New automatic defect classification algorithm based on a classification-after-segmentation framework

Sang-Hak Lee, Hyung-Il Koo, and Nam-Ik Cho
Seoul National University, Department of Electrical Engineering and Computer Science, San 56-1, Shilim-Dong, Kwanak-Gu, Seoul, Korea
E-mail: crane@ispl.snu.ac.kr

Abstract. *We propose a new method that classifies wafer images according to their defect types for automatic defect classification in semiconductor fabrication processes. Conventional image classifiers using global properties cannot be used in this problem, because the defects usually occupy very small regions in the images. Hence, the defects should first be segmented, and the shape of the segment and the features extracted from the region are used for classification. In other words, we need to develop a classification-after-segmentation approach for the use of features from the small regions corresponding to the defects. However, the segmentation of scratch defects is not easy due to the shrinking bias problem when using conventional methods. We propose a new Markov random field-based method for the segmentation of wafer images. Then we design an AdaBoost-based classifier that uses the features extracted from the segmented local regions. © 2010 SPIE and IS&T.* [DOI: 10.1117/1.3429116]

1 Introduction

Automatic defect classification (ADC) is a wafer fabrication process that classifies defects into predefined types, e.g., particle, scratch, etc. Figures 1(a) and 1(b) show examples of a particle and scratches, respectively. By correctly classifying the defects, the cause of the defects can be analyzed, and this information is used for improving the process and consequently the yield. Hence, there have been many studies on ADC. For example, Kameyama and Kosugi:[1] proposed a method that exploits a hyperellipsoid clustering network (HCN) with radial basis function (RBF) and model switching. Also, smart beam search (SBS) using a support vector machine (SVM) was proposed for feature selection.[2] However, there are some difficulties in applying conventional appearance-based pattern classification methods (e.g., techniques used in face detection[3]) to ADC, because defects in the same class have too many variants in their shapes. Also, since the defect regions occupy very small portions of the image, the global feature statistics (e.g., frequency or filter bank responses) are useless. To resolve these difficulties, we propose a new method based on classification-after-segmentation.

The most essential part of the process may be the correct segmentation of defects. Due to the shrinking bias

problem,[7] however, the conventional state of the art segmentation methods based on maximum a posterior Markov random fields (MAP-MRF) with a Potts model data term[4–6] often oversegment a scratch (thin object) into several objects, and thus a scratch can appear as particles. For dealing with this problem, we propose a new MAP-MRF based segmentation method. To be specific, we develop a new energy function for the MAP-MRF scheme based on the Retinex theory[8] for handling the blurry scratches and the color inconsistency. After segmentation, we define several features (shape, intensity, and so on) for each defect, and develop an AdaBoost classifier to tell whether the patch corresponds to a particle or other defects (including scratch).

In the experiments, we show that our segmentation method reduces the oversegmentation of scratches and thus keeps the false alarm rate low, even for high particle detection rates.

2 Segmentation

Among the many segmentation methods, the state of the art MAP-MRF approach is adopted here.[4–6] That is, segmentation is achieved by minimizing an energy function:

$$E(f) = \sum_{p \in P} V_p(f_p) + \sum_{p \in P} \sum_{q \in N(p)} V_{p,q}(f_p, f_q), \quad (1)$$

where P is a set of sites, $N(p)$ is an eight-neighborhood system, $f_p \in \{1, \ldots, N\}$ is a label of the site p, and $f = \{f_p\}_{p \in P}$. The label f_p indicates that the site p belongs to the segment whose label is f_p. Specifically, the data term $V_p(f_p)$ is designed so that pixels having similar intensity values are clustered into a single segment:

$$V_p(f_p) = |I(p) - M \times f_p|, \quad (2)$$

where $I(p)$ is the intensity of a given pixel p, and M denotes the difference of intensity between the adjacent labels. $V_{p,q}(f_p, f_q)$ is called the smoothness term, defined as

$$V_{p,q}^C(f_p, f_q) = \exp\left[-\frac{|I(p) - I(q)|^2}{\sigma^2}\right]\delta(f_p, f_q), \quad (3)$$

where σ is a constant and $\delta(f_p, f_q)$ is defined as

$$\delta(f_p, f_q) = \begin{cases} 1 & \text{if } f_p \neq f_q \\ 0 & \text{otherwise} \end{cases}, \quad (4)$$

in conventional works.[5,6] $V_{p,q}^C(f_p, f_q)$ enforces the continuity of labels by penalizing label discontinuities. However, it is well known that since the $V_{p,q}^C(f_p, f_q)$ tries to minimize the length of boundary[9] (shrinking bias), this method often segments a long and thin object into several parts. Also, blurred boundaries deteriorate the performance of $V_p(\cdot)$ (especially when capturing thin objects). Hence, many long and thin objects in wafer images are frequently segmented into several parts, and they are often confused with particles.

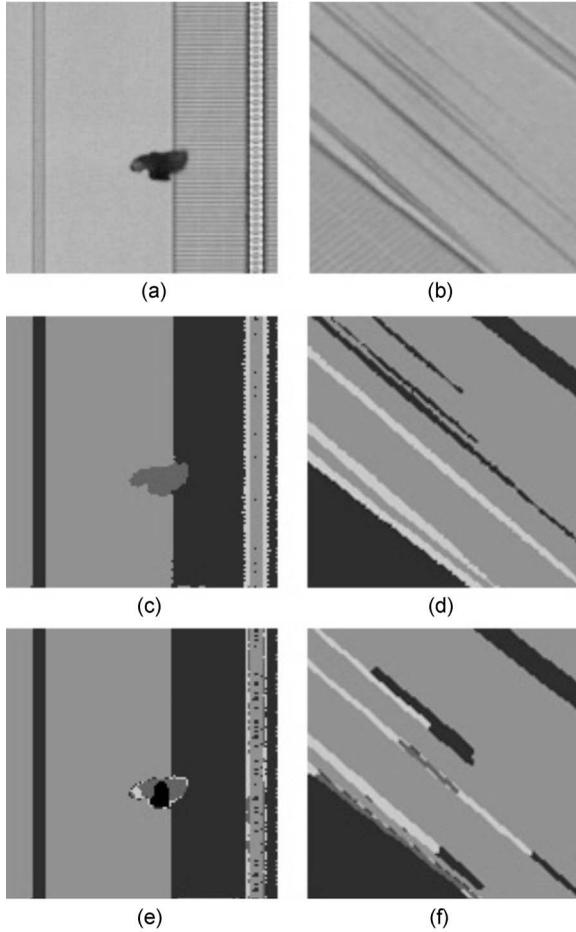To prevent oversegmentation, we develop a new function to be included into the data term:

Fig. 1 Example images and segmentation results: (a) particle image, (b) scratch image, (c) proposed segmentation for particle image, (d) proposed segmentation for scratch image, (e) Potts model-based segmentation for particle image, and (f) Potts model-based segmentation for scratch image.

$$V_{p,q}^R(f_p,f_q) = \{(1 - \delta[L(p),L(q)])\} \times \delta(f_p,f_q), \tag{5}$$

where $L(p)$ is a Retinex filtering of input.[8] To be specific, the $L(p)$ for our purpose is defined as

$$L(p) = \begin{cases} 1, & \text{if } \dfrac{I(p)}{G(p)} < T \\ 0, & \text{otherwise} \end{cases}, \tag{6}$$

where $G(p)$ is a Gaussian blurred image at pixel p, and T is a predefined parameter (in our implementation, T is 0.9). Note that the Retinex filtering compensates for the blurring at the image formation process, and improves color consistency caused by illumination change.[8] Therefore, $L(p)=L(q)$ and $p \in N(q)$ means that p and q have the same intrinsic colors [even if $|I(p)-I(q)| \gg 0$], and the labels satisfying $f_p \neq f_q$ should be penalized. In summary, $V_{p,q}^R(f_p,f_q)$ alleviates the shrinking bias by penalizing the discontinuities occurring on the pixels having the same intrinsic colors. A similar idea can also be found in the binarization of document images.[10] Finally, the smoothness term in Eq. (1) is given by

$$V_{p,q}(f_p,f_q) = \lambda_1 V_{p,q}^C(f_p,f_q) + \lambda_2 V_{p,q}^R(f_p,f_q), \tag{7}$$

where λ_1 and λ_2 are two balancing parameters.

3 Classification

Since each particle and scratch is segmented into a single region by the proposed segmentation algorithm, we can use several features extracted from the segments to determine whether it is a particle defect or not.

3.1 Features

Let S be a set of positional vectors in a given segment, $|S|$ be the size of S, and $I(x,y)$ be the pixel intensity at $(x,y) \in S$. Then features can be summarized as follows.

Mean intensity:

$$\frac{1}{|S|} \sum_{(x,y) \in S} I(x,y). \tag{8}$$

Shape descriptor:

$$\frac{\lambda_{\max}}{\lambda_{\min}}, \tag{9}$$

where $\lambda_{\max}$ and $\lambda_{\min}$ are two eigenvalues of the covariance matrix of S (when the value is close to unity, the shape has no directional preference, and $\lambda_{\max}/\lambda_{\min} \gg 1$ means that the shape is a thin and elongated one).

Texture measure:

$$\sum_{i=1}^{4} \lambda_i^2, \tag{10}$$

where λ_i's are eigenvalues of the covariance matrix of the vectors at points $(x,y) \in S$,

$$\left(\left| \frac{\partial I}{\partial x} \right|, \left| \frac{\partial I}{\partial y} \right|, \left| \frac{\partial^2 I}{\partial x^2} \right|, \left| \frac{\partial^2 I}{\partial y^2} \right| \right). \tag{11}$$

The vector in Eq. (11) is computed over $(x,y) \in S$ and the covariance matrix of these vectors is found. Then, Eq. (10) is defined as the measure of texture, which has a large value when there exists a particular directional texture.

Measure of orientation bias (of edges): obtained from the histogram of

$$\arctan\left(\frac{\partial I}{\partial y} \Big/ \frac{\partial I}{\partial x} \right). \tag{12}$$

By computing the orientation histogram and summing the sizes of four dominant bins, we can measure the bias of orientation distribution.

3.2 AdaBoost

For machine learning using the extracted features, we use the AdaBoost algorithm, where each weak classifier is based on the log-likelihood ratio test
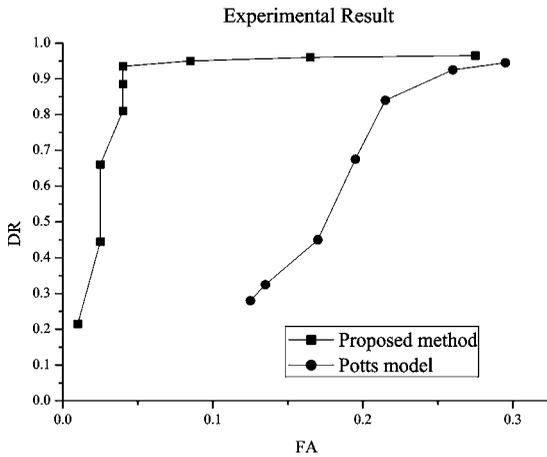
Fig. 2 ROC curve comparing our method to the Potts model-based method.

$$h_t(x) = \ln \frac{p_t^+(x)}{p_t^-(x)}, \qquad (13)$$

where x indicates the output of segmentation in the feature space, $p_t^+(x)$ denotes the weighted histogram of the t'th feature for a positive sample, and $p_t^-(x)$ is similarly defined. Then, after the training process, a strong classifier $F(x)$ is given by

$$F(x) = \sum_{t=1}^{T} \alpha_t h_t(x), \qquad (14)$$

where, from the standard AdaBoost algorithm,[11] α_t is determined by the function of current error rate e:

$$\alpha_t = \log \frac{1-e}{e}. \qquad (15)$$

4 Experimental Results

The dataset for the experiment consists of defected images acquired by a 266-nm bright field inspection instrument (12-in. $\langle 100 \rangle$ oriented silicon wafer, magnification of more than 10^4 times). Among the dataset, we use 380 images including particle defects and 150 images including scratch defects as training samples. Then we test 200 images containing particle defects and 200 images having no particle defects (of course they contain other defects such as scratch defects). As can be seen in Fig. 1(c)–1(f), the proposed term improves segmentation performance. The main purpose of ADC is to automatically classify the particle from the other kinds of defects (mostly scratches), hence we evaluate the performance of ADC using the detection ratio (DR) of particle defects and the false alarm (FA) of other defects considered as particle, which are defined as

$$\text{DR} = \frac{\text{the number of correctly detected particle defects}}{\text{the number of all particle defects}},$$
$$(16)$$

$$\text{FA} = \frac{\text{the number of defects misclassified as particle}}{\text{the number of all scratch defects}}. \quad (17)$$

In typical detection problems, if we try to increase the DR, the FA also increases and vice versa. Hence the performance of classifier or detector can be measured by the receiver operating characteristic (ROC) that shows the DR versus FA. When the FA is kept low for the high DRs, the detector is considered to be a good one. We compare the ROCs before and after applying our new method in Fig. 2, where it can be observed that our method keeps the DR very high even for very low (down to 0.05) FA rates.

5 Conclusion

In this work, we propose a new approach to ADC based on the classification-after-segmentation framework. The wafer image is first segmented based on the MAP-MRF approach, where a new energy function is designed to prevent the degeneration of scratch into several regions. Then, an Ada-Boost classifier is trained using the features extracted from the segments. According to the experimental results on wide variants of particles, the proposed approach shows good classification performance.

References

1. K. Kameyama and Y. Kosugi, "Semiconductor defect classification using hyperellipsoid clustering neural networks and model switching," *Proc. Int. Joint Conf. on Neural Networks (IJCNN'99)*, Washington, DC, July 10–16, 1999, pp. 3505–3510 (1999).
2. P. Gupta, D. Doermann, and D. DeMenthon, "Beam search for feature selection in automatic svm defect classification," *Proc. 16th Int. Conf. on Pattern Recognition*, vol. 2, pp. 212–215, IEEE, Piscataway, NJ (2002).
3. P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proc. Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 511–518, IEEE, Piscataway, NJ (2001).
4. Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(11), 1222–1239 (2001).
5. Y. Boykov and M. Jolly, "Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images," *Proc. Int. Conf. on Computer Vision*, pp. 105–112, IEEE, Piscataway, NJ (2001).
6. C. Rother, V. Kolmogorov, and A. Blake, "grabcut: interactive foreground extraction using iterated graph cuts," in *SIGGRAPH '04*, pp. 309–314, ACM, Washington, DC (2004).
7. S. Vicente, V. Kolmogorov, and C. Rother, "Graph cut based image segmentation with connectivity priors," *Proc. Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8, IEEE, Piscataway, NJ (2008).
8. E. L. Land, "The retinex theory of colour vision," *Scientific Am.* **237**(6), 108–128 (1977).
9. V. Kolmogorov and Y. Boykov, "What metrics can be approximated by geo-cuts, or global optimization of length/area and flux," *Proc. Int. Conf. on Computer Vision*, vol. 1, pp. 564–571, IEEE Computer Society, Washington, DC (2005).
10. M. Pilu and S. Pollard, "A light-weight text image processing method for handheld embedded camera," *Proc. Brit. Mach. Vision Conf.*, pp. 547–556, British Machine Vision Association (BMVA), Malvern, UK (2002).
11. R. E. Schapire, "The boosting approach to machine learning: an overview," *Proc. MSRI Workshop Nonlinear Est. Class*, pp. 149–172, Springer, New York (2002).