Research on Image-Based Automatic Wafer Surface Defect Detection Algorithm

Yefan Zhou
School of Information Science and Engineering, Southeast University, Nanjing, China
Email: yefan__zhou@126.com

***Abstract*—Defects on the wafer surface usually reflect the abnormal problems in the semiconductor manufacturing process. By detecting and identifying the wafer surface defect mode, it can diagnose fault source in time and adjust online. In this paper, an online detection and adaptive recognition model for wafer surface defect mode is presented. Firstly, the model is used to extract the feature of the wafer surface defect mode. The Hidden Markov Model (HMM) is constructed for each wafer mode based on the feature set, and an on-line detection and recognition method based on HMM dynamic integration is proposed. The proposed model is successfully applied to the wafer defect detection and recognition in WM-811K database. The experimental results fully demonstrate the validity and practicability of the model.**

***Index Terms*—wafer, defect, automatic detection, mode recognition**

I. INTRODUCTION

Semiconductor manufacturing is a very complex dynamic manufacturing process, and wafer manufacturing is the most critical part. Any abnormality in any process of wafer production may lead to the occurrence of wafer defects [1], [2]. Therefore, the accurate recognition of the wafer defect mode can not only avoid the cost loss caused by product defects, but also assist the identification of abnormal sources in the manufacturing process, improve the production efficiency and product quality, and further improve the market competitiveness of enterprises.

Some statistical methods have been applied to the wafer defect recognition successively. For example, Hwang [3], Yuan [4] and Wang [5] respectively proposed different probability models to describe the wafer defect mode. Hwang [3] used two-step algorithm, and then used binary normal distribution and principal curve to describe the defect cluster. By comparing the size of the likelihood probability value of each cluster in the two distributions, he determined whether the shape of the cluster was a curve. Yuan [4] improved the two-step algorithm, and proposed a mixed model based on Poisson distribution, normal distribution and main curve, which could detect the curve shape, line character and ellipsoidal defect mode at the same time. Wang [5] introduced the spherical shell algorithm [6] to distinguish circular defect mode. Although these methods can detect the defect shape on the wafer surface, they cannot well identify the specific defect mode [7]. In order to accurately predict the defect mode, some data mining methods are applied to the wafer defect recognition, such as neural network [8], clustering [9], supervised, machine learning (such as K-NearestNeighbor [10], KNN), Support Vector Machine [10], [11], SVM)). These methods often require large sets of training data. Although they are good at recognizing defect modes, new defect modes appear on the production line because semiconductor manufacturing is a complex and dynamic process. Many classifiers cannot solve the problem of detection and recognition of new defect modes because they have fixed recognition models in the off-line learning stage.

In this paper, an on-line detection and adaptive recognition model of wafer surface defect mode is proposed to improve the accuracy. The model firstly extracts the features of defect mode based on the segmented defect image, converts the high-dimensional defect data based on pixels into low-dimensional defect feature data, and builds the feature set corresponding to each mode. Then, a specific Hidden Markov Model (HMM) is constructed for the online defect detection and recognition of each wafer mode. At the same time, the dynamic integrated recognition model based on HMM proposed in this paper can detect the new defect mode generated on the line, build the new HMM recognition model by collecting the new defect mode, update the HMM recognition library, and identify the new defect mode in the subsequent production and manufacturing process. Experimental results show that the model presented in this paper can be effectively applied to the detection and recognition of wafer defects in the actual manufacturing process.

II. WAFER DEFECT DETECTION AND RECOGNITION BASED ON HMM DYNAMIC INTEGRATION

A. The Basic Concept of HMM

1) Definition of HMM

HMM is a double random process. In this random process, the state inside the random process cannot be directly observed, only the sequence of observed values can be observed, and the existence and characteristics of the state can be observed through this double random process [12], [13]. HMM is represented by a five-tuple flow ((N, M, $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{\pi}$), where:

Manuscript received January 10, 2019; revised March 17, 2019.

doi: 10.18178/joig.7.1.26-31

(1) Set $\theta_1, \theta_2, \ldots, \theta_N$ as N state of Markov chain, where the state of time t is recorded as q_t ($q_t \in (\theta_1, \theta_2, \ldots, \theta_N)$).

(2) M: Corresponding the observation number of each state, the observation value are respectively $V_1, V_2, \ldots$ and V_M, Set the observation value at time t is O_t ($O_t \in (V_1, V_2, \ldots, V_M)$).

(3) $\boldsymbol{A} = (a_{ij})_{N\times N}$: State transition probability matrix, where $a_{ij} = P(q_{t+1} = \theta_j | q_t = \theta_i)$, $1 \le i, j \le N$.

(4) $\boldsymbol{B} = (b_{jk})_{N\times N}$: Observation probability matrix, where $b_{jk} = P(O_t = V_k | q_t = \theta_i)$, $1 \le j \le N$, $1 \le k \le N$.

(5) $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_N)$: Initial state probability vector, where $\pi_i = P(q_t = \theta_i)$, $1 \le i \le N$.

2) Three basic problems of HMM

(1) Probability calculation problem: Given model λ (N, M, $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{\pi}$), and observation sequence $O(o_1, o_2, \ldots, o_T)$, calculate the probability $P(O|\lambda)$ of O under model λ.

(2) Learning problem: How do we adjust model parameter λ (N, M, $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{\pi}$) to maximize $P(O|\lambda)$).

(3) Prediction problem: Given model λ (N, M, $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{\pi}$), and observation sequence O ($o_1, o_2, \ldots, o_T$), calculate the maximum state sequence $I = (i_1, i_2, \ldots, i_T)$ [12], [13] of conditional probability $P(I|O)$.

Problem (1) solution: use forward and backward algorithm to calculate $P(O|\lambda)$.

The forward algorithm is described here:

Define partial observation sequence at time t as o_1, o_2, …, o_t and the probability of state q_t is forward probability as:

$$\alpha_t(i) = P(o_1, o_2, \ldots, o_t, i_t = q_t | \lambda) \tag{1}$$

(1) Initial value:

$$\alpha_t(i) = \pi_i b_i(o_1),\ i = 1, 2, \ldots, N \tag{2}$$

(2) Make recurrence to t=1,2,…,T - 1,

$$\alpha_{t+1}(i) = \left[\sum_{j=1}^{N} \alpha_t(j) a_{ji}\right] b_i(o_{t+1}),\ i = 1, 2, \ldots, N \tag{3}$$

(3) Terminate

$$P(O|\lambda) = \sum_{t=1}^{N} \alpha_t(i) \tag{4}$$

Problem (2) solution: use Baum-Welch model to train algorithm, then determine the largest model λ of $P(O|\lambda)$.

Problem (3) solution: use Viterbi [14] algorithm to calculate the largest state sequence $I = (i_1, i_2, \ldots, i_T)$ of conditional probability $P(I|O)$.

3) Defect detection and recognition method based on HMM

Firstly, the HMM recognition model is constructed for each wafer mode (*HMM0* is the hidden Markov model of normal wafer mode, and the hidden Markov model corresponding to n defective wafer modes is *HMM1~HMMn*). For the input wafer characteristic data x_t, HMM corresponds to the output P (x_t|λ). The output of HMM is Negative Log Likelihood Probability (NLLP), as shown in formula (5). In order to detect whether there are defects on the wafer surface, this paper adopts the monitoring method based on NLLP control chart, taking the output NLLP of each wafer mode in corresponding HMM as the monitoring variable, setting the control rate at 99.73%, and confirming the Upper control limit *UCL0 ~UCLn* on the control chart.

$$NLLP = -\ln P(x_t | \lambda) \tag{5}$$

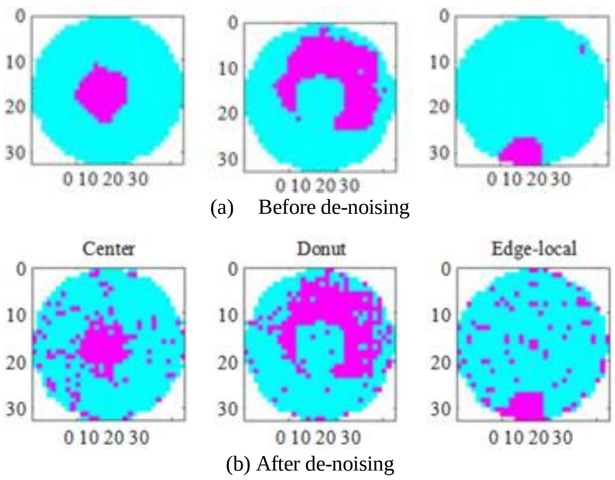
Defect detection: input the collected wafer characteristic data into HMM_0. If $NLLP_0 > UCL_0$, the defective wafer will be determined. The defective wafers are input into $HMM_1 \sim HMM_n$, respectively. If for all recognition models, it satisfies $NLLP_i > UCL_i$ (i=1, 2, …, n), the defect wafer is determined as a new defect mode, HMM_{new} is constructed, and the HMM recognition library is updated.

Defect pattern recognition: input defect mode data into HMM1~HMMn, and select class c* corresponding to the minimum value of NLLP as the discrimination result, as shown in formula (6).

$$c^* = \arg\min_{1 \le c \le n} [-\ln(P(x_t | \lambda_c))] \tag{6}$$

B. Wafer Image Preprocessing

The semiconductor manufacturing process is susceptible to many random factors, and various noises are often mixed on the surface of wafer image, thus covering up the defect mode in the wafer image, as shown in Fig. 1(a). In this paper, median de-noising method [9] is adopted to preprocess the wafer image, as shown in Fig. 1(b). It can be seen that the defect mode in the wafer image is retained by de-noising, so the defect characteristics are extracted effectively.



(a) Before de-noising

(b) After de-noising

Figure 1. Median de-noising result in wafer image

C. Feature Extraction

Effective defect features can reflect the characteristics of defect mode, reflect the differences between different modes, and improve the mode recognition rate. The wafer defect mode and normal wafer are shown in Fig. 2.

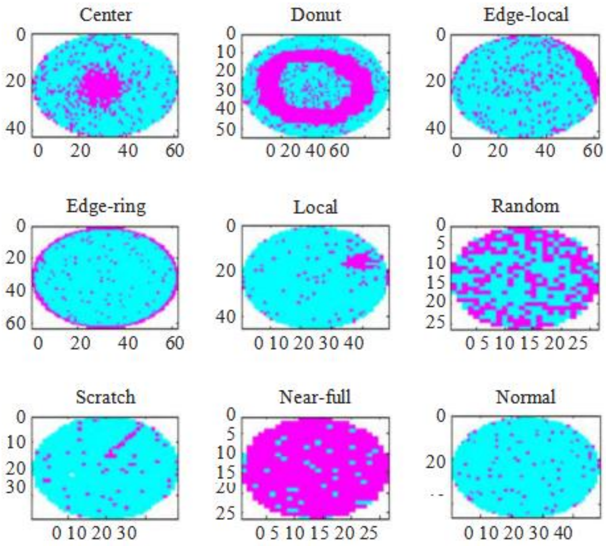


Figure 2. Eight wafer defect modes and normal wafer

Due to the randomness of defect size and distribution on the wafer surface, in order to accurately describe the defect characteristics, this paper extracts the defect characteristics on the wafer surface from the following three aspects [15]. The total dimension of the final wafer feature data is 47, in which 18 geometric feature data and 24 projection feature data are extracted.

(1) Geometric features: including regional features and linear features. Regional features [16] include defect boundary area, circumference, defect compactness, rectangle degree, center of gravity coordinate, and regional duty ratio, length of axis and eccentricity of ellipse with the same standard second-order center distance as the region. Linear features are represented by the number of lines in the defect region, and the extraction of the number of lines in the defect target is completed by Hough Transform [17].

(2) Texture features: in this paper, the correlation texture features of defect images are extracted using the gray co-occurrence matrix [16], [18].

(3) projection features: in this paper, the defect area are calculated respectively in four directions of projection, namely 0 ° and 90 °, 45 ° and 135 ° direction. The extracted features include mean amplitude, root mean square amplitude, projection waveform feature, projection pulse and projection peak value.

D. Wafer Defect Detection and Adaptive Recognition

The scheme of online detection and adaptive recognition model for wafer surface defect proposed in this paper is shown in Fig. 3, which consists of two stages, offline modeling and online detection. When the off-line modeling is conducted, the wafer image is firstly denoised by median filtering technology, and the geometry, texture and projection features of the wafer image are extracted from the feature extraction module, and the corresponding HMM recognition model is constructed for each feature set of defect mode respectively. During online detection, the collected wafer images are processed by image preprocessing and feature extraction module, and the extracted features are input into the HMM recognition library. Firstly, the defect of the wafer is detected. If it is determined to be a defective wafer, it is detected whether it is a new defect mode. If the defect wafer is a new defect mode, a certain number of new defect mode wafers are collected online. The new HMM recognition model is built for each new defect mode and the HMM recognition library is updated. Otherwise, the defect mode recognition is carried out directly into the HMM recognition library.

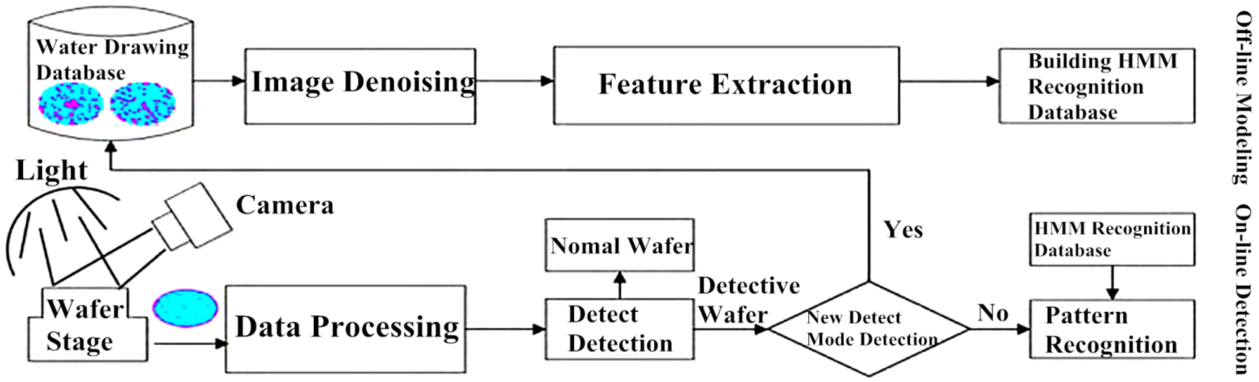


Figure 3. Wafer surface defect detection and adaptive recognition model

III. Experiment and Result Analysis

A. Test Data

The data used in this paper is the actual wafer image data (WM-811Kdata [19]) of the semiconductor manufacturing industry, which is used to test the performance of the HMM dynamic integration method.

First, the selected representative five wafer defect modes (Edge-local/Edge-ring/Center/Donut/Local) and normal wafer modes are selected from WM-811K wafer image data set. There are 1400 training samples in total, among which 400 samples are selected for normal wafer and 200 samples for each defect mode. A total of 700 samples are tested, among which 200 samples are selected for normal wafer and 100 samples are selected for each defect mode.

B. Wafer Defect Detection

NLLP control chart is used to detect defects on the wafer surface online. Fig. 4 shows the detection of defect wafer based on NLLP control chart, and Table I also lists the defect detection rate of each defect wafer. It can be seen that all four defect modes including Center, Donut, Edge-ring and Edge-local can be detected, while the misjudgment rate of normal wafer is only 4.5%. During semiconductor production, some noises will be mixed on the surface of normal wafers, and some normal wafers will be judged as defective by the HMM dynamic recognition model. During the experiment, 9 normal wafer samples (all of which are misdiagnosed as defective) are input into the HMM recognition library for defect recognition, among which 8 samples are identified as Edge-local defect mode and 1 sample is identified as Center defect mode.

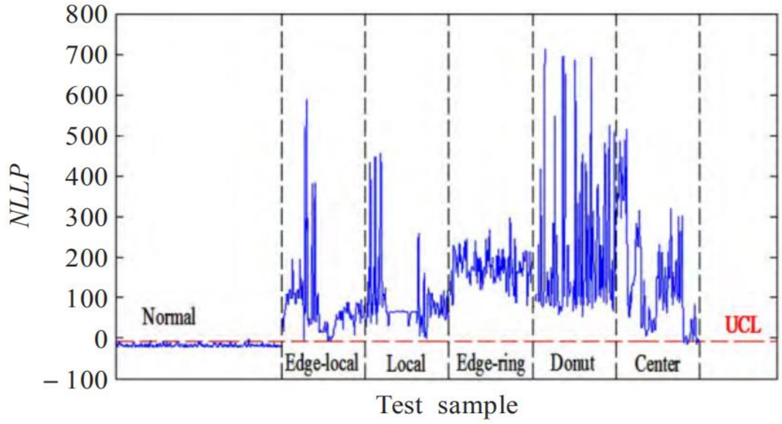
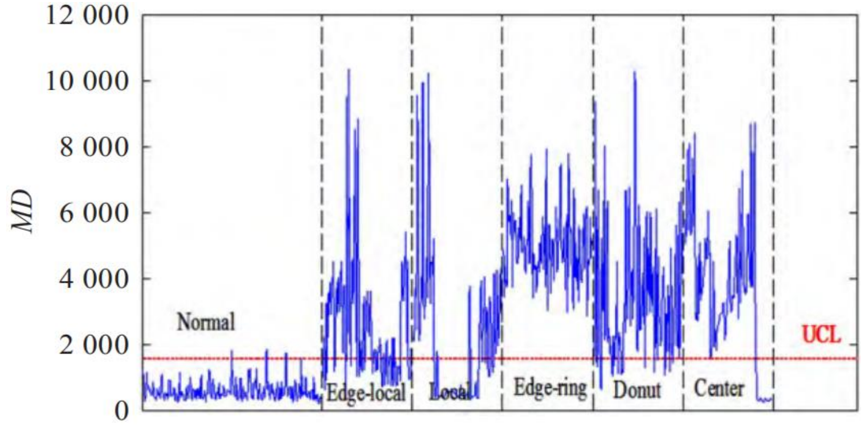


Figure 4. Wafer defect NLLP monitoring chart

TABLE I. WAFER DEFECT DETECTION RATE (%)

Defect mode		Center	Donut	Edge-ring	Edge-local	Local
Detection rate	NLLP	100	100	100	100	94
	MD	80	84	100	66	46

To verify the effectiveness of the method proposed in this paper to detect the wafer defect based on NLLP control chart, Mahalanobis Distance (MD) is selected as the detection index. Since a sample outputs multiple MD in an integrated HMM, the monitored variable is average MD. UCL is set as $\bar{\bar{x}} + 3\sigma_{\bar{x}}$, according to the 3σ criterion. Fig. 5 shows the monitoring of defect wafer based on MD control chart. In Table I, MD-based test sample wafer defect detection is also listed, and the overall detection rate is 75.2%. However, the NLLP control chart defect detection rate based on the HMM dynamic recognition model proposed in this paper can reach 98.80%, which fully demonstrates the effectiveness of the wafer defect detection method.



Test sample

Figure 5. Wafer defect MD monitoring chart

C. New Defect Mode Detection of Wafer

Due to the influence of multi-source random factors in the online manufacturing process, new defect modes may appear in the production of wafers. This paper establishes the control trust limit corresponding to various defect modes in the HMM recognition library. In the experiment, Edge-local is used as a new defect mode, and the HMM recognition library and corresponding trust limit are established for the other four defect modes. Fig. 6 shows the detection of Edge-local based on the known trust limit of defect mode, with a detection rate of 99.50%. It can be seen that the HMM recognition library can detect the new defect mode online very well.

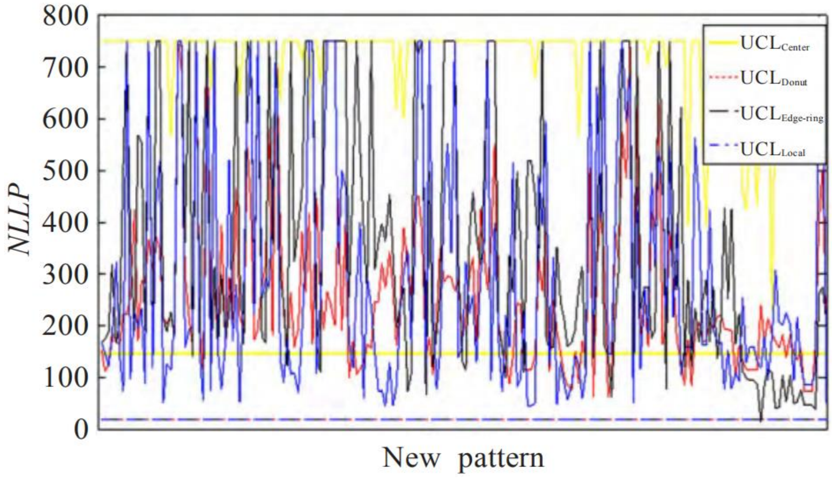


Figure 6. Wafer new defect mode (Edge-local) NLLP detection chart

D. Wafer Defect Mode Recognition

In the process of wafer production, once the defect wafer is detected, it is input into the HMM recognition library for defect mode recognition. The experimental results are presented by Confusion Matrix, as shown in Table II. In this paper, the online detection and adaptive recognition model based on HMM can well identify the two defect modes of Edge-local and Donut, and the overall recognition rate is up to 82.60%. For the Local defect mode with poor recognition effect, Fig. 7 is the four Local defect wafer images identified as Edge- Local mode. When the surface noise of the defect wafer is serious, the defect mode cannot be separated well even by the de-noising technology, so the model will produce misjudgment.

TABLE II. WAFER DEFECT MODE RECOGNITION RATE (%)

Defect mode	Center	Donut	Edge-ring	Local	Edge-local
Center	71	9	0	14	6
Donut	0	98	0	2	0
Edge-ring	0	0	75	0	14
Local	0	11	0	69	20
Edge-local	0	5	1	4	90

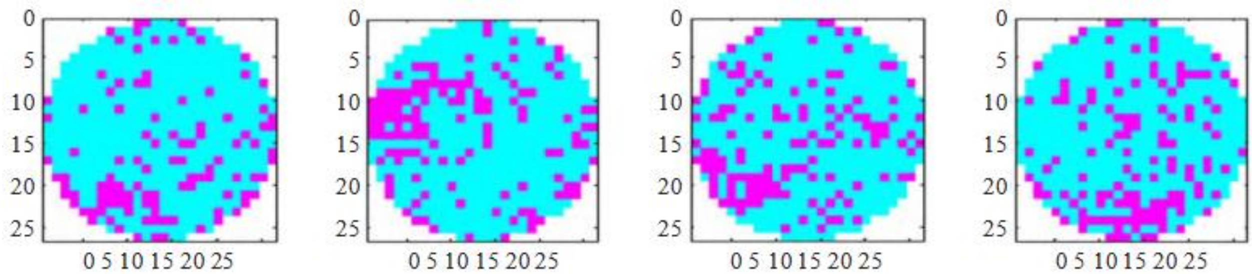


Figure 7. Local defect wafer image misrecognized as Edge-local mode classifier (%)

In order to test the recognition effect of the HMM dynamic integration model, the three classifiers for experimental comparison in this paper are back propagation neural network (BPN), KNN and SVM. Among them, BPN adopts a three-layer network with 15 neurons in the middle layer, the maximum number of iterations is set as 100, and the learning rate is set as 0.01. SVM chooses LIBSVM [20] classifier, and the parameters are set as the default. The nearest neighbor coefficient of KNN is 5. The experimental results are shown in Table III. It can be seen that the overall recognition rate of HMM is higher than that of other classifiers. At the same time, it is found that each recognizer has higher recognition rate for a certain defect. For example, BPN's recognition rate of Edge-ring is 95%, and SVM's recognition rate of Donut is 98%. In this paper, a classifier fusion experiment is conducted and Ensemble Learning is adopted [21]. Recognizer integration is one of the most commonly used recognizer fusion techniques, which can generally achieve better performance. Four weak classifiers are trained: HMM, SVM, KNN and BPN. By Majority Voting [22], all four classifiers are binded into an integrated classifier, the classification results are given in Table III. Although the recognition rate of integrated classifier is 85.80% and the recognition performance is higher than HMM, the unknown wafer defect mode is often generated in the online detection process, so it is difficult to determine the type of classifier. Therefore, classifiers with higher comprehensive recognition rate are generally preferred.

TABLE III. WAFER DEFECT RECOGNITION RATE OF EACH

Classifier	HMM	KNN	SVM	BPN	Ensemble
Center	71	82	52	89	89
Donut	98	80	74	97	97
Edge-ring	75	87	72	95	91
Local	69	60	66	53	61
Edge-local	90	59	98	65	91
Average	82.60	73.60	72.40	79.80	85.80

IV. CONCLUSION

In this paper, a model based on hidden Markov dynamic integration is proposed, which can be applied to the on-line wafer surface defect detection and adaptive recognition. In the experimental results of WM-811K data, the overall defect detection rate of the model proposed in this paper reaches 98.80%. Moreover, the model is able to detect the new defect mode generated online with a detection rate of 99.50% and a defect mode recognition rate of 82.60%, which fully verifies the effectiveness of the model. The follow-up work will focus on the feature selection of wafer defects and further improve the accuracy of recognition.

REFERENCES

[1] C. H. Wang, "Recognition of semiconductor defect patterns using spatial filtering and spectral clustering," *Expert Systems with Applications*, vol. 34, no. 3, pp. 1914-1923, 2008.
[2] C. R. Pan and N. Q. Wu, "Study on scheduling of automatic integrated equipment for wafer production," *Computer Integrated Manufacturing System*, vol. 15, no. 3, pp. 522-528, 2009.
[3] J. Y. Hhang and W. Kuo, "Model-based clustering for integrated circuit yield enhancement," *European Journal of Operational Research*, vol. 178, no. 1, pp. 143-153, 2007.
[4] T. Yuan and W. Kuo, "A model-based clustering approach to the recognition of the spatial defect patterns produced during semiconductor fabrication," *IIE Transactions*, vol. 40, no. 2, pp. 93-101, 2007.
[5] C. H. Wang, W. Kuo, and H. Bensmail, "Detection and classification of defect patterns on semiconductor wafers," *IIE Transactions*, vol. 38, no. 12, pp. 1059-1068, 2006.
[6] R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering," *IEEE Transactions on Fuzzy Systems*, vol. 1, no. 2, pp. 98-110, 1993.
[7] Q. Zhou, L. Zeng, and S. Zhou, "Statistical detection of defect patterns using Hough transform," *IEEE Transactions on Semiconductor Manufacturing*, vol. 23, no. 3, pp. 370-380, 2010.
[8] F. L. Chen and S. F. Liu, "A neural-network approach to recognize defect spatial pattern in semiconductor fabrication," *IEEE Transactions on Semiconductor Manufacturing*, vol. 13, no. 3, pp. 366-373, 2000.
[9] C. H. Wang, "Recognition of semiconductor defect patterns using spatial filtering and spectral clustering," *Expert Systems with Applications*, vol. 34, no. 3, pp. 1914-1923, 2008.
[10] B. V. Dasarathy, *Nearest Neighbor ({NN}) Norms: {NN} Pattern Classification Techniques*, IEEE Computer Society Press, 1991.
[11] R. Baly and H. Hajj, "Wafer classification using support vector machines," *IEEE Transactions on Semiconductor Manufacturing*, vol. 25, no. 3, pp. 373-383, 2012.
[12] L. R. Rabiner and B. H. Juang, "An introduction to hidden Markov models," *IEEE ASSP Magazine*, vol. 3, no. 1, pp. 4-16, 1986.
[13] L. H. Yu, H. Wang, and H. S. Zhong, "Face recognition based on hidden Markov model," *Computer Technology and Development*, vol. 22, no. 2, pp. 25-28, 2012.
[14] J. G. D. Forney, "The viterbi algorithm," *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268-278, 1973.
[15] F. F. Han, "Research on key technology of online surface defect detection," Tianjin University, Tianjin, 2011.
[16] J. B. Yu, X. L. Lu, and W. Z. Zong, "Detection and recognition of wafer surface defect based on local and non-local linear discriminant analysis and dynamic integration of Gaussian mixture model," *Journal of Automation*, vol. 42, no. 1, pp. 47-59, 2016.
[17] J. K. P. White, B. Kundu, and C. M. Mastrangelo, "Classification of defect clusters on semiconductor wafers via the Hough transformation," *IEEE Transactions on Semiconductor Manufacturing*, vol. 21, no. 2, pp. 272-278, 2008.
[18] D. C. He and L. Wang, "Texture feature extraction from texture spectrum," in *Proc. 10th Annual International Geoscience and Remote Sensing Symposium*, 1990, pp. 987-1990.
[19] M. J. Wu, J. S. R. Jiang, and J. L. Chen, "Wafer map failure pattern recognition and similarity ranking for large-scale data sets," *IEEE Transactions on Semiconductor Manufacturing*, vol. 28, no. 1, pp. 1-12, 2015.
[20] C. C. Chang and C. J. Lin. Libsvm: A library for support vector machine [EB/OL]. [Online]. Available: http: //www.csie.ntu.edu.tw/_cjlin/
[21] A. Krohg and J. Vedelsby, "Neural network ensembles, cross validation, and active learning," *Advances in Neural Information Processing Systems*, vol. 7, pp. 231-238, 1995.
[22] Z. H. Zhou, J. X. Wu, and W. Tang, "Ensembling neural networks: Many could be better than all," *Artificial Intelligence*, vol. 137, no. 1-2, pp. 239-263, 2002.

Yefan Zhou was born on July 26, 1998, in Jiangsu, China. He is a third-year undergraduate student at Southeast University, Nanjing, China. He is a major in Information Engineering.

He started research internship in the Artificial Intelligence and Multimedia Lab in Southeast University from March 2018. He participated in the research of Improvement of Object Detection SSD Algorithm and later joined the development team of Object Detection RFCN Algorithm in National key project Gezhou Dame Loose Nuts Detection. Then He started the research on chip defect detection from September 2018. His primary research interests are in the fields of computer graphic, computer vision, machine learning.

Mr. Zhou is the student of Professor Luxi Yang, IEEE author. He was an internship in the Artificial Intelligence and Multimedia Lab and accepted the academic guidance of Prof. Yang.