Industry 4.0: Mining Physical Defects in Production of Surface-Mount Devices

Farshid Tavakolizadeh[1], José Ángel Carvajal Soto[1], Dávid Gyulai[2], and Christian Beecks[1]

[1] Fraunhofer Institute for Applied Information Technology
Schloss Birlinghoven, Sankt Augustin, Germany
{farshid.tavakolizadeh,angel.carvajal,christian.beecks}@fit.fraunhofer.de
[2] Institute for Computer Science and Control, MTA SZTAKI
Kende u. 13, Budapest, Hungary
david.gyulai@sztaki.mta.hu

Abstract. With the advent of Industry 4.0, production processes have been endowed with intelligent cyber-physical systems generating massive amounts of streaming sensor data. Internet of Things technologies have enabled capturing, managing, and processing production data at a large scale in order to utilize this data as an asset for the optimization of production processes. In this work, we focus on the automatic detection of physical defects in the production of surface-mount devices. We show how to build a classification model based on random forests that efficiently detects defect products with a high degree of precision. In fact, the results of our preliminary experimental analysis indicate that our approach is able to correctly determine defects in a simulated production environment of surface-mount devices with a MCC score of 0.96. We investigate the feasibility of utilizing this approach in realistic settings. We believe that our approach will help to advance the production of surface-mount devices.

Keywords: industry 4.0, surface-mount technology, data mining, classification

1 Introduction

A *surface-mount device* is an electronic device with components mounted on the surface of printed circuit boards (PCBs). The method for production of surface-mount devices is called the *surface-mount technology (SMT)*. This method consists of three steps: solder pasting, component pick and place, and solder reflow. SMT has enabled manufacturers to reduce the volume, weight, and cost of electronic devices.

The increasing demand in surface-mount devices has resulted in considerable advancement in SMT production systems [1, 2]. Yet, the generation of defects remains inevitable due to a number of factors such as machinery failure, bad parts, and imperfect production process. Common physical defects are missing solder paste and solder bridge in the solder printing step; missing and falsely oriented parts during pick and place phase; and faulty solder joints and component shift in the solder reflow step [1]. Occurrence of defects increases costs and deteriorates the SMT manufacturing process [3].

Defect detection of surface-mount devices is mostly attributed to physical properties. These properties can be inspected with a combination of vision-based and data mining methods. In this work, we present the SMT process through a realistic simulation. Considering the real-time constraints and unlikelihood of defects, we evaluate the effectiveness of random forests in mining defects. Furthermore, we study the feasibility of utilizing the algorithm in a closed production feedback loop to save manufacturing costs.

This paper is structured as follows. Section 2 outlines related work. Section 3 details the utilized production environment. We propose our classification approach in Section 4. The results of the performance analysis are given in Section 5 before the paper is concluded with an outlook on future work in Section 6.

2 Related Work

Several works focus on detecting surface-mount defects during and after the production. Kim et al. [4] present a system that extracts solder joints from images and trains neural networks classifying defects with 98% accuracy. Ko et al. [5] propose a X-ray system using linear vector quantization and multi-layer perceptions to detect defective *ball grid array* solder joints. Acciani et al. [1] extract regions of interests from images of surface-mount boards and use similar algorithms to detect solder joint defects with an accuracy of 98.8%. Tsai and Yang [6] present an eigen-based similarity measure for finding deviations from a reference image to detect flaws in PCBs and components placement. Hui and Pang [2] propose a method to detect solder paste defects by 3D reconstruction of solder profiles, geometric segmentation, and classification with full accuracy. In another work, authors present DMAIC [3] to improve the solder pasting process and reduce defects by 44%.

Maier et al. [7] propose a general purpose system that uses probabilistic deterministic timed automata to learn the topology of production components and their behavior. Subsequently, it detects anomalies by comparing new entries with the learned models. Another work [8] presents a distance-based classification method for anomaly detection in real-time after an offline training phase.

Previous works present systems suitable for detection of different kinds of defects on surface-mount devices. While [1, 2, 4–6, 8] show outstanding results, none offer a methodology that adapts to changes in product specifications. These systems typically rely on a training set acquired at an offline phase and assume that the training set applies to all new products. In addition, [1, 4–6] solely consider images taken from products and do not take into account data collected by other means (e.g. non-visual sensor measurements). On the other hand, [7] learns adaptively from production data and performs accurate defect detection. However, because the approach only models process information, it is unlikely

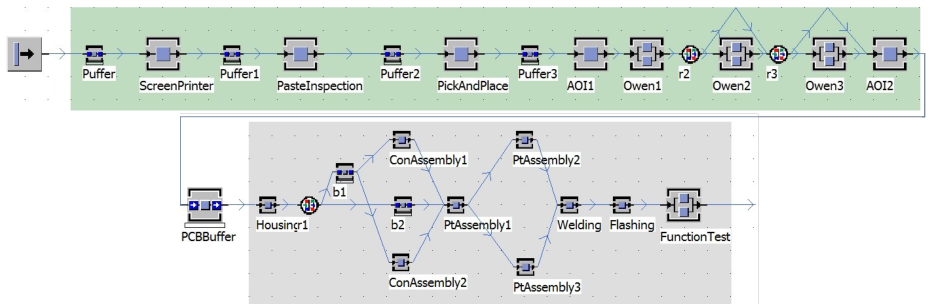

Fig. 1. Layout of the SMT (top) and Non-SMT (bottom) assembly lines.

that it will be successful in SMT classification. In production of surface-mount devices, most defects occur due to deviations in physical properties of devices [1, 2, 4–6]. In this paper, we focus on physical properties of surface-mount devices and present elements of an adaptive classification system.

3 SMT Production Environment

We created an environment using Siemens Tecnomatix Plant Simulation [9] that mimics the material flow and production processes performed in an actual SMT production line. The use of simulation enables us to understand the SMT process control without accessing and interfering with a running factory.

The environment consists of two assembly lines interconnected via an in-process buffer. The first line is fully automated, responsible for all steps of the SMT method. The second line performs operations related to assembly of machanical components (e.g. springs, screws, and connectors) and functional inspections. In this line, several stations are operated by humans and parallelized for an optimized throughput. Figure 1 shows the layout of both assembly lines. On the first line, production items are transported from one station to the other using a belt conveyor system, whereas they are conveyed manually in the second line. Furthermore, a buffering mechanism is in place between several stations to compensate for variations of the time needed for different operations. Table 1 describes the stations.

The environment manufactures six types of surface-mount devices $P_1, \ldots, P_6$ with similar product layouts. In this way, each product P_i is mapped to a mathematical feature vector $x_i \in \mathbb{R}^{53}$ as follows:

$$P_i \rightarrow x_i = (d_1, \ldots, d_{42}, \tau_1, \ldots, \tau_7, t_1, \ldots, t_3, \nu) \in \mathbb{R}^{53} \tag{1}$$

Each feature vector x_i comprises four types of measurements that are recorded for each product: the spatial coordinates of the solder joints and the surface-mount components $\{d_j\}_{j=1}^{42}$, the torques applied for tightening screws $\{\tau_j\}_{j=1}^{7}$, the temperatures in reflow ovens $\{t_j\}_{j=1}^{3}$, and the welding frequency for mounting the housing components ν. Moreover, each product is given a label, based

Table 1. Description and total measurements produced by each of the stations.

Station	**Description**	**M.**
ScreenPrinter	Pastes solder on designated areas of the PCB using a stencil	2
PasteInspection	Optically checks if solder correctly covers the solder pads	12
PickAndPlace	Picks components and places them on specific positions	4
AOI	Automated optical inspection: checks position of components	12
Oven	Reflows the solder paste so that components mount to the PCB	1
Housing	Placement of the surface-mount device in a plastic housing	1
ConAssembly	Mounting Non-SMT connectors on the PCB and/or housing	2
PtAssembly	Assembly of other components (e.g. springs, auxiliary modules)	2
Welding	Closing the lid of the housing with high-frequency welding	1
Flashing	Flashing firmware on the Chip(s)	0
FunctionTest	Performing final quality inspections on products	1

on the functional test performed at the end of the production. On average, 5% of products fail the functional test.

4 Random Forests Classifier

Random forests [10] is an ensemble learning method that works by constructing multiple decision trees during the training and taking an aggregation of their outputs for each classification task. The use of multiple decision trees makes random forests less prone to overfitting and more robust on imbalanced datasets.

We simulated the production of 30768 products of type P_1 and utilized [11] to aggregate sensor measurements during the production. For each generated product, we extracted a feature vector $x \in \mathbb{R}^{53}$ with a label and stored them together in a database. We performed simple dimensionality reduction to map identical measurements from parallel production station into one feature. This enabled us to make the dimension of feature vectors uniform regardless of the production process. All other numeric measurements were used directly as features, in total resulting in feature vector $x' \in \mathbb{R}^{51}$.

We tried various parametrization of the algorithm to find the optimal configurations for high accuracy and short training time. The result was an ensemble with 100 trees and max depth of 90 for each tree. In addition, as in [10] we set the maximum number of features used for training of each tree to $\lfloor log_2 M + 1 \rfloor$ where M is the dimension of x'. Other parameters such as minimum number of samples required to split an internal node and a leaf node were set to 2 and 1 respectively per [12]'s recommendation. During the training we allow the growth of a tree until all its nodes reach an impurity of below 10^{-7} or the tree exceeds the max depth threshold.

To evaluate the effectiveness of the ensemble with the suggested settings, we trained a model with 90% of the dataset selected at random. The model was trained in 26 seconds and achieved a Matthews correlation coefficient (MCC) score of 0.96 on the remaining 10% of the data. We chose MCC for evaluating the

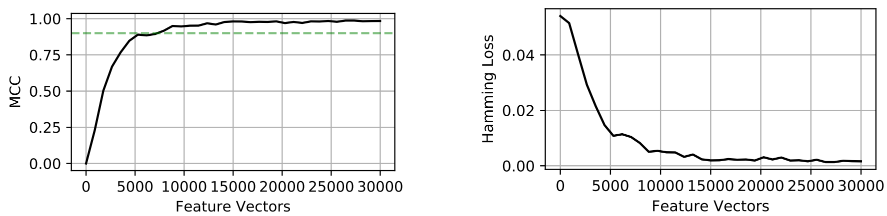

Fig. 2. The performance of the classifier with respect to total feature vectors available during the training.

performance because it considers the weight of classes and also reflects true/false positives and negatives. In a class-wise comparison, the classifier detects 94% of defects and wrongly classifies non-defects at a rate of 3×10^{-4}.

5 Experiments

We showed that a random forests model trained with a large dataset can accurately classify most defects. However, in realistic settings, it is not always feasible to manufacture that number of products in a reasonable amount of time. In another experiment, we continuously re-train the model over a growing dataset; see Figure 2. The MCC score is close to 0.9 when trained with about 5000 vectors and well above 0.9 after 8000 vectors. The fraction of falsely classified items (Hamming Loss) goes below 0.01 with more than 7500 vectors.

During the production, the items go through subsequent stations responsible for different tasks. A classification mechanism is practical only if it is able to detect a defect in a short time window after leaving a station and before it is placed on a track toward the next station. We closely monitored the production process and realized that this time window is at least 6.57 seconds. On average, the classification using the presented approach takes 110 milliseconds to complete (Powered by Intel W3565 @3.20GHz and 8GB of RAM). This leaves us with a remaining 6.46 seconds for transfer of data from and back to the production line.

The suggested classifier achieves an Accuracy (ACC) of 98.4% when trained with a balanced dataset of 2000 feature vectors. We balanced the data because ACC measure does not consider the weight of classes but it is required for comparison with the related work.

6 Conclusions and Future Work

In this work, we defined and presented a realistic production process. We simulated the process with state-of-the-art technologies, creating a data mining testbed. We showed that it is feasible to train accurate models in such an environment while fulfilling real-time constraints. Altogether, those parts constitute the building blocks for constructing and proving the feasibility of an adaptable stream mining framework applying state-of-the-art machine leaning techniques in real-time and online. To do so, the following future work is foreseen:

1. Integrating the classification system with the production environment using existing protocols and standards.
2. Collecting the data in an efficient and scalable manner.
3. Providing an adaptable framework that fits into Industry 4.0 needs.
4. Investigating the application of interactive machine learning techniques [13].
5. Offering a flexible, integrated, and manageable solution to data scientists allowing them to work in industrial settings effectively and realistically.
6. Evaluating the benefits of such systems from a manufacturing perspective.
7. Finally, validating the approach in real-world use cases.

Acknowledgements. This work is supported by the European Commission through the H2020 projects EXCELL and COMPOSITION under grant no. 691829 and 723145 respectively.

References

1. Acciani, G., Brunetti, G., Fornarelli, G.: Application of neural networks in optical inspection and classification of solder joints in surface mount technology. IEEE Trans. on Industrial Informatics. 2(3), 200–209 (2006)
2. Hui, T., Pang, G.: Solder paste inspection using region-based defect detection. The International J. of Advanced Manufacturing Technology. 42(7-8), 725–734 (2009)
3. Li, M., Al-Refaie, A., Yang, C.: DMAIC approach to improve the capability of SMT solder printing process. IEEE Trans. on Electronics Packaging Manufacturing. 31(2), 126–133, (2008)
4. Kim, T., Cho, T., Moon, Y., Park, S.: Visual inspection system for the classification of solder joints. J. Pattern Recognition. 32(4), 565–575 (1999)
5. Ko, K., Roh, Y., Cho, H., Kimn, H.: A neural network approach to the inspection of ball grid array solder joints on printed circuit boards. IJCNN. 5, 233–238 (2000)
6. Tsai, D., Yang, R.: An eigenvalue-based similarity measure and its application in defect detection. Image and Vision Computing. 23(12), 1094–1101 (2005)
7. Maier, A., Vodencarevic, A., Niggemann, O., Just, R., Jaeger, M.: Anomaly detection in production plants using timed automata. ICINCO. 363–369 (2011)
8. Toth, D., Aach, T.: Improved minimum distance classification with Gaussian outlier detection for industrial inspection. Image Analysis and Processing. 584–588 (2001)
9. Siemens PLM: Plant Simulation. `https://www.plm.automation.siemens.com/en/about_us/index.shtml` (Visited on 2016-12-10)
10. Breiman, L.: Random forests. J. Machine learning. 45(1), 5–32 (2001)
11. Soto, J.A.C., Jentsch, M., Preuveneers, D., Ilie-Zudor, E.: CEML: Mixing and moving complex event processing and machine learning to the edge of the network for IoT applications. IoT 2016. 103–110 (2016)
12. Louppe, G., Holt, B., Arnaud, J., Hedayati, H.: Random Forests. Scikit-learn: Machine learning in Python. `http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html` (Visited on 2017-03-15)
13. Saket, B., Kim, H., Brown, E. T., Endert, A.: Visualization by Demonstration: An Interaction Paradigm for Visual Data Exploration. IEEE Trans. on Visualization and Computer Graphics. 23(1), 331-340 (2017)