2014 IEEE 28-th Convention of Electrical and Electronics Engineers in Israel

Multi-Channel Wafer Defect Detection Using Diffusion Maps

Gal Mishne and Israel Cohen
Technion - Israel Institute of Technology, Electrical Engineering Dept.
Haifa 32000, Israel
{galga@tx, icohen@ee}.technion.ac.il

***Abstract*—Detection of defects on patterned semiconductor wafers is a critical step in wafer production. Many inspection methods and apparatus have been developed for this purpose. We recently presented an anomaly detection approach based on geometric manifold learning techniques. This approach is data-driven, with the separation of the anomaly from the background arising from the intrinsic geometry of the image, revealed through the use of diffusion maps. In this paper, we extend our algorithm to 3D data in multichannel wafer defect detection. We test our algorithm on a set of semiconductor wafers and demonstrate that our multiscale multi-channel algorithm has superior performance when compared to single-scale and single-channel approaches.**

***Index Terms*—Wafer defect detection, anomaly detection, diffusion maps, dimensionality reduction, multiscale representation**

I. Introduction

Defect detection is critical to the manufacturing of semiconductor wafers, yet relying on manual detection is time consuming, expensive and may cause yield ratio loss. A robust automated solution to this problem is essential, as the user will be shown only suspicious regions, thus saving valuable time. Defect detection is challenging as there are no precise characteristics of the possible defects and they may include particles, open lines, shorts between lines or other problems. Defects may belong to the wafer background or to its pattern, and may be predominant or scarcely noticeable. This variety makes it very difficult to perform template matching based on some a-priori features or training database of detects, and therefore encourages the development of unsupervised, data-driven methods.

Various image processing techniques have been applied to automatic defect detection in wafers. A common approach for wafer defect detection utilizes a defect-free reference image and detection is performed on the difference between the reference and inspection images [1], [2], [3]. Calculation of the difference image is very sensitive to image registration between the reference and inspection images, and can affect the performance of reference-based methods. Zontak and Cohen [4], [5], introduced a method which avoids image registration and is robust to pattern variations, based on anisotropic kernel reconstruction of the source image using its reference image. Defect regions are identified since they cannot be properly reconstructed from the reference image. A semiconductor wafer typically contains many copies of the same electrical component laid out in a matrix pattern, and this repetitive pattern can be utilized for detection without a reference image. Guan et al. [6] proposed to generate a golden-block database from the wafer image itself, and then modify and refine its content when used in further inspections of the same pattern. Gleason et al. [7] employ fractal image encoding and active contours for defect detection based on self-similarities within the inspection image. Unsupervised neural networks which do not require reference images have been proposed in [8], [9].

Recently, we presented a multiscale algorithm for anomaly detection based on spectral dimensionality reduction [10], [11]. This approach is data-driven, with the separation of the anomaly from the background arising from the intrinsic geometry of the image, revealed through the use of diffusion maps [12]. The algorithm is unsupervised, requires no prior knowledge regarding the anomaly or the background, and does not rely on training data or reference images. When applying diffusion maps to images, it is common practice to perform sampling and out-of-sample extension due to the large size of the dataset. We showed [10] that this process can limit the success of the dimensionality reduction in revealing the presence of anomalies in the data. To overcome these limitations, we proposed a multiscale approach, which drives the sampling process to ensure separability of the anomaly from the background clutter. In [11], we presented an anomaly detection score inspired by the saliency map developed by Goferman, Zelnik-Manor and Tal [13]. We adapted the color-based dissimilarity measure used in [13] to the diffusion map feature space used in our algorithm, as a noise-robust feature space is more appropriate for the task of anomaly detection in noisy images. In this paper, we apply our anomaly detection algorithm to the problem of defect detection in patterned wafers, specifically multichannel defect detection in Scanning Electron Microscope (SEM) images. The application of diffusion maps to defect detection is extremely appropriate as our algorithm provides a natural extension from single channel to multichannel images, the diffusion distance is robust to the noise in the SEM images and the diffusion map provides a compact representation of the repetitive geometry of the wafer.

The paper is organized as follows. Sec. II reviews the diffusion map framework for dimensionality reduction and Sec. III describes out-of-sample extension methods and their limitations in anomaly detection. In Sec. IV, the proposed multiscale algorithm is presented. Finally, Sec. V demonstrates the application of the proposed algorithm to defect detection in SEM images of patterned wafers.

II. Diffusion Maps

Real world data is usually represented with features of high dimensionality, yet can be shown to lie on low-dimensional manifolds. Finding a low-dimensional representation of the data is necessary to efficiently handle it and the representation usually reveals meaningful structures within the data. In recent years, a large number of nonlinear techniques for dimensionality reduction have been proposed, some of which are spectral methods, based on the eigenvectors of adjacency matrices of graphs on the data [14], [12]. These methods take into account the geometry of the dataset and the representation they yield preserves local neighborhood information. Diffusion maps [12] is one such technique, based on the construction of the graph Laplacian of the dataset.

Let $\Gamma = \{x_1, .., x_n\}$ be a high-dimensional set of n data points. A weighted graph is constructed with the data points as nodes and the weights of the edges connecting two nodes is a measure of the similarity between the two data points. The affinity matrix

This research was supported by the Israel Science Foundation (grant no. 1130/11).

$\mathbf{W} = w(x_i, x_j)$, $x_i, x_j \in \Gamma$ is required to be symmetric and non-negative. The choice of the weight function should be determined by the application, since it conveys the local geometry of the dataset. A common choice is a Radial basis function (RBF) kernel $w(x_i, x_j) = \exp\left\{-\|x_i - x_j\|^2/\sigma^2\right\}$, where $\sigma > 0$ is a scale parameter. A random walk is created on the dataset by normalizing the kernel:

$$\mathbf{P} = \mathbf{D}^{-1}\mathbf{W}, \tag{1}$$

where $\mathbf{D}(i,i) = \sum_{j\in\Gamma} w(x_i, x_j)$. The row-stochastic matrix $\mathbf{P}$ satisfies $p(x_i, x_j) \geq 0$ and $\sum_{j\in\Gamma} p(x_i, x_j) = 1$ and can be viewed as the transition matrix of a Markov chain on the dataset Γ. The spectral decomposition of $\mathbf{P}$ yields that t steps of the Markov chain can be presented as

$$p_t(x_i, x_j) = \sum_{l\geq 0} \lambda_l^t \psi_l(x_i)\phi_l(x_j), \tag{2}$$

where ϕ_l and ψ_l are the biorthogonal left and right eigenvectors, respectively, and $|\lambda_0| \geq |\lambda_1| \geq ... \geq 0$ are the sequence of eigenvalues.

A diffusion distance $d_{\text{DM}}(x_i, x_j; t)$ between two points $x_i, x_j \in \Gamma$ is defined using the eigen-decomposition of $\mathbf{P}$ (2) by

$$\begin{aligned} d_{\text{DM}}(x_i, x_j; t) &= \sum_{x_k\in\Gamma} \frac{\big(p_t(x_i, x_k)) - p_t(x_j, x_k)\big)^2}{\phi_0(x_k)} \\ &= \sum_{l\geq 1} \lambda_l^{2t}(\psi_l(x_i) - \psi_l(x_j))^2. \end{aligned} \tag{3}$$

This measures the similarity of two points according to the evolution of their probability distributions in the Markov chain. This metric is robust to noise, since the distance between two points depends on all possible paths of length t between the points. Due to the spectrum decay, the diffusion distance can be approximated using only the first ℓ eigenvectors. Thus, the computational complexity of the diffusion distance is low given the eigen-decomposition of $\mathbf{P}$.

Equation (3) implies that a mapping can be defined between the original space and the eigenvectors ψ_l, defining a new set of coordinates for the dataset Γ, such that the diffusion distance is equal to the Euclidean distance in this new embedding. Retaining only the first ℓ eigenvectors, the diffusion map embeds the dataset Γ into the Euclidean space $\mathbb{R}^\ell$:

$$\Psi_t : x_i \to \left(\lambda_1^t\psi_1(x_i), \lambda_2^t\psi_2(x_i), ..., \lambda_\ell^t\psi_\ell(x_i)\right)^T. \tag{4}$$

Note that ψ_0 is not used because it is a constant vector. We set $t = 1$ since running the Markov chain forward can join the defect with the background.

The scale parameter σ is of great significance in constructing the weighted graph. Specifically in the setting of anomaly detection, setting σ to be too large will connect the anomalies with the cluttered background. Assuming the anomaly to be in a low density neighborhood and the background to belong to a high density neighborhood, a local scale factor is beneficial, such as the one proposed by Zelnik-Manor and Perona [15]. The scale σ is calculated for each point x_i based on the local statistics of its neighborhood:

$$\sigma_i = \|x_i - x_K\|^2 \tag{5}$$

where x_K is the K-th nearest neighbor. The similarity kernel is then calculated as $w(x_i, x_j) = \exp\left\{-\|x_i - x_j\|^2/\sigma_i\sigma_j\right\}$.

III. Function Extension

The size of the dataset for images is very large. Therefore, it can be computationally inefficient to construct a diffusion map using all the pixels in the image, especially for high-resolution images. Instead, it is a common approach [16], [17] to construct the diffusion map for an image using a subset of random samples, $\Gamma \subseteq \overline{\Gamma}$, and then the diffusion map coordinates Ψ are extended to the set of all patches in the image $\overline{\Gamma}$ using an out-of-sample extension method.

The Nyström method is a common method for extending functions from a given training set to new samples. Different methods have been proposed to approximate the Nyström method or improve upon it, such as the Geometric Harmonics method [18]. Recently, a new algorithm was presented for out-of-sample function extension using the multiscale Laplacian pyramid [19]. At each iteration, the Laplacian pyramid algorithm constructs a coarse approximation of a function f for a given scale. Then, the difference between f and the coarse approximation is used as input for the next iteration. The difference is approximated at each level using a Gaussian kernel with increasingly finer scales. For more details, see [19]. We perform this extension method for each diffusion coordinate $f = \Psi_l, l \in \{1, ..., \ell\}$ separately. The number of levels in the pyramid extension can differ between coordinates, dependent on their smoothness over Γ. A smooth function can be extended using coarse scales, i.e. will not require many levels of the pyramid. An oscillating function on the other hand will require finer and finer levels of the pyramid to enable an accurate extension.

As discussed in [10], out-of-sample extension methods can cause anomaly detection to fail, depending on the set of random samples $\Gamma \subseteq \overline{\Gamma}$ used to construct the diffusion map. In a case where there are no anomalies, or defects, in Γ and it consists only of samples from a single n-dimensional cluster (the background), then the eigenvectors capture only the relaxation process within this cluster [20]. In such a case, the diffusion map will not capture the difference between the defect and the background, and the out-of-sample extension of the diffusion map to the pixels in the defect region will not succeed in assigning them new coordinates which separate them from the background. Thus, defect detection when the samples from the defect are not included in the initial diffusion map requires extrapolation of the diffusion coordinates and not interpolation. However it is not clear how to perform extrapolation on the low-dimensional manifold, if at all possible. This is a "chicken and egg" problem in which it is necessary to sample the defect for the purpose of detecting it. To overcome this limitation of the out-of-sample extension, we propose a multiscale method which drives the sampling process and ensures the inclusion of samples from the defect region in Γ.

IV. Multiscale Anomaly Detection

To overcome the limitations of random sampling, we propose a multiscale approach. The defects in the inspection image are larger than a single pixel. Hence, they can be detected at several resolutions of the image. At a lower resolution, it is computationally possible to sample a larger percentage of the image for the construction of the diffusion map. Thus, detecting a defect at a lower resolution is less likely to fail due to sampling. In addition, performing detection at different resolutions of the image increases the robustness to miss-detections, since even if the defect is missed on a coarse level due to its size, it can still be detected on the following finer levels. In addition, it is possible to lower the threshold for detection on the coarser levels, since this will not harm the false alarm rate as a decision is only reached at the full-scale level. Thus, we are able

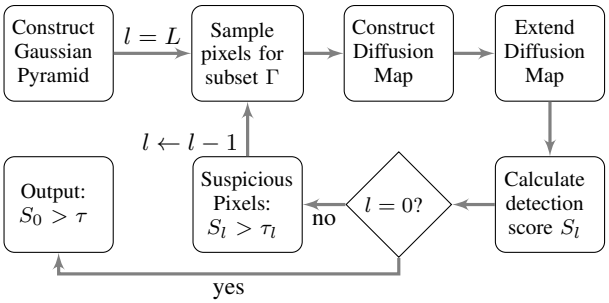


Fig. 1: Flowchart of the multiscale algorithm.

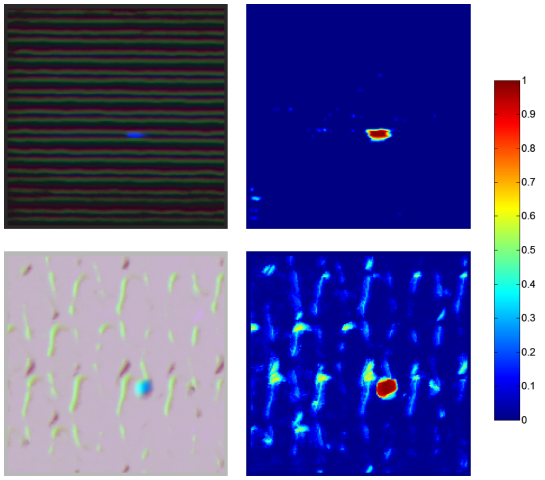
to detect defects on the higher levels, even at the cost of detecting more false alarms, since these false alarms will be removed at the final level.

Our multiscale approach is based on constructing a Gaussian pyramid representation of the image [21], yielding $\{G_l\}_{l=0}^{L}$, where G_0 is the original image and G_L is the coarsest resolution. Starting with G_L, a subset Γ_L of random pixels is sampled from the image. Since the image at this level is at very low resolution, the subset can include all pixels, depending on memory constraints. Using the diffusion coordinates, a detection score S_L is calculated for all pixels and then a threshold τ_l on the score determines suspicious pixels. Proceeding to the image G_{L-1}, pixels which correspond to the suspicious pixels found in G_L are included in Γ_{L-1}. The rest of the pixels in Γ_{L-1} are sampled randomly from the image. The threshold τ_l used at the output of each level is chosen to be the 95^{th} percentile of the detection score for that level. If the image does not hold a defect this will result in random samples with the highest scores. If the image holds a defect, the defect will have a high score compared to the rest of the image and it will be sampled more densely in the next level.

The process of sampling, dimensionality reduction and anomaly detection continues from level to level, with each previous level providing prior information on which samples of the dataset will be used in Γ_l to construct the diffusion map. At the full-scale level G_0, the detection score for each pixel determines the existence of defects in the image. We use a hard threshold τ on S_0 and then smooth the resulting image. Defects have a high score, close to 1. Figure 1 presents a flowchart of the algorithm. This approach greatly increases the detection rate of the diffusion-based defect detector, compared with a single scale approach.

A. Multi-Channel Defect Detection

In wafer inspection, the analyzed wafer is illuminated with electrons, which causes interactions on the surface of the wafer. These interactions lead to subsequent emission of electrons that provides information about the edges and the material of the inspected wafer. This information is rendered into a two-dimensional intensity distribution that can be stored as a digital image and analyzed for defect detection. An SEM tool that is manufactured by Applied Materials can simultaneously produce three different images for a given sample, namely External_1, External_2 and Internal images. 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), and are more noisy than the Internal image. The Internal image provides information about edges and material of the sample. Spatial alignment of the three images is a byproduct of the imaging process. Defects may appear more salient in one



(a) (b)

Fig. 2: Defects in two patterned wafers (top row and bottom row). (a) Image pixels colored according to RGB color associated with the first three coordinates of the diffusion map. (b) Anomaly score for multichannel multiscale (MCMS) method. Only the defects receive a high score.

channel than in others, and may not even be noticeable in the other channels. However, given an arbitrary defect, it is impossible to know in advance in which channel the defect is distinct. Hence, information from all three channels should be incorporated.

Our algorithm provides a natural extension from two-dimensional images to three-dimensional data as in this application. For a multichannel image, the feature vector associated with each pixel is the $N_p \times N_p \times n_c$ patch surrounding the pixel, ordered as a column, where $n_c = 3$ is the number of channels. The resulting diffusion map for multichannel images is dominated by both the geometry of the data, revealed by integration of the three channels, and the presence of the defect, which is distinct in at least one channel. Fig. 2 demonstrates the compact and denoised representation achieved by diffusion maps when applied to two multi-channel SEM wafer images (top row and bottom row). For display purposes, the first three coordinates of the diffusion map (Ψ_1, Ψ_2, Ψ_3) are mapped to RGB values $[0, 255] \times [0, 255] \times [0, 255]$. Then, each pixel in the image is colored (Fig. 2(a)) according to the RGB value assigned to its diffusion coordinates. The representation obtained by the diffusion map provides both a denoising of the image and a compact representation of the intrinsic geometry of the patterned wafer, integrating information from all three channels, and assigning repetitive components such as edges distinct values. Also, as expected, the defect is easily separable from the background in the diffusion embedding, as demonstrated by the high detection score of the defects in Fig. 2(b). Note that this display is only of the first three coordinates of the diffusion map and additional coordinates provide additional information on the geometry of the patterns and the defect.

The dimension of the diffusion map (4) depends only on the random walk (1) and is independent of the length of the feature vector used in the original representation of the data. Thus, using a multichannel feature vector does not entail a higher diffusion map dimension compared to using a single channel feature vector.

The detection problem can be viewed as a clustering problem in which we are separating between background clusters and the anomaly cluster. Using a combined multichannel feature vector

translates into adding dimensions in which the anomaly is distinct from the background. This increases the distance between the cluster centers, which leads to fewer misidentifications [22].

B. Saliency-based Detection Score

In [11], we introduced a saliency-based anomaly detection score. This score was inspired by the local-global dissimilarity measure defined between two image patches by Goferman et al. [13]:

$$d(p_i, p_j) = \frac{d_{\text{color}}(p_i, p_j)}{1 + c \cdot d_{\text{position}}(p_i, p_j)} \tag{6}$$

where d_{color} is a color distance between patches and d_{position} is the Euclidean distance between the image positions of patches p_i and p_j, normalized by the larger image dimension. This measure realizes the authors' observations that background pixels are similar to both near and far pixels, whereas salient patches are grouped together, therefore similar only to nearby patches. Normalizing the color distance by the spatial distance penalizes background pixels by assigning them a low distinctness value. In addition, the authors note that in order to evaluate the distinctness of a patch it is sufficient to consider its K most similar patches $\{q_k\}_{k=1}^{K}$ ($K = 64$), and not calculate its dissimilarity to all image patches.

These observations holds for anomaly detection as well. The problems of anomaly detection and saliency in images are closely related, where an anomaly can be viewed as a salient object in the image. However, while saliency is important in natural images, anomaly detection is usually performed in images which are not natural and suffer from noise, such as in the given defect detection problem in SEM images. Thus, instead of using d_{color}, the color distance between patches, we proposed using the diffusion distance, which is preferable in our application as it is robust to noise. Also, the diffusion embedding better separates the anomaly from the background, compared to using image patches. This requires normalizing the diffusion distance such that most values are spread out in the range $[0, 1]$ and therefore comparable to d_{position}. The detection score we proposed is given by

$$S(i)_{\text{DM}} = 1 - \frac{1}{K} \sum_{k=1}^{K} \exp\left\{ -\frac{d_{\text{DM}}(p_i, q_k)/2\sigma_K}{1 + c \cdot d_{\text{position}}(p_i, q_k)} \right\}, \tag{7}$$

where $c = 3$ and σ_K is a normalizing factor given by the standard deviation of the distances to the K^{th} neighbor: $\sigma_K = \underset{i \in \Gamma}{\text{std}}\{d_{\text{DM}}(p_i, q_K)\}$. This score has the following advantages:

- Background regions which have similar diffusion coordinates, yet are spatially distant from one another in the image are suppressed. On the other hand, pixels from defects lie close together, yet are distinct from all other regions, thus they receive a high score.
- No prior knowledge is required regarding the defect size.
- This score requires very little fine-tuning, only K and c need to be determined.

For more information on this score and its performance in other applications, see [11]. Our code is available at [23].

V. Experimental Results

We demonstrate the proposed algorithm on SEM patterned wafer images, achieving a high detection rate with a low rate of false-alarms. We treat the defects in the images as anomalies and the patterned wafer is considered normal background clutter. We evaluated our algorithm on a set of 36 images of size 200×200 pixels. The parameters of the multiscale detector are given in Table I, for a Gaussian pyramid of $L = 3$ levels. Note that the size of the images enables denser sampling of the image than what we used, however we intentionally use a small percentage of the pixels to demonstrate that this framework is also applicable for larger images.

TABLE I: Parameters Used in Multiscale Detector.

Pyramid Level	Patch size ($N_p \times N_p$)	Embedding Dimension	Percentage of pixels in subset
0	8x8	6	0.10
1	4x4	6	0.33
2	2x2	3	0.5

Detections are found by applying a threshold to the detection score image, resulting in a binary image. A detection is a connected component (CC) in the binary image. A CC containing the defect is a true positive (TP) and any other CCs are false alarms (FA). The size of the CC can be used to reject noisy detections, where small CCs are discarded. We compare two thresholds on the area of the CC: 5 pixels and 20 pixels. Using a larger threshold on the size rejects more FAs, but can also result in a decreased amount of TPs, for small sized defects. We compared the percentage of TPs for each method for a given FA rate. Results are given in Table II.

We compare our proposed multi-channel multiscale algorithm (MCMS) with four other methods. The first three methods perform single-channel multiscale detection on each of the three channels: External_1, External_2 and Internal, separately, using the parameters given in Table I. The fourth method is MCSS, a single-scale method using the full-size image. The parameters of MCSS are the same as those for pyramid level 0 in Table I, however 20% of the image is randomly sampled to construct the diffusion map. For all methods, the detection score is the one described in Sec. IV-B.

Our new MCMS approach has the highest TP rate. Comparison to the three single-channel methods demonstrates that the defects are usually not apparent in all three channels, and combining information yields the best results. We can see that the Internal channel is usually the least informative, and although both External images hold similar topographical information, their performance is not identical. Also, as in our previous work, using a multiscale approach results in increased performance compared to performing detection only on the full-scale image. The poor performance of the single-scale detector, MCSS, reveals the limitations of sampling the image compared with the multiscale detectors, which have a significantly better detection rate. Indeed, this shows that having more information on the presence of a defect by combining the three channels as in MCSS is less beneficial to defect detection than properly sampling a single channel, as in the single-channel multiscale detectors.

We present an example of all detection methods for a given patterned image in Fig. 3. For External_2-MS, Internal-MS and MCSS, Fig. 3 (f),(g) and (h) respectively, there are no detections with high score. Thresholding the image with a low threshold will reveal the defect for Internal-MS and MCSS, but will also cause many FAs. External_1-MS, Fig. 3 (e), detects the defect with a high score but also has a FA with similar score. Only in MCMS, Fig. 3 (d), the defect receives a high score, while all other regions receive low scores.

VI. Conclusions

We have extended our multiscale anomaly detection algorithm to 3D data in multichannel wafer defect detection. The proposed algorithm is tested on a dataset of semiconductor wafer SEM images and demonstrated superior performance when compared to single-scale and single-channel approaches. Our method, based on diffusion

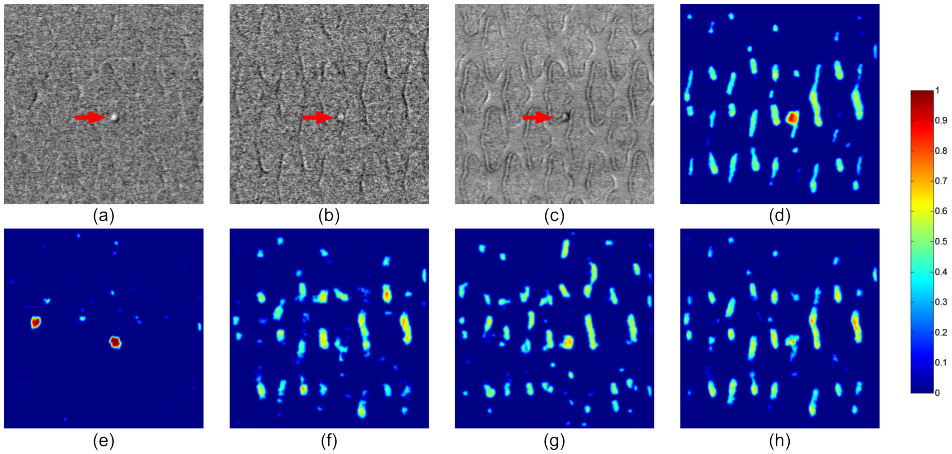


Fig. 3: Defect in a patterned wafer. (a)-(c) Source images in different channels: External$_1$, External$_2$ and Internal respectively, defect marked with a red arrow. Anomaly scores for multichannel multiscale (MCMS) method (d), single channel multiscale method in External$_1$ (e), External$_2$ (f) and Internal (g) channels, and for multichannel single-scale (MCSS) method (h).

TABLE II: Percentage of True Positives for Given Numbers of False Alarms.

	size=5			size=20		
# of FA	10	3	0	10	3	0
MCMS	97%	97%	97%	97%	97%	92%
MCSS	70%	70%	61%	70%	67%	58%
External$_1$-MS	89%	89%	78%	92%	89%	83%
External$_2$-MS	92%	92%	89%	94%	92%	89%
Internal-MS	75%	75%	75%	75%	75%	75%

map embedding, is especially attractive in this application as the diffusion map provides a compact representation of the repetitive geometry in the wafer image, obviating the need for a reference image, and in addition is robust to the dominant noise in these images. This algorithm was previously applied to side-scan sonar images in the detection of sea-mines. This paper demonstrates the robustness of the proposed algorithm and its robustness to the imaging sensor, background and noise models.

REFERENCES

[1] N. Shankar and Z. Zhong, "Defect detection on semiconductor wafer surfaces," *Microelectronic Engineering*, vol. 77, no. 3–4, pp. 337 – 346, 2005.

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

[3] H. W. Kim and S. I. Yoo, "Defect detection using feature point matching for non-repetitive patterned images," *Pattern Anal. Applic.*, vol. 17, no. 2, pp. 415–429, 2014.

[4] M. Zontak and I. Cohen, "Defect detection in patterned wafers using anisotropic kernels," *Machine Vision and Applications*, vol. 21, no. 2, pp. 129–141, June 2008.

[5] ——, "Defect detection in patterned wafers using multichannel scanning electron microscope," *Signal Process.*, vol. 89, no. 8, pp. 1511–1520, Aug. 2009.

[6] S. U. Guan, P. Xie, and H. Li, "A golden-block-based self-refining scheme for repetitive patterned wafer inspections," *Mach. Vis. Appl.*, vol. 13, no. 5, pp. 314–321, 2003.

[7] S. S. Gleason, R. K. Ferrell, T. P. Karnowski, and K. W. Tobin, Jr., "Detection of semiconductor defects using a novel fractal encoding algorithm," in *Proceedings of SPIE, Design, Process Integration, and Characterization for Microelectronics*, vol. 4692, 2002, pp. 61–71.

[8] C.-Y. Chang, S.-Y. Lin, and M. Jeng, "Using a two-layer competitive hopfield neural network for semiconductor wafer defect detection." in *Proc. IEEE International Conference on Automation Science and Engineering, no. 5*, 2005, pp. 301–306.

[9] C.-Y. Chang, C. Li, J.-W. Chang, and M. Jeng, "An unsupervised neural network approach for automatic semiconductor wafer defect inspection," *Expert Systems with Applications*, vol. 36, no. 1, pp. 950 – 958, 2009.

[10] G. Mishne and I. Cohen, "Multiscale anomaly detection using diffusion maps," *IEEE J. Sel. Topics Signal Process.*, vol. 7, pp. 111 – 123, Feb. 2013.

[11] ——, "Multiscale anomaly detection using diffusion maps and saliency score," in *Proc. 39th IEEE Internat. Conf. Acoust. Speech and Signal Process. (ICASSP-2014)*, 2014.

[12] R. R. Coifman and S. Lafon, "Diffusion maps," *Appl. Comput. Harmon. Anal.*, vol. 21, no. 1, pp. 5–30, July 2006.

[13] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," in *In CVPR*, 2010, pp. 2376–2383.

[14] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003.

[15] L. Zelnik-Manor and P. Perona, "Self-tuning spectral clustering," in *NIPS 17*, 2005, pp. 1601–1608.

[16] Z. Farbman, R. Fattal, and D. Lischinski, "Diffusion maps for edge-aware image editing," *ACM Trans. Graph.*, vol. 29, no. 6, pp. 145:1–145:10, Dec. 2010.

[17] J. He, L. Zhang, Q. Wang, and Z. Li, "Using Diffusion Geometric Coordinates for Hyperspectral Imagery Representation," *IEEE Geosci. Remote Sens. Letters*, vol. 6, no. 4, pp. 767–771, Oct. 2009.

[18] R. R. Coifman and S. Lafon, "Geometric harmonics: a novel tool for multiscale out-of-sample extension of empirical functions," *Appl. Comput. Harmon. Anal.*, vol. 21, pp. 31–52, 2006.

[19] N. Rabin and R. R. Coifman, "Heterogeneous datasets representation and learning using diffusion maps and Laplacian pyramids," in *Proc. 12th SIAM International Conference on Data Mining*, 2012.

[20] B. Nadler, S. Lafon, R. R. Coifman, and I. G. Kevrekidis, "Diffusion maps - a probabilistic interpretation for spectral embedding and clustering algorithms," in *Principal Manifolds for Data Visualization and Dimension Reduction*. Springer, 2007.

[21] P. Burt and E. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Communications*, vol. 31, no. 4, pp. 532–540, 1983.

[22] A. Singer, Y. Shkolnisky, and B. Nadler, "Diffusion interpretation of nonlocal neighborhood filters for signal denoising," *SIAM Journal Imaging Sciences*, vol. 2, no. 1, pp. 118–139, Jan. 2009.

[23] [Online]. Available: http://webee.technion.ac.il/Sites/People/IsraelCohen/