# **A Clustering Technique for Defect Inspection**

Z.W. ZHONG<sup>1</sup> and N.G. SHANKAR<sup>2</sup>

<sup>1</sup>School of Mechanical & Aerospace Engineering, Nanyang Technological University, Singapore <sup>2</sup>Euro Technology Pte Ltd, Singapore

*Abstract:* - A system of rules was developed to join disconnected clusters based on the location of the defects for semiconductor defect inspection. The clusters are evaluated on a pair-wise basis using the rules and are joined or not joined based on a threshold. The system continuously re-evaluates the clusters under consideration as the rules change with each joining action. The technique to measure the features and the methods to improve the system speed are developed. The technique proved very effective in field tests for semiconductor inspection applications.

Key-Words: - Features, rules, defect clustering, semiconductor inspection

# **1** Introduction

In semiconductor assemblies, the visual inspection process of wafer surfaces depends on manual review by human experts. The decision instability of an inspector can be quite large against various defect classes, and inspectors rely on different features and strategies [1].

A key component of clustering techniques that step beyond common proximity-based clustering methods has been discussed [2], which utilizes a gravitational force analogy to connect clusters together. Feature classification techniques benefit today from many years of research performed on different domains such as life science, medicine, astronomy, earth science and engineering [3].

Choosing many features to be sure that all the information is encoded may be a neither efficient nor correct strategy, because even the most significant feature may contain noise. By integrating redundant features, the total amount of noise increases, while significant information saturates. This phenomenon is referred to as curse of dimensionality problem [4]. Several filters are available to reduce the amount of noise. Their choice and parameters need to be selected. General filtering approaches exist, such as median and Gaussian filters, anisotropic diffusion filters [5] and morphological filters [6], but better performance may be obtained if the information is preserved by reducing the nature of the noise. Another alternative to avoid the curse of dimensionality problem is to vary the relative weight of the features according to an estimation of their reliability.

Two categories of the methods to perform clustering of a digital image are defined: spatially constrained and unconstrained clustering techniques. The first group performs clustering in the image coordinate domain. Well-known techniques are region growing [7], split and merge [8] and pyramidal approaches [9]. The second group performs clustering directly in the feature space domain. There are simple thresholding techniques [10] and well-known partitional clustering procedures such as ISODATA [11,12] and fuzzy algorithms [13].

However, engineering an efficient clustering procedure for a specific application remains a complex task. It requires the attentive analysis of several issues. In this paper, we discuss the major issues that we have taken into account in our work and that should be carefully evaluated whenever a feature classification problem needs to be solved. Clusters are evaluated on a pair-wise basis using rules and are joined or not joined based on a threshold. The system continuously re-evaluates the clusters under consideration as rules change with each joining action. Examples are shown using semiconductor products obtained from manufacturers. The superior clustering method offered through rule-based systems enables more accurate feature measurements and process characterization for the classification process.

# 2 Diagnostic Rule

### 2.1 Clustering Algorithm

According to the diagnostic rule, the visual error data or defective images should be grouped in areas with uniform visual features. This is the hypothesis for which a multi-feature clustering technique is applied to group homogeneous classes of the error elements of the input data. Several multi-feature classification techniques are available in the literature. The partitional clustering algorithms described in the form of ISODATA [11,12] were adopted in this study. Flexibility is an important characteristic for our approach because it is necessary to combine both local and global analyses of the input data to emulate the principles of perceptual organization. The clustering algorithm adopted in this work is as follows:

#### 2.1.1 Initialization procedure

Choose arbitrary initial estimates of the cluster representatives  $C_j(0)$  with  $j = 1, \dots, M$ . Repeat until no change in  $C_j$ 's occurs between two successive iterations or a maximum number of iterations is performed.

#### 2.1.2 Association step of the algorithm

The first issue is the definition of a distance measure between a feature point,  $\overline{p}$ , and a cluster representative *C*.

It is proposed to adopt proximity measure, the distance defined by the weighted sum of distances:

$$D(\vec{p}, C) = \frac{1}{T} \sum_{i=1}^{T} W_i D_i(\vec{p}, C)$$
(1)

Here,  $D(\vec{p}, C)$  is a normalized linear combination of T distance measures,  $D(\vec{p}, C)$ . Each of these measures with real values between zero and one describes the distance for a specific group of features. The weighting coefficients  $W_i$  are responsible for associating an absolute weight for each group of features contributing to the final distance,  $D(\vec{p}, C)$ . The weights values are real values between zero and one. Equation (1) is a distance measure generalized for an arbitrary number T of groups of features. In our experiments, we have limited this number to four groups, described as the low-level chip out, scratch, metallization and bridging. Thus, in our experiments, Equation (1) becomes:

$$D(\vec{p},C) = \frac{1}{4} \begin{pmatrix} W_C \times D_C(\vec{p},C) + W_S \times D_S(\vec{p},C) + \\ W_M \times D_M(\vec{p},C) + W_B \times D_B(\vec{p},C) \end{pmatrix} (2)$$

where the distances for each group of features are normalized to real values of [0, ..., 1] according to the following expression; for example the distance for chip out is given as

$$D_{C}(\vec{p},C) = \frac{\left|f_{C}(\vec{p}) - f_{C}(C)\right|}{500}$$
(3)

Similarly for scratch, metallization and bridging.

#### **Association Step:**

For all the input data points  $\vec{p}_k \in I$  with  $k = 1, \dots, P$ .

->Determine the closest representative, say  $C_j$  for  $\vec{p}_k$ .

->Determine the mean distance to all the representatives,  $\overline{D}$ .

If the distance between  $C_j$  and  $\vec{p}_k$  is close to  $\overline{D}$ ,

**Then**  $\vec{p}_k$  is a new cluster representative. M is increased of 1. Label  $l_k$  of  $\vec{p}_k$  is set to M. **Else** set label  $l_k$  of  $\vec{p}_k$  to j.

#### **End For**

#### 2.1.3 Update step of the algorithm

The result of the association step is that each feature point  $\vec{p}_i$  in I is now associated to a label  $l_i$ corresponding to one of the cluster representatives  $C_i$ with  $j = 1, \dots, M$ . If we represent in the image coordinate system the label values associated to each feature point, we obtain a  $(H \times W)$  matrix O of labels. This representation puts in correspondence each pixel of the input image with the most similar cluster. The main role of the update step is to use the label information to update the feature properties of each cluster. This is obtained in computing the mean properties of all the points belonging to the same cluster. The result of this update is that each cluster  $C_i$  is associated to a new feature vector. Each element of this feature vector is computed as the mean value of the corresponding element in the feature vectors  $\vec{p}_i$  associated to  $C_i$ .

The update of the input feature points is computationally expensive. Moreover, new similarity relationships are generated between feature points and existing clusters that need new iterations to find stable convergences. For these reasons, it is important to update the feature values only when it is known that they are not correctly defined. In our experiments, the update is performed first at the beginning of the clustering. A second update is performed after the convergence to a stable classification has been obtained to verify the consistence of the updated features.

#### **Update Step:**

For all defined  $C_i$  with  $j = 1, \dots, M$ ,

->Update  $C_j$  according to the feature properties of the data points labeled with j.

#### **End For**

#### 2.2 Similarity Rule or Similarity Level

The role of the diagnostic level is to group the pixels of the input error image into M perceptually uniform clusters  $C_j$ . This is achieved through diagnostic rules that perform a low-level analysis of the input data. Each cluster  $C_j$  can be represented in the image coordinate system as a region  $R_j$ . The role of the similarity level is to combine the regions  $R_j$  into objects according to what are defined as similarity rules in analogy with the clinical language.

The context of similarity rules is different from the context of diagnostic rules. The P pixels of the error image are grouped in a number  $M \ll P$  of regions. This representation offers two important advantages.

The regions defined by the diagnostic level are important supports to be used in the computation of features. More accurate models can be used to reduce the probability of generating artificial features. Moreover, the neighborhood relationship between regions provides a new tool for measuring the border similarity (contrast) between them.

Similarity rules are applied to regions and not to pixels as the diagnostic rules. Even if a large number of descriptors are used to characterize the visual properties of regions, the relationship ( $M \ll P$ ) guarantees a low computational analysis, i.e., reduction in the amount of data to be evaluated.

### **3. Results**

The diagnostic rule performs the decomposition of images into separate objects. After the objects have been constructed, they are handled as independent entities. Geometric features, such as area and width, can be computed. Then the objects of interest can be selected by means of their positions or computed features. The selection criteria can be combined incrementally. The desired objects selected can be listed and inquired about their geometric characteristics, and sorted using the values of the geometric features.

The rule that determines what constitutes an object is simply the grouping of neighboring pixels of the same gray level range using the clustering algorithm. Fig. 1 shows the actual real data used in our experiments. In image processing parlance, thresholding followed by connected components labeling is used.

A two-step process is applied. First, the gray level images are decomposed into homogenous light/medium/dark areas by using the double thresholding technique: the values below the lower threshold belong to the black class; the values above the upper threshold belong to the white class. The remaining pixels belong to the neutral class.

Fig. 2 shows the clustering results of the visual error image generated from Fig. 1 using the proposed clustering algorithm. The thresholds can be adaptive, i.e. they can be defined pixel-wise rather than remaining constant over the whole image. In this case, the local threshold values must be provided as user-defined images. One possible way to prepare suitable threshold images is to take a reference image and add tolerance margins to it.



Fig. 1. Real data used for experiments (a) reference image, (b) test image, and (c) visual error data or image



Fig. 2. Clustered image contains chip out, metallization, and bridging defects

After pixel classification using diagnostic rules, the objects are built by grouping neighboring pixels of the same class. The objects belonging to the black/neutral/white classes can be built or not, depending on the needs. The objects may have holes. Pixels can touch each other along an edge or by a corner. Pixels touching by an edge are considered neighbors, if one speaks of 4connectedness. Pixels touching by a corner are also considered neighbors, if one speaks of 8connectedness.

# 4. Conclusions

We have shown the applicability of rule-based clustering to the problem of connecting clustered defects in semiconductor inspection. We have presented the background of the rule-based clustering and how it pertains to our solution. Due to the nature of our problem, we require pair-wise rule evaluation and feature measurement that can be computationally expensive. We have discussed techniques to improve the method, and showed clustering results demonstrating the ability of our system to connect clusters. While no single clustering method can satisfy every human observer, these results show the effectiveness of our system in joining disconnected clusters. The technique proved very effective in field tests for semiconductor inspection applications.

References:

- L. Breaux and B. Singh, "Automatic defect classification system for patterned semiconductor wafers," Proceedings of International Symposium on Semiconductor Manufacturing, pp. 68-73, 1995.
- [2] S.S., Gleason and K.W., Tobin, "Directional-based Dilation for Connection of Piece-Wise Objects: A Semiconductor Manufacturing Case Study", International Conference on Image Processing, Lausanne, Switzerland, 1996.
- [3] M. Anderberg, Cluster Analysis for Applications. Academic Press, New York, 1973.
- [4] J.R. Duin and J. Mao, Statistical Pattern Recognition: A Review, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, pp. 4-37, 2000.
- [5] M.J. Black, G. Sapiro, D.H. Marimont and D. Heeger, Robust anisotropic diffusion, IEEE Transactions on Image Processing, Vol. 7, pp. 421-432, 1998.
- [6] C.H. Chen and D.L. Yang, Realization of morphological operations, IEEE Proceedings on circuits devices and systems, Vol. 142, pp. 364-368, 1995.
- [7] A. Mehnert and P. Jackway, An improved seeded region-growing algorithm, Pattern Recognition Letters, Vol. 18, pp. 1065-1071, 1997.
- [8] R.H. Laprade, Split-and-merge segmentation of aerial photographs, Computer Vision, Graphics, and Image Processing, Vol. 44, pp. 77-86, 1988.
- [9] F. Ziliani and B. Jensen, Unsupervised segmentation using modified pyramidal linking approach, Proceedings of IEEE International Conference on Image Processing, October 1998.
- [10] L. Liyuan, G. Jian and C. Winan, Gray-level image thresholding based on Fisher linear projection of two-dimensional histogram, Pattern Recognition, Vol. 30, pp. 743-749, 1997.
- [11] V.F.R. Dias, Thresholding using the ISODATA clustering algorithm, IEEE Transactions on Systems, Man and Cybernetics, Vol. 10, pp. 771-774, 1980.
- [12] A.K. Jain and R.C. Dubes, Algorithms for Clustering Data. Prentice Hall, Englewood Cliffs, NJ, 1988.
- [13] M.J. Sabin, Convergence and consistency of Fuzzy C-means/ISODATA algorithms, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 9, pp. 661-668, 1987.