International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:13, No:11, 2019

Automatic Adjustment of Thresholds via Closed-Loop Feedback Mechanism for Solder Paste Inspection

Chia-Chen Wei, Pack Hsieh, Jeffrey Chen

***Abstract*—**Surface Mount Technology (SMT) is widely used in the area of the electronic assembly in which the electronic components are mounted to the surface of the printed circuit board (PCB). Most of the defects in the SMT process are mainly related to the quality of solder paste printing. These defects lead to considerable manufacturing costs in the electronics assembly industry. Therefore, the solder paste inspection (SPI) machine for controlling and monitoring the amount of solder paste printing has become an important part of the production process. So far, the setting of the SPI threshold is based on statistical analysis and experts' experiences to determine the appropriate threshold settings. Because the production data are not normal distribution and there are various variations in the production processes, defects related to solder paste printing still occur. In order to solve this problem, this paper proposes an online machine learning algorithm, called the automatic threshold adjustment (ATA) algorithm, and closed-loop architecture in the SMT process to determine the best threshold settings. Simulation experiments prove that our proposed threshold settings improve the accuracy from 99.85% to 100%.

***Keywords*—**Big data analytics, Industry 4.0, SPI threshold setting, surface mount technology.

I. Introduction

WITH the rapid development of technology and market demand, electronic components on the PCB have become smaller, denser and more complex. Moreover, frequent changes to component specifications and combinations and rapid development of new products, production flexibility and production capacity have become essential. Electronics manufacturers are constantly striving to develop robust process methods to improve process stability and advanced process technology. SMT technology is a widely used method in the electronics assembly industry. Its advantages are not only the mass production of consumer electronic components in a low-cost way, but also the rapid adaptation and flexibility of small and medium-sized production and prototyping. A traditional SMT process is shown in Fig. 1 (a).

According to industry reports, approximately 52% to 71% of SMT defect problems are related to solder paste printing [5]-[6]. In the actual solder paste printing process, two situations often occur (i) insufficient solder, which is easy to cause some defects such as improper solder joints, open, or offset after reflow oven; (ii) over solder, which can cause a solder-bridge and result in a short circuit, or offset defect. If the solder paste on the PCB has the above defects, it is necessary to clean the PCB and reprint the solder paste, or it must be sent to the repair station for repair. Thereby, it affects the overall production efficiency and production quality. However, the above mentioned situations can be avoided by installing the SPI machine after solder paste printing and setting the appropriate threshold for the volume, area, and height of the solder paste for each PCB pad. SPI machine not only can set an acceptable solder paste volume, area and height range according to each of the aperture of the stencil but also provides a statistical analysis tool called SPC to help process engineers identify soldering related problems. For the volume, area, and height of the solder paste on each PCB pad, in addition to the customer's special requirements threshold setting, the threshold setting is basically set by senior engineers' knowledge and experiences, using the trial-and-error method or the threshold provided in the SPC statistical tool. However, there are many factors that will affect the judgment of the engineers and the accuracy of the threshold provided by the SPC statistical tool will be affected by whether the data are approximately normally distributed. Therefore, this paper proposes an ATA algorithm through a closed-loop feedback mechanism to find the best threshold settings for volume, area, and height of the solder paste for each PCB pad, as shown in Fig. 1 (b). The experimental results indicate that the proposed threshold settings improve the accuracy from 99.85% to 100%.

The remainder of this paper is organized as follows. Section II introduces the definitions and notations used throughout the paper. Section III presents a proposed algorithm. Simulation experiment result is presented in Section IV. Finally, Section V offers some concluding remarks.

II. Preliminaries

The SPI machine and AOI machine play the role of measurement and monitor in the SMT process. This paper will propose the SPI thresholds of the solder paste in volume, area, height based on different component types because different component types may have a different number of pins of an electronic component and corresponding pads. Whenever SPI inspects one or a batch of PCBs, it will produce volume, area, and height measurements which are denoted by $\hat{V}(C_i(j))$, $\hat{A}(C_i(j)), \hat{H}(C_i(j))$, respectively, and $C_i(j)$ represents the jth pad in the ith component type, where j = 1, 2, ..., $k(i)$. Thus, if $\hat{V}_{ij}$= 1, then the volume of solder paste reaches the target value. If $\hat{V}_{ij} \ll 1$ (or $\hat{V}_{ij} \gg 1$), then the volume of the solder paste is insufficient (or too much).

Chia-Chen Wei and Pack Hsieh are with the Software Solution Development Division of Central Engineering of USI, Taiwan (e-mail: stanley_wei@usiglobal.com, shinan_xie@usiglobal.com).

Jeffrey Chen is the AVP of Central Engineering of Universal Scientific Industry now, Taiwan (e-mail: OJ_Chen@usiglobal.com).

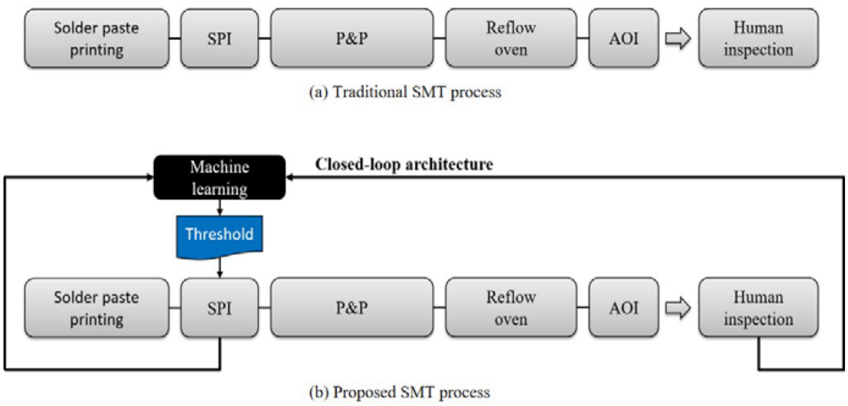


Fig. 1 SMT process architecture diagram

After going through reflow soldering process, the assembled boards are conveyed into AOI machine and optically inspected whether there is any surface-related defect. Because the AOI misjudgment rate is still high, this paper only considers the result of the human inspection as the correct judgment. It is worth noting that AOI machines usually give inspection results in units of components. Therefore, when the human inspection result of a certain component is insufficient solder (or over solder), we will regard the minimum (or maximum) solder paste measurement of the component as a representative of the insufficient solder defect.

For the jth PCB pad of the component type i, the measured values obtained from SPI machine can be considered as a point $(\hat{V}_{ij}, \hat{A}_{ij}, \hat{H}_{ij})$ in a three-dimensional space. Because the amount of the solder paste in the solder paste printing will decrease as the times of printing increases, the engineer will regularly monitor them. Thus, the scattered plot of the set $\{(\hat{V}_{ij}, \hat{A}_{ij}, \hat{H}_{ij}) \in \mathcal{R}^3 | 1 \le j \le k(i)\}$ collected from the SPI machine can be various elliptical or other different shapes, rather than being restricted to circles. In this paper, we consider the method called DBSCAN (Density-based spatial clustering of applications with noise) [1] which is a density-based clustering algorithm. The advantage of this method is that it can discover clusters with arbitrary shape and with minimal number of input parameters as follows: (i) Eps: neighborhood of a point; (ii) MinPts: minimum number of points to create a cluster. The following definitions are the basic idea behind this DBSCAN method. For definitions of density-reachable and density-connected not being mentioned here, please refer to [4].

III. Inspection Quality Regression

This paper is based on the assumption that we do not know the correct SPI thresholds (upper & lower) at the beginning of production. The initial value of the upper and lower thresholds of volume, area, and height are set to 1.

In the actual production process, it takes at least 15~20 minutes for the PCB to pass from the SPI inspection to the human inspection. Because we use the results from human inspection as the correct answer to determine whether the threshold settings are appropriate, we use the unsupervised clustering method to find the appropriate thresholds before getting the human inspection results. Once we have the results of the human inspection, we will label those corresponding data and adopt a supervised classification to find the optimal thresholds. Therefore, the proposed ATA algorithm is divided into two stages as follows: (i) Shape-based threshold algorithm (Algorithm 1) which is based on the unsupervised clustering method - DBSCAN; (ii) AOI feedback-based threshold algorithm (Algorithm 2) which is based on supervised classification – Support Vector Machine (SVM).

A. Shape-Based Threshold

Whenever we collect data from an SPI machine, we will do basic data preprocessing to convert the raw data into a clean and meaningful data set according to (1). Ideally, for each component type i, the data set $\{(\hat{V}_{ij}, \hat{A}_{ij}, \hat{H}_{ij}) \in \mathcal{R}^3 | 1 \le j \le k(i)\}$ collected from the SPI machine should be aggregated into one cluster, but the actual production process may experience the amount of solder paste decrease, just like mentioned in Section II; or it may have solder paste material problems so that the data distribution will become elliptical or other shapes, rather than being restricted to circles. However, in the solder paste printing process, if we consider the pads' neighbor relationships and/or its symmetry, the difference in the amount of solder paste in volume, area, and height of these red PCB pads will be relatively small. In addition, under the normal operation of the solder paste printing, the difference in the amount of solder paste will be less than ε, where $\varepsilon > 0$.

According to mentioned above, it is better to use the DBSCAN clustering method which is mentioned in Section II, we set Eps = ε which can be obtained by history data and set MinPts = 2 ~ 6. After executing DBSCAN on the data set, it will output the clusters and noise data points according to Definition 1 and Definition 2 in [4], respectively. In general, if the parameter settings are appropriate, there are 1~3 resulting clusters and only one of them is the largest. In our algorithm we will set the largest cluster as the main cluster C (called core cluster) and the others as noise data points. Then, we set the thresholds as follows:

$$(L_{V_i}, U_{V_i}) = \left(\min_{\hat{V}_{ij} \in C} S_{\hat{V}_{ij}}, \max_{\hat{V}_{ij} \in C} S_{\hat{V}_{ij}}\right),$$

$$(L_{A_i}, U_{A_i}) = \left(\min_{\hat{A}_{ij}\in C} S_{\hat{A}_{ij}}, \max_{\hat{A}_{ij}\in C} S_{\hat{A}_{ij}}\right)$$
$$(L_{H_i}, U_{H_i}) = \left(\min_{\hat{H}_{ij}\in C} S_{\hat{H}_{ij}}, \max_{\hat{H}_{ij}\in C} S_{\hat{H}_{ij}}\right) \quad (1)$$

where $S_{\hat{V}_{ij}} = \{\hat{V}_{ij} | 1 \le j \le k(i)\}$, $S_{\hat{A}_{ij}} = \{\hat{A}_{ij} | 1 \le j \le k(i)\}$, $S_{\hat{H}_{ij}} = \{\hat{H}_{ij} | 1 \le j \le k(i)\}$ and L_{V_i}, L_{A_i} , L_{H_i} represent the lower threshold in volume, area, and height, respectively, and U_{V_i}, U_{A_i}, U_{H_i} represent the upper threshold in volume, area, and height, respectively. According to the mentioned above, we propose Algorithm 1 in Table I and the main idea of Algorithm 1 is to classify the core cluster and noise data points.

TABLE I
ALGORITHM 1: SHAPE BASED THRESHOLD

Input: Raw data from SPI machine for each component type i, ε, n_0.
Output: $L_{V_i}, U_{V_i}, L_{A_i}, U_{A_i}, L_{H_i}, U_{H_i}$
Step 1. Collect raw data set $S_i = \{(\hat{V}_{ij}, \hat{A}_{ij}, \hat{H}_{ij}) \in \mathcal{R}^3 | 1 \le j \le k(i)\}$ from SPI machine;
Step 2. $C_1, C_2, \dots, C_m = DBSCAN(S_i, Eps = \varepsilon, MinPts = n_0)$, where m is depend on ε and n_0;
Step 3. $C_{i_0} \equiv \mathrm{argmax}_{C\in\{C_1,C_2,\dots,C_m\}}(\|C\|)$ and $N_{i_0} \equiv S_i \backslash C_{i_0}$;
Step 4. **return** the thresholds $L_{V_i}, U_{V_i}, L_{A_i}, U_{A_i}, L_{H_i}, U_{H_i}$ according to (1)

B. AOI Feedback-Based Threshold

This section is mainly considered when we get feedback from the human inspection, as shown in Fig. 1 (b). This paper considers a binary classification problem where we want to predict label y ∈ {0, 1} based on observation $\vec{X}$. A SVM [3], [7] is a type of supervised machine learning algorithm which is mostly used in classification problems and applied in a variety of fields. The main idea of SVM is to find a hyperplane (classifier) such that there exists the largest margin where we can separate the labelled data into two categories well. However, there are extremely low number of defect samples in the actual SMT process, therefore, we adopt LIBSVM tool [2] which is a simple, and efficient software for SVM classification and regression on imbalanced dataset by choosing the weights for different classes.

Let $\vec{x}_{i1}, \vec{x}_{i2}, \dots, \vec{x}_{in_i}$ be the data points in the set S_i which is mentioned in Algorithm 1, where $\vec{x}_{ij} = (\hat{V}_{ij}, \hat{A}_{ij}, \hat{H}_{ij}) \in \mathcal{R}^3$ and $y_{i1}, y_{i2}, \dots, y_{in_i}$ denote the corresponding labels and $y_{ij} \in \{1, -1\}$, where the value 1 represents "re-pass" and the -1 represents the defect of "insufficient solder" or "over solder". We are given a training dataset of n_i points of the form $(\vec{x}_{i1}, y_{i1}), (\vec{x}_{i2}, y_{i2}), \dots, (\vec{x}_{in_i}, y_{in_i})$, and for any hyperplane can be written as the set of points $\vec{x}$ satisfying $\vec{w} \cdot \vec{x} - b = 0$, where $\vec{w}$ is the weight vector and b is the bias. In order to individually find the thresholds of volume, area and height, we divided the training dataset $\{(\vec{x}_{i1}, y_{i1}), (\vec{x}_{i2}, y_{i2}), \dots, (\vec{x}_{in_i}, y_{in_i})\}$ into three training datasets as follows:

$$\boldsymbol{V_i} = \{(\hat{V}_{i1}, y_{i1}), (\hat{V}_{i2}, y_{i2}), \dots, (\hat{V}_{in_i}, y_{in_i})\}, \quad (2)$$

$$\boldsymbol{A_i} = \{(\hat{A}_{i1}, y_{i1}), (\hat{A}_{i2}, y_{i2}), \dots, (\hat{A}_{in_i}, y_{in_i})\}, \quad (3)$$

$$\boldsymbol{H_i} = \{(\hat{H}_{i1}, y_{i1}), (\hat{H}_{i2}, y_{i2}), \dots, (\hat{H}_{in_i}, y_{in_i})\}. \quad (4)$$

For finding the optimal lower threshold of the volume of the solder paste, we first consider the training dataset $\boldsymbol{V_i}$ in (2) and define the following sets:

$$INSU_{V_i} = \{\hat{V}_{ij} | \, y_{ij} = -1\}, \quad (5)$$

$$Core_{V_i} = \{ \hat{V}_{ij} | (\hat{V}_{ij}, \hat{A}_{ij}, \hat{H}_{ij}) \in C_{i_0}\}, \quad (6)$$

$$Noise_{V_i} = \{ \hat{V}_{ij} | (\hat{V}_{ij}, \hat{A}_{ij}, \hat{H}_{ij}) \in N_{i_0}\}, \quad (7)$$

where $INSU_{V_i}$ is a set of collecting all of the insufficient solder data in the volume of the solder paste inspected by the human inspection, $Core_{V_i}$ and $Noise_{V_i}$ are the sets of collecting all of the data in the volume of the solder paste in cluster C_{i_0} and Noise N_{i_0}, respectively. We consider two sets $N_{INSU_{V_i}}$ and $C_{INSU_{V_i}}$ as $N_{INSU_{V_i}} = Noise_{V_i} \cap INSU_{V_i}$ and $C_{INSU_{V_i}} = Core_{V_i} \cap INSU_{V_i}$, respectively, where the set $N_{INSU_{V_i}}$ represents the insufficient solder data belonging to the noise data and the set $C_{INSU_{V_i}}$ represents insufficient solder data belonging to the core data. It is worth noting that the amount of solder paste of the data in the set $C_{INSU_{V_i}}$ is normal under the SPI inspection, which means that the material of the solder paste, improper reflow soldering, or other factors which are independent with the solder paste printing may cause the insufficient solder results. Because the data in set $C_{INSU_{V_i}}$ is independent with the solder paste printing and they cannot be controlled by the SPI threshold, we relabel those data as "re-pass" (i.e., $y_{ij} \equiv 1$, where $\hat{V}_{ij} \in C_{INSU_{V_i}}$) unless those data $\hat{V}_{ij}$ in $C_{INSU_{V_i}}$ satisfies the condition that $\hat{V}_{ij} \le \hat{V}$, $\forall \hat{V} \in Core_{V_i} \backslash C_{INSU_{V_i}}$. In addition, in order to avoid the AOI machine for the missed inspection on the solder paste, we relabel the data in $Noise_{V_i}$ when $Noise_{V_i} \ne \emptyset$ and $N_{INSU_{V_i}} = \emptyset$. According to the mentioned above, we propose the main idea of Algorithm 2 (Table II) is to find the SVM classifiers based on the training dataset $\boldsymbol{V_i}$, $\boldsymbol{A_i}$, and $\boldsymbol{H_i}$, respectively.

TABLE II
ALGORITHM 2: AOI FEEDBACK BASED THRESHOLD

Input: Training dataset $\boldsymbol{V_i}$ ($\boldsymbol{A_i}$ *or* $\boldsymbol{H_i}$)
Output: $\hat{L}_{V_i}$ ($\hat{L}_{A_i}$ or $\hat{L}_{H_i}$)
Step 1.Collect the set $INSU_{V_i}$, $Core_{V_i}$, $Noise_{V_i}$ according to (5), (6), and (7), respectively;
Step 2. $N_{INSU_{V_i}} \equiv Noise_{V_i} \cap INSU_{V_i}$ & $C_{INSU_{V_i}} \equiv Core_{V_i} \cap INSU_{V_i}$;
Step 3. **if** $N_{INSU_{V_i}} \ne \emptyset$ and $C_{INSU_{V_i}} \ne \emptyset$ **then**
relabel the $y_{ij} \equiv 1$, where $\hat{V}_{ij} \in C_{INSU_{V_i}}$ and
$\exists \hat{V} \in Core_{V_i} \backslash C_{INSU_{V_i}}$ such that $\hat{V}_{ij} > \hat{V}$;
else if $N_{INSU_{V_i}} = \emptyset$ **then**
relabel the data in $N_{INSU_{V_i}}$ according to the ratio of the distance $d(N_{INSU_{V_i}}, Core_{V_i})$ to the density of $Core_{V_i}$;
if $C_{INSU_{V_i}} \ne \emptyset$ **then**
relabel the $y_{ij} \equiv 1$, where $\hat{V}_{ij} \in C_{INSU_{V_i}}$;
Step 4. Find the classifier $\hat{L}_{V_i}$ by using imbalanced-SVM [2];
Step 5. **return** $\hat{L}_{V_i}$

The method for finding the upper thresholds is similar to the method for finding the lower thresholds mentioned above, so the description will not be repeated here.

C. ATA Algorithm

When the product is starting to be produced, the initial thresholds $\tilde{L}_{V_i}, \tilde{U}_{V_i}, \tilde{L}_{A_i}, \tilde{U}_{A_i}, \tilde{L}_{H_i}, \tilde{U}_{H_i}$ are set to 1. The SPI machine inspects one or a batch of PCB each time and outputs the measured results for volume, area, and height of the solder paste for each PCB pad. Because there is no correct answer to those results whether there are defects of insufficient solder or over solder, we use Algorithm 1 (unsupervised learning) at the beginning. After obtaining the values $L_{V_i}, U_{V_i}, L_{A_i}, U_{A_i}, L_{H_i}, U_{H_i}$, we update the thresholds as follows:

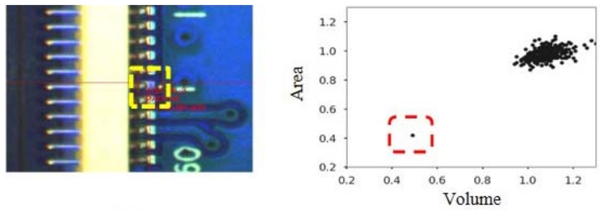
$$(\tilde{L}_{V_i}, \tilde{U}_{V_i}) = \left(\min(\tilde{L}_{V_i}, L_{V_i}), \max(\tilde{U}_{V_i}, U_{V_i})\right)$$
$$(\tilde{L}_{A_i}, \tilde{U}_{A_i}) = \left(\min(\tilde{L}_{A_i}, L_{A_i}), \max(\tilde{U}_{A_i}, U_{A_i})\right)$$
$$(\tilde{L}_{H_i}, \tilde{U}_{H_i}) = \left(\min(\tilde{L}_{H_i}, L_{H_i}), \max(\tilde{U}_{H_i}, U_{H_i})\right)$$

If we still do not have a correct answer to those results, we will repeat the above steps for measured results each time until we get the results of the human inspection. Once we have the results of the human inspection correspond to previous SPI measured values, we use Algorithm 2 (supervised learning) and obtained the values $\hat{L}_{V_i}, \hat{U}_{V_i}$, $\hat{L}_{A_i}, \hat{U}_{A_i}, \hat{L}_{H_i}$, and $\hat{U}_{H_i}$. Then we update the thresholds as follows:

$$(\tilde{L}_{V_i}, \tilde{U}_{V_i}) = \left(min(\tilde{L}_{V_i}, \hat{L}_{V_i}), max(\tilde{U}_{V_i}, \hat{U}_{V_i})\right)$$
$$(\tilde{L}_{A_i}, \tilde{U}_{A_i}) = \left(min(\tilde{L}_{A_i}, \hat{L}_{A_i}), max(\tilde{U}_{A_i}, \hat{U}_{A_i})\right)$$
$$(\tilde{L}_{H_i}, \tilde{U}_{H_i}) = \left(min(\tilde{L}_{H_i}, \hat{L}_{H_i}), max(\tilde{U}_{H_i}, \hat{U}_{H_i})\right)$$

We repeat the above process until the number of updates is greater than a certain value (N_0) or the difference between the updated values is less than a certain value (ε_0), and the update process is stopped.

IV. Experimental Results



(a) (b)

Fig. 2 Illustration of defect of insufficient solder in the connector

In this section, we collect the data from SPI and its corresponding results from human inspection on one of the component types (called connector). There are a total of 28,872 connectors, 12 connectors inspected each time, and each connector has corresponding 62 PCB pads (i.e., a total of 744 PCB pads are inspected each time). In the results of human inspection, there are totally 7 connectors with the defect and all of them are insufficient solder, one of them as shown in Fig. 2 (a). In addition, three of 7 defects are independent with the solder paste printing because their amounts of the solder paste are normal, as shown in the red region of Fig. 3. We simulate the production process and inspect 2406 times, each time we obtain 744 measured values for the connector from the SPI machine, and the corresponding results of the human inspection will be obtained after 20 minutes. In Fig. 2 (b), it is a scattered plot containing the data of Fig. 2 (a) at a certain timestamp. It is clear that the data in the red region of Fig. 2 (b) correspond to the defect of insufficient solder in the yellow region of Fig. 2 (a).

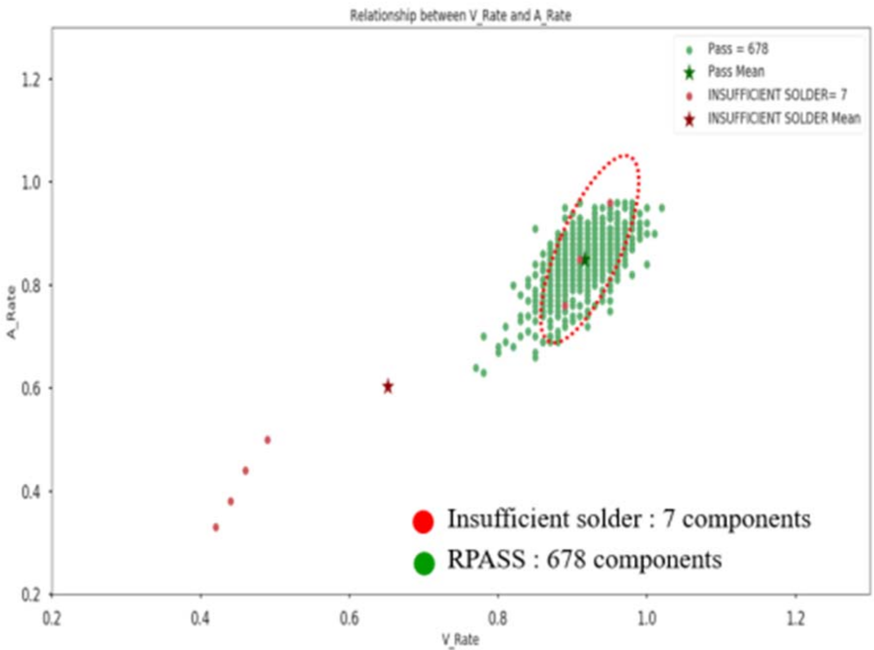


Fig. 3 Illustration of the scattered plot of all results of the human inspection

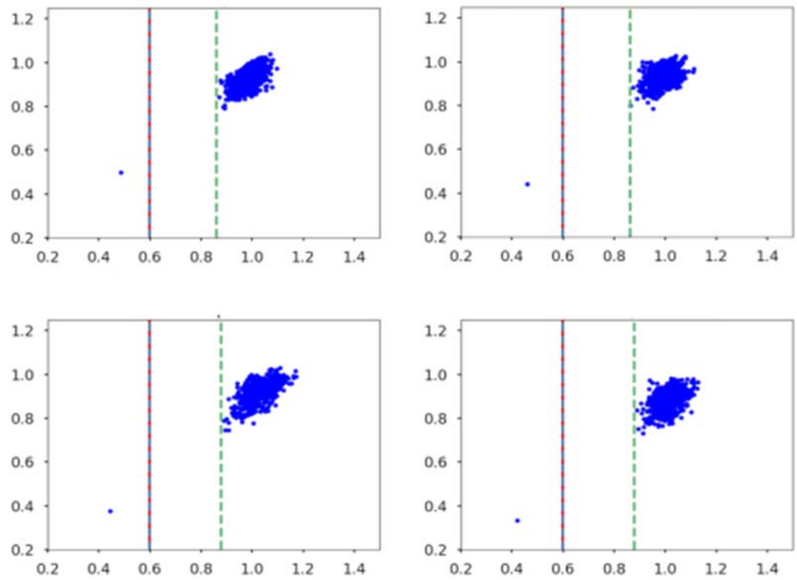


Fig. 4 Illustration of the defects of insufficient solder in four timestamps

For the initial parameter settings, we set $\varepsilon = 0.04$, $n_0 = 2$, N_0 =700, $\varepsilon_0 = 0.01$. After executing the ATA algorithm, we obtain the thresholds as follows: $\tilde{L}_{V_i} = 0.60, \tilde{U}_{V_i} = 1.34, \tilde{L}_{A_i} = 0.48, \tilde{U}_{A_i} = 1.16, \tilde{L}_{H_i} = 0.85$, and $\tilde{U}_{H_i} = 1.48$. In addition, in the timestamps of the four real defects of insufficient solder, our proposed thresholds can filter out these defects, as shown in Fig. 4. But in the original threshold setting will miss filtering one of the defects. Thus, our proposed threshold setting improves the accuracy from 99.85% $\left(\frac{(3+678)}{(4+678)} = 99.85\%\right)$ to 100%.

V. CONCLUDING REMARKS

In this paper, we propose ATA algorithm with closed-loop architecture in the SMT process to determine the best threshold setting. Our proposed threshold settings improve the accuracy from 99.85% to 100% in our simulated experiment. In addition, the ATA algorithm not only provides more accurate threshold settings than the threshold settings provided by traditional method but also plays an important role in automated manufacturing in the future.

REFERENCES

[1] Azamathulla, H. M., and Wu, F.C., “Support vector machine approach for longitudinal dispersion coefficients in natural streams,” Applied Soft Computing Journal, vol. 11, no. 2, 2011.
[2] Chang, C. C., and Lin, C. J., LIBSVM: A Library for Support Vector Machines,” (Online). Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm, 2001.
[3] Cortes, C., and Vapnik, V., “Support-vector networks,” Machine Learning, vol. 20, pp. 273-297, 1995.
[4] Ester, M., Kriegel, H. P., Sander, J., Xu, X., ”A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise”, 2nd International conference on Knowledge Discovery and Data Mining (KDD), 1996.
[5] Huang, J.C.Y., ”Reducing solder paste inspection in surfacemount assembly through Mahalanobis-Taguchi analysis,” IEEE Transactions on Electronics Packaging Manufacturing, vol. 33, no. 4, pp. 265–274, 2010.
[6] Pan, J., Tonkay, G., Storer, R., Sallade R., and Leandri, D., ”Critical variables of solder paste stencil printing for micro-BGA and fine-pitch QFP,” IEEE Transactions on Electronics Packaging Manufacturing, vol. 27, no. 2, pp. 125–132, 2004.
[7] Vapnik, V. N., “The Nature of Satistical Learning Theory,” Springer, New York, 1995.