Automated Framework for Unsupervised Counterfeit Integrated Circuit Detection by Physical Inspection

Pallabi Ghosh, Fatemeh Ganji, Domenic Forte, Damon L. Woodard, Rajat Subhra Chakraborty

***Abstract*— Undetected counterfeit ICs, if incorporated in deployed systems, have far reaching, long lasting and sometimes, devastating consequences for critical infrastructure and national security. Although physical inspection currently constitutes a promising way of identifying counterfeit ICs, the state-of-the-art techniques often require manual intervention by subject matter experts (SMEs). Automated image processing and analysis techniques for counterfeit IC detection can be employed, but are currently underexplored. Lack of sufficiently large counterfeit IC image datasets exacerbate the difficulties of intelligent image processing based counterfeit IC detection. In this paper, we propose a comprehensive, counterfeit IC detection methodology, which uses block-by-block analysis of IC images, and is capable of simultaneously identifying multiple external defects, while optionally employing defective IC image dataset. In the first step, cluster analysis is performed to detect a cluster of probable counterfeit ICs. In the second optional step, if a cluster of defective ICs is detected, further image analysis is performed to identify the exact defect based on a standard counterfeit IC taxonomy. Our experiment explores mainly infra-red (IR) images of IC packages and our experimental results show that the proposed methodology has high detection accuracy.**

I. Introduction

Counterfeit ICs are fake components made to fraudulently pass as genuine in the global market. Such industry emerged due to the transferable value of electronics parts. Any company in the world requiring components to manufacture printed circuit boards (PCBs) is at risk, and may have already received batches of old and used ICs, reducing the longevity of the product. There are mainly seven major categories of electronic counterfeits [1], based on how the components are misrepresented. They are **recycled, remarked, overproduced, out-of-spec/defective, cloned, forged documentation, and tampered**. Among them recycled and remarked counterfeit ICs constitute over 80% of counterfeit ICs, and their detection has been the most widely studied problem in the current state-of-the-art. Unfortunately, not many methods are available to swiftly and reliably detect counterfeits. Sometimes, it is not possible to differentiate counterfeit components until they are placed on a PCB, and the production team runs the first tests on the complete assembly. In most cases, however, the counterfeit might not be identified until it causes a failure in the system. This is particularly costly because finished products need to be recalled by the original equipment manufacturer (OEM). In either case, identifying counterfeit components requires inspecting them for physical defects. This requires a subject matter expert (SME) to go through each and every chip manually, thus making the process both time consuming and costly. Also, being human beings, the SMEs are error prone, and often cannot keep up with the new methods that counterfeiters improvise to avoid detection.

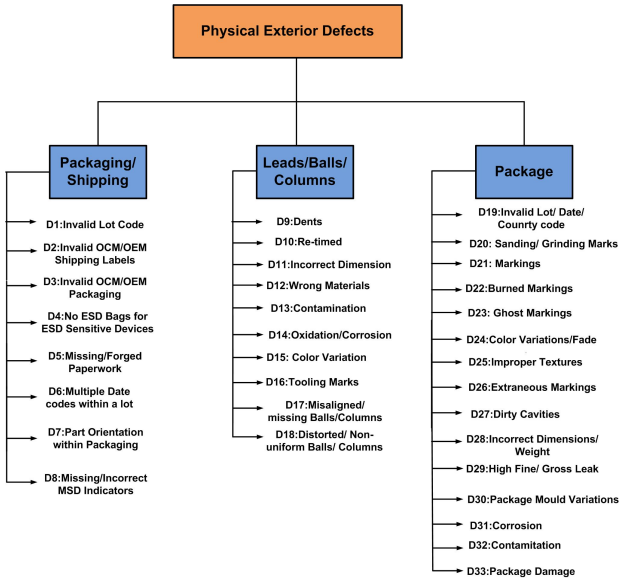

Fig. 1: Taxonomy of common external physical defects in Counterfeit ICs [2].

Recently, several techniques have been proposed in the literature to automatically identify physical defects. These require expensive image acquisition infrastructure and are often aimed at identifying one defect at a time [2]–[4]. A significant work was [5], in which the authors used advanced image acquisition techniques such as *Scanning Electron Microscope* (SEM), X-Ray microscopy and *Energy Dispersive Spectroscopy* (EDS). Here 3-D reconstruction of the SEM images was performed using stereo-photogrammetry. Another recent work [6] applies *Artificial Neural Networks* (ANN) with image processing to automate the process of detecting counterfeit ICs, based on scratch defects present on its surface. For training the machine learning classifier, the authors have used images collected from an extremely useful counterfeit image database that they have themselves designed and maintain [7].

Over the past few years a taxonomy of counterfeit IC defects has been created, as shown in Fig. 1 [1], [8]. In this paper, we have examined each of these physical defects and developed a automated framework to detect multiple such defects simultaneously. For this purpose we have used Near-Infrared (NIR) imaging. These type of images are useful for our technique as there is no significant reflection which may

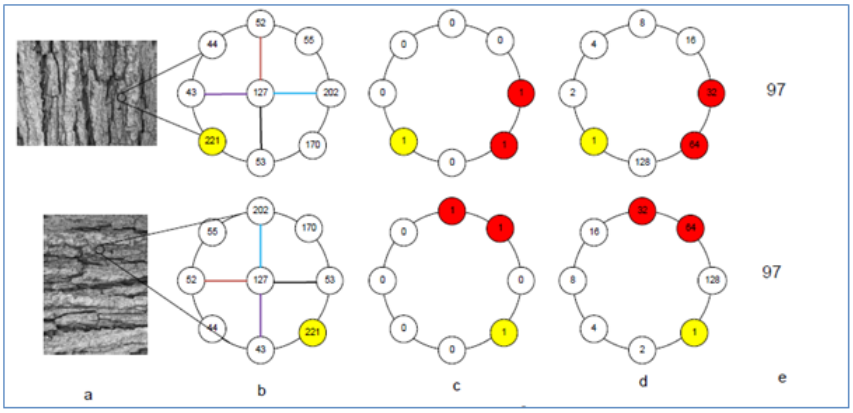

Fig. 2: Effect of rotation on RLBP operator (a) The image (top) and 900 counter-clockwise rotated image (bottom), (b) Yellow colour pixel indicate the dominant direction, (c) Values above threshold are shown in red colour, (d) The weights are circularly shifted with respect to dominant direction, (e) RLBP [11].

result in false positives.

II. Background

In this paper we have primarily used image texture features for the clustering and identification. Texture is a measure which gives information about the spatial arrangement of the color or intensities in an image [9] [10]. There are several techniques to extract textural features from an image. In this paper we have used two such methods. They are *Local Binary Pattern* and *Laws' Texture Energy Measure*.

A. Local Binary Pattern

Local Binary Pattern is one of the most widely used textural feature extraction technique [10], [12]. In this method, the following steps are performed:

1) For each pixel p of an image, set an 8-bit bitstring $b_1b_2b_3b_4b_5b_6b_7b_8$ to zero.
2) Examine its eight neighbors, and set $b_i = 1$ iff the intensity of the i-th neighbor is less than equal to that of p, and 0 otherwise.
3) Considering these bit strings to be 8-bit unsigned numbers, construct a histogram of these numbers.

In our work we have used a rotation invariant Local Binary Pattern algorithm [11]. The main advantage of this version of Local Binary Pattern is that it is completely independent of alignment as shown in Figure 2. In the figure one image is the rotated version of the other image, still the final value for both is 97. When applied on all the pixels of an image a set of values are obtained which is usually represented in the form of histograms. The consecutive values of the histograms are grouped in forms of bins for ease of representation.

B. Laws' Texture Energy Measure

Laws' Texture Energy Measure is another textural feature extraction technique [10], [13]. This "texture-energy" approach measures the amount of variation within a fixed-size window (a typical window size is 15×15). In this method, four vectors are used to form nine 5 convolution masks (L5E5'/E5L5', L5S5'/S5L5', L5R5'/R5L5', E5E5', E5S5'/S5E5', E5R5'/R5E5', S5S5', S5R5'/R5S5' and R5R5'). Each of the vectors are chosen to detect particular features, as follows:

$$\begin{aligned} \text{L5 (Level)} &= \begin{bmatrix} 1 & 4 & 6 & 4 & 1 \end{bmatrix} \\ \text{E5 (Edge)} &= \begin{bmatrix} -1 & -2 & 0 & 2 & 1 \end{bmatrix} \\ \text{S5 (Spot)} &= \begin{bmatrix} -1 & 0 & 2 & 0 & -1 \end{bmatrix} \\ \text{R5 (Ripple)} &= \begin{bmatrix} 1 & -4 & 6 & -4 & 1 \end{bmatrix} \end{aligned} \quad (1)$$

In this paper, we have estimated the horizontal and vertical edges (L5E5'+ E5L5') present on the IC surface using this feature.

III. Methodology

The proposed methodology has two steps. The first step is texture analysis of the images of IC packages and clustering based on that, and the second phase includes analysis of each cluster to detect the nature of defect. Figure 3 shows the overall proposed flow of automated image processing based counterfeit IC detection. The acquired images undergo a pre-processing step where they are aligned with each other. For alignment we have used image registration [14]. The advantage of IR images is that there is no visible reflection and noise. We ensure that all images are of the same size and resolution. In our dataset we have included one golden IC image set to compare different parameters of the IC set under test. The golden IC set contains a group of similar ICs which need not be of the same type as the ICs under test.

A. Phase-I of Counterfeit IC Detection

In this phase, the actual phase of counterfeit detection is initiated. Here each image is divided into a series of non-overlapping sub-blocks. From each sub-block image LBP based features are extracted and a histogram is formed. All these histograms are then concatenated together in a row-major order to form a single histogram. Thus for each image one single concatenated histogram is formed. This histogram is the feature vector for each image where each histogram bins represents one histogram. Clustering of the IC images

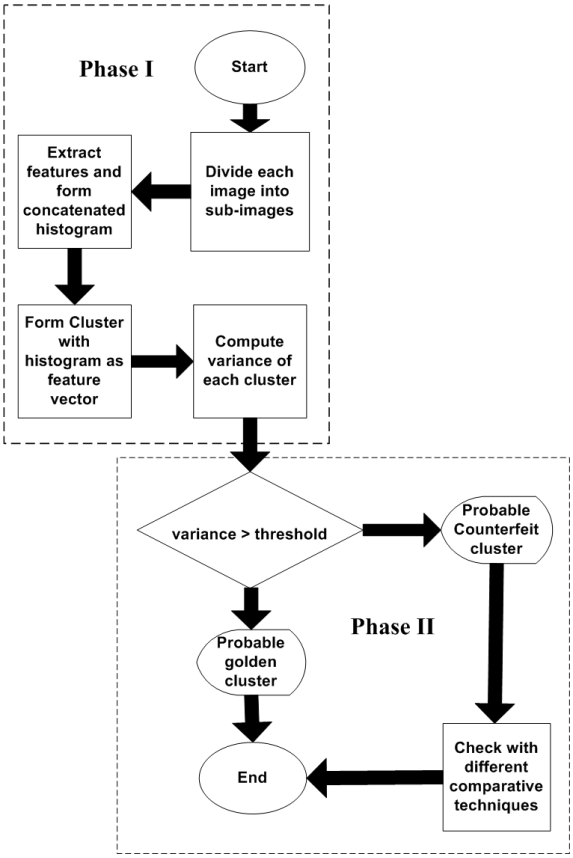

Fig. 3: Flowchart showing the entire technique.

are done based on these histograms. The analysis of the variance of each such clusters along with the distance from cluster centres to other points facilitates the identification of the probable counterfeit cluster.

B. Phase-II of Counterfeit IC Detection

Once the probable counterfeit cluster is detected, the second phase of the counterfeit detection is initiated. Note that although the second phase is optional, it helps to establish the true identity (counterfeit/non-counterfeit) of a clustered IC from phase-I, and thus helps to minimize error. In this phase, each of the IC is passed through a series of tests based on features like indents, scratches, ghost marking, etc. As of now, this phase is not fully automated, but we are trying to automate it and hope that in future this phase will not require any human intervention. For each of the defect identification, part of the process is done manually by SMEs. For instance the amount of edges present on the surface can be quantified automatically but to differentiate between scratches and ghost-marking human intervention is required. Similarly the indent size, shape and location can be identified automatically but differentiation between a golden IC indent and a counterfeit IC in terms of texture and other factors is done manually.

1) Scratches: Scratches can be quantified by the measure calculated using the edge feature of Laws' Texture Energy measure [13]. If this amount is more than some set threshold

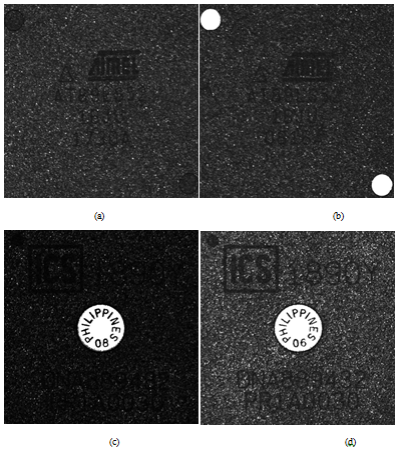

Fig. 4: Types of images present in the dataset.

for that IC then it can be concluded the IC surface contains scratches. When ICs with scratches are found they can be further passed for ghost marking detection.

2) Ghost Marking: The marking of a scratched IC surface is compared with the markings of the golden IC surface manually. Also the coating applied to hide the previous markings are usually of low grade and gets removed easily.

3) Indent Mismatch: Active contour method [15] is used to identify the indents present on the surface of the ICs [4]. Once the indents are located, information like size, position and texture of the indents are extracted. They are then compared with the golden IC. In our dataset, the points which showed anomaly had different indent texture.

4) Marking Imperfection: The marking positions on all ICs are constant. From the concatenated histograms, bins belonging to those blocks are extracted and compared. If the distance between them is more than the set threshold found in golden ICs, then the marking of that IC is not perfect. Further manual inspection can help in knowing more details of the marking imperfection.

IV. Experimental Results

There are 40 IC images in our dataset. This data points are obtained using Phemos 1000 microscope [16]. They are similar structured IC surface images. The results obtained by following the methodology in Section III are described in detail in this section.

A. Data Organization

The dataset contains mainly four varieties of data points as shown in Figure 4, where (a) and (b) are of type IC1 and (c) and (d) are of type IC2. Here IC2 is the known golden IC. We have used this set of golden ICs to compare the characteristics of the other set. Ideally, after clustering, each IC should be clustered with other IC samples of the same type. All the images are of same size (747×747 pixels) and similar IC types are registered with each other. Since the images were acquired with the same configuration and alignment in an automated system, so that they are always

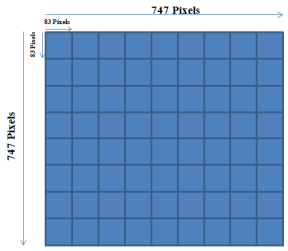

Fig. 5: Block division of each image.

TABLE I: Average Silhouette Values

No. of Clusters	RLBP (8,1)	RLBP (16,2)	RLBP (24,3)
1	0.0000	0.0000	0.0000
2	**0.8909**	**0.8831**	**0.8713**
3	0.8751	0.8720	0.8045
4	0.7434	0.7919	0.8278
5	0.7301	0.5846	0.5053
6	0.6513	0.5334	0.5354
7	0.3109	0.2555	0.2201
8	0.4622	0.5236	0.4948
9	0.3082	0.1638	0.1727
10	0.3619	0.3066	0.1947

aligned with high accuracy with each other. The contrast of (a), (b) and (d) are similar. The contrast of type (c) is made different to investigate the effect of contrast on the results.

B. Phase-I of Counterfeit IC Detection

1) Image Block Division: Each image in our dataset is of size (747×747 pixels) and is divided into 83×83 non overlapping sub-images as shown in Figure 5. This division is completely data dependent. Total number of blocks obtained from each image is 81. Since no image registration is 100% accurate, this image division helps in comparing regions instead of direct pixel positions.

2) Features Extraction: Textural features are extracted in this phase. For textural feature extraction different variations in radius of rotation invariant Local Binary Pattern (RLBP)is used, as described in Section II. The different radii for RLBP used in our work are 1, 2 and 3 pixels which are referred as predicate. So the number of neighbouring pixel positions for each of them is respectively 8, 16 and 24. The pixel values are bi-linearly interpolated whenever the sampling point is not in the centre of a pixel. This technique is applied on each pixel of each sub-block of an image. A set of values are obtained for each block from which a corresponding LBP histogram is formed for that block. So for each image in our dataset 81 such histograms are obtained. The consecutive values of the histograms are grouped in forms of bins. Number of bins for predicate 1, 2 and 3 are respectively 10, 18 and 26.

3) Histogram Concatenation: The histograms of each sub-image are concatenated sequentially in a fixed row-major order. One concatenated histogram is obtained for each full image which is used as feature vector for clustering. This concatenated histogram is such that the value of the K-th bin of the histogram of sub-image number b is same as the value of bin number $[K + (b-1) * N]$ of the concatenated histogram.

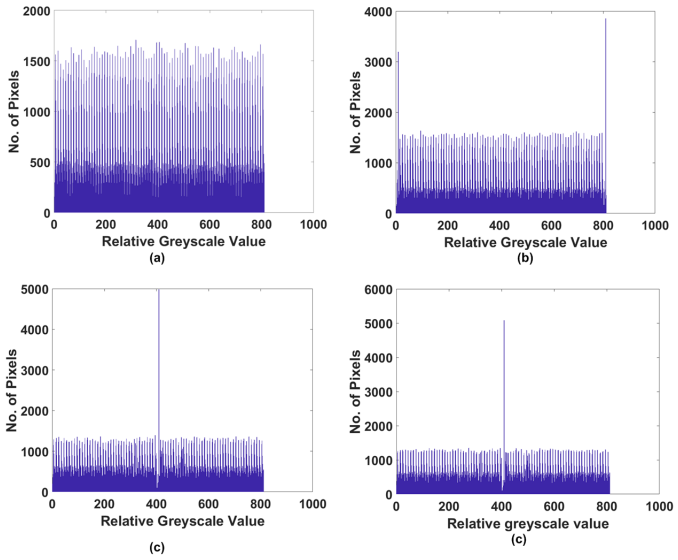

Fig. 6: (a), (b), (c) and (d) are the corresponding concatenated LBP histograms, for predicate value 1, of the different types of IC1 and IC2 shown in Figure 4.

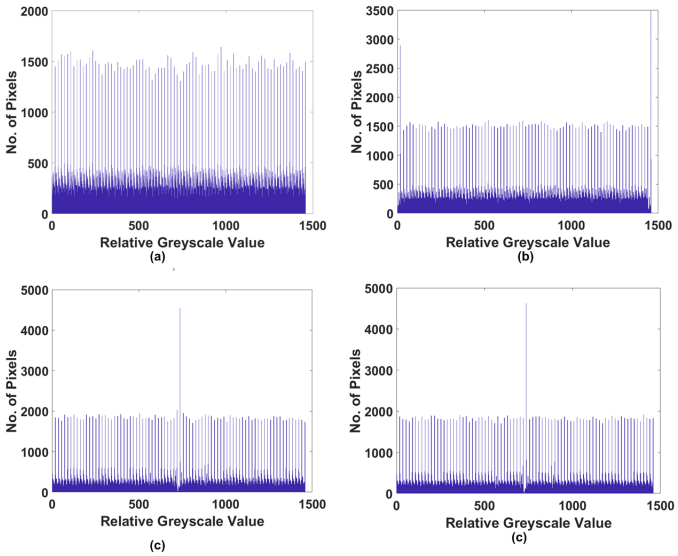

Fig. 7: (a), (b), (c) and (d) are the corresponding concatenated LBP histograms, for predicate value 2, of the different types of IC1 and IC2 shown in Figure 4.

In this work, each bin of a histogram of an image is compared with the corresponding bin of another histogram of another image to form clusters. That is, here each bin of the concatenated histogram is one feature and the overall concatenated histogram is the feature vector. The histograms obtained for different predicate values of LBP for the four different sample images (Figure 4) are shown in Figure 6, 7 and 8. In all the three sets of predicate values, it is observed that, inspite of having different contrast in between the two varieties of IC2, there is not much difference in their LBP pattern. However for the two varieties of IC1, dissimilarities are observed in the first and last blocks of their concatenated

TABLE II: Variance and silhouette values for different clustering levels

	RLBP(8,1) (Level 1)	Silhouette Value (L1)		RLBP(8,1) (Level 2)	Silhouette Value (L2)	RLBP(16,2) (Level 1)	Silhouette Value (L1)		RLBP(16,2) (Level 2)	Silhouette Value (L2)	RLBP(24,3) (Level 1)	Silhouette Value (L1)		RLBP(24,3) (Level 2)	Silhouette Value (L2)
Cluster 1	4.38e3	0.89	**Cluster 11**	1.32e3	0.91	1.96e3	0.88	**Cluster 11**	732.84	0.88	1.27e3	0.87	**Cluster 11**	538.61	0.85
			Cluster 12	807.20				**Cluster 12**	514.21				**Cluster 12**	393.33	
Cluster 2	2.88e3		**Cluster 21**	1.41e3	0.63	1.48e3		**Cluster 21**	610.57	0.67	1.09e3		**Cluster 21**	429.46	0.78
			Cluster 22	4.72e3				**Cluster 22**	2.54e3				**Cluster 22**	1.84e3	

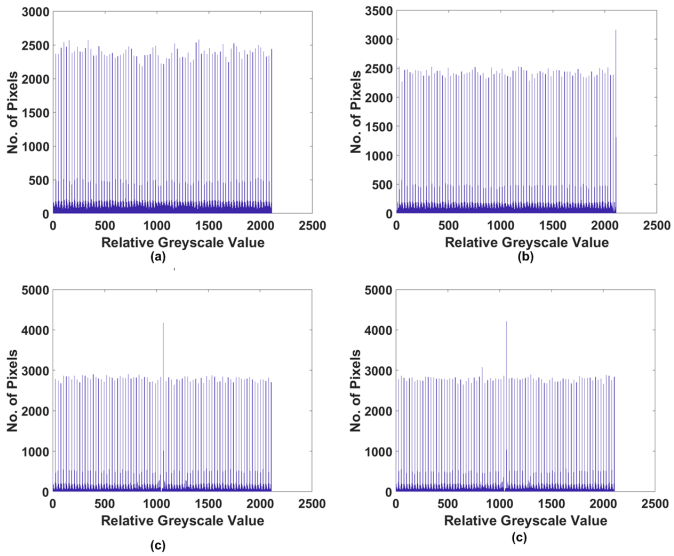

Fig. 8: (a), (b), (c) and (d) are the corresponding concatenated LBP histograms, for predicate value 3, of the different types of IC1 and IC2 shown in Figure 4.

histograms.

4) Clustering: The concatenated histograms are given as input to a clustering algorithms. The clustering algorithm used to find the clusters is K-Means [17]. The distance used for K-Means clustering is Euclidean. The value of K is determined with the help of *silhouette index* [18]. In our dataset we have mixed one set of golden IC (IC2) having similar structure to compare the variance. It is observed in Table I that for all the algorithms, the silhouette value for K equals to 2 is highest. Also for all the LBP variations, K-means separated all the IC1 types in one class and IC2 types in a separate class. A second phase of clustering is also done for each clusters. It is observed that again the silhouette value is maximum for 2 clusters for both IC1 and IC2. The second phase of clustering on cluster-1 divided the set into two clusters having different indent textures, but it failed to show any proper categorization for the golden IC cluster. From this we can conclude that since we are using texture feature, difference in contrast of images does not affect the clustering result. The different silhouette values for the second phase of clustering is shown in Table II

5) Variance Calculation: After clustering, to calculate the spread of each cluster, the variance is calculated with respect to each bin of the concatenated histograms, corresponding to the same cluster, which in our work represents one feature. For ease of comparison, average of these variance values is taken as the variance for the corresponding cluster. Thus for each cluster one variance value is obtained. Variance among

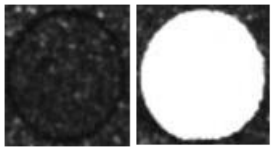

Fig. 9: Indents of the sub-clusters.

the golden IC samples are taken as the threshold value. The different variance values obtained are shown in Table II. For all the predicate values of LBP, it is observed that all the golden samples in our dataset are clustered in one cluster and the points under test are grouped in a separate cluster. Also the variance of cluster-1, which is the cluster formed by images of ICs under test, is more than cluster-2, which belongs to the golden cluster. Hence in our dataset, cluster-1 is the probable counterfeit cluster. Another set of clustering and variance calculation for each cluster is also done to observe the way the data is scattered. After the second phase, it is observed that the silhouette index for the clustering of the golden cluster is much less than the probable counterfeit cluster. This is an indication that the clustering of the golden cluster is not so beneficial compared to the probable counterfeit cluster. The results obtained after second phase of clustering is also shown in Table II. Now, to find the exact anomalous points in cluster-1, distances from the cluster centre are also calculated. Two different values of distances are found in the output. These outputs are exactly the same points that were clustered separately after the second phase of clustering. The points are noted and compared using the other comparative techniques discussed in Section III.

C. Phase-II of Counterfeit IC Detection

The probable counterfeit cluster and the sub-clusters corresponding to the same cluster obtained in Phase-I is passed to Phase-II for further inspection based on other image features.

1) Scratches and Ghost-Markings: In Table III the rows in bold font differentiates one set of data from another which were clustered after second time clustering in Phase-I. This measure is applied on the whole surface image. Since not much difference is observed in the total amount of edges, a next set of experiments are performed to test the size, position and texture of indents.

2) Indent Size and Markings: The indents are located with the help of Active Contour techniques [4]. Based on that location, the block containing the indent is identified and the texture map of that block is extracted. Now, using these indent size, position and texture, a second time K-means clustering is performed. Here also we found for K

TABLE III: Laws' Texture Energy Measure for Cluster-1 values

IC No.	L5E5T + E5L5T	L5S5T + S5L5T	L5R5T + R5L5T	E5S5T + S5E5T	E5E5T	E5R5T + R5E5T	S5R5T + R5S5T	S5S5T	R5R5T
1	46.75	9.81	1.37	0.71	7.83	13.27	3.97	3.16	2.43
2	21.79	8.68	1.08	0.67	6.93	6.99	2.09	2.95	2.26
3	29.94	9.16	1.15	0.68	7.60	8.70	2.56	3.08	2.38
4	32.03	9.21	1.16	0.68	7.53	9.15	2.69	3.09	2.38
5	**37.45**	**9.49**	**1.23**	**0.70**	**7.75**	**10.46**	**2.94**	**3.15**	**2.44**
6	34.44	9.30	1.19	0.68	7.66	9.75	2.89	3.10	2.39
7	35.23	9.35	1.21	0.69	7.74	10.11	2.99	3.12	2.42
8	**32.24**	**9.26**	**1.16**	**0.69**	**7.68**	**8.78**	**2.49**	**3.10**	**2.40**
9	25.87	8.97	1.11	0.68	7.34	7.77	2.28	3.04	2.34
10	**32.13**	**9.31**	**1.16**	**0.69**	**7.64**	**9.01**	**2.53**	**3.12**	**2.41**
11	27.28	9.01	1.12	0.68	7.45	8.07	2.38	3.04	2.35
12	44.71	9.75	1.33	0.70	7.81	12.46	3.71	3.16	2.43
13	44.05	9.72	1.32	0.70	7.82	12.47	3.68	3.15	2.43
14	41.04	9.56	1.27	0.70	7.80	11.32	3.36	3.14	2.43
15	43.03	9.62	1.30	0.70	7.79	11.87	3.54	3.15	2.43
16	31.57	9.21	1.17	0.68	7.49	9.19	2.71	3.06	2.36
17	31.90	9.20	1.17	0.68	7.55	9.24	2.72	3.08	2.37
18	47.40	9.84	1.37	0.71	7.81	13.29	3.99	3.15	2.43
19	44.08	9.67	1.31	0.70	7.76	12.12	3.63	3.16	2.43
20	42.24	9.62	1.30	0.70	7.76	11.86	3.52	3.14	2.42
21	39.36	9.51	1.25	0.69	7.75	10.93	3.22	3.13	2.42
22	42.53	9.62	1.30	0.70	7.83	11.83	3.53	3.14	2.43
23	44.38	9.71	1.32	0.70	7.82	12.41	3.70	3.16	2.43

value-2 the silhouette index is maximum. The elements of the two clusters are then compared manually. It is found that the texture of the indents are completely different for two different clusters, one smooth and another of matte type, as shown in (a) and (b) of Fig. 4 and Fig. 9. No visible difference was observed in the manual inspection of the markings on the IC surface. Comparing these observations with golden IC information it is found that the probable counterfeit ICs of our dataset are the ones having matte indents. This information matched exactly with the previously known information about the ICs of our dataset. So we can conclude, for this dataset, this methodology gave 100% accuracy.

V. Conclusion

In this paper, a comprehensive technique capable of detecting multiple defects simultaneously has been developed. This technique uses previously discussed individual defect detection techniques and forms a comprehensive framework. Our experimental result gave 100% accuracy for the dataset we have used. Our future work would be directed towards fully automating the proposed methodology.

References

[1] U. Guin, K. Huang, D. DiMase, J. M. Carulli, M. Tehranipoor, and Y. Makris, "Counterfeit Integrated Circuits: A Rising Threat in the Global Semiconductor Supply Chain," *Proceedings of the IEEE*, vol. 102, no. 8, pp. 1207–1228, Aug 2014.

[2] G. P, F. D, W. DL, and C. RS, "Automated Detection of Pin Defects on Counterfeit Microelectronics," in *2018 International Symposium for Testing and Failure Analysis (ISTFA)*, 2018.

[3] P. Ghosh and R. S. Chakraborty, "Counterfeit IC Detection By Image Texture Analysis," in *2017 Euromicro Conference on Digital System Design (DSD)*, 2017, pp. 283–286.

[4] P. Ghosh and R. S. Chakraborty, "Recycled and remarked counterfeit integrated circuit detection by image-processing-based package texture and indent analysis," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 1966–1974, April 2019.

[5] S. Shahbazmohamadi, D. Forte, and M. Tehranipoor, "Advanced Physical Inspection Methods for Counterfeit Detection," in *Proceedings of International Symposium for Testing and Failure Analysis (ISFTA)*, 2014, pp. 55–64.

[6] N. Asadizanjani, M. Tehranipoor, and D. Forte, "Counterfeit Electronics Detection Using Image Processing and Machine Learning," *Journal of Physics: Conference Series*, vol. 787, no. 1, pp. 1–6, 2017. [Online]. Available: http://stacks.iop.org/1742-6596/787/i=1/a=012023

[7] ——, "A Database for Counterfeit Electronics and Automatic Defect Detection Based on Image Processing and Machine Learning," 2016, Accessed: 2017-07-04. [Online]. Available: http://www.counterfeit-ic.org/

[8] U. Guin, D. DiMase, and M. Tehranipoor, "A Comprehensive Framework for Counterfeit Defect Coverage Analysis and Detection Assessment," *Journal of Electronic Testing*, vol. 30, no. 1, pp. 25–40, 2014.

[9] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Boston, MA, USA: Pearson, 2007.

[10] G. Stockman and L. G. Shapiro, *Computer Vision*, 1st ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2001.

[11] R. Mehta and K. O. Egiazarian, "Rotated local binary pattern (rlbp)-rotation invariant texture descriptor." in *ICPRAM*, 2013, pp. 497–502.

[12] T. Ojala, M. Pietikäinen, and D. Harwood, "Performance Evaluation of Texture Measures with Classification based on Kullback Discrimination of Distributions," in *Proceedings of International Conference on Pattern Recognition (ICPR)*, 1994, pp. 582–585 vol.1.

[13] K. I. Laws, "Textured Image Segmentation," Ph.D. dissertation, University of Southern California, Jan 1980, Accessed: 2017-07-04. [Online]. Available: http://cdm15799.contentdm.oclc.org/cdm/ref/collection/p15799coll3/id/498876

[14] B. Zitova and J. Flusser, "Image registration methods: a survey," *Image and vision computing*, vol. 21, no. 11, pp. 977–1000, 2003.

[15] S. Lankton and A. Tannenbaum, "Localizing region-based active contours," *Image Processing, IEEE Transactions on*, vol. 17, no. 11, pp. 1–11, Nov. 2008.

[16] "Hamamatsu phemos-1000 emission microscope." [Online]. Available: "https://www.hamamatsu.com/us/en/product/type/C11222-16/index.html"

[17] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: Analysis and implementation," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 7, pp. 881–892, 2002.

[18] P. J. Rousseeuw, "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.