Computer Communication & Collaboration (Vol. 3, Issue 3, 2015)
ISSN 2292-1028(Print) 2292-1036(Online)
Submitted on Jul. 6, 2015
DOIC: 2292-1036-2015-03-003-58

Reference-based Defect Detection on *OLED* Images using Local Gaussian Distribution

Suil Son (Correspondence author)
Department of Computer Science and Engineering,
Seoul National University, Seoul, Republic of Korea
E-mail: likerivers12@snu.ac.kr

Suk I. Yoo
Department of Computer Science and Engineering,
Seoul National University, Seoul, Republic of Korea
E-mail: sukinyoo@snu.ac.kr

Abstract: The difficulty of detecting defects on images usually stems from their small deviation of features from normal data. However, when the direction of the deviation of such data is considered with respect to the local distribution, even the small deviation can cause large difference in the distribution. Therefore, to detect defects not easily differentiated, we propose a novel measure of defect reflecting local Gaussian distribution of data points. We form the local Gaussian distributions from local data sets of data points and measure the difference of those distributions. To compute the distance between two Gaussian distributions, we utilize the Fisher information distance. Since the proposed method has the scale invariant property, it also attains stable detection results on data sets with different scale. We evaluated the performance of our method through experiments of defect detection on images of organic light-emitting diode (*OLED*). Compared to traditional detection methods, our method showed better detection results.

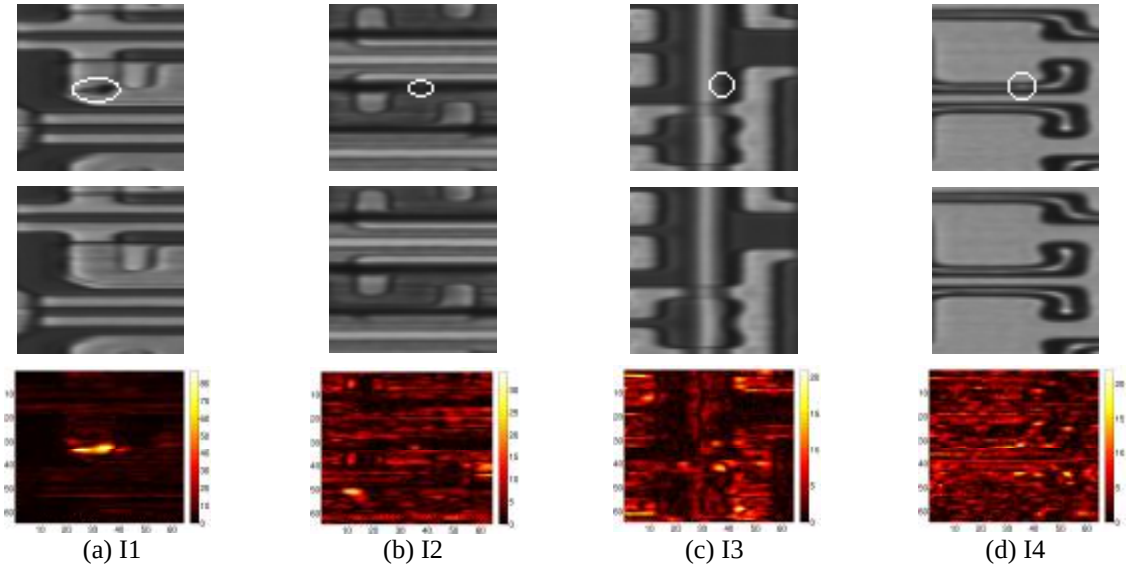
Keywords: Defect Detection, Outlier Detection, Local Gaussian Distribution, Fisher Information Distance

1 Introduction

Defect detection on images has been applied to various manufacturing process such as semiconductors, liquid crystal displays (*LCD*), and organic light-emitting diode (*OLED*). The presence of defects is determined by analyzing whether the features of images are probable or not. Since this analysis contributes to high quality and a low cost in production, various defect detection methods have been suggested.

Defect detection methods can be classified into two types of approaches: detection with reference images [1,2,3] and detection without reference images [4,5,6,7]. To detect defects, the

first approach compares features of an inspection image with those of associated reference images. The second approach examines whether features of an inspection image satisfy the conditions of defect-free images or not. When reference images are used, methods comparing features strongly



(a) I1 (b) I2 (c) I3 (d) I4

Figure 1. Examples of defects in *OLED* images. The first row: inspection images with defects. Defects are in the white ellipses. The second row: the associated defect free reference images. The third row: difference of intensity between each the inspection images and the reference images.

influence the performance of defect detection. For this comparison, various methods have been suggested, but they often suffered from false positive detections on images with much noise and high variations of objects.

Outlier detection research has also studied similar detection problems. According to Hawkins [8], ”an outlier is an observation that deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism.” This definition of an outlier can be regarded as an interpretation of defect's feature appearing together with normal features. Therefore, outlier detection methods can be applied to the defect detection with reference images.

Based on the definition of an outlier, various outlier detection methods have been suggested so as to differentiate the deviation of an outlier. Real world data sets are usually composed of various data generated by different mechanisms. In these data sets, a data point has little relation to data points located far away from it. Therefore, local outlier models [9,10,11,12,13] that consider data points only in the vicinity of an interested data point have been suggested. In these models, a data point with small deviation still has a possibility of an outlier if its neighbors have even smaller deviations.

However, it is still difficult to differentiate some outliers from normal data points because they have relatively small differences from other data points. This is easily observed in the problem of defect detection on *OLED* images. For example, images of *OLED*s are illustrated in Figure 1. In this figure, the images in the first row are defective images, the images in the second row are the associated normal images, and the figures in the third row represent the differences between the the defective images and the normal images. As can be seen in the third row, the defect in Figure 1 (a) is easily detected, but the defects in Figure 1 (b), (c) and (d) are not because their differences are not noticeable, compared to those of some normal pixels.

Though an outlier or a defect has small difference, it can be differentiated by considering

direction of its deviation. A data point deviating in a different direction from others causes a larger difference in its local distribution than a data point deviating in the same direction. To reflect this characteristic, we propose a method reflecting the distribution of local data points. In our approach, the local distribution of a data point is obtained from a data set derived from its *k*-nearest neighbors (*k-NN*), and outlierness of it is computed from differences between several local distributions, one of which is for the data point, and the others are for its *k-NN*. Since local data sets have simple patterns, their distributions are assumed to follow the Gaussian distributions. Then the distance between distributions is computed by the Fisher information distance [14,15,16]. We call the proposed measure of outlier or defect as *the local Gaussian distribution difference* (*LGDD*). Since our *LGDD* has the scale invariant property, it can also attain stable detection results on data sets with different scale.

The remainder of this paper is organized as follows. In Section 2, related works are reviewed. We define and explain the method to compute *LGDD* in Section 3. Then we describe a process of application of our method for defect detection on images in Section 4. In Section 5, the experimental results on *OLED* images are illustrated. Finally, we conclude in Section 6.

2 Related works

In defect detection research, various comparison-based approaches have been suggested. A simple method is to subtract the intensity of an inspection image from that of the associated reference image. Xie and Guan [17] generated a golden-template from the wafer image, and then they calculated the difference between the golden-template and an inspection image. Since the subtraction-based methods are vulnerable to illumination change, Tsai [18] proposed a method using the normalized cross correlation (*NCC*) between two images. The *NCC*-based method detects defects by computing the similarity of the sub-regions of two images, but this method was sensitive to noise and alignment. Lee and Yoo [3] formulated the defect detection problem as a binary labeling problem. In this approach, probabilistic model of a normal image was constructed by using the kernel density estimation (*KDE*), and then it was used as the likelihood of feature value. After this likelihood was combined with prior distribution given by uniform distribution, the label of each pixel was determined by the graph-cut algorithm [20].

In order to detect outliers, various outlier detection methods have been suggested. In statistics, Gaussian modeling for a given data set was suggested. In this method, the mechanism generating data was modeled as a Gaussian distribution that fits the data well, and then data points were determined to be outliers if they had too large deviation from the mean of the distribution. The guideline for the deviation was 3σ, where σ denotes the standard deviation of the Gaussian distribution. However, when the data contain several mechanisms, the Gaussian distribution is inappropriate to model them. Therefore, the Gaussian mixture model [21] or *KDE* [19,22,23] was used to model such complex data, and then the data points with low probability density were determined to be outliers. These approaches, however, suffered from the curse of dimensionality problem in high dimensional data which cause low density for most of the data.

In data mining research, Knorr *et al*. [24] suggested a global distance-based outlier method which labels a data point as an outlier if a particular fraction of all data points in a data set has larger distance from the data point than the specified distance. Like the Gaussian modeling in statistics, this method had difficulty in detecting outliers for a data set with several mechanisms. To overcome this difficulty, Breunig *et al*. [9] proposed the local outlier factor (*LOF*) computed from local reachability density (*lrd*), which is obtained from *k-NN*. By computing the outlierness of a data point as a relative value of the *lrd* with respect to *lrd*s of its neighbors, this method could detect an outlier with relatively low density in a local dense area. After the work, *LOF*, various methods to reflect local structures of data have been suggested. *COF* [10] represented the local structure by using the minimum spanning tree for *k-NN* so as to detect an outlier with quite high density but with an isolated pattern. *LOCI* [11] introduced a multi-granularity deviation factor from ε-neighbors

rather than *k-NN*. To avoid the problem caused by asymmetry of neighborhoods, *INFLO* [12] included the reverse neighbors of a data point which count the point as their *k-NN*, while the reverse does not. *LDOF* [13] used the relative location of a data point to its neighbors to determine the degree of deviation to each neighbor.

(a-1) (a-2) (b-1) (b-2)

Figure 2. Comparing local Gaussian distributions. (a-1) and (b-1): Each ellipse stands for the local Gaussian distribution formed by the associated data point with the same color and its neighbors. (a-2) and (b-2): The ellipses are rearranged by translating the position of the associated data points to the coordinate origin. They are obtained from (a-1) and (b-1), respectively. The mark + stands for the coordinate origin.

Methods applying the information theory also have been suggested. Lee *et al*. [25] presented various information-theoretic measures for anomaly detection. In [26,27], the information content of a data point was assessed by the Kullbak-Leibler divergence (*KL*-divergence) or changes of entropy between two data sets, one of which includes a data point of interest, and the other does not.

Though various defect detection and outlier detection methods were developed, a method comparing distributions of local data sets was not suggested. In this paper, the distance between local Gaussian distributions is reflected in outlierness or defectness.

3 Local Gaussian distribution difference (*LGDD*)

We compute the degree of outlier by comparing local Gaussian distributions of data points. The concept of comparing local Gaussian distributions is illustrated in Figure 2, where data points are represented by bullets. Here, it is examined whether the red data point is an outlier or not. For the examination, we consider the local distribution formed by the data point and its neighbors where the number of neighbors is assumed to be two for simplicity. Since local data sets do not have complex distribution, the distributions are represented by the Gaussian distributions. The distributions are visualized by the ellipses with the same color to the associated data points because the level sets of the Gaussian distributions are ellipses. In order to compare the distributed pattern of neighbors of a data point of interest, these ellipses are rearranged as in Figure 2 (a-2) and (b-2) by moving the associated data point of each of the ellipses to the coordinate origin. Then, as can be seen in Figure 2 (b-2), the ellipses have similar shapes. The shape of the red ellipse in Figure 2 (a-2), however, is different from others because the red data point deviates in different direction from other data points. Therefore, the deviation of the red data point in Figure 2 (a-1) can be differentiated. This concept is reflected in our approach.

To compute outlierness, we firstly find *k*-nearest neighbors of data points. Given a data set $D \subset \quad^{M}$, let $d(\mathbf{p},\mathbf{q})$ denote the Euclidean distance between $\mathbf{p}$ and $\mathbf{q}$ in D, and let *k-dist*($\mathbf{p}$) denote the distance from $\mathbf{p} \in D$ to $\mathbf{p}$'s k^{th}-nearest neighbor in D in terms of the Euclidean distance. *The k-nearest neighbors* (*k-NN*) of $\mathbf{p}$ is then defined as

$$N_k(\mathbf{p};D) := \{\mathbf{q} \in D \mid d(\mathbf{p},\mathbf{q}) \le k\text{-}dist(\mathbf{p}), \mathbf{q} \ne \mathbf{p}\}. \tag{1}$$

Using *k-NN*, we define the outlierness of a data point $\mathbf{p}$ as *the local Gaussian distribution difference* (*LGDD*):

$$LGDD(\mathbf{p};D,k) := \frac{\sum_{\mathbf{q} \in N_k(\mathbf{p};D)} d_{ld}(ld(\mathbf{p}), ld(\mathbf{q}))}{|N_k(\mathbf{p};D)|}, \tag{2}$$

where $ld(\cdot)$ denotes the local distribution of the given data point, and d_{ld} computes the distance between two distributions. The methods to obtain the local distribution and to compute the distance are explained in the following sub-sections.

3.1 Centered local data set

The local distribution $ld(\mathbf{p})$ is obtained from a data set which is constructed from $N_k(\mathbf{p}; D)$. Since the local distribution is for representing the distributed pattern of $N_k(\mathbf{p}; D)$ around the data point $\mathbf{p}$, for the computation of $ld(\mathbf{p})$, we use *the centered local data set* (*CLDS*) of $\mathbf{p}$, which is defined to be

$$CLDS(\mathbf{p}; D, k) := \{\mathbf{x} \mid \mathbf{x} = (\mathbf{q} - \mathbf{p}), \mathbf{q} \in N_k(\mathbf{p}; D)\} \cup \{\mathbf{0}\}. \tag{3}$$

In *CLDS*, the coordinates of the data points in $N_k(\mathbf{p};D)$ are translated by the coordinate of $\mathbf{p}$. As a results, the set $CLDS(\mathbf{p};D,k)$ reflects relative positions of the neighbors of $\mathbf{p}$ with respect to $\mathbf{p}$ regardless of absolute coordinates of the original data points. The point zero is included so that this forms quite different distribution from that without it in the case where the data point $\mathbf{p}$ is an outlier.

3.2 Local Gaussian distributions and the Fisher information distance

To compute *LGDD*, a method to measure distance between local distributions is necessary. We compute the local distributions from *CLDSs* derived from neighbors of data points. If neighbors are formed in a sufficiently small local region, a local distribution can be approximated by a Gaussian distribution. Therefore, we utilize the method measuring distance between Gaussian distributions.

Though there are many other methods measuring distance between Gaussian distributions [28] such as *KL*-divergence and Hellinger distance, it is known that with some arithmetic, these distance measures converge to the Fisher information distance [29]. Also, according to the theory of information geometry, the Fisher information distance is a geodesic distance between two distributions on the statistical manifold [14,30]. Thus, we consider the Fisher information distance.

In information geometry, which studies geometrical structures on manifolds of probability distributions [31], much research has been devoted to distance between distributions. In the early study of the information geometry, Rao [32] introduced a method for measuring distance between two distributions over the parametric space in the Riemannian geometry based on the Fisher information matrix. The Riemannian metric using the Fisher information matrix is known as the Fisher information metric or the Fisher information distance. Atkinson and Mitchell [14] derived the Fisher information distance for Gaussian distribution with the same covariance as

$$d_F((\boldsymbol{\mu}_1, \boldsymbol{\Sigma}), (\boldsymbol{\mu}_2, \boldsymbol{\Sigma})) = \left((\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\right)^{1/2} \tag{4}$$

where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are mean vectors, and $\boldsymbol{\Sigma}$ is the same covariance matrix; or with the same mean as

$$d_F^2((\boldsymbol{\mu}, \boldsymbol{\Sigma}_1), (\boldsymbol{\mu}, \boldsymbol{\Sigma}_2)) = \frac{1}{2}\sum_{j=1}^{M}(\ln l_j)^2, \tag{5}$$

where $\boldsymbol{\mu}$ is the same mean vector, $\boldsymbol{\Sigma}_1$ and Σ_2 are covariance matrices, and λ_j are the eigenvalues of matrix $(\boldsymbol{\Sigma}_1)^{-1}\boldsymbol{\Sigma}_2$. Costa *et al.* [15] derived the Fisher information distance for single variable Gaussian distribution. They further developed the Fisher information distance for multivariable Gaussian distribution with diagonal covariance matrix [16]:

$$
\begin{aligned}
& d_F(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \\
& = d_F((m_{11}, s_{11}, ..., m_{1M}, s_{1M}), (m_{21}, s_{21}, ..., m_{2M}, s_{2M})) \\
& = \sqrt{2 \sum_{i=1}^{M} \left(\ln \frac{\left\| \left(\frac{m_{1i}}{\sqrt{2}}, s_{1i} \right) - \left(\frac{m_{2i}}{\sqrt{2}}, -s_{2i} \right) \right\| + \left\| \left(\frac{m_{1i}}{\sqrt{2}}, s_{1i} \right) - \left(\frac{m_{2i}}{\sqrt{2}}, s_{2i} \right) \right\|}{\left\| \left(\frac{m_{1i}}{\sqrt{2}}, s_{1i} \right) - \left(\frac{m_{2i}}{\sqrt{2}}, -s_{2i} \right) \right\| - \left\| \left(\frac{m_{1i}}{\sqrt{2}}, s_{1i} \right) - \left(\frac{m_{2i}}{\sqrt{2}}, s_{2i} \right) \right\|} \right)^2}
\end{aligned}
\tag{6}
$$

where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are the parameters of the Gaussian distributions, μ_{1i} and μ_{2i}, $i \in \{1,2,\ldots,M\}$ are the elements of the mean vectors, and σ_{1i}^2 and σ_{2i}^2 are the diagonal elements of the covariance matrices, respectively.

In general, we cannot assume that two distributions have the same mean or the same covariance. Therefore, we utilize the Fisher information distance given by Eq. (6). This Fisher information distance also has the good property of scale invariance.

Theorem 1. (d_F is scale invariant.) Given that $\boldsymbol{\theta}_1 = (\mu_{11}, \sigma_{11}, ..., \mu_{1M}, \sigma_{1M})$ and $\boldsymbol{\theta}_2 = (\mu_{11}, \sigma_{11}, ..., \mu_{1M}, \sigma_{1M})$ where μ_{1i} and μ_{2i}, $i \in \{1,2,\ldots,M\}$ are the means, and σ_{1i} and σ_{2i} are the standard deviations of the given data, then $d_F(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = d_F(c \cdot \boldsymbol{\theta}_1, c \cdot \boldsymbol{\theta}_2)$, where $c > 0$.

Proof. See Appendix A.□

Using this Fisher information distance, we compute distance between two local Gaussian distributions as follows. Given two data points $\mathbf{p}, \mathbf{q} \in D \subset \mathbb{R}^M$, their local distributions are represented by

$$
\begin{aligned}
ld(\mathbf{p}) &= \boldsymbol{\theta}_{\mathbf{p}} = (m_{\mathbf{p}1}, s_{\mathbf{p}1}, ..., m_{\mathbf{p}M}, s_{\mathbf{p}M}), \\
ld(\mathbf{q}) &= \boldsymbol{\theta}_{\mathbf{q}} = (m_{\mathbf{q}1}, s_{\mathbf{q}1}, ..., m_{\mathbf{q}M}, s_{\mathbf{q}M}),
\end{aligned}
\tag{7}
$$

each of which is a parameter of a Gaussian distribution obtained from $CLDS(\mathbf{p};D,k)$ and $CLDS(\mathbf{q};D,k)$, respectively. Then we compute the distance between $ld(\mathbf{p})$ and $ld(\mathbf{q})$ as

$$
d_{ld}(ld(\mathbf{p}), ld(\mathbf{q})) = \frac{1}{\sqrt{M}} d_F(\boldsymbol{\theta}_{\mathbf{p}}, \boldsymbol{\theta}_{\mathbf{q}}). \tag{8}
$$

where d_F is the Fisher information distance given in Eq. (6). The division by square root of M is for preventing the distance from being too large when the dimension of data is very high. If we use this distance, our *LGDD* is also scale invariant.

Theorem 2. (*LGDD* is scale invariant.) Given a data set $D_1 = \{\mathbf{p}_1, \ldots, \mathbf{p}_N\}$, let a data set $D_2 = \{\mathbf{o}_1, \ldots, \mathbf{o}_N\} = cD_1 = \{c\mathbf{p}_1, \ldots, c\mathbf{p}_N\}$, where $\mathbf{o}_i = c\mathbf{p}_i$ and $c > 0$. Then $LGDD(\mathbf{p}_i;D_1,k) = LGDD(\mathbf{o}_i;D_2,k)$.

Proof. See Appendix B.□

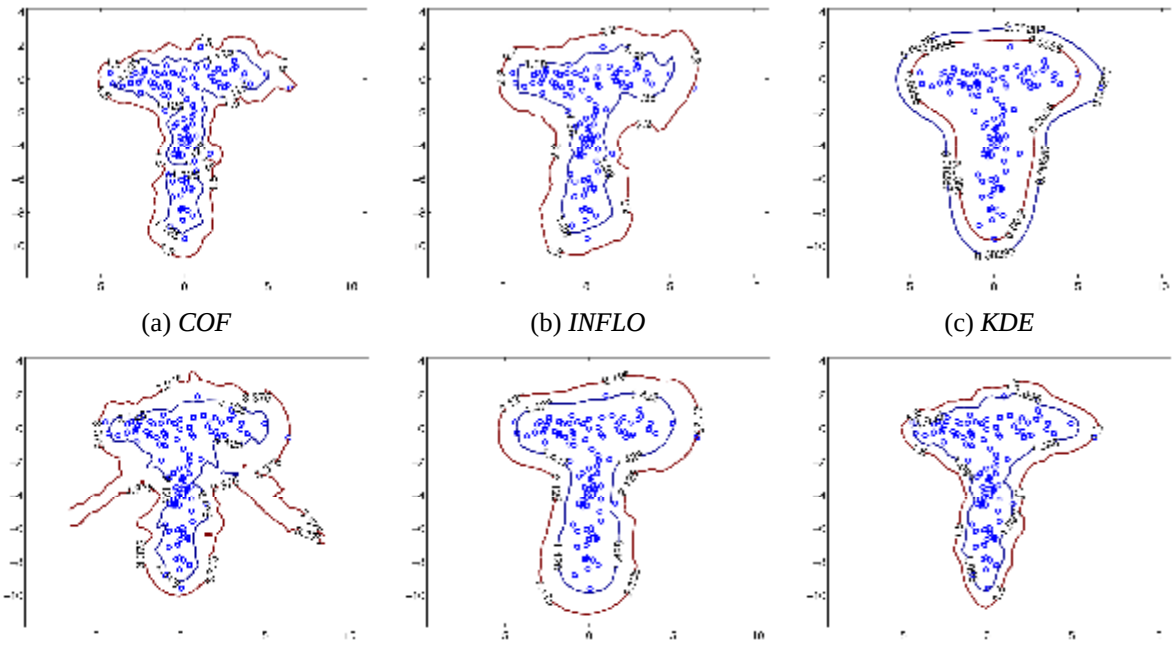
This scale invariance plays an important role for stable detection of outlier or defect. For several data points at different positions, their local data sets may be in different scale. Depending on the scale of data, the degree of outlier of many methods varies in a wide range. Nevertheless, we usually have to determine whether data points are outliers or not with a common threshold. In our approach, the effect of the different scale of data is mitigated by the scale invariant property.

3.3 Analysis

To analyze the characteristics of the proposed method, we computed the outlierness of a data set with *T*-shape. This data set was constructed by random samples generated from two Gaussian

distributions, one of which is spreading horizontally, and the other is spreading vertically. We plotted the degree of outlier from each method in Figure 3. The number of neighbors was set to be k =8.

The contours in Figure 3 show the same degrees of outlier. When the value of a contour is selected as a threshold, the inside of the contour is the region where normal data points lie, and the outside is where outliers occur. Thus, the shapes of contours show the characteristics of each method which determine whether some data points are outliers or not. The contours were deliberately selected so that the outer contour crosses the rightmost data point. Compared to contours of other method, that of our method shows narrower *T*-shape. This implies that in our *LGDD*, small deviation of outliers in different direction can be differentiated, while large deviation of normal data in the same direction to neighbors is allowed.



(a) *COF* (b) *INFLO* (c) *KDE*

(d) *LDOF* (e) *LOF* (f) *LGDD*

Figure 3. Contour of the degree of outlier from each method for a data set with *T*-shape.

4 Detecting defects on images

In this section, we explain the process of defect detection on images with reference images including construction of a data set for the detection. The definition of outlier can be regarded as an interpretation of defect's feature appearing together with normal features. Outlier detection methods, therefore, can be applied to the defect detection with reference images. In this application, after features of an inspection image and its reference images are combined into a data set, a defect is determined by the degree of outlier of the feature extracted from the inspection image.

Given an inspection image and N reference images, each image is denoted by a function $g_i : X \times Y \rightarrow Z$, where $X = \{1,2,...,W\}$, $Y = \{1,2,...,H\}$, $Z = \{0,1,...,255\}$, and $i \in I = \{0, 1, ..., N\}$. Here, W is the width and H is the height of the image, g_0 denotes the inspection image, and $g_1,...,g_N$ denote each of the reference images. Under these notations, $g_i(x,y)$ gives the intensity of the pixel at position (x,y) in the i^{th} image.

Defect detection of an image is composed of pixel-wise outlier detection. Therefore, a data set should be constructed for each of the pixel-wise outlier detection. Let $D_{(x,y)}$ be the data set for the

defect detection of the pixel at (x,y) in an inspection image. $D_{(x,y)}$ is constructed from a set of pixels, $P_{(x,y)}$. For each pixel in $P_{(x,y)}$, a feature vector is extracted, and it composes $D_{(x,y)}$. The construction of $P_{(x,y)}$ is performed as follows. First, the pixel at (x,y) position in an inspection image is selected. Second, a B B window is located at (x,y) position in each reference images, and then all pixels inside this window are selected. Finally, those selected pixels are combined into the set $P_{(x,y)}$. Since any pixel in the given images can be represented by its index of images and its position in those images, this construction of a set of pixels is expressed as

$$P_{(x,y)} = \{(i,u,v) \mid i \in I \setminus \{0\}, u \in \{x-B,\ldots,x,\ldots,x+B\}, v \in \{y-B,\ldots,y,\ldots,y+B\}\} \cup \{(0,x,y)\} \quad (9)$$

This $P_{(x,y)}$ also contains a few pixels in reference images at positions which are different from the pixel $(0,x,y)$. This allows slight shape variation of objects in images.

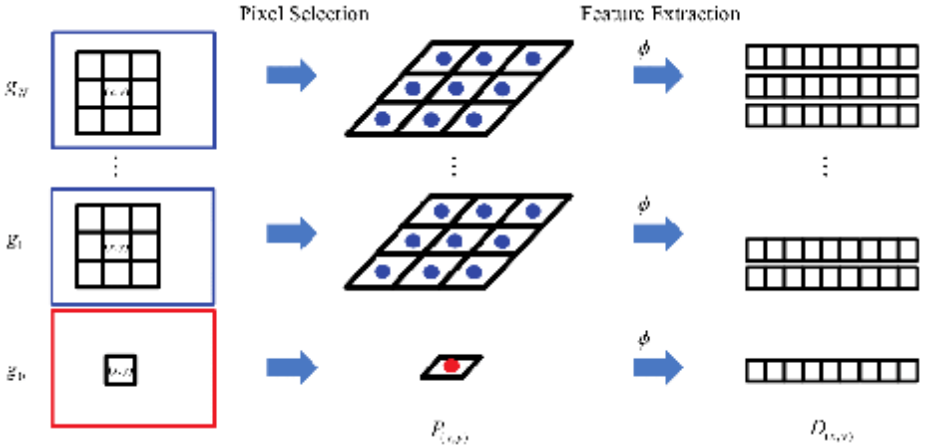


Figure 4. Construction of a data set for defect detection of a pixel on images.

Algorithm 1: Defect detection on an image with associated reference images.

Data:

g_0: an inspection image. g_i, $i \in \{1, \ldots, N\}$: associated N reference images.

H: height of the image. W: width of the image.

k: the number of neighbors.

T: threshold.

ϕ: feature extraction function in Eq. (10).

Result:

R: Result image

begin

for $y = 1$ **to** H **do**

for $x = 1$ **to** W **do**

construct $P_{(x,y)}$ in Eq. (9).

construct $D_{(x,y)}$ in Eq. (10) from g_j, $j \in \{0,1, \ldots, N\}$.

construct $N_k(F_\mathbf{p};D_{(x,y)})$ in Eq. (1), where $F_\mathbf{p} = \phi(0,x,y)$.

construct $CLDS(F_\mathbf{p}; D_{(x,y)}, k)$ in Eq. (3).

compute $ld(F_\mathbf{p})$ from $CLDS(F_\mathbf{p}; D_{(x,y)}, k)$.

foreach $F_{\mathbf{q}}$ **in** $N_k(F_{\mathbf{p}}; D_{(x,y)})$ **do**
construct $N_k(F_{\mathbf{q}};D_{(x,y)})$ in Eq. (1).
construct $CLDS(F_{\mathbf{q}}; D_{(x,y)}, k)$ in Eq. (3).
compute $ld(F_{\mathbf{q}})$ from $CLDS(F_{\mathbf{q}}; D_{(x,y)}, k)$.
end
compute $LGDD(F_{\mathbf{p}};D_{(x,y)},k)$ via Eq. (2) by using $F_{(\cdot)}$s.
label pixel (x,y) of R via $L(x,y;T)$ in Eq. (11).
end
end
end

After pixel selections are completed, from each pixel in $P_{(x,y)}$, we extract 9 1 feature vector: intensity, x and y positions, two first order intensity differences, and four second order intensity differences. Finally, the combination of these feature vectors constructs the data set $D_{(x,y)}$. This is formulated as

$$D_{(x,y)} = \{f(i,u,v) \mid f(i,u,v) = (g_i(u,v), u, v, \nabla g_i(u,v), \mathbf{vec}(\nabla^2 g_i(u,v))), (i,u,v) \in P_{(x,y)}\} \tag{10}$$

where ϕ is a feature extraction function for a given pixel, and **vec** means the vectorization of a matrix. This data set is used to compute the *LGDD* for determining whether pixel $(0,x,y)$ is defective or not. The overall process of a data set construction is illustrated in Figure 4, where the value of B is 3.

Finally, a decision on each pixel in an inspection image is determined by a labeling function,

$$L(x,y;T) = \begin{cases} 1, \text{ if } LGDD(F_{\mathbf{p}}; D_{(x,y)}, k)) > T, \\ 0, \text{ otherwise,} \end{cases} \tag{11}$$

where $F_{\mathbf{p}} = \phi(0,x,y)$, T is a threshold, 1 means that the pixel $(0,x,y)$ is defective, and 0 means that it is normal.

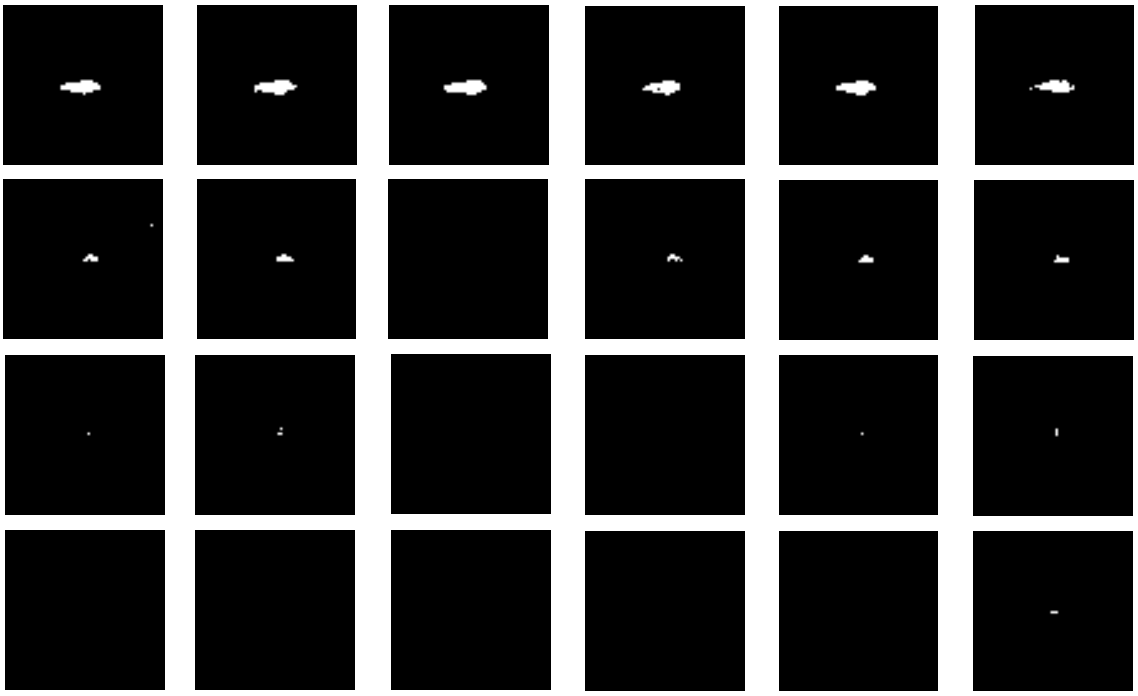
We summarized this overall process in Algorithm 1. In the computation of *LGDD* in Eq. (8) and Eq. (6), (μ_{1i}, σ_{1i}) and (μ_{2i}, σ_{2i}) could have the same value. In this case, we ignore the value of the dimension i and reduce M by one.

5 Experimental results

We conducted defect detection experiment on images of *OLED* displays as following the process in Section 4. The results from the proposed method, *LGDD*, were compared to those from several outlier or defect detection methods including *COF* [10], *INFLO* [12], *KDE* [3], *LDOF* [13], and *LOF* [9]. The images are taken from manufacturing process of *OLED* displays. The defect detection was conducted on 87 inspection images with 111 defects in total. For an inspection image, 6 reference images were used. The images in Figure 1 are the examples of those images: the images in the first row are the inspection images and those in the second row are one of the reference images. The size of an image is 64 64. The defect detection of an inspection image is composed of pixel-wise outlier detection as described in Section 4. Each data set for pixel-wise outlier detection

was constructed as in Eq. (10). The window size was set to B=3. All methods including ours except *KDE* have a parameter k, the number of neighbors, because they are local outlier models. The number of neighbors, k, was set to values from 6 to 10. Since *KDE* is a global model, it does not need the parameter k.

The defect detection results on the inspection images in Figure 1 are shown in Figure 5. The images from the first row to the fourth row in Figure 5 are for the images in Figure 1 (a)-(d), respectively. In Figure 5, pixels identified as defect are indicated by white. These results were determined by the parameters attaining the highest overall performance for each method. As shown in Figure 5, only our *LGDD* detected all defects. The defect of the inspection image in the fourth row of Figure 5, which all of the other methods failed to detect, is slightly different from normal pixels. This result shows that the small difference can be differentiated by our *LGDD*.



(a) *COF* (b) *INFLO* (c) *KDE* (d) *LDOF* (e) *LOF* (f) *LGDD*

Figure 5. Defect detection on *OLED* images in Figure 1.

To analyze characteristics of the methods, for the inspection images in Figure 5, we illustrated the outlierness maps, which represent degrees of a defect for each pixel, in Figure 6. The outlierness of *KDE* was taken by applying *-log* to the raw densities for better visualization in Figure 6. As can be seen in Figure 6, the outlierness maps of *COF*, *INFLO*, *KDE*, *LDOF*, and *LOF* showed very strong contrast between defective pixels and normal pixels in the first row. However, the contrasts were weak in the outlierness maps from the second row to the fourth row. On the other hand, our method consistently showed moderate contrast in all of the outlierness maps although their contrast in the first row was relatively weaker than other methods.

In Table 1, we listed the maximum value of each outlierness map in Figure 6 and computed the ratio of the highest value to the lowest one in each outlier detection method. As shown in Table 1, the ratios of ours showed the smallest value. This small value means that our method is stable in detecting defect over various images. This stability stems from the scale invariance of our *LGDD*. Due to this stability of outlierness, our *LGDD* could achieve better detection results by selecting

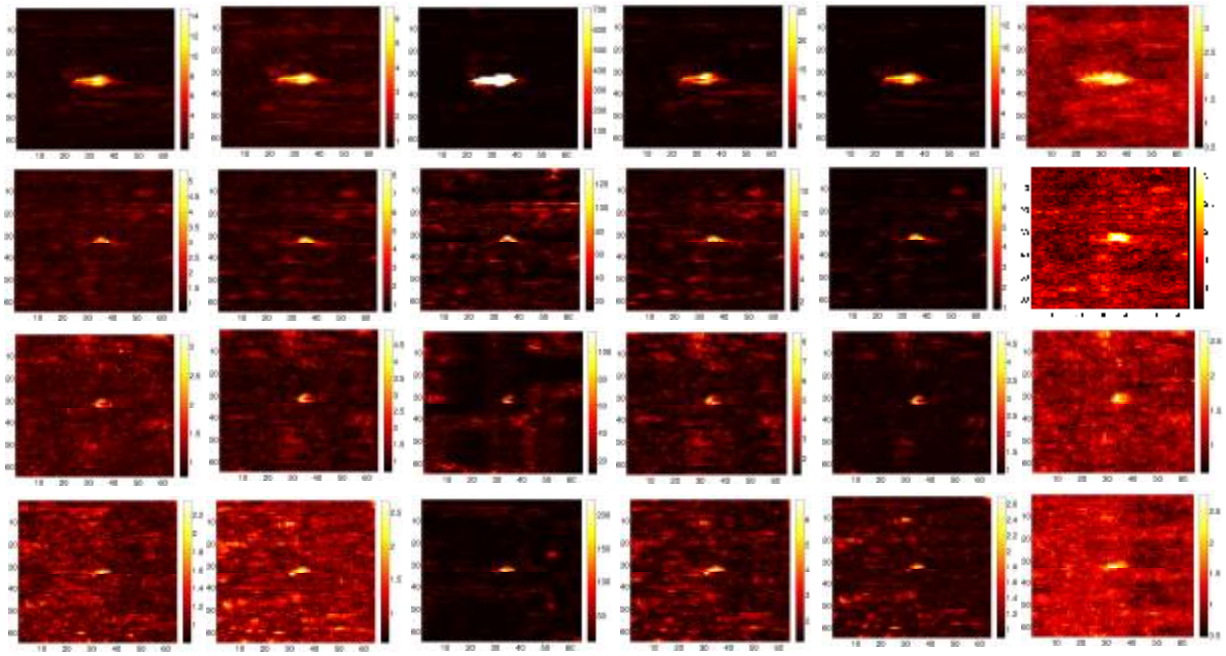
appropriate threshold. On the contrary, the thresholds could not be lowered in the traditional detection method because a few degrees at normal pixels in other images were close to the threshold or sometimes exceeded it. Hence, the lowering would cause false positive detections in other images.

To evaluate the quantitative performance of each method, *precision*, *recall*, and *F-measure* are computed as

$$precision = \frac{TP}{TP + FP}, \tag{12}$$

$$recall = \frac{TP}{TP + FN}, \tag{13}$$

$$F\text{-}measure = 2 \cdot \frac{precision \cdot recall}{precision + recall}, \tag{14}$$



(a) *COF* (b) *INFLO* (c) *KDE* (d) *LDOF* (e) *LOF* (f) *LGDD*

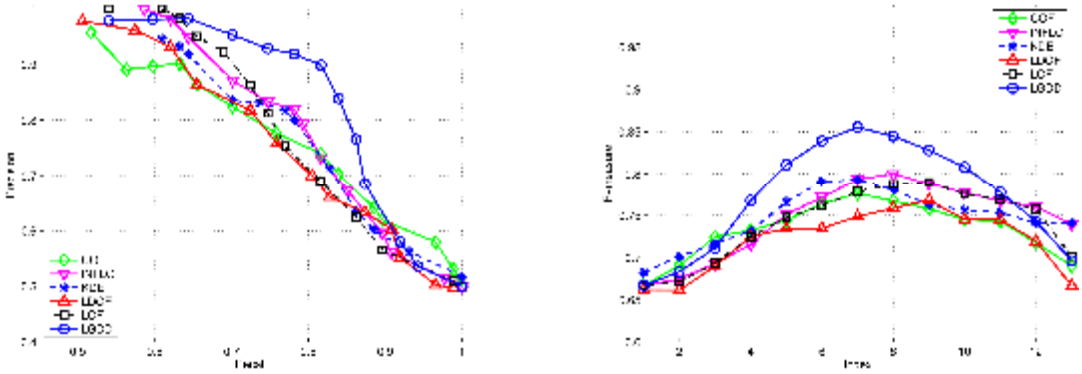
Figure 6. Outlierness map which represent degree of a defect for each pixel.

Table 1. The maximum value of each outlierness map in Figure 6 and the ratio of the highest value to the lowest one in each method.

	COF	*INFLO*	*KDE*	*LDOF*	*LOF*	*LGDD*
I1	**6.47**	**14.72**	**708.4**	**25.98**	**13.74**	**3.46**
I2	5.36	8.28	132.76	13.98	7.98	3.15
I3	3.2	4.95	114.84	8.46	4.83	2.65
I4	2.35	2.68	114.84	6.79	2.69	2.76
Ratio	2.75	5.49	6.17	3.83	5.11	1.31

where *TP* is true positive detection, *FP* is false positive detection, and *FN* is false negative detection. High *precision* means high correctness in positive decisions, and high *recall* means high detection rate for defects. There is a trade-off between them; therefore, the overall performance is determined by *F-measure* which simultaneously counts both *precision* and *recall*.

The change of performances is examined by varying thresholds, and the results are represented with *precision-recall* curves and with *F-measure* curves in Figure 7. The evaluation was performed by using various values of *k* from 6 to 10; the result attaining the best *F-measure* was selected. To compute the performance, *TP*, *FP*, and *FN* are counted for each inspection images. Positive detection for each defect is determined if any pixels in a segment of a defect are identified, and *TP* for each image is counted if all defects in that image are identified as positive detection. *FP* for each image is counted if any normal pixels are determined as defective pixels. Negative detection for each defect is determined if no detection was attained for that defect, and *FN* for each image is counted if any defects in that image are determined as negative detection.



(a) *precision-recall* curve. (b) *F-measure* curve.

Figure 7. Performances of defect detection by varying thresholds.

A *precision-recall* curve shows the change between *precision* and *recall* by varying threshold. *F-measure* shows overall performance considering both *precision* and *recall*. As shown in Figure 7, our *LGDD* showed better *precision* and *recall* than those of other methods. Consequently, our *LGDD* showed the highest *F-measure*. This result indicates that *LGDD* achieves better performance than traditional outlier detection methods for defect detection on *OLED* images.

6 Conclusion and future works

To detect defects not easily differentiated, we suggested a new measure of defect, called local Gaussian distribution difference (*LGDD*), which is computed from distance between local Gaussian distributions of data points. We computed the distance of local Gaussian distributions by using the Fisher information distance. Our *LGDD* has ability to differentiate the small deviation of a data point whose direction is different from those of its neighbors. We showed that our *LGDD* has a scale invariant property, which enables stable defect detection. The performance of our method was evaluated by experiments on images of organic light-emitting diode (*OLED*). Compared to several traditional outlier detection methods, our approach showed better performances in detecting defects not easily differentiated by the traditional detection methods.

In spite of the superiority of our method, there is still room for improvement. First, like other methods, our method is also vulnerable to noise. Most outlier detection methods depend on l_2-norm, *i.e.*, the Euclidean distance. The l_2-norm is the fundamental element of the Gaussian distribution. It is well-known that the l_1-norm is more robust to noise than the l_2-norm. The distribution related to

the l_1-norm is the Laplace distribution. Therefore, the distance between Laplace distributions may be a good alternative. Second, our method may not be good at detecting defects in low contrast images. Our method was devised for detecting defects on *OLED* images with high resolution. In practice, methods for low contrast images usually utilize transformed domains such as a discrete cosine transform (*DCT*) or a discrete wavelet transform (*DWT*). The combination of such approaches is another future direction.

Appendix

A. Proof of Theorem 1

Proof. By definition of d_F in Eq. (6)

$$
\begin{aligned}
& d_F(c\cdot\boldsymbol{\theta}_1, c\cdot\boldsymbol{\theta}_2) \\
& = d_F((c\cdot m_{11}, c\cdot s_{11}, ..., c\cdot m_{1M}, c\cdot s_{1M}), (c\cdot m_{21}, c\cdot s_{21}, ..., c\cdot m_{2M}, c\cdot s_{2M})) \\
& = \sqrt{2\sum_{i=1}^{M}\left(\ln\frac{\left|\left(\frac{c\cdot m_{1i}}{\sqrt{2}}, c\cdot s_{1i}\right)-\left(\frac{c\cdot m_{2i}}{\sqrt{2}}, -c\cdot s_{2i}\right)\right|+\left|\left(\frac{c\cdot m_{1i}}{\sqrt{2}}, c\cdot s_{1i}\right)-\left(\frac{c\cdot m_{2i}}{\sqrt{2}}, c\cdot s_{2i}\right)\right|}{\left|\left(\frac{c\cdot m_{1i}}{\sqrt{2}}, c\cdot s_{1i}\right)-\left(\frac{c\cdot m_{2i}}{\sqrt{2}}, -c\cdot s_{2i}\right)\right|-\left|\left(\frac{c\cdot m_{1i}}{\sqrt{2}}, c\cdot s_{1i}\right)-\left(\frac{c\cdot m_{2i}}{\sqrt{2}}, c\cdot s_{2i}\right)\right|}\right)^2} \\
& = \sqrt{2\sum_{i=1}^{M}\left(\ln\frac{c\cdot\left|\left(\frac{m_{1i}}{\sqrt{2}}, s_{1i}\right)-\left(\frac{m_{2i}}{\sqrt{2}}, -s_{2i}\right)\right|+c\cdot\left|\left(\frac{m_{1i}}{\sqrt{2}}, s_{1i}\right)-\left(\frac{m_{2i}}{\sqrt{2}}, s_{2i}\right)\right|}{c\cdot\left|\left(\frac{m_{1i}}{\sqrt{2}}, s_{1i}\right)-\left(\frac{m_{2i}}{\sqrt{2}}, -s_{2i}\right)\right|-c\cdot\left|\left(\frac{m_{1i}}{\sqrt{2}}, s_{1i}\right)-\left(\frac{m_{2i}}{\sqrt{2}}, s_{2i}\right)\right|}\right)^2} \\
& = \sqrt{2\sum_{i=1}^{M}\left(\ln\frac{\left|\left(\frac{m_{1i}}{\sqrt{2}}, s_{1i}\right)-\left(\frac{m_{2i}}{\sqrt{2}}, -s_{2i}\right)\right|+\left|\left(\frac{m_{1i}}{\sqrt{2}}, s_{1i}\right)-\left(\frac{m_{2i}}{\sqrt{2}}, s_{2i}\right)\right|}{\left|\left(\frac{m_{1i}}{\sqrt{2}}, s_{1i}\right)-\left(\frac{m_{2i}}{\sqrt{2}}, -s_{2i}\right)\right|-\left|\left(\frac{m_{1i}}{\sqrt{2}}, s_{1i}\right)-\left(\frac{m_{2i}}{\sqrt{2}}, s_{2i}\right)\right|}\right)^2} \\
& = d_F((m_{11}, s_{11}, ..., m_{1M}, s_{1M}), (m_{21}, s_{21}, ..., m_{2M}, s_{2M})) \\
& = d_F(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)
\end{aligned}
\tag{15}
$$

□

B. Proof of Theorem 2

Proof. The scaling by c does not change the relative distance between data points in a data set. Therefore,

$$N_k(\mathbf{o}_i; D_2) = c\cdot N_k(\mathbf{p}_i; D_1). \tag{16}$$

The set cardinality is not changed by the scaling; thus

$$
\begin{aligned}
|N_k(\mathbf{o}_i; D_2)| &= |c\cdot N_k(\mathbf{p}_i; D_1)| \\
&= |N_k(\mathbf{p}_i; D_1)|.
\end{aligned}
\tag{17}
$$

Since *CLDS* is obtained by translation of N_k and union with $\{\mathbf{0}\}$, from Eq. (16) it also holds that

$$CLDS(\mathbf{o}_i; D_2, k) = c \cdot CLDS(\mathbf{p}_i; D_1, k). \tag{18}$$

Let

$$\begin{aligned} ld(\mathbf{o}_i) &= \boldsymbol{\theta}_{\mathbf{o}_i} = (m_{\mathbf{o}_i 1}, s_{\mathbf{o}_i 1}, ..., m_{\mathbf{o}_i M}, s_{\mathbf{o}_i M}), \\ ld(\mathbf{p}_i) &= \boldsymbol{\theta}_{\mathbf{p}_i} = (m_{\mathbf{p}_i 1}, s_{\mathbf{p}_i 1}, ..., m_{\mathbf{p}_i M}, s_{\mathbf{p}_i M}), \end{aligned} \tag{19}$$

each of which is a parameter of a Gaussian distribution obtained from $CLDS(\mathbf{o}_i; D_2, k)$ and $CLDS(\mathbf{p}_i; D_1, k)$, respectively. Then since both mean and standard deviation are linearly proportional to a scaling factor, it follows that

$$\begin{aligned} ld(\mathbf{o}_i) &= \boldsymbol{\theta}_{\mathbf{o}_i} = (m_{\mathbf{o}_i 1}, s_{\mathbf{o}_i 1}, ..., m_{\mathbf{o}_i M}, s_{\mathbf{o}_i M}) \\ &= (c \cdot m_{\mathbf{p}_i 1}, c \cdot s_{\mathbf{p}_i 1}, ..., c \cdot m_{\mathbf{p}_i M}, c \cdot s_{\mathbf{p}_i M}) \\ &= c \cdot (m_{\mathbf{p}_i 1}, s_{\mathbf{p}_i 1}, ..., m_{\mathbf{p}_i M}, s_{\mathbf{p}_i M}) \\ &= c \cdot \boldsymbol{\theta}_{\mathbf{p}_i} = c \cdot ld(\mathbf{p}_i). \end{aligned} \tag{20}$$

By definition of d_{ld} in Eq. (8) and Theorem 1, it is trivially true that d_{ld} is also scale invariant.

Then, from Eqs. (16) – (20), it follows that

$$\begin{aligned} LGDD(\mathbf{o}_i; D_2, k) &= \frac{\sum_{\mathbf{q} \in N_k(\mathbf{o}_i; D_2)} d_{ld}(ld(\mathbf{o}_i), ld(\mathbf{q}))}{|N_k(\mathbf{o}_i; D_2)|} \\ &= \frac{\sum_{\mathbf{q} \in c \cdot N_k(\mathbf{p}; D_1)} d_{ld}(c \cdot ld(\mathbf{p}_i), ld(\mathbf{q}))}{|c \cdot N_k(\mathbf{p}_i; D_1)|} \\ &= \frac{\sum_{\mathbf{q} \in N_k(\mathbf{p}_i; D_1)} d_{ld}(c \cdot ld(\mathbf{p}_i), c \cdot ld(\mathbf{q}))}{|N_k(\mathbf{p}_i; D_1)|} \\ &= \frac{\sum_{\mathbf{q} \in N_k(\mathbf{p}; D)} d_{ld}(ld(\mathbf{p}_i), ld(\mathbf{q}))}{|N_k(\mathbf{p}_i; D_1)|} = LGDD(\mathbf{p}_i; D_1, k) \end{aligned} \tag{21}$$

□

References

[1]. Xie, X. and Mirmehdi, M. (2007), "TEXEMS: Texture exemplars for defect detection on random textured surfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(8): 1454-1464.

[2]. Zontak, M. and Cohen, I. (2009), "Defect detection in patterned wafers using multi-channel scanning electron microscope," *Signal Processing*, 89(8): 1511-1520.

[3]. Lee, J. and Yoo, S. I., (2012), "Defect detection on images using multiple reference images: solving a binary labeling problem using graph-cuts algorithm," *Journal of Electronic Imaging*, 21(3), 033014-1.

[4]. Tsai, D.-M. and Lin, C.-P. (2002), "Fast defect detection in textured surface using 1D Gabor filters," *The international Journal of Advanced Manufacturing Technology*, 20(9): 664-675.

[5]. Lee, J. Y. and Yoo, S. I. (2004), "Automatic detection of region-mura defect in TFT-LCD," *IEICE Transactions on Information and Systems*, 87(10): 2371-2378.

[6]. Lu, C.-J. and Tsai, D.-M. (2005), "Automatic defect inspection for LCDs using singular value decomposition," *The International Journal of Advanced Manufacturing Technology*, 25, 53-61.

[7]. Kim, H. W. and Yoo, S. I. (2014), "Defect detection using feature point matching for non-repetitive patterned images," *Pattern Analysis and Applications*, 17(2): 415-429.

[8]. Hawkins, D. M., (1980), *Identification of outliers*, 11.

[9]. Breunig, M. M., Kriegel, H. P., Ng, R. T., Sander, J. (2000), "LOF: Identifying density-based local outliers,", in *ACM Sigmod Record*, 29, 93-104.

[10]. Tang, J., Chen, Z., Fu, A. W. C., and Cheung, D. W., (2002), "Enhancing effectiveness of outlier detections for low density patterns,", in *Proceedings of PAKDD*, 535-548.

[11]. Papadimitriou, S., Kitagawa, H., Gibbons, P. B., Faoutsos, C. (2003) "LOCI: Fast outlier detection using the local correlation integral," in *Proceedings of IEEE International Conference on Data Engineering*, 315-326.

[12]. Jin, W., Tung, K., Han, J., and Wang, W. (2006), "Ranking outliers using symmetric neighborhood relationship," in *Advances in Knowledge Discovery and Data Mining*, 577-593.

[13]. Zhang, K., Hutter, M., and Jin, H. (2009), "A new local distance-based outlier detection approach for scattered real-world data," in *Proceedings of PAKDD*, 813-822

[14]. Atkinson, C. and Mitchell, A. F. (1981), "Rao's distance measure," *The Indian Journal of Statistics, Series A*, 345-365.

[15]. Costa, S. I., Santos, S. A., and Strapasson, J. E (2005), "Fisher information matrix and hyperbolic geometry," in *IEEE Information Theory Workshop*.

[16]. Costa, S. I., Santos, A., and Strapasson, J. E. (2014), "Fisher information distance: a geometrical reading," *Discrete Applied Mathematics*.

[17]. Xie, P. and Guan, S.-U. (2000), "A golden-template self-generating method for patterned wafer inspection," *Machine Vision and Applications*, 12(3): 149-156.

[18]. Tsai, D.-M. and Chen, J.-F. (2003), "The evaluation of normalized cross correlations for defect detection," *Pattern Recognition Letters*, 24(15): 2525-2535.

[19]. Parzen, E. (1962), "On estimation of a probability density function and mode," *Annals of mathematical statistics*, 33(3), 1065-1076.

[20]. Boykov, Y., Veksler, O., and Zabih, R. (2001), "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11): 1222-1239.

[21]. Stauffer, C. and Grimson, W. E. (2000) "Learning patterns of activity using real time tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8), 747-757.

[22]. Elgammal, A., Harwood, D., and Davis, L. (2000), "Non-parametric model for background subtraction," in *Proceedings of ECCV,* 751-767.

[23]. Sheikh, Y. and Shah, M. (2005), "Bayesian modeling of dynamic scenes for object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(11), 1778-1792.

[24]. Knorr, E. M. and Ng, R. T. (1998), "Algorithms for mining distance-based outliers in large datasets," in *Proceedings of the International Conference on Very Large Data Bases*, 392-403.

[25]. Lee, W. and Xiang, D. (2001), "Information-theoretic measures for anomaly detection," in *Proceedings of IEEE Symposium on Security and Privacy*, 130-143.

[26]. Filippone, M. and Sanguinetti, G. (2010), "Information theoretic novelty detection," *Pattern Recognition*, 43(3), 805-814.

[27]. Wu, S. and Wang, S. (2013), "Information-theoretic outlier detection for large-scale categorical data," *IEEE Transactions on Knowledge and Data Engineering*, 25(3), 589-602.

[28]. Zhou, S. K. and Chellappa, R. (2006), "From sample similarity to ensemble similarity: Probabilistic distance measures in reproducing kernel Hilbert space," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(6): 917-929.

[29]. Kass, R. E. and Vos, P. W. (2011), "Geometrical foundations of asymptotic inference," *John Wiley & Sons.*

[30]. Carter, K. M. (2009), "Dimensionality reduction on statistical manifolds," *ProQuest.*

[31]. Amari, S. I. and Nagaoka, H. (2007), "Methods of information geometry," *American Mathematical Society*, 191.

[32]. Rao, C. R. (1945), "Information and accuracy attainable in the estimation of statistical parameters," *Bulletin of the Calcutta Mathematical Society*, 37(3), 81-91.