature spaces

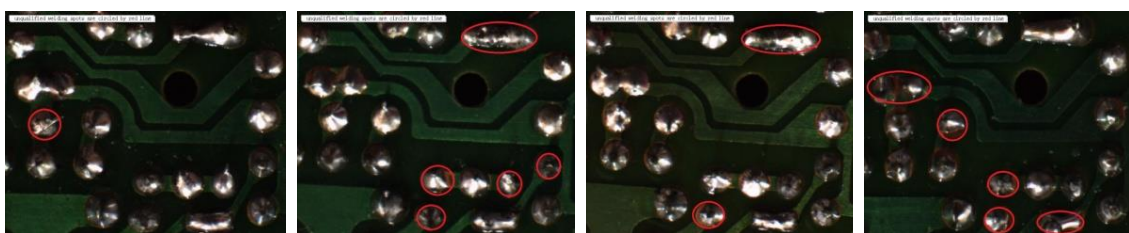


Figure 8. The detection results based on the improved K-nearest neighbor method

From Fig. 7, we can see that solder joints are roughly judged by circumference and area. And most of the solder joints can be distinguished by the features of hydraulic radius and sphericity. Mean value can discriminate the types of the solder joints well, but variance only can discriminate a part of solder joints. Two-moment can distinguish the more solder joints better, but it is less accurate to distinguish less and normal solder joints. From Fig. 8, we see that the improved K-nearest neighbor method can accurately detect unqualified solder joints.

### The Analysis of Experiment Results

In the article, the improved K- nearest neighbor method is compared with K- nearest neighbor method, AOI system and BP neural net. This experiment aimed at the three types of more solder,

normal solder and less solder. There are 90 training solder joints and 209 testing solder joints, as shown as Table 1.

Table 1 Solder joint samples

Type	Sample	Training samples	Testing samples
	More solder	26	32
	Less solder	23	37
	Normal solder	41	140
	Amount	90	209

The methods of improved K- nearest neighbor, K- nearest neighbor, AOI system and BP neural net are compared to detect the solder joints, the results are shown in table 2. The improved K- nearest neighbor method has stronger adaptability and higher precision rate than K- nearest neighbor method. The improved K- nearest neighbor method not only can decline the rates of mistake and miss, but also avoid falling into local minimum and being inadequate for small samples. Therefore, the improved K- nearest neighbor method has investigative and applicative values in the aspect of solder joint detection.

Table 2 The comparison of detection methods

Detection method	Type	Precision rate	Mistake rate	Miss rate
The improved K-nearest neighbor method(k=3)	More solder	96.875%	3.125%	0%
	Less solder	100%	0%	0%
	Normal solder	99.286%	0.714%	0%
	Comprehensive	99.043%	0.957%	0%
AOI system	More solder	93.75%	6.25%	3.125%
	Less solder	94.595%	5.405%	0%
	Normal solder	96.428%	2.143%	0%
	Comprehensive	95.694%	3.35%	0.478%
K-nearest neighbor method(k=3)	More solder	90.625%	3.125%	0%
	Less solder	94.595%	5.405%	2.703%
	Normal solder	97.857%	1.429%	0%
	Comprehensive	96.172%	2.392%	0.479%
BP neural Net method	More solder	96.875%	3.125%	3.125%
	Less solder	91.892%	5.405%	0%
	Normal solder	98.571%	2.857%	0.714%
	Comprehensive	97.129%	3.349%	0.957%

## Conclusion

The methods of improved K- nearest neighbor, K- nearest neighbor, AOI system and BP neural net are compared to detect the solder joints, the results are shown in table 1. The improved K- nearest neighbor method has stronger adaptability and higher precision rate than K- nearest neighbor method. The improved K- nearest neighbor method not only can decline the rates of mistake and miss, but also avoid falling into local minimum and being inadequate for small samples. Therefore, the improved K- nearest neighbor method has investigative and applicative values in the aspect of solder joint detection.

## Acknowledgements

The Project of Liaoning Province's Education Department (L2013225) ; National Natural Science Found Project of China (61272253) .

## References

- [1] Qiang G, Shanshan Z, Yang Z, et al. Detection method of PCB component based on automatic optical stitching algorithm [J]. Circuit World, 2015, 41(4): 133-136.
- [2] Liu Y, Yu F. Automatic inspection system of surface defects on optical IR-CUT filter based on machine vision [J]. Optics and Lasers in Engineering, 2014, 55: 243-257.
- [3] Chen S H, Perng D B. Automatic optical inspection system for IC molding surface[J]. Journal of Intelligent Manufacturing, 2014: 1-12.
- [4] Lin C S, Tzeng G A, Lay Y L, et al. An automatic optical inspection system for assembly and evaluation of an optical encoder [J]. J. Test. Eval, 2014, 42(4): 819-827.
- [5] Ma X, He X, Shi X. A variant of K nearest neighbor quantile regression [J]. Journal of Applied Statistics, 2015: 1-12.
- [6] Lu W, Shen Y, Chen S, et al. Efficient processing of k nearest neighbor joins using mapreduce[J]. Proceedings of the VLDB Endowment, 2012, 5(10): 1016-1027.
- [7] Steenwijk M D, Pouwels P J W, Daams M, et al. Accurate white matter lesion segmentation by k nearest neighbor classification with tissue type priors (kNN-TTPs) [J]. NeuroImage: Clinical, 2013, 3: 462-469.
- [8] Yesilbudak M, Sagioglu S, Colak I. A new approach to very short term wind speed prediction using k-nearest neighbor classification [J]. Energy Conversion and Management, 2013, 69: 77-86.
- [9] Garcia-Pineda O, MacDonald I R, Li X, et al. Oil spill mapping and measurement in the Gulf of Mexico with Textural Classifier Neural Network Algorithm (TCNNA) [J]. Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of, 2013, 6(6): 2517-2525.
- [10] Li J, Cheng J, Shi J, et al. Brief introduction of back propagation (BP) neural network algorithm and its improvement[M]//Advances in Computer Science and Information Engineering. Springer Berlin Heidelberg, 2012: 553-558.
- [11] Cao B Q, Liu J X, Wen B. Currency Characteristic Extraction and Identification Research Based on PCA and BP Neural Network [J]. Journal of Convergence Information Technology, 2012, 7(2).
- [12] Jakati J S, Matad S S. PCB defect detection based on pattern matching and segmentation algorithm [J]. International Journal of Engineering and Innovative Technology (IJEIT) ,2015, 4(11):188-192.