Semiconductor Defect Classification

Terence Sweeney, Sonya Coleman, Dermot Kerr
School of Computing, Engineering and Intelligent Systems,
Ulster University
Londonderry, Northern Ireland
Email: Sweeney-t4@ulster.ac.uk, sa.coleman@ulster.ac.uk, d.kerr@ulster.ac.uk

***Abstract*—Automated inspection has become a vital part of quality control during semiconductor wafer production. Current processes are focussed on finding defects via variation from a 'golden' image using pixel to pixel comparisons or utilization of opaque neural network-based approaches. We present a novel approach, which uses the Bag of Visual Words technique to determine local features that correspond to specific defects within a wafer image, known as a custom vocabulary, as a way to begin creation of a more transparent system for automated defect detection and classification. We demonstrate that the custom vocabularies, combined with machine learning algorithms, result in high performance accuracies with efficient computational run-times.**

Keywords*— *Defect Detection; Defect Classification; Bag of Visual Words; Local Features; Semiconductor wafers; Image Processing.

I. Introduction

Semiconductor wafers are a component used in products such as processors and hard drive media. Inspection is vital during the manufacturing process in order to detect defects and ensure quality control. Several methods have been proposed for defect detection on semiconductor wafers, however the majority of techniques focus on defect detection across the wafer as a whole. When defects are detected they are marked on the wafer bin map in order to identify the total number found. This is a useful approach when looking for systematic defects across a product line and removing a defective product earlier in the production line. However, it is sometimes desirable to detect not only the location of a defect but also the type of defect as some of the product may still be commercially viable. The goal of this research is to use images of a single chip on the wafer, known as die images, to detect and classify defects. An example of a wafer bin map and a die image is presented in Figure 1.

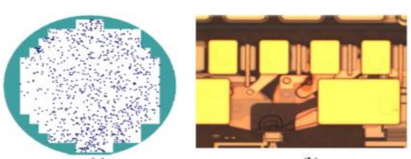


Figure 1. (a) Wafer bin map with detected defects coloured blue and (b) Single die image

Production of semiconductor wafers involves multiple stages and many different components are used during this process. Due to the varying size and criticality of these components, many different inspection techniques are used throughout manufacturing to ensure quality control. Inspection techniques include using electrical input and microwave testing along with optical cameras that can inspect to pico-meter level. The difference in these types of inspection systems has resulted in many interpretations of how to best detect and classify defects [10][21]. One widely used approach is to observe the overall frequency and location of defects using the wafer map in order to detect systematic or widespread damage over the complete wafer, such as a scratch or tear, as shown in Figure 2. Whilst this solution [9][11]has been proven to be useful for finding systematic or clustered defects across the whole wafer, it does not consider the type of defect, and consequently whether the product is still viable.

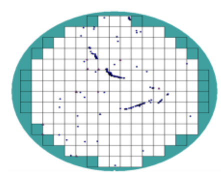


Figure 2. Example of a Scratch defect using the wafer bin map

When considering automated visual inspection of semiconductor wafers, die images are used and examples of defects upon these die images are given in Figure 3. Most previous work is based on the use of global features with Tobin's content-based image retrieval golden image comparison method [25] being the most popular. This inspection representation is commonly found in most Automated Defect Classification (ADC) machines [1][21]. However, other methods have been used to detect specific types of defects across the industry. Chou [6] uses the Hough transformation to detect scratches or gouges on a wafer surface while Park [23] detailed an approach using the Histogram of Gradient (HOG) operators to great effect. However, there has been little work on the use of local image features for automated inspection using techniques, such as Sobel [14], Scale Invariant Feature Transform (SIFT) [8], Oriented FAST and Rotated BREIF (ORB) [13], and SURF [2]. In [16], we proposed the use of SIFT and SURF local image features for wafer defect detection and concluded that whilst both techniques could identify wafer defects, the use of SURF resulted in improved detection accuracy.

Copyright (c) IARIA, 2020. ISBN: 978-1-61208-815-0

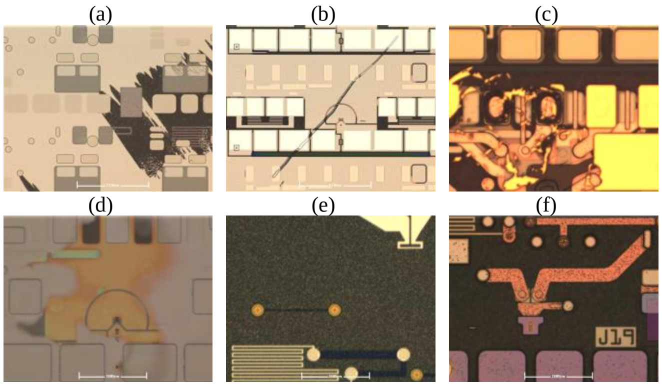


Figure 3. Examples of defects present on semiconductor die images: (a) Rip, (b) Scratch, (c) Warp, (d) Delamination, (e) Incomplete liftoff, (f) Corrosion

In addition to detecting defects, defect classification is also necessary. In some cases, products that have been identified with non-critical defects, which would otherwise be removed due to the presence of a defect, can continue on the production line. Additionally, identifying the type of defect and the stage at which it occurs in the production process can help in improving overall quality, yield and production processes. The majority of defect classification approaches to date focus on the use of neural networks and deep learning. For example, Reza [12] used an artificial neural network with a back-propagation algorithm to observe contamination defects on wafers, a method also applied by Chou [6]. The work in [11][19][20][21] uses Convolutional neural networks for classification. While these deep learning methods return good results, the black box nature of neural networks can be a problem in industry, such as semiconductor manufacturers, since the designs are frequently updated and changed and although the neural network could be trained to work well with current designs, new designs could cause system failure as we are currently seeing in domains such as self-driving cars [26] and image recognition [27]. Thus, we have developed a novel approach based on transparent local features to create a more understandable system.

A well-known feature extraction technique is the Bag of Visual Words (BoVW) method, which extends the Bag of Words (BoW) method from the text retrieval domain to the visual classification domain and can be used as an alternative to global image features. When using the BoW technique on a text document, a normalized histogram of word counts is computed as well as a sparse term vector where each bin corresponds to a term in the vocabulary. The BoVW technique [29] enables the generalisation of local image feature descriptors in a similar manner and has been used for image classification [3][15][17]. Improving further on BoVW, a custom vocabulary [20] or codebook is a concept in which specific subsets of visual words are selected, which represent the most important features of the images, rather than using the complete vocabulary created from a set of training images. One example of this approach is the dual vocabulary approach [17] where two vocabularies are trained on different training set classes before being run on its testing data in order to observe, which returns the highest accuracy for each testing class and therefore which features are most important for detection and classification of these classes. Custom Vocabularies take this a step further by observing which visual words contain the most important information for a given task and utilize only these visual words in order to increase overall accuracy and also reduce overall computation time.

It is also possible to combine BoVW with machine learning classifiers. For example, Hentschel [4] evaluated several different classification methods such as AdaBoost [5], Support Vector Machines (SVM) and decision trees on an image classification problem utilizing BoVW and found several methods that achieve high accuracy when combined with local feature methods. Two popular image classification approaches that are widely used across many different fields of automated visual inspection are multi-class SVM [7][24] and Random forest [22].

Building on previous work [16], this paper proposes a novel approach to defect detection and classification in semiconductor wafers. We identify specific visual words that correspond to a defect descriptor, *a custom vocabulary*, and use these for classifying a defect within an image as close to real time speed as possible, whilst still retaining high levels of accuracy. The remainder of this paper is organized as follows: Section 2 introduces the current industry inspection process used by our industry partner and its problems. Section 3 covers the proposed *custom vocabulary* and Section 4 discusses the performance evaluation of the approach. Finally, Section 5 details the conclusion and further work.

II. CURRENT INDUSTRY INSPECTION PROCESS

There are around 600 stages in the production of a single semiconductor wafer. In order for the wafer to fully function it needs to be kept free of defects which can be caused in many ways, including particle damage, atmospheric changes as well as human- and machine-error. Thus, a typical semiconductor production line will have many in-line inspection tools at various manufacturing stages in order to ensure quality control. Due to the size of critical parts on the wafer, some as small as 7nm, specialised inspection equipment must be used. The inspection process can be conducted in various ways, for example using electrical fault detection and x-rays, however the most time-affordable systems are visual inspection systems.

Current industry practice for defect detection and classification is a global image matching approach where a direct pixel-to-pixel comparison is performed using a database of control images which are directly compared with the current product passing through the inspection system. This is commonly known in the industry as a 'golden image' approach. In order to prevent false detection of defects, a defect reduction factor is used where pixel intensities within a 3x3 pixel neighborhood are compared before any area is regarded as a defect.

Copyright (c) IARIA, 2020. ISBN: 978-1-61208-815-0

The Rudolph NSX105 [1] is a commonly used industry standard inspection device which uses the golden image approach. The NSX105 inspection system uses its initial stage camera to strobe over the wafer comparing captured images with the corresponding database of golden images. The golden images in the database are initially manually pre-programmed. Hence, when a product is developed or updated, a new set of golden images must be created. If a defect is detected, its coordinates are saved into a reference file and then additional high-resolution images of the defect on the die are captured using a second inspection camera for subsequent manual inspection. The number of defects for which high resolution images are captured is capped at a level according to parameters set manually, typically 80 images per wafer. A critical problem with the NSX105's inspection detection is that while it can determine a problem at a specific location, it cannot determine the type of defect that has been found on the die. Hence the severity of the defect is unknown, and this may result in more serious defect types, such as corrosion damage on critical parts, going unnoticed until later in production.

We seek to improve on this by developing an automated inspection system, which uses the existing inspection equipment output, and is focussed specifically on classifying high resolution defect images from the die rather than the defect identification stage which creates the wafer bin map.

III. Custom Vocabulary

In the proposed methodology, the SURF interest point detector is used to obtain key-points k_n and corresponding SURF descriptors d_n where $i = 1 \ldots n$ such that a keypoint is represented as:

$$k_i = (x_i, y_i, d_i) \quad (1)$$

where x and y are the coordinates of a point in an image. The SURF keypoint descriptors are of 64 dimensions. An image feature set S can be represented by the set of local keypoint descriptors such that

$$S_I = \{k_1, k_2, \ldots, k_n\} \quad (2)$$

where $I = 1 \ldots m$ and m is the number of images in the image set. The BoVW algorithm B is considered to quantize the descriptor $d \in R^I$

$$B: R^I \to [1, K] d \to B(d). \quad (3)$$

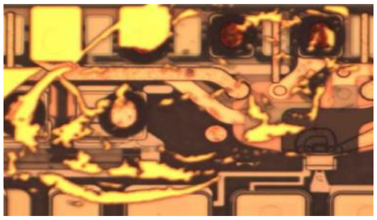
The B assigns descriptor $d \in R^I$ to the appropriate cluster K, where each cluster represents a visual word and the set of visual words is the initial defect vocabulary.

We can further refine the initial vocabulary to form a *custom vocabulary* through manual inspection of the defect images where only visual words that represent wafer defect features are retained and that is the approach used here.

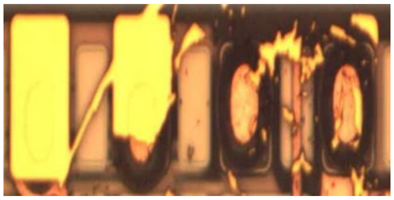
IV. Performance Evaluation

There are various defects that can occur in semi-conductor wafers, such as splatter, warp, scratch, rip, delamination and corrosion. To evaluate the proposed approach for defect detection, we focus on the warp defect. The warp defect occurs for various reasons including temperature changes, rise in atmospheric pressure, or human and machine error. Its main feature is that parts of the golden resist (or paint), also called the gold pad, are removed or warped in some way. Examples of warp die images are presented in Figure 4 where Figure 4(a) illustrates the gold resist in various stages of damage from the warp defect and Figure 4(b) illustrates complete removal of the resist. All images are captured by the Rudolph NSX105 from one layer of one product, and all images are 648x494 pixels. All experiments are run using Python OpenCV and Sklearn on an Intel Xeon CPU E5-120 0@ 3.60 GHZ with 16 GB of RAM.

In the initial experiment, we evaluate the proposed approach using a vocabulary of 1000 visual words and various well-known machine learning algorithms including AdaBoost, Random Forest, Support Vector Machines (SVM) with a range of kernel functions (Linear, Polynomial and Radial). We use sets of warp images (Figure 4) and control images (Figure 5). Both the warp and control classes contain 100 images each (200 in total), split 80/20 for training and testing. The machine learning algorithms have been optimised via a grid search and a summary of the results is displayed in Table I. Using 1000 visual words, the results vary across the different machine learning approaches with the SVM using a Radial Basis Function (RBF) and C=10 providing the highest accuracy.



(a) Gold resist in various stages of damage



(b) Complete removal of the resist

Figure 4. Examples of the warp defect

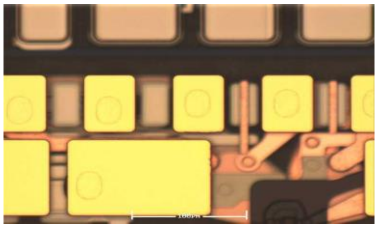


Figure 5. Control image

While the accuracy results are promising, the system takes significant time to process 1000 visual words, with the training alone taking 7 minutes and 37 seconds. Additionally, although we can determine from the BoVW histogram which visual words occur most often for each class, it is not possible to determine what initial local features make up each visual word. From the 1000 visual words, it was possible to isolate 106 visual words that corresponded solely to the warp defect, with no key points detected on the image background. The experiments were conducted again using this refined set of visual words, a *custom vocabulary*, and the results are presented in the last column of Table I. In this scenario, several machine learning approaches, combined with the *custom vocabulary*, provide an accuracy of 97%, hence the use of a custom vocabulary is more consistent and less dependent on the machine learning algorithm it is combined with, and the training time using this vocabulary is also much closer to a real time system, taking only 28 seconds, approximately 15x faster than using 1000 words.

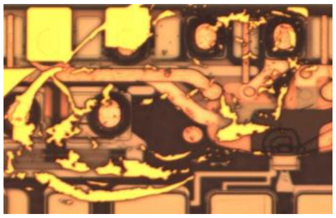
TABLE I– EXPERIMENTAL RESULT

	Accuracy **1000 visual words**	**Accuracy** **106 visual words (custom vocabulary)**
AdaBoost	**95%**	87%
Random Forest	**79%**	75%
SVM - Linear C=1	51%	**51%**
SVM – Linear C=10	53%	**75%**
SVM – Linear C=100	90%	**97%**
SVM – Poly C=1	**51%**	50%
SVM – Poly C=10	51%	**80%**
SVM – Poly C=100	56%	**97%**
SVM – RBF C=1	56%	**65%**
SVM – RBF C=10	**100%**	97%
SVM – RBF C=100	97%	**97%**

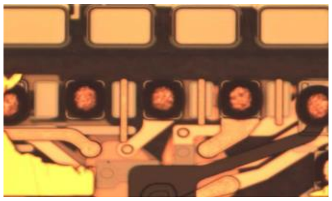
The ability to identify a defect in a die image is important in automated inspection, and it is possible to further define the warp defect into 3 sub-classes. This has important consequences as some sub-classes of warp defect have more impact on the wafer production than others. The first sub-class denoted as Warp 1 contains erratic shapes and sharp-edged resist pieces that appear across the wafer image. The second sub-class, denoted Warp 2, focusses on the circles that appear as the resist is wiped away from the wafer. The third sub-class, denoted as Warp 3, has circular blobs or scratches through the resist. An example of each warp sub-class is illustrated in Figure 6.

Using the custom vocabulary of 106 visual words, we create a new custom vocabulary for each sub-class where Warp 1 requires 63 visual words, Warp 2 requires 48 visual words and Warp 3 requires 29 visual words. The custom vocabulary for Warp 1 contains the most unique visual words whereas the custom vocabulary for Warp 2 has overlap with both Warp 1 and Warp 3. The experiments were conducted again using the custom vocabularies. As the results in Table I demonstrated that AdaBoost and Random Forest do not perform as well as SVM, we present results only for SVM in Table II.

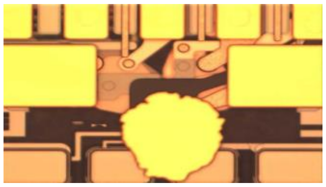
As shown in Table II, the linear SVM performs similar to the results presented in Table I, and hence it remains the worst performing SVM. The polynomial kernel SVM has increased accuracy compared with the linear SVM, however the RBF kernel SVM retains the highest accuracy for all SVMs across the three sub-classes. The key significance of the results in Table II is that the classification accuracy is high with an improvement in computational efficiency due to the reduced feature set, *the custom vocabulary*.



(a) Example of Warp 1 image



(b) Example of Warp 2 image



(c) Example of Warp 3 image

Figure 6. Examples of warp sub-classes

TABLE II - SVM ACCURACY RESULTS

SVM			
	Warp 1	**Warp 2**	**Warp 3**
SVM Linear C-1	51%	51%	51%
SVM Linear C-10	92%	85%	68%
SVM Linear C-100	97%	97%	87%
SVM Poly C-1	70%	73%	78%
SVM Poly C-10	85%	95%	87%
SVM PolyC-100	**100%**	97%	70%
SVM RBF C-1	87%	90%	70%
SVM RBF C-10	97%	**100%**	92%
SVM RBF C-100	**100%**	97%	95%

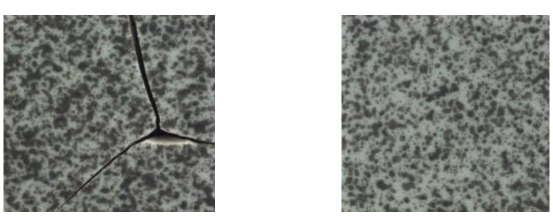
TABLE III - SPEED TEST RESULTS

Computational speed Test	Training Time	Prediction Time	Highest Classification Accuracy
1000 Words	7m 37s	13s	100%
106 words	28s	9s	97%
Subclass	**25s**	**6s**	**100%**

Another important consideration is the speed of this system, as it is required to operate with in-line inspection tools and should therefore be as close to real time as possible whilst still retaining a high degree of accuracy. Table III shows that the proposed approach, based on the custom vocabulary, achieves the fastest run-time compared with the use of a larger vocabulary, as well as high accuracy.

V. MVTEC EVALUATION

The results presentenced in the previous section demonstrate that the custom vocabulary that corresponds to a specific defect provides high classification accuracies. In order to further validate this system, we use the MVTEC anomaly detection dataset [28]. From the dataset, we selected the Tile Crack image set which contains 20 images, 10 for training and 10 for testing, along with a control class, again using 10 for training and 10 for testing. Examples of these images are given in Figure 7 .



(a) (b)

Figure 7. Examples of (a) Tile Crack Defect image and (b) Tile Control Image

In line with the previous experiment, as the SVM performed best, we use only an SVM with all 1000 visual words and the defect only visual words, for which 69 were detected for this dataset.

TABLE IV – TILE CRACK RESULTS

SVM		
	Full 1000 Words	**69 Defect only Words (Custom Vocabulary)**
SVM Linear C-1	72%	80%
SVM Linear C-10	72%	80%
SVM Linear C-100	72%	80%
SVM Poly C-1	72%	80%
SVM Poly C-10	72%	80%
SVM Poly C-100	72%	**97%**
SVM RBF C-1	72%	80%
SVM RBF C-10	72%	80%
SVM RBF C-100	82%	**97%**

As illustrated in Table IV, a maximum accuracy for this dataset, when using 1000 visual words was 82% using an SVM, with the RBF kernel and C=100. However, this is reproved significantly by using a custom vocabulary that corresponds to the defect only features present in the images. We can see an increase to 97% using both the polynomial and RBF kernels with C=100. This is excellent performance accuracy given the small dataset used and would be difficult to achieve using deep learning which requires a significant volume of data. This demonstrates the robustness of the proposed approach across industrial datasets.

VI. CONCLUSION AND FURTHER WORK

We have presented an approach to semi-conductor wafer defect classification by utilizing the bag of visual words method with a *custom vocabulary* formed from a reduced set of visual words. We have demonstrated that this novel approach achieves competitive accuracies when compared with the use of a larger set of visual words (1000) but is much more computationally efficient as demonstrated by the presented run-times.

As the proposed approach works well, both on our industrial dataset and the MVTEC anomaly dataset, future work will investigate the design of custom vocabularies for other defect types, namely splatter, scratch, rip, delamination, and corrosion. Additionally, we will explore the ability to accurately characterise and hence classify the warp defect images using only the custom vocabulary without additional machine learning. The motivation for this is that, within the production line, if there is a design change then a neural network focused automated inspection system will require retraining. However, if we can accurately classify defects without the use of deep learning and by using the custom vocabulary approach, this will enable the system to be readily adaptable to product

Copyright (c) IARIA, 2020. ISBN: 978-1-61208-815-0

changes and developments creating a more open and understandable system.

ACKNOWLEGMENT

This work was funded by a DfE CAST scholarship in collaboration with Seagate Technology. We would also like to thank Seagate Technology for providing the image dataset used in the research.

REFERENCES

[1] Tarasemi, "Rudolph August NSX 105 Automated Defect Inspection," 2019.[Online].Available: https://www.tarasemi.com/product/rudolph-august-nsx-105-automated-defect-inspection/. [Accessed 2 September 2020].

[2] H. Bay, "Surf: Speeded up robust features," in *European Conference on computer vision, pp. 404-417*, Berlin, 2006.

[3] G. Csurka, "Visual categorization with bags of keypoints," in *Workshop on Statistical learning in computer vision, pp. 1-2*, 2004.

[4] C. Hentschel and H. Sack, "Does one size really fit all? Evaluating classifiers in Bag-of-Visual-Words classification," in *14th International Conference on Knowledge Technologies and Data-driven Business.*, pp. 1-8, Graz, Austria, 2014.

[5] T. Hastie, S. Rosset, J. Zhu and H. Zou, "Multi-class AdaBoost" *Statistics and its Interface,* pp. 349-360, 2009.

[6] C. J. Huang, C. F. Wu and C. C. Wang, "Image processing techniques for wafer defect cluster identification," *IEEE Design & Test of Computers,* pp. 44-48, 2002.

[7] C. Li-Chang and T. Lee-Ing, "Wafer defect pattern recognition by multi-class support vector machines," *Expert Systems with Applications,* pp. 10158-10167, 2009.

[8] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision,* pp. 91-110, 2004.

[9] D. P. Mital and E. K. Teoh, "Computer based wafer inspection system," in *Proceeding of international conference on industrial electronics, control and instrumentation*, Kobe, 1991.

[10] N.G. Shankar and Z.W. Zhong, "Defect detection on semiconductor wafer surfaces," *Microelectronic Engineering,* vol. 77, no. 3, pp. 337-346, 2005.

[11] M. P.-l. Ooi, "Defect cluster recognition system for fabricated semiconductor wafers," *Engineering Applications of Artificial Intelligence,* pp. 1029-1043, 2013.

[12] A. Z. Reza, T. Hisashi, T. Y. S. Shinichi , and S. Kazuma, "Automated inspection of IC wafer contamination," *Pattern Recognition,* pp. 1307-1317, 2001.

[13] E. Rublee, V.Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," *ICCV,* p. 8, 2011.

[14] I. Sobel and G. Feldman, "A 3x3 Isotropic Gradient Operator for Image Processing"," in *Stanford Artificial Intelligence Project*, Stanford, 1968.

[15] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," in *Proceedings Ninth IEEE International Conference on Computer Vision*, Nice, 2003.

[16] T. Sweeney, S. Coleman, and D. Kerr, "A Machine Learning Approach to Wafer Defect Classification using Bag of Visual Words," in *Irish Machine Vision and Image Processing Conference. 2019*, Dublin, 2019.

[17] K. S. Sujatha, P. Keerthana, S. Suga Priya, and E. Kaavya "Fuzzy based multiple dictionary bag of words for image classification," *Procedia Engineering,* pp. 2196-2206, 2012.

[18] F.-C. Tien, "Reclaim wafer defect classification using large scale BPNS with Tensorflow," *National Academy of Managerial Staff of Culture and Arts Herald,* pp. 641-644, 2018.

[19] S. Varsha Devi et al., "Better object recognition using bag of visual word model with compact vocabulary," in *13th International Conference on Emerging Technologies (ICET),* Islamabad, 2017.

[20] S. Zeng, S. Dai, and P. Mu, "Wafer Defects Detecting and Classifying System Based on Machine Vision," in *8th International Conference on Electronic Measurement and Instruments*, Xi'an, 2007.

[21] P. Chou et al., "Automatic defect classification for semiconductor manufacturing," *Machine Vision and Applications,* vol. 9, no. 4, pp. 201-214, 1997.

[22] L. Puggini, J. Doyle, and S. McLoone, "Fault Detection using Random Forest Similarity Distance," in *9th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes,* Paris, 2015.

[23] J.-K. Park et al., "Machine Learning-Based Imaging System for Surface Defect Inspection," *International Journal of Precision Engineering and Manufacturing Green Technology,* vol. 3, no. 3, pp. 303-310, 2016.

[24] C. Pu et al., "Recognition and classification of coating film defects on automobile body based on image processing," in *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics*, Shanghai, 2017.

[25] K. Tobin, T. Karnowski, and F. Lakhani, "Integrated applications of inspection data in the semiconductor manufacturing environment," *Metrology-based Control for Micro-Manufacturing,* pp. 31-41, 2001.

[26] K. Eykholt, et al. "Robust physical-world attacks on deep learning visual classification." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

[27] N. Anh, J. Yosinski, and J. Clune. "Deep neural networks are easily fooled: High confidence predictions for unrecognizable images." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

[28] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger. MVTec AD - A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection; in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2019.

[29] L. Fei-Fei and P. Perona, "A Bayesian hierarchical model for learning natural scene categories," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* San Diego, 2005.