KERNEL-BASED DETECTION OF DEFECTS ON SEMICONDUCTOR WAFERS

Maria Zontak and Israel Cohen

Department of Electrical Engineering, Technion - Israel Institute of Technology
Technion City, Haifa 32000, Israel
{zontakm@tx,icohen@ee}.technion.ac.il

ABSTRACT

Recent computational methods of wafer defect detection often rely on the difference image between an inspected image and its reference image, and highly depend on registration accuracy. In this paper, we present a novel method for defect detection in patterned wafers, based on reconstruction of the inspected image from the reference image using anisotropic kernels. This method avoids registration between the inspected and reference image and compensates for pattern variations, thus reducing the false detection rate. Experimental results demonstrate the advantages and robustness of the proposed method. Efficient implementation of the algorithm makes it be suitable for industrial use. We also demonstrate extension of the kernel-based similarity concept to the multichannel Scanning Electron Microscope (SEM) images.

Index Terms— Semiconductor defect detection, anomaly detection, anisotropic kernels, image reconstruction, similarity measure.

1. INTRODUCTION

Defect detection in wafers is a critical component of wafers manufacturing process. A common approach for automatic wafer defect detection utilizes a reference image. A semiconductor wafer typically contains many copies of the same electrical component (denoted as "dies") laid out in a matrix pattern. A reference image for one die is obtained by acquiring an image of the neighboring die, which is verified to be clear of defects. A common detection procedure is based on the difference between the inspected image (further referred to as the "source image") and the reference image, which are spatially aligned [1–3]. A major drawback of this approach is that the detection performance is very sensitive to image registration inaccuracies between the source and reference images [4, 5]. Moreover, printed patterns on the source and reference dies may differ slightly, particularly in the neighborhood of their edges. These pattern variations obscure the defects in the difference image and may yield high false detection rate.

We propose a defect detection procedure, which avoids image registration and is robust to pattern variations. The method is based on anisotropic kernel reconstruction of the source image using its reference image [6]. Anisotropic kernels were successfully used in recent edge-preserving de-noising application [7]. The source and reference images are mapped into a feature space, where every feature from the source image is estimated by a weighted sum of neighboring features from the reference image. We use patches around pixels as features and show that patches originating from defect regions are not reconstructible from the reference image, and hence can be identified [6]. Here, we demonstrate the advantages of the kernel-based method over other methods and discuss the implementation issue. We also extend the kernel-based similarity to simultaneous detection in the multichannel Scanning Electron Microscope (SEM) images, based on the consistency criterion developed in [8].

This paper is organized as follows. In Section 2, we explain the procedure of the source image reconstruction from the reference image, which is based on a novel kernel-based similarity measure. In Section 3, we present experimental results that demonstrate the advantage of our kernel-based detection over two other competitive methods, and discuss computational complexity and efficient implementation of the algorithm. Finally in Section 4, we briefly describe a kernel-based detection using multichannel SEM images, and demonstrate the advantage of the proposed method over simple union or intersection of single channels results.

2. KERNEL-BASED RECONSTRUCTION

Let us pick a d-vector $G = (g_1, .., g_d)$ of filters and map pixels of source image, I_{src}, and the reference image, I_{ref}, into $\mathbb{R}^d$ features space ξ_G:

$$\begin{aligned} \mathbf{s} \rightarrow \xi_G(\mathbf{s}) &= \{I_{\text{src}} * g_1(\mathbf{s}), ..., I_{\text{src}} * g_d(\mathbf{s})\} \\ \mathbf{s}' \rightarrow \xi_G(\mathbf{s}') &= \{I_{\text{ref}} * g_1(\mathbf{s}'), ..., I_{\text{ref}} * g_d(\mathbf{s}')\} \ , \end{aligned} \qquad (1)$$

where $\mathbf{s}, \mathbf{s}' \in \Omega$ and Ω is a general set of indices in the image space, the indices $\mathbf{s}$ are associated with features from the source image and $\mathbf{s}'$ are associated with features from the reference image. The construction presented in (1) is flexible, and there are many interesting choices for the filters. According to Szalm *et al.* [7], we construct G from non local means filters (NL-means) of Morel [9], where $g_{m,n}$ is an $[s_x \times s_y]$ matrix with one in (m, n) position and zeros elsewhere. Thus, ξ_G is the set of overlapping patches of the source and reference images embedded in $d = s_x \times s_y$ dimensions.

Given $\xi_G(\mathbf{s}')$ for all $\mathbf{s}' \in \mathcal{N}_{\mathbf{s}}$, a reconstructed source image is obtained by

$$\hat{I}_{\text{src}}(\mathbf{s}) = \frac{1}{\sum_{\mathbf{s}' \in \mathcal{N}_{\mathbf{s}}} W(\mathbf{s}, \mathbf{s}')} \sum_{\mathbf{s}' \in \mathcal{N}_{\mathbf{s}}} W(\mathbf{s}, \mathbf{s}') I_{\text{ref}}(\mathbf{s}') \ , \qquad (2)$$

where $\mathcal{N}_{\mathbf{s}} = \{\mathbf{s}' \,|\, \mathbf{s}' \in n_k(\mathbf{s})\}$ and $n_k(\mathbf{s})$ is the set of k nearest

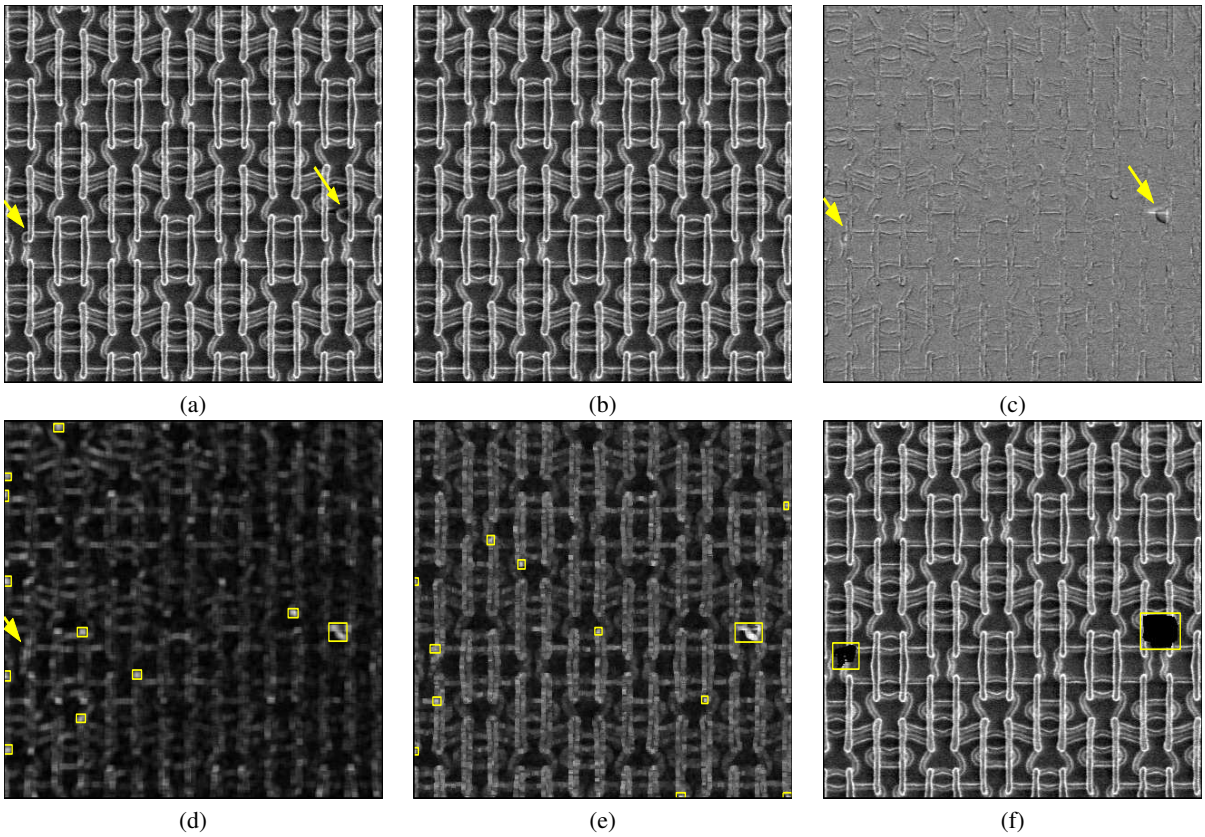


Fig. 1. (a)-(c) Source, reference and difference images, respectively (images' **(!**FOV) is equal to 4μ, arrows point at defects); Respective detections are overlaid (light rectangles) on (d) Single Hypothesis Test (SHT) of the Mahalanobis distance of the difference image (c) (arrow points at a missed defect); (e) Improved difference image, based on Onishi algorithm; (f) Kernel-based reconstructed source image, using $[29 \times 29]$ patches. High false detection rate in (d) and (e) results from sensitivity to pattern variations, whereas the reconstruction in (f) is robust to this disturbance.

neighbors of $\mathbf{s}$ in spatial domain. According to [7], we choose

$$W(\mathbf{s},\mathbf{s}') = \exp^{-\rho(\mathbf{s},\mathbf{s}')^2/\varepsilon} , \tag{3}$$

where ρ is a metric in our feature space and ε is a similarity parameter. The similarity $W(\mathbf{s},\mathbf{s}')$ is measured as a decreasing function of the Euclidean distance $\rho^2(\mathbf{s},\mathbf{s}') = \|\xi_G(\mathbf{s}) - \xi_G(\mathbf{s}')\|_2^2$. The similarity parameter ε controls the decay of the exponential function and therefore the decay of the weights as a function of the Euclidean distances.

Let $p(x)$ denote a probability density function of a random variable X, and let $\{x_i\}_{i=1}^m$ represent samples of X. A nonparametric estimate of $p(x)$ can be obtained by Parzen method [10] and is given by

$$\hat{p}_\varepsilon(x) = \frac{1}{m}\sum_{i=1}^{m} b_\varepsilon(\|x - x_i\|) , \tag{4}$$

where b_ε is a normal density with zero mean and variance ε. Denoting $x = \xi_G(\mathbf{s})$, $x_i = \xi_G(\mathbf{s}')$ for $\mathbf{s}' \in \mathcal{N}_\mathbf{s}$ and $k(x,x_i) = W(\mathbf{s},\mathbf{s}')$ (*i.e.* $k(\cdot,\cdot)$ is a Gaussian kernel function), a Parzen estimator can be related [11] to the total similarity measure $\sum_{\forall \mathbf{s}' \in \mathcal{N}_\mathbf{s}} W(\mathbf{s},\mathbf{s}')$ between the source patch and representative reference patches by

$$\hat{p}_\varepsilon(x) = x_X^T \cdot \vec{1}_m , \tag{5}$$

where $X_x = (k(x_1,x), k(x_2,x), ..., k(x_m,x))^T$ and $\vec{1}_m$ denotes an m-dimensional column vector with all components equal to m^{-1}.

Hence, the kernel-based similarity measure can be considered as the likelihood of the source feature to arise from the pattern statistics, which is represented by reference features from the search region. Features originated form defects will have low similarity to the pattern and hence can be identified. That is, a pattern-originated patch from the source image is reconstructible from similar patches from the reference image. On contrary, there are no similar patches in the reference image for a patch that contains a defect, and therefore this patch cannot be reconstructed. Thus, the detection is obtained according to the total similarity measure:

$$\begin{aligned} \sum_{\forall \mathbf{s}' \in \mathcal{N}_\mathbf{s}} W(\mathbf{s},\mathbf{s}') \neq 0 \quad &\Rightarrow \quad \mathbf{s} \notin \mathcal{A} \\ \sum_{\forall \mathbf{s}' \in \mathcal{N}_\mathbf{s}} W(\mathbf{s},\mathbf{s}') \to 0 \quad &\Rightarrow \quad \mathbf{s} \in \mathcal{A} . \end{aligned} \tag{6}$$

where $\mathcal{A}$ denotes a set of defect regions.

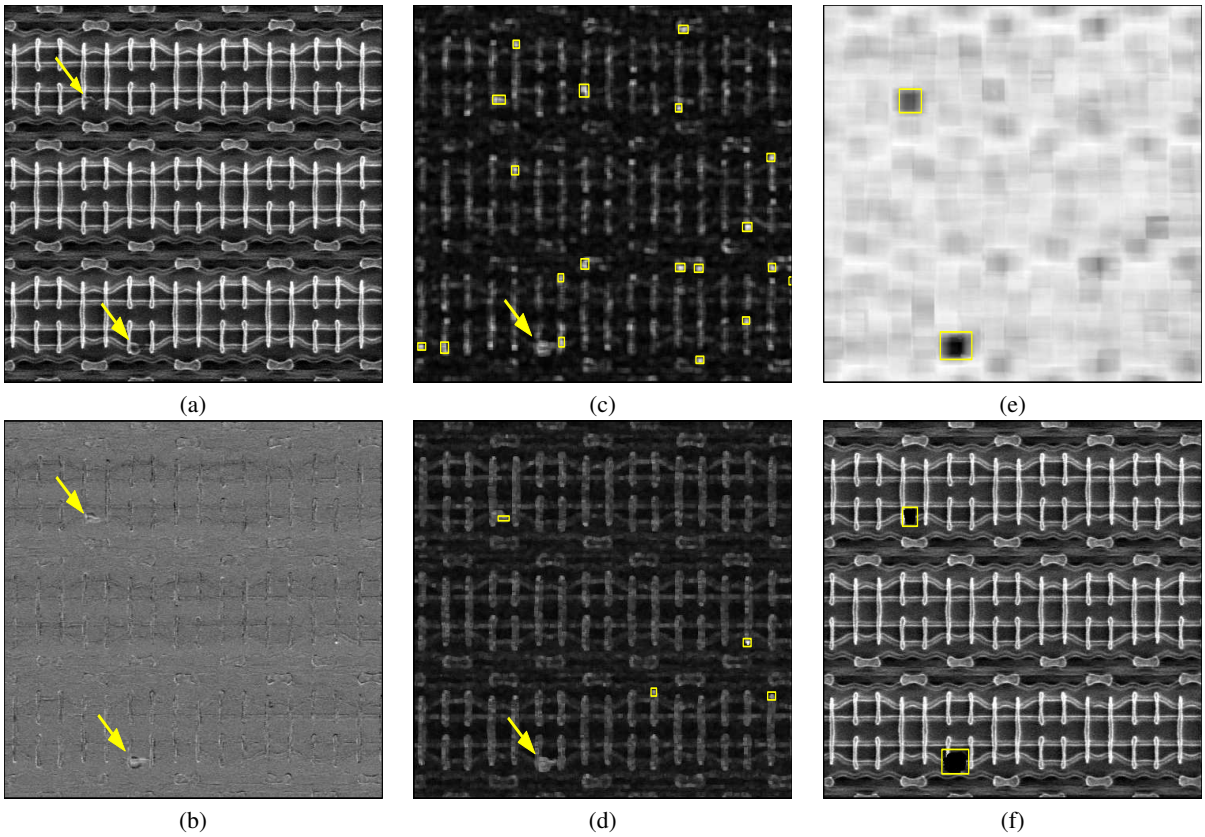


Fig. 2. (a)-(b) Source and difference images (images' **(!**FOV) is equal to 4μ, arrows point at defects); (c)-(d) Single Hypothesis Test (SHT) and Onishi-based detection results (rectangles denote the detected region, arrows point at missed defects); (e)-(f) Kernel-based similarity measure (in logarithmic representation) and respectively reconstructed source image ($[29 \times 29]$ patches are used). In case of (c) and (d) detections, the defects could not be detected without additional false detections, due to the pattern variations. The kernel-based similarity measure allows robustness to pattern variations and achieves high separation ability and ideal detection, compared to the other presented methods.

3. EXPERIMENTAL RESULTS

In this section we demonstrate examples of the defect detection in patterned wafers using the proposed algorithm and compare it to two other methods. The first method applies generally anomaly detection algorithm, Single Hypothesis Test (SHT) [12], on the difference image $D(\mathbf{s}) = I_{\text{ref}}(\mathbf{s}) - I_{\text{src}}(\mathbf{s})$, after spatial alignment of the images[1]. Anomalies are often associated with localized groups of pixels, hence it is common for the anomaly decision at a given pixel $\mathbf{s}$ to be based on a small block of pixels in the neighborhood of $\mathbf{s}$ in the image. Accordingly, a data set is constructed from overlapping patches formed around every pixel in the difference image $D(\mathbf{s})$. Given the expected vector M and the covariance matrix Σ of the constructed data set, the SHT, applied to the Mahalanobis distance of any vector $\mathbf{X}$ from M, is given by $d^2(\mathbf{X}) = (\mathbf{X} - M)^T \Sigma^{-1} (\mathbf{X} - M) \underset{H_1}{\overset{H_0}{\lessgtr}} \eta^2$, where H_1 and H_0 represent hypotheses of anomaly presence and absence, respectively. η is a distance threshold, which could be determined statistically according to the desired false detection rate [13].

A simple pixel by pixel difference image handles only the translation differences between the images. However, the source and reference image patterns are not identical and pattern variations may occur. These differences could be as intense as the differences caused by defects, which violate the required assumption about statistics of the data set, $\mathbf{X}$ and may cause false detections. Onishi *et al.* [14] tried to overcome the problem of slight distortion or local rotation misalignment between the source and reference patterns by using gray-scale morphological dilation of the reference and inspected images. The difference image is calculated according to the minimal distance between the reference and inspected images in the dilation range. However, this technique can only manipulate slight misregistration and pattern variations.

The first example is presented in Fig. 1: a source image (a) contains two defects, marked by arrows: a big, conspicuous defect at the right and smaller and less noticeable defect at the left of the image. The detection results of the aforementioned SHT algorithm (based on the difference image (c)) and Onishi algorithm are brought in (d) and (e). These algorithms don't succeed to handle existing pattern variations, and the small defect could not be detected without false

[1]The exact registration of the source and reference images is required only for the SHT-based algorithm, however for convenience the same registered reference image is used in all the three procedures.

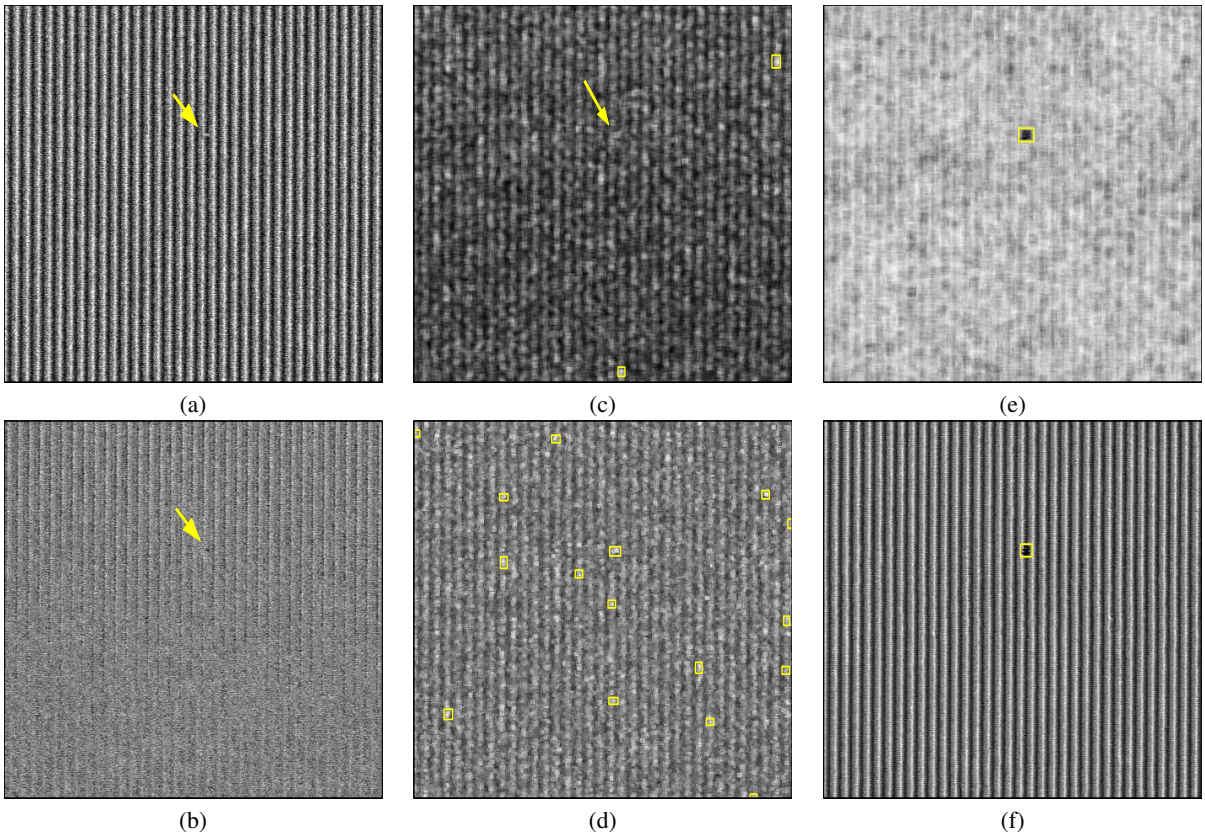
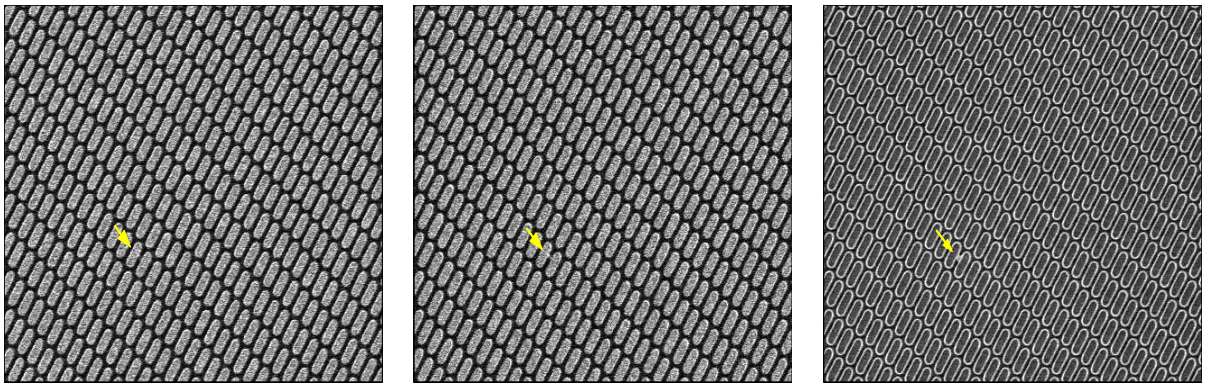


Fig. 3. (a)-(b) Source and difference images (images' FOV is equal to 5μ, arrow point at defect); (c)-(d) SHT and Onishi-based detection results (rectangles denote the detected region, arrows point at missed defect); (e)-(f) Kernel-based similarity measure (in logarithmic representation) and respectively reconstructed source image ($[11 \times 11]$ patches are used). Detections presented in (c) and (d) demonstrate high false detection rate and miss detection. The kernel-based similarity measure allows robustness for pattern variations and achieves high separation ability and ideal detection, compared to the other presented methods.

detections. The kernel-based reconstruction, presented in Section 2, is robust to the pattern variations and achieves ideal detection in (f). Additional example is brought in Fig. 2, where one of the two defects is missed and false detections occur in both SHT (c) and Onishi (d) detection procedures. Fig. 2(e) presents a likelihood image (total similarity to the pattern of every pixel), which demonstrates high separation ability of the kernel-based algorithm and ideal detection of the defects. The third example in Fig. 3 presents less complicated pattern with one defect. In this example the references image could not be ideally registered to the source image, due to slight rotation. The detection ability of the presented kernel-based algorithm outperforms other two methods in this case as well.

Finally, we discuss the computational efficiency of the proposed algorithm. The computational complexity is O($n \cdot m \cdot d$) (n denotes the number of pixels in the image, m denotes the number of reference patches and d is a column-stacked patch size), hence implementation on typical home computer results in high computational load. In order to improve the performance a highly parallel Graphical Processing Unit (GPU) implementation of the similarity measure calculation was developed[2]. The implementation runs a grid of thread blocks that perform parallel calculation for all pixels. Each block has 16×16 threads and it calculates a single pixel result. To get best performance, each block loads a source patch, which represents a specific pixel, and corresponding reference search region into block shared memory and performs parallel exponents calculation and sum up in the block shared memory. Generally the block shared memory is smaller than the size of source patch and reference search region, the data is loaded and processed in several steps and each block performs a number of load-calculate iterations. The computation rate of the implementation of the detection procedure for $d = 11 \times 11$ and $m = 49 \times 49$ (sizes used in our third example) executed on GeForce 8800 GT [15] was ~ 75 KPPS (Kilo Pixels per Second). These results can be improved by implementation optimization and by using state of the art hardware like GeForce 200 Series which has x4 processor cores and almost x4 memory bandwidth than the hardware we have used.

[2]The authors thank Mr. Yuri Pekelny for providing a CUDA implementation of the algorithm.



(a) (b) (c)

Fig. 4. (a)-(c) Source images in different channels : *External*$_1$, *External*$_2$, *Internal* respectively; The external images indicate the topography of the sample by light and shadows as if a "light source" is directed to a sample from top-left (*External*$_1$) or top-right (*External*$_2$). The internal image provides information about edges and material of the sample. In the presented sample the defect is more salient in the *Internal* channel.

4. MULTI-CHANNEL EXTENSION

Novel SEM tool, manufactured by Applied Materials, can simultaneously produce three different images for a given sample, namely *External*$_1$, *External*$_2$ and *Internal* images, as for example demonstrated in Fig. 4. The external images indicate the topography of the sample by light and shadows as if a "light source" is directed to a sample from top-left (*External*$_1$) or top-right (*External*$_2$). The internal image provides information about edges and material of the sample. Spatial alignment between images from the three channels is a byproduct of the capturing process. Defects may look more salient in one channel than in others. However, given an arbitrary defect, it is impossible to know in which channel the detection should be hold. Hence, the detection results from all the channels should be incorporated. We propose to use a consistency criterion [8], which allows simultaneous detection based on the information from all the channels. We assume that if a pattern-originated region in the source wafer is similar to certain regions in the reference wafer, then this similarity is maintained across the three SEM images. Accordingly, the similarity between a source patch and its reference patches in the three channels is constrained by a consistency criterion that the locations of reference patches, which are most similar to the source patch, are identical in the three channels. Reader is referred to [8] for more profound explanation of the consistency criterion and its statistical interpretation.

Let us denote by x, y, z corresponding features of the pixel $\mathbf{s}$ in the three channels (*External*$_1$, *External*$_2$ and *Internal*), the joint similarity measure of the three channels then is

$$\hat{p}(x,y,z) = \frac{1}{m}\sum_{i=1}^{m} k(x_i,x)k(y_i,y)k(z_i,z)\ . \qquad (7)$$

Using Gaussian kernel the above joint similarity measure could be viewed as a single similarity measure with a combined feature:

$$\mathbf{s} \rightarrow v = \begin{bmatrix} \sqrt{\varepsilon_y\varepsilon_z}x \\ \sqrt{\varepsilon_x\varepsilon_z}y \\ \sqrt{\varepsilon_x\varepsilon_y}z \end{bmatrix} \qquad (8)$$

and combined similarity parameter $\varepsilon = \varepsilon_x\varepsilon_y\varepsilon_z$ Figure 5 demonstrates that combing the single channel detection results (overlaid on single similarity measures in Fig. 5(a)-(c)) by 'AND' or 'OR' operations over the three channels may either decrease the detection rate or increase the false detection rate, whilst the multi-channel algorithm (Fig. 5(d)) enables perfect detection.

5. CONCLUSIONS

We have presented a kernel-based defect detection procedure, which exploits the periodic nature of the wafer pattern and compensates for pattern variations and miss-registration. Registration of a reference image relative to the source image is not required, as long as reference patches are taken from a wide search region that covers at least one pattern period. Furthermore, a source patch does not have to be identical to one reference patch, but could be a combination of several patches to overcome the problem of pattern variations. The presented procedure not only compares the grey level in a single point but also incorporates the information of the neighborhood, using patches as features. Experimental results have demonstrated the advantages of the presented method over competitive methods. Efficient implementation of the proposed algorithm enables its use in industrial applications.

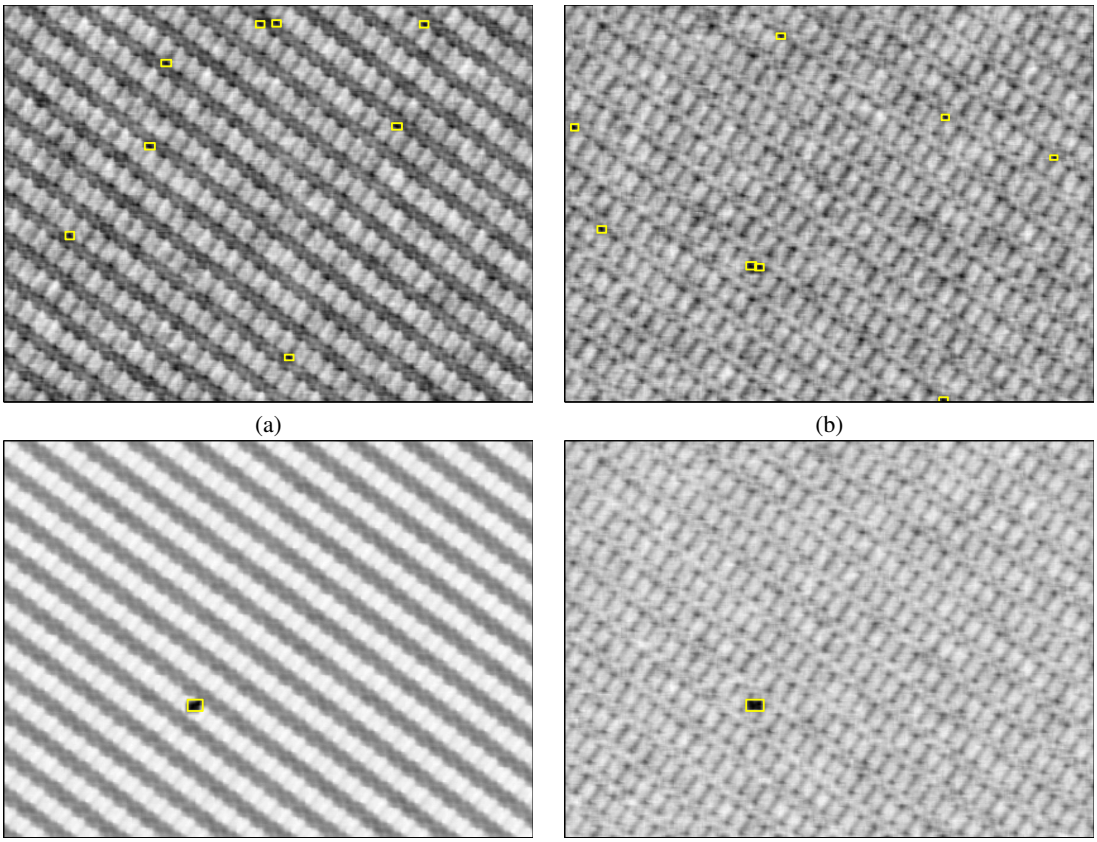
6. REFERENCES

[1] N. Shankar and Z. Zhong, "Defect detection on semiconductor wafer surfaces," *Microelectronic Engineering*, vol. 77, pp. 337–346, Apr. 2005.

[2] D. M. Tsai and C. H. Yang, "A quantile-quantile plot based pattern matching for defect detection," *Pattern Recognition Letters*, vol. 26, no. 13, pp. 1948–1962, Oct. 2005.

[3] D.-M. Tsai and C.-H. Yang, "An eigenvalue-based similarity measure and its application in defect detection," *Image and Vision Computing*, vol. 23, no. 12, pp. 1094–1101, Nov. 2005.

[4] T. Hiroi, C. Shishido, and M. Watanabe., "Pattern alignment method based on consistency among local registration



(a) (b) (c) (d)

Fig. 5. (a)-(c) Single channel similarity measures (in logarithmic representation) of $External_1$, $External_2$, *Internal* channels respectively with overlaid detection, (d) Joint similarity measure (in logarithmic representation) of the three channels with overlaid detection. The defect is separable from the pattern variations noise only in the *Internal* channel and the joint similarity measure allows perfect detection, whilst 'AND' or 'OR' operations over the three channels detection results may either decrease the detection rate or increase the false detection rate. The same threshold was used for all the cases above to detect outliers.

candidates for LSI wafer pattern inspection," in *Proc. 6th IEEE Workshop on Applications of Computer Vision*, Orlando, Florida, USA, 2002, pp. 257–263.

[5] C. Costa and M. Petrou, "Automatic registration of ceramic tiles for the purpose of fault detection," *Machine Vision and Applications*, vol. 11, no. 5, pp. 225–230, Feb. 2000.

[6] M. Zontak and I. Cohen, "Defect detection in patterned wafers using anisotropic kernels," Machine Vision and Applications, to be published. [Online]. Available: http://springerlink.com/content/c8p3724522221277

[7] A. Szlam, "Non-stationary analysis on datasets and applications," Ph.D. dissertation, Yale University, New Haven, Connecticut, USA, May 2006.

[8] M. Zontak and I. Cohen, "Defect detection in patterned wafers using multichannel scanning electron microsope," Signal Processing, to be published.

[9] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Modeling and Simulation*, vol. 4, no. 2, pp. 490–530, 2005.

[10] E. Parzen, "On estimation of a probability density function and mode," *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 1065–1076, Sep. 1962.

[11] A. Ruiz and P. E. Lopez-de-Teruel, "Nonlinear kernel-based statistical pattern analysis," *IEEE Transactions on Neural Networks*, vol. 12, no. 1, pp. 16–32, Jan. 2001.

[12] A. Goldman and I. Cohen, "Anomaly detection based on an iterative local statistics approach," *Signal Processing*, vol. 84, no. 7, pp. 1225–1229, Jul. 2004.

[13] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. San Diego, CA, USA: Academic Press, 1990.

[14] H. Onishi, Y. Sasa, K. Nagai, and S. Tatsumi, "A pattern defect inspection method by parallel grayscale image comparison without precise image alignment," in *Proc. 28th IEEE Annual Conference of the Industrial Electronics Society*, vol. 3, Santa Clara, CA, Nov. 2002, pp. 2208 – 2213.

[15] NVIDIA. [Online]. Available: http://www.nvidia.com