Granulometric Analysis of Spots in DNA Microarray Images

Behara Latha[1]* and Balasubramanian Venkatesh[2]*

[1] *Faculty of Engineering and Technology, Multimedia University, 75450 Melaka, Malaysia;* [2] *Department of Electrical and Computer Engineering, University of New Brunswick, Fredericton, NB E3B 5A3, Canada.*

As the topological properties of each spot in DNA microarray images may vary from one another, we employed granulometries to understand the shape-size content contributed due to a significant intensity value within a spot. Analysis was performed on the microarray image that consisted of 240 spots by using concepts from mathematical morphology. In order to find out indices for each spot and to further classify them, we adopted morphological multiscale openings, which provided microarrays at multiple scales. Successive opened microarrays were subtracted to identify the protrusions that were smaller than the size of structuring element. Spot-wise details, in terms of probability of these observed protrusions, were computed by placing a regularly spaced grid on microarray such that each spot was centered in each grid. Based on the probability of size distribution functions of these protrusions isolated at each level, we estimated the mean size and texture index for each spot. With these characteristics, we classified the spots in a microarray image into bright and dull categories through pattern spectrum and shape-size complexity measures. These segregated spots can be compared with those of hybridization levels.

Key words: microarray, spot, mathematical morphology, image

Introduction

Recent advances in human genetics have enormous implications for the future (*1*, *2*). The new knowledge has created several scientific fields such as bioinformatics, pharmacogenomics, and others. Bioinformatics, also called biological computing, includes the storage, retrieval, and comparison of DNA sequences within the human genome and between genomes of different species. It is a field that offers various tools and techniques to deal with, in particular, data analysis problems, genomics, proteomics, medical informatics, computational biology, and many others (*3*). The tools of bioinformatics are powerful computers and sophisticated software used to manage and analyze the biological gene data. One of the recent advances in the field of bioinformatics is the analysis of gene expression levels in microarrays using image processing techniques. DNA microarrays allow researchers to compare the activities of thousands of genes in normal and cancer cells. It is known that cancer disease exhibits altered patterns of gene expression. DNA microarrays have the potential to identify genes that can be targeted by therapeutic drugs. A typical microarray image consists of a few hundreds to several thousands of spots. The extent of hybridization of these spots determines the level of gene expression in the sample. DNA array forms an orderly arrangement of samples for examining gene expression and is known as gene expression array. These gene chips are paradoxically referred to as microarrays.

Microarray chips consist of various spots with varied degrees of luminescence. The motivation for the present investigation is from the fact that the luminescence property of each spot explains the hybridization level. Present investigation aims to classify DNA microarray images based on various structural and morphological characteristics. DNA microarray image analysis is an important topic from the point of view of bioinformatics. Through applications of image analysis techniques, several studies addressing microarray image classifications were carried out earlier (*3–10*). A microarray consisting of several spots with varied gene expression levels is a gray level image. The gene expression level of a spot can be observed

*** Corresponding authors.**
E-mail: m3160300@mmu.edu.my;
behara_latha@yahoo.co.in;
venkat@unb.ca

through its luminescence. In other words, the brightness, which is expressed in terms of gray level of each spot, needs to be investigated by image processing tools. However, a microarray has spots of several categories ranging from very dull to very bright luminescence. The higher the brightness, the higher the gene expression level, and *vice versa*. To classify these spots automatically, one of the potential techniques is the pattern spectrum procedure. It is proposed to implement the pattern spectrum procedure on an available DNA microarray image for the purpose of classifying its spots.

In this work, we investigated DNA microarray images based upon the luminescence characteristics using morphological tools. These processed images were analyzed for changes during the processing phases. The changes were quantified using parameters that reflect the shape and size of the spots being analyzed. They exhibited a unique pattern during processing for spots with different luminescence.

Basic Morphological Transformations

In order to understand the granulometric analysis, we briefly explain the basic morphological transformations such as erosion, dilation, opening, and closing. Morphological operations will be performed on the grayscale image (f) by means of a binary structuring element (B). The structuring element (B) is like a matrix of size $n \times n$ ($n \in$integers).

$$\text{Erosion of an image } f \text{ by octagon } B = (f \ominus B) \quad (1)$$

$$\text{Dilation of an image } f \text{ by octagon } B = (f \oplus B) \quad (2)$$

$$\text{Opening of an image } f \text{ by octagon } B = (f \circ B) = [(f \ominus B) \oplus B] \quad (3)$$

$$\text{Closing of an image } f \text{ by octagon } B = (f \bullet B) = [(f \oplus B) \ominus B] \quad (4)$$

$$\text{Octagon of size } n = B_n = B \oplus B \oplus B \oplus \ldots \oplus B \ (n \text{ times}) \quad (5)$$

Multiscale transformations of the four basic types given in (1)–(4) are expressed as:

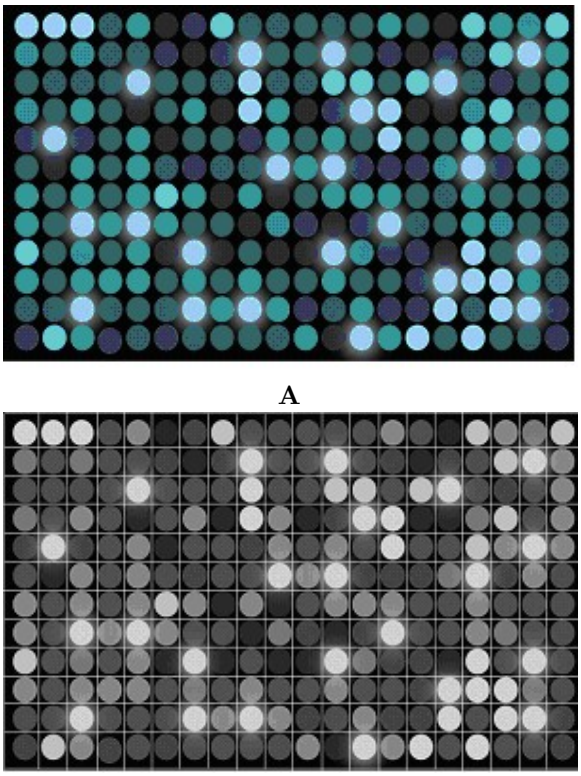
$$\left.\begin{array}{l} \text{Erosion} : (f \ominus B_n) \\ \text{Dilation} : (f \oplus B_n) \\ \text{Opening} : [(f \ominus B_n) \oplus B_n)] \\ \text{Closing} : [(f \oplus B_n) \ominus B_n]] \end{array}\right\} \quad (6)$$

By following a specific criterion, image (f) will be transformed by convoluting it with the structuring element (B). Morphological erosion is nothing but replacing a center pixel with the minimum intensity value from the neighborhood positions where there will be 20 neighborhood values. We consider an octagonal type of element of size 5×5 (Figure 1C). Performing this operation twice with similar characteristic information of the structuring element gives the effect of morphological erosion of the image by structuring element of size 9×9. The reason behind choosing an octagonal template is that it is a circular disk in eight-connectivity grid (*11*). This transformation can be iteratively performed to achieve multi-scale morphological erosion. In contrast to this transformation, for morphological dilation, the central value within the specified sub-image in terms of structuring element will be replaced with maximum value occurring in the neighborhood positions. By following this minimum and maximum criteria through convolution, an image can be transformed by desired morphological transformation. Morphological erosion and dilation will make the image darker and brighter respectively. Increasing the levels of iteration can increase the area extents of these zones. The combination of morphological erosion followed by dilation forms a transformation called opening. This opening transformation can be performed with increasing size of structuring element to have the effect of multi-scale opening. By systematically increasing the size of the structuring element with equal intervals, granulometries can be performed. For more details on the basic morphological transformations and their wide-ranging applications, readers may refer to Serra (*12*).

To unravel several important spot characteristics, we employed these morphological transformations in a systematic way on a sample microarray image (Figure 1A) with an aim to compute spot-wise pattern spectrum (*4*), in other words, gray level granulometries and shape-size complexity measures. With these characteristics, we classified the spots in a microarray. In the section that follows we provide a simple but elegant framework, based on mathematical morphology, to unravel various morphological characteristics of microarray spots.

Proposed Method of Analysis

We recorded the spatial coordinates of each spot embedded within a grid that we placed on the microarray



A

B

	1	1	1	
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
	1	1	1	

C

Fig. 1 A. DNA microarray image. Spots with different luminescence expressing varied hybridization or gene expression levels are conspicuous; **B**. Spots with dividing grid lines superposed to understand characteristics of each spot forming a sub-image; **C**. primitive octagon template of size 5×5, which is considered as B_n, $n = 1$.

image such that each spot was centered in the respective grid (Figure 1B). Further, we computed general statistics of each spot within the specified grid to understand the dynamic range. Shape-size complexity measures, such as average size and roughness of each sub-image consisting of a spot, quantified the level of gene expression.

We considered a microarray data that consisted of 240 spots (Figure 1A). It is obvious from these data that the spots with varied levels of gene expression were conspicuous in terms of gray levels. Let $f(\mathrm{x}, \mathrm{y})$ denote a gray level microarray image consisting of regular spots with varied dynamic ranges that reflect the hybridization level. From the spot-wise histograms, we understood that the higher the dynamic range, the higher the hybridization level. Each spot in this image (Figure 1B) was appropriately indexed from the spot starting from the left to the right and from the top to the bottom. In other words, an equally spaced square grid was placed on the spots such that each spot was symmetrically embedded within the grid. The coordinates of each grid were considered as the spatial coordinates of a sub-image consisting of a spot. This exercise facilitated to estimate spot-wise gray level distribution and other measures. We performed gray scale granulometries on the entire image by employing an octagonal type of structuring element (Figure 1C). Probabilistic size distribution and its density function were computed to estimate the *contributing area* due to a significant gray level.

To perform granulometries, the opening transformation is employed. Structuring element (B) in this work is a matrix of primitive size 5×5 (Figure 1C). B is of octagon shape and is symmetric with respect to the origin.

The emphasis in this section is to develop a framework to understand the gene expression levels that are depicted in the form of varied luminescence values in each sub-image. To avoid extensive computations where it may require the analysis of 256 gray levels since we considered an 8-bit microarray chip, the microarray image of size 810×470 pixels was transformed into multiscale images by performing iterative openings up to 20 cycles by means of a binary octagonal structuring element. The reason for performing opening up to 20 cycles is due to the fact that after 20 cycles of iterative openings, the spots in the microarray image are vanished. To compute the pattern spectrum for each spot and the whole microarray image, we subtracted the succeeding levels of opened images by taking the algebraic difference among these images as below:

$$(f \circ B_{n-1}) - (f \circ B_n) \tag{7}$$

The non-zero pixels that are lost from successive levels of opened microarray images provide areas (A). These areas provide the basis to construct pattern spectrum values by means of octagonal element, which we express as

$$A\left[(f \circ B_{n-1}) - (f \circ B_n)\right] \tag{8}$$

Let f represent a gray level image of DNA microarray consisting of several spots with different luminescence. To compute probabilistic size distribu-

tion and probability size density functions, we performed multi-scale opening by increasing the size of the binary structuring element iteratively up to "n" number of times (Figures 2A–U). In these figures, the evolution of spots under the influence of multiscale opening is obvious. Figures 2I–U show significant different information from that of Figures 2A–H, which is due to the fact that certain spots that possess varied gray level compositions are filtered out. This process is also called granulometry that is shown as:

$$f \circ B_n, \tag{9}$$

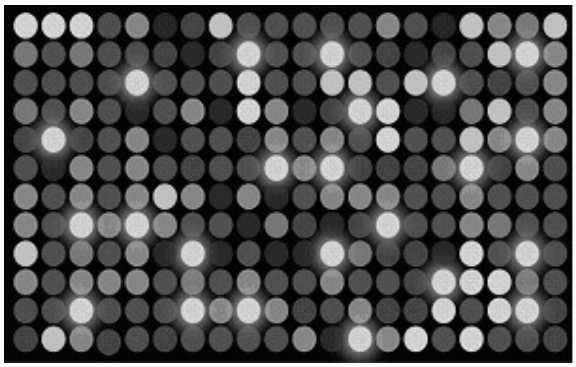
where n ranges from 0 to N.

From these granulometries, we compute pattern spectrum (PS) as

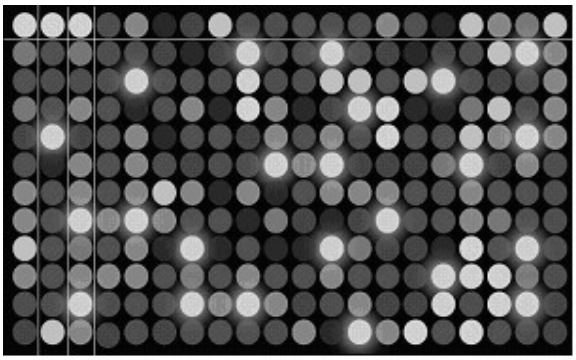
$$PS_f = A\left[(f \circ B_{n-1}) - (f \circ B_n)\right], \tag{10}$$

where n is greater than or equal to 0.

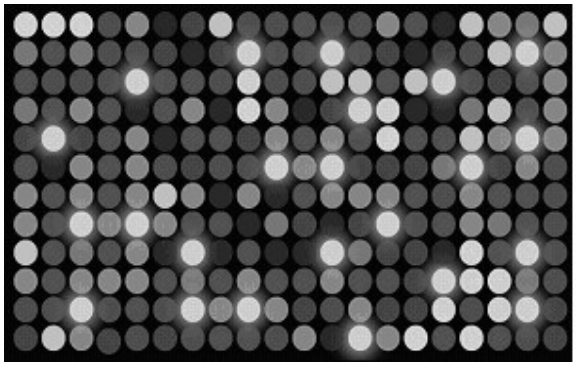
Eq. (10) implies that the PS equals the area occupied by the image obtained by subtracting the image opened using B_{n+1} from the one opened using B_n. This subtraction is nothing but the algebraic difference between the two images. We then computed pattern spectrum values for each spot (Figure 1B).



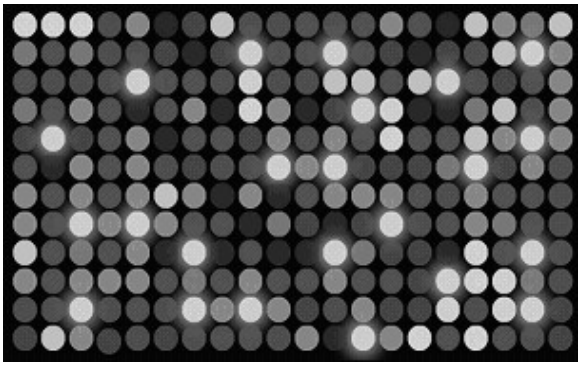
A. Original microarray image



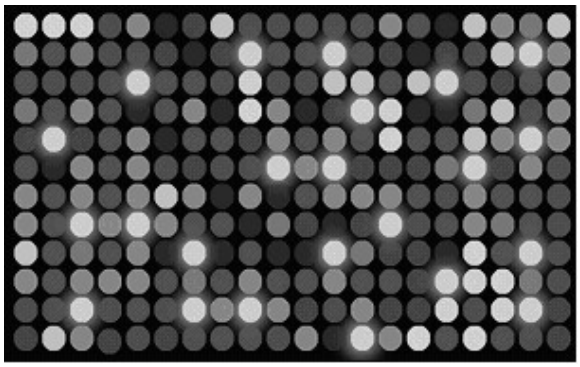
B. Microarray image after 1 opening



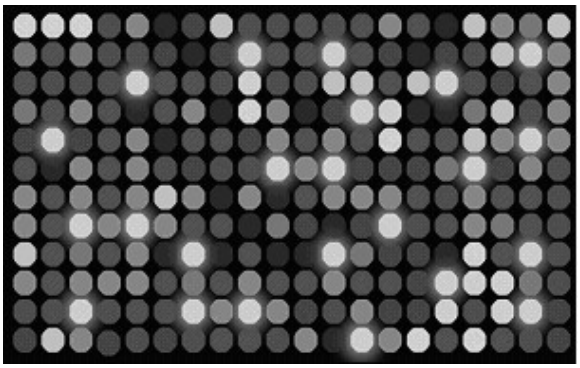
C. Microarray image after 2 openings



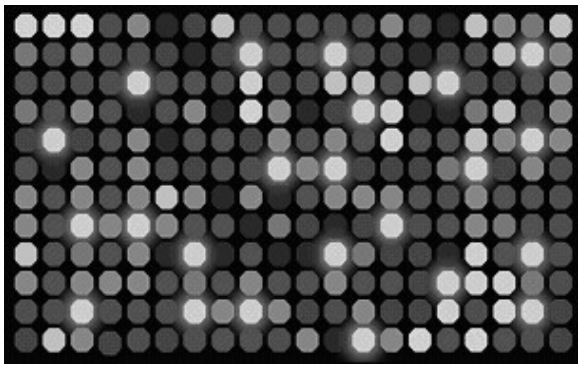
D. Microarray image after 3 openings



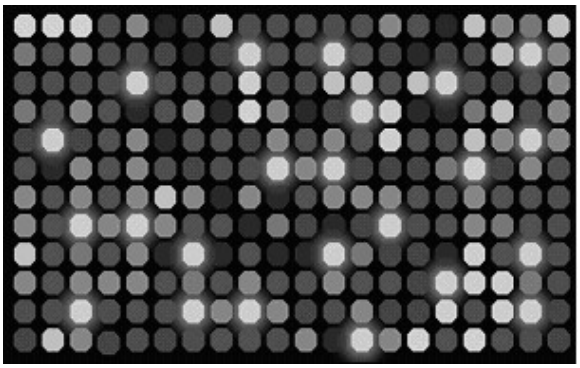
E. Microarray image after 4 openings



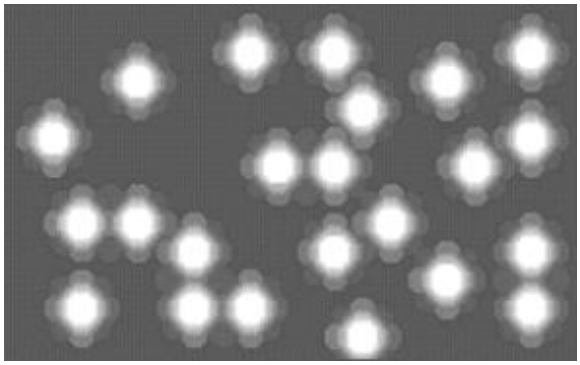
F. Microarray image after 5 openings



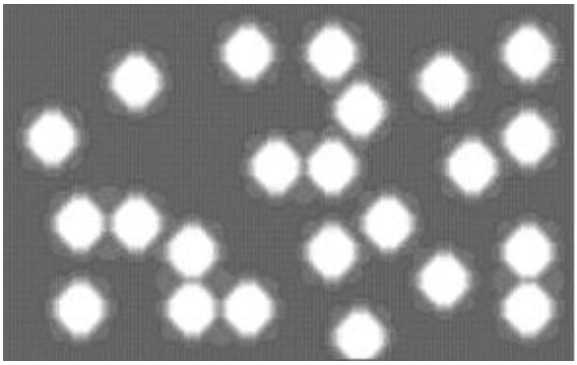
G. Microarray image after 6 openings



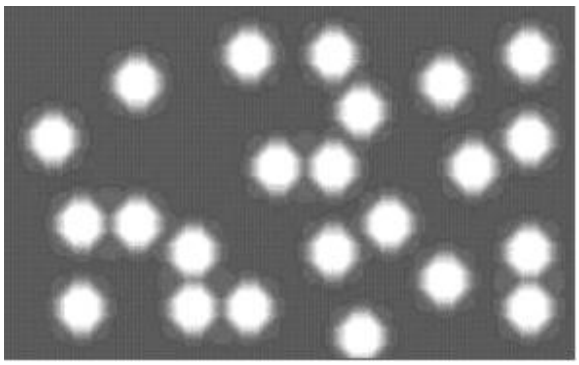
H. Microarray image after 7 openings



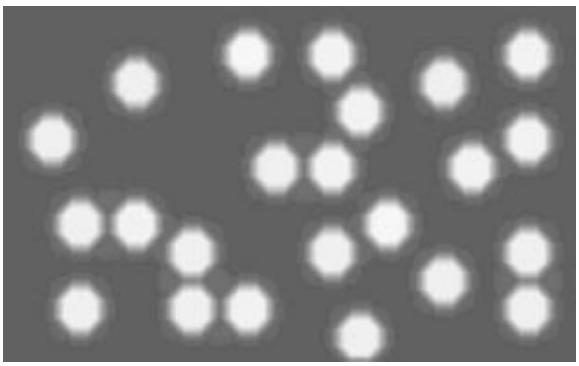
I. Microarray image after 8 openings



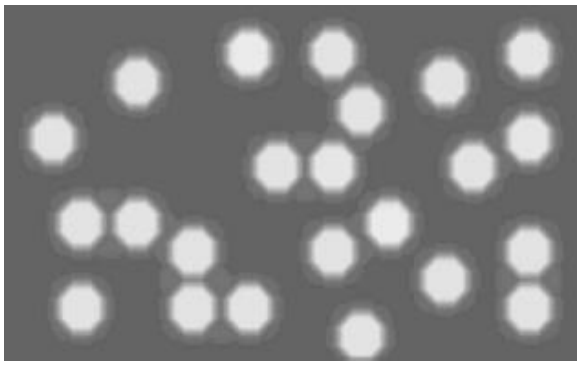
J. Microarray image after 9 openings



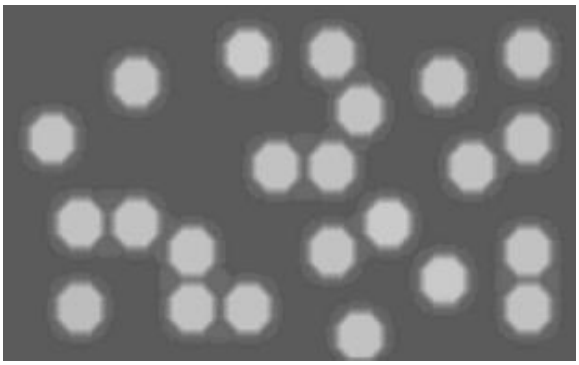
K. Microarray image after 10 openings



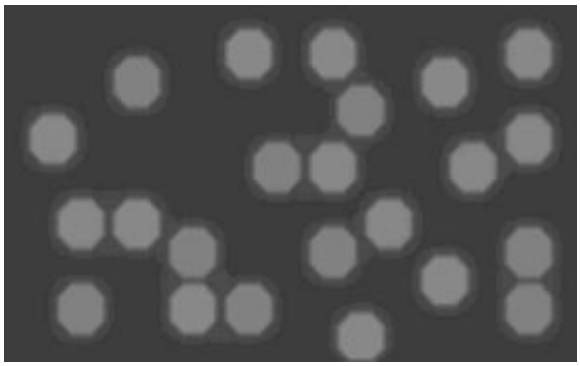
L. Microarray image after 11 openings



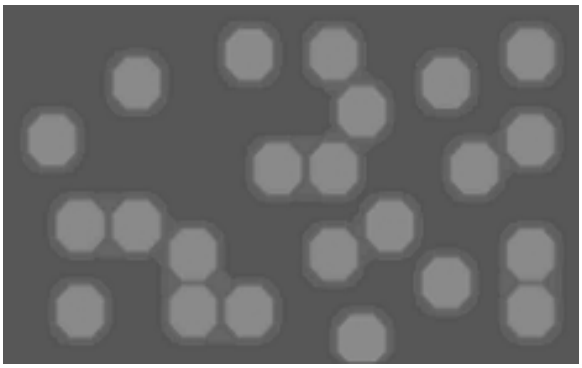
M. Microarray image after 12 openings



N. Microarray image after 13 openings



O. Microarray image after 14 openings



P. Microarray image after 15 openings

Q. Microarray image after 16 openings

R. Microarray image after 17 openings

S. Microarray image after 18 openings

T. Microarray image after 19 openings

U. Microarray image after 20 openings

Fig. 2 A–U. Granulometric analysis of spots after respective cycle of gray level openings performed by means of octagonal structuring element. The brightness and contrast properties of I–U are readjusted for better legibility.

The four parameters that we computed for each spot include pattern spectrum, probability function, average size, and average roughness. Probability function is an important estimate that we computed by following a simple framework based on multiscale opening, algebraic difference between the two functions, total area of the sub-image consisting of a spot, and pattern spectrum. The algebraic difference between two functions is the difference between the image opened by n^{th} size of structuring element (B_n) and the one opened by $(n+1)^{\text{th}}$ size. The information that is lost will be attained through this algebraic difference. The information loss was quantified through number of pixels (Figures 3A–T). We term this algebraic difference as pattern spectrum. The pattern spectrum at respective opening cycles is employed to compute the probability function as below:

$$p_\lambda = \frac{p_x(k, B)}{A(X)}, \tag{11}$$

where p_λ = probability function,
p_x = pattern spectrum,
A(X) = area of the original function (original image size),
B = structuring element,
K = size of the structuring element.

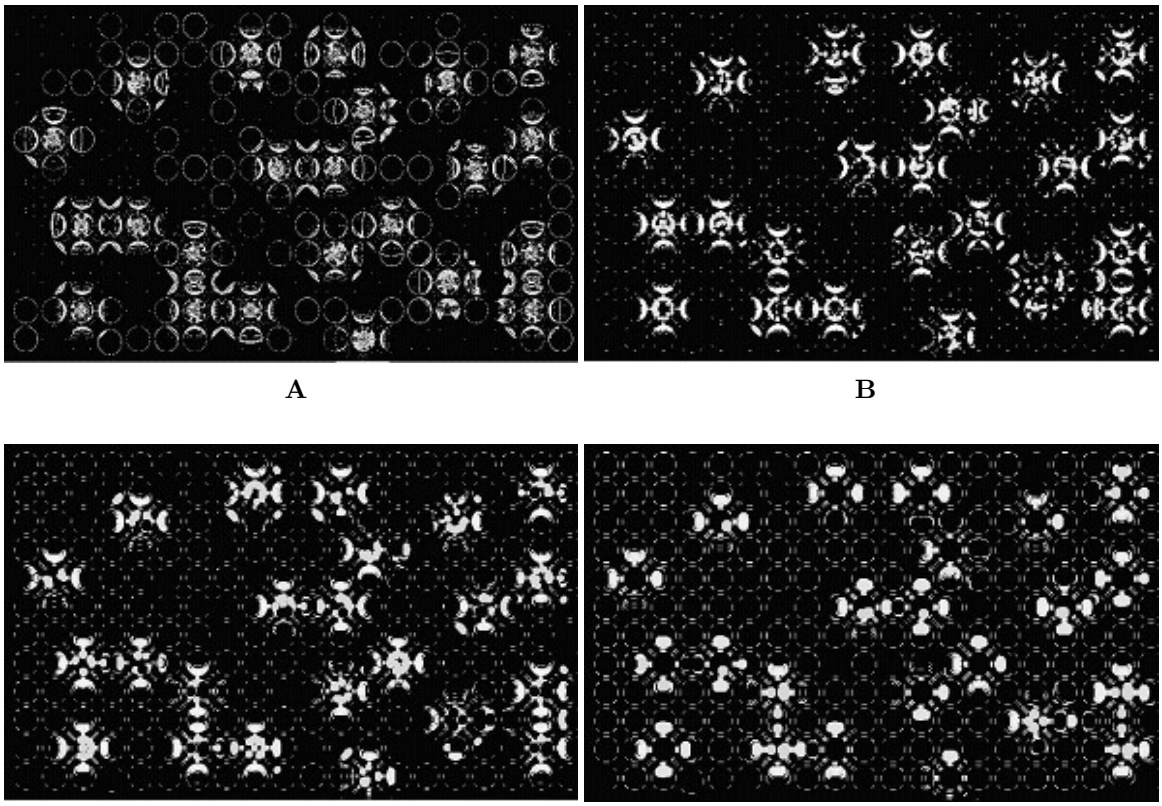
Further, we estimate average size and average roughness for each spot by employing these probability functions as follows:

Average size:

$$\overline{n}\,(f/B) = \sum_{n=0}^{N} n p_\lambda\,(n)\,, \tag{12}$$

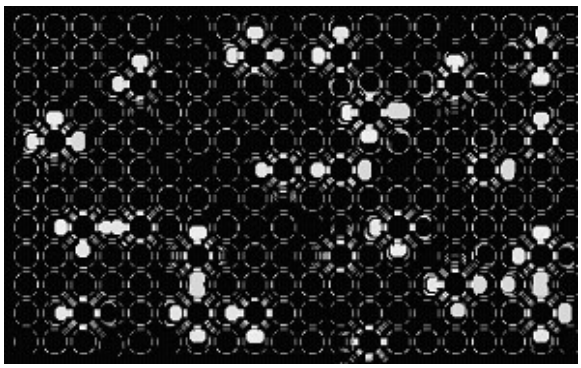
Average roughness:

$$H\,(f/B) = -\sum_{n=0}^{N} p_\lambda\,(n) \log\left[p_\lambda\,(n)\right] \tag{13}$$

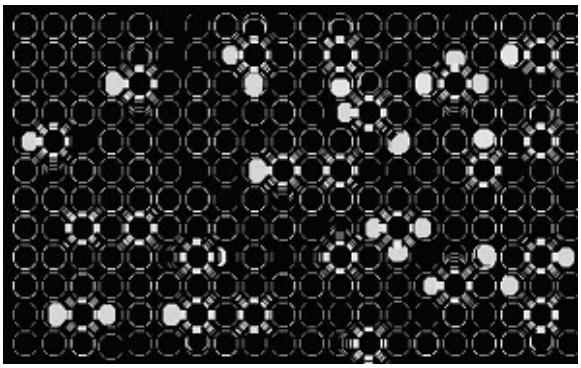


A **B**

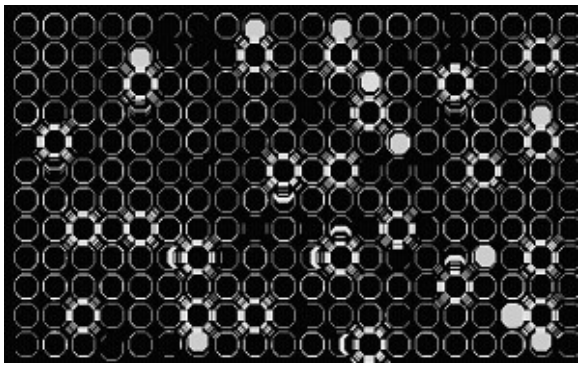
C **D**



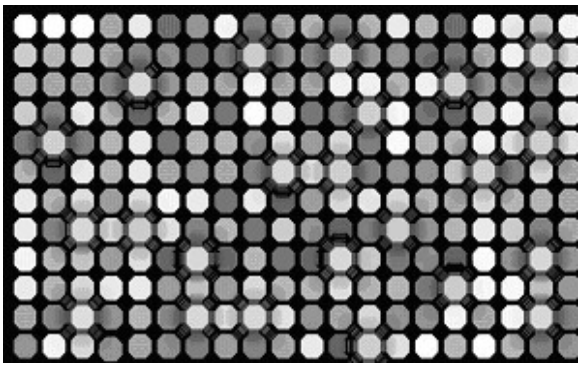
E



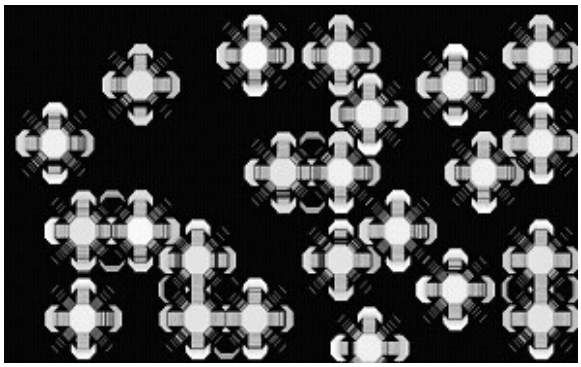
F



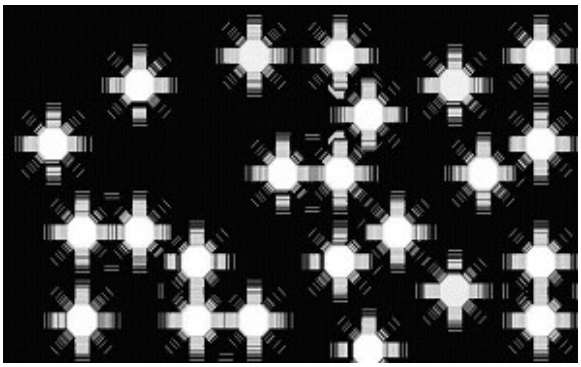
G



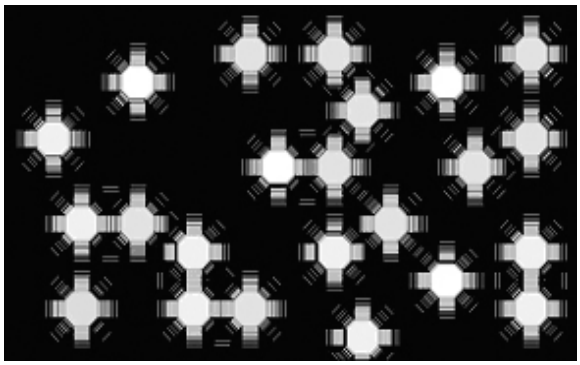
H



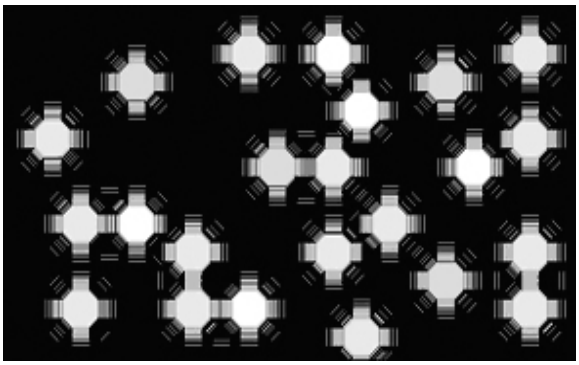
I



J



K



L

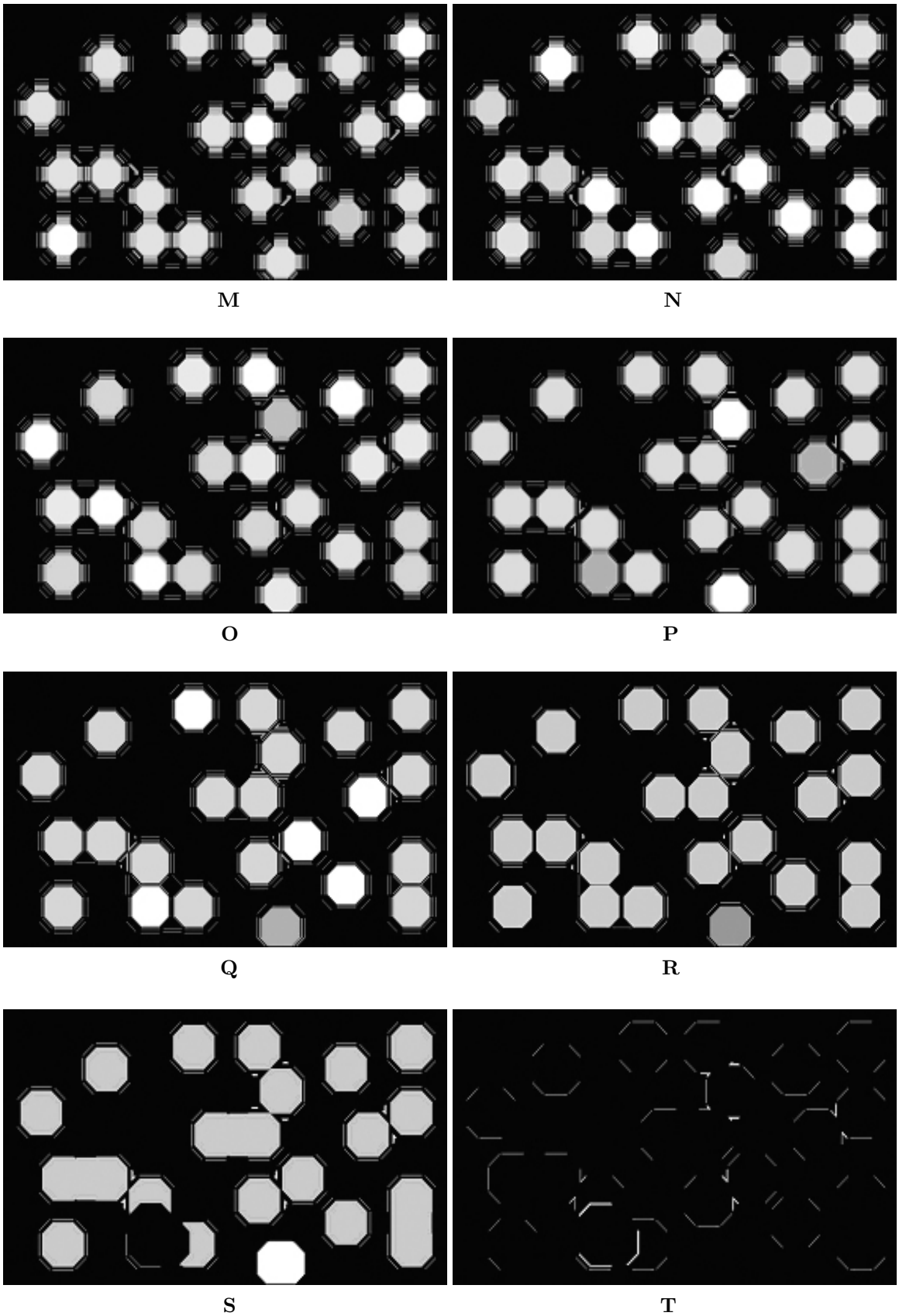


Fig. 3 A–T. The images that were obtained by subtracting the respective level of opened spot images from the preceding level. The different images were achieved by simple algebraic difference.

To summarize, this whole study provides a simple framework to classify DNA microarray spots into several categories. The steps of the proposed method are given in Figure 4.

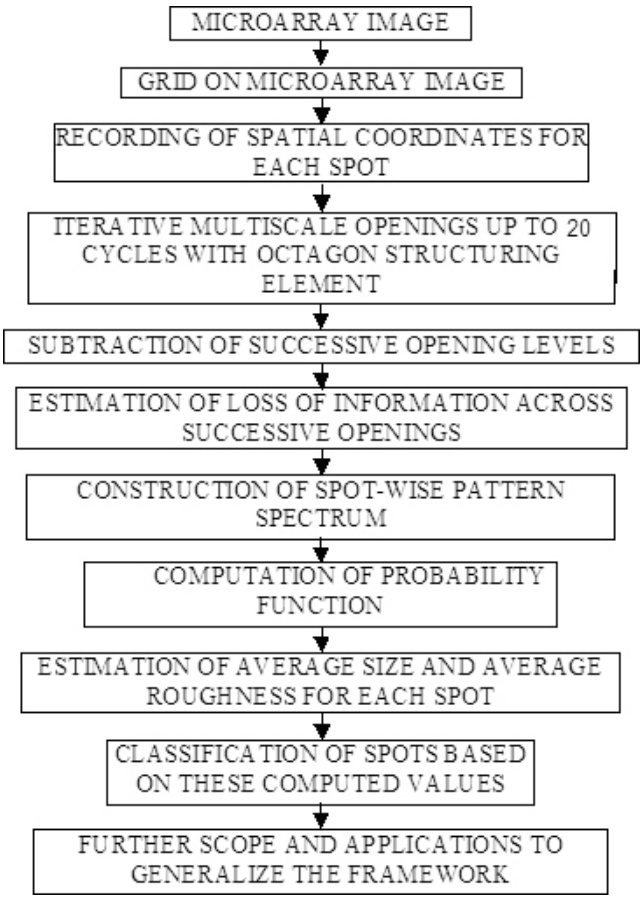


Fig. 4 Steps of the proposed method of analysis.

Results and Discussion

Spot-wise (Sub-image-wise) pattern spectrum values were computed for all 240 spots in the microarray image by computing the area of non-zero pixels (A) that existed in the respective subtracted images. Equation (8) enables this step. A sequence of these subtracted images is illustrated in Figure 3 (A–T). To have a better visibility, the subtracted images are equalized histograms. Table 1 depicts spot-wise spatial coordinates, average size, and average roughness. By employing equations (11)–(13), we estimated respectively the probability functions, average size, and average roughness for each spot.

On physical correlation, it is observed that the spots with higher luminescence have higher roughness. It may be inferred that the higher the roughness of a spot, the higher the luminescence and thus the hybridization or gene expression level. Spot-wise average size and roughness values are graphically represented for better legibility (Figures 5A and 5B). A simple double-logarithmic graph between average size and average roughness for all the 240 spots is shown in Figure 5C. It is observed that brighter spots form a cluster on positive side (above the horizontal axis) of the graph. By employing these roughness and size values, one can classify these spots using a threshold limit in an automated manner. From Table 1, we can infer that spots with roughness values between 0.8 and 1.1 represent bright ones.

From this analysis, the larger values of average size estimated by the pattern spectrum procedure indicate spots with higher luminescence characters. Meanwhile, roughness values would provide a basis to understand the topology-based classification. The bright spots that are obvious from Figure 1A evidently possess average roughness values in the range of 0.8 and 1.1. Other spots possessing the values beyond this

Table 1 Spot-wise Average Size and Average Roughness Parameters Computed Through the Pattern Spectrum Procedure

No.	Spatial Coordinates				Average	Average
	Xmin	Xmax	Ymin	Ymax	Size	Roughness
1	4	44	7	49	4.73	0.49
2	4	44	48	89	5.11	0.51
3	4	44	88	127	5.12	0.52
4	5	43	129	166	4.91	0.39
5	5	43	166	207	5.13	0.52
6	7	43	207	247	4.75	0.42
7	4	44	246	286	4.39	0.42
8	4	44	286	326	10.00	1.04
9	3	45	325	367	59.60	2.83
10	4	43	365	405	12.20	1.22
11	4	43	405	445	9.73	1.06
12	4	44	445	485	61.90	2.84
13	4	43	485	523	11.60	1.09
14	4	44	523	563	5.24	0.51
15	5	44	563	602	5.25	0.53
16	7	41	603	642	7.43	0.36
17	5	44	641	683	4.18	0.45
18	3	45	682	722	5.12	0.51
19	2	43	723	763	5.00	0.51
20	2	43	761	801	4.79	0.46
21	44	81	8	49	5.40	0.52
22	44	81	49	87	5.77	0.52
23	44	82	87	127	5.72	0.56
24	44	82	126	168	11.00	1.18
25	44	84	167	207	69.30	2.82
26	42	82	207	248	12.70	1.20
27	46	80	248	285	7.44	0.35
28	43	81	285	325	59.00	2.75
29	45	82	326	365	181.00	0.87
30	44	81	366	405	75.00	2.75
31	44	81	405	444	60.00	2.74
32	44	82	444	485	180.00	0.92
33	43	82	485	523	76.40	2.68
34	44	82	523	566	4.76	0.47
35	45	80	566	602	14.30	1.15
36	43	81	602	642	58.20	2.72
37	44	80	645	682	4.87	0.46
38	44	81	682	722	5.86	0.54
39	44	80	723	761	5.70	0.53
40	45	83	762	802	8.87	0.99
41	83	120	9	48	5.54	0.53
42	82	119	49	88	5.41	0.51
43	83	119	88	129	5.02	0.49
44	81	118	128	167	61.90	2.79
45	82	120	167	207	181.00	0.94
46	82	118	207	246	81.10	2.64
47	81	119	246	285	5.78	0.55
48	82	121	286	326	15.60	1.47
49	82	118	326	365	93.30	2.52
50	81	118	365	404	24.50	1.81
51	80	118	405	444	17.40	1.51
52	82	118	445	484	98.90	2.41
53	81	118	485	523	82.60	2.46
54	83	119	525	563	14.20	1.28
55	80	121	564	605	65.60	2.58
56	81	120	602	644	146.00	1.41
57	81	120	643	681	4.45	0.49
58	82	118	683	724	5.57	0.50
59	83	119	722	762	5.32	0.48
60	81	118	763	803	47.10	2.64
61	119	156	8	47	9.90	1.00
62	118	157	48	87	62.60	2.77
63	120	155	88	128	13.30	1.23
64	119	153	128	169	19.20	1.59
65	121	158	167	206	87.30	2.49
66	120	157	205	247	23.10	1.75
67	118	158	246	288	5.28	0.56
68	120	158	285	326	5.02	0.51
69	120	156	323	369	5.18	0.54
70	119	157	364	407	5.35	0.56
71	123	155	407	443	5.65	0.35
72	120	157	442	487	62.10	2.83
73	120	156	487	524	182.00	0.81
74	119	157	523	566	77.30	2.74
75	119	153	563	598	12.90	1.18
76	124	154	605	639	86.80	2.41
77	118	158	640	681	8.46	0.97
78	119	161	681	723	69.70	2.77
79	119	159	721	762	27.20	1.98
80	118	156	758	804	12.00	1.16
81	157	193	7	48	59.80	2.76
82	157	196	48	89	181.00	0.91
83	156	193	89	126	78.10	2.69
84	156	192	127	165	5.80	0.54
85	156	195	168	209	5.32	0.55
86	157	192	208	245	5.36	0.45
87	157	195	248	283	5.94	0.56
88	158	194	285	327	5.35	0.53
89	157	193	326	365	10.30	1.09
90	157	196	365	406	67.80	2.70
91	154	195	404	446	24.00	1.72
92	154	194	444	488	75.10	2.59

Table 1 ***Continued***

No.	Spatial Coordinates				Average	Average
	Xmin	Xmax	Ymin	Ymax	Size	Roughness
93	157	193	483	523	103.00	2.53
94	159	194	523	565	24.80	1.81
95	157	194	564	602	5.72	0.53
96	155	193	603	643	9.19	1.04
97	156	194	642	683	41.40	2.52
98	156	194	682	721	178.00	1.06
99	156	195	721	762	99.70	2.51
100	156	195	761	803	6.36	0.68
101	195	232	8	47	13.60	1.28
102	197	232	48	83	74.90	2.38
103	195	232	88	127	21.60	1.59
104	196	232	127	168	5.66	0.55
105	195	232	166	208	5.54	0.58
106	192	231	206	246	5.23	0.48
107	194	233	246	285	4.95	0.49
108	194	233	288	325	5.32	0.50
109	194	231	326	364	59.30	2.74
110	195	233	367	407	64.70	1.07
111	194	231	404	446	141.00	2.08
112	195	231	444	488	181.00	0.95
113	195	232	484	524	77.60	2.70
114	195	231	524	566	4.67	0.45
115	193	230	564	603	5.11	0.47
116	192	232	603	645	45.50	2.52
117	194	231	644	685	11.00	1.13
118	193	231	682	721	78.80	2.55
119	193	231	720	760	37.50	2.16
120	194	232	763	801	5.54	0.56
121	232	269	8	49	5.49	0.55
122	229	268	49	87	10.10	1.07
123	233	269	110	89	125.00	1.85
124	232	269	130	165	23.00	1.61
125	232	271	168	208	69.00	2.69
126	232	268	205	247	14.50	1.28
127	232	268	249	287	5.74	0.55
128	232	266	287	322	5.35	0.44
129	231	269	324	367	17.70	1.48
130	233	268	265	403	30.50	2.04
131	231	269	408	444	36.60	1.94
132	232	269	444	481	102.00	2.36
133	230	268	483	524	30.70	1.97
134	231	272	524	563	68.60	2.65
135	233	267	563	605	12.60	1.21
136	233	268	602	643	12.50	1.24
137	232	270	641	685	4.48	0.50
138	231	270	683	724	9.27	1.02
139	233	267	720	759	58.10	2.65
140	232	267	761	805	28.10	1.87
141	269	305	7	48	5.44	0.52
142	269	306	51	86	59.40	2.77
143	271	307	89	126	163.00	0.83
144	270	307	128	167	122.00	2.04
145	267	307	166	203	180.00	0.97
146	269	307	207	247	82.50	2.59
147	269	307	247	284	65.90	2.78
148	269	307	286	327	14.40	1.36
149	270	300	326	360	5.14	0.41
150	270	307	365	407	5.31	0.55
151	269	307	406	442	9.95	1.03
152	270	301	446	485	47.80	2.44
153	269	307	485	524	74.00	2.83
154	270	307	523	566	76.00	1.07
155	269	306	562	601	83.70	2.67
156	270	305	602	643	5.23	0.51
157	271	307	640	681	4.55	0.48
158	270	306	682	790	106.00	2.41
159	268	307	760	780	174.00	1.22
160	307	288	760	801	127.00	1.93
161	309	344	6	50	5.09	0.50
162	307	342	48	88	16.10	1.44
163	309	345	88	129	81.50	2.67
164	307	343	129	165	35.70	1.91
165	307	344	166	207	84.90	2.59
166	307	342	206	237	63.80	2.52
167	307	344	243	288	179.00	1.04
168	308	342	288	326	75.20	2.53
169	305	345	324	367	5.19	0.55
170	307	343	367	405	5.36	0.46
171	308	343	404	436	41.90	2.31
172	304	344	444	485	158.00	1.05
173	306	344	484	524	90.20	2.60
174	305	344	524	563	94.20	2.59
175	306	344	564	603	27.30	1.91
176	306	341	604	646	46.30	2.41
177	307	345	641	684	4.37	0.47
178	308	344	683	720	8.64	0.92
179	307	343	720	761	79.00	2.71
180	305	342	761	805	45.20	2.23
181	345	382	7	50	5.14	0.53
182	344	379	50	85	9.40	0.97
183	345	382	89	128	68.00	2.62
184	346	382	128	168	15.10	1.28
185	345	380	168	207	5.70	0.53
186	342	381	208	247	22.40	1.78

Table 1 ***Continued***

No.	Spatial Coordinates				Average	Average
	Xmin	Xmax	Ymin	Ymax	Size	Roughness
187	344	382	246	285	116.00	1.82
188	342	380	289	326	21.70	1.65
189	346	385	326	366	69.80	2.69
190	342	381	367	404	12.90	1.18
191	342	379	404	443	17.10	1.48
192	343	383	443	483	87.70	2.59
193	345	379	484	520	24.40	1.73
194	343	382	525	564	5.47	0.54
195	346	380	563	602	64.70	2.77
196	340	381	602	643	144.00	1.43
197	345	382	642	683	4.58	0.47
198	344	381	682	721	8.15	0.89
199	346	381	722	763	66.50	2.63
200	343	382	761	802	24.20	1.75
201	382	418	9	46	5.24	0.46
202	382	417	15	89	38.20	2.47
203	382	419	88	129	160.00	0.94
204	382	419	128	166	79.50	2.69
205	382	419	169	204	5.92	0.52
206	381	418	206	246	52.90	2.67
207	384	418	247	286	163.00	0.86
208	382	418	288	327	104.00	2.05
209	381	420	325	364	155.00	1.16
210	381	419	365	404	78.50	2.57
211	381	419	406	445	4.97	0.46
212	384	419	445	485	14.70	1.28
213	383	419	486	524	72.80	2.58
214	383	422	525	561	19.60	1.40
215	381	419	563	601	14.40	1.33
216	385	418	603	643	78.20	2.68
217	381	418	643	683	4.46	0.43
218	382	420	682	725	46.80	2.64
219	382	418	720	761	157.00	1.11
220	380	416	762	800	103.00	2.46
221	418	457	8	46	4.81	0.42
222	420	459	50	90	16.2	1.37
223	420	458	87	126	85.40	2.57
224	422	458	126	165	6.42	0.63
225	420	456	167	205	18.90	1.24
226	421	457	208	245	12.50	1.17
227	419	457	246	286	80.90	2.66
228	418	456	286	327	30.40	2.02
229	419	455	326	362	93.40	2.69
230	419	458	367	407	18.70	1.58
231	420	458	402	445	5.40	0.56
232	422	459	446	483	78.50	2.22
233	421	458	483	526	160.00	0.97
234	418	457	524	565	78.20	2.51
235	422	458	563	600	5.86	0.56
236	419	457	602	641	5.49	0.53
237	419	457	644	682	4.18	0.40
238	418	457	683	722	10.10	1.02
239	419	458	721	767	76.10	2.63
240	420	460	761	804	25.90	1.75

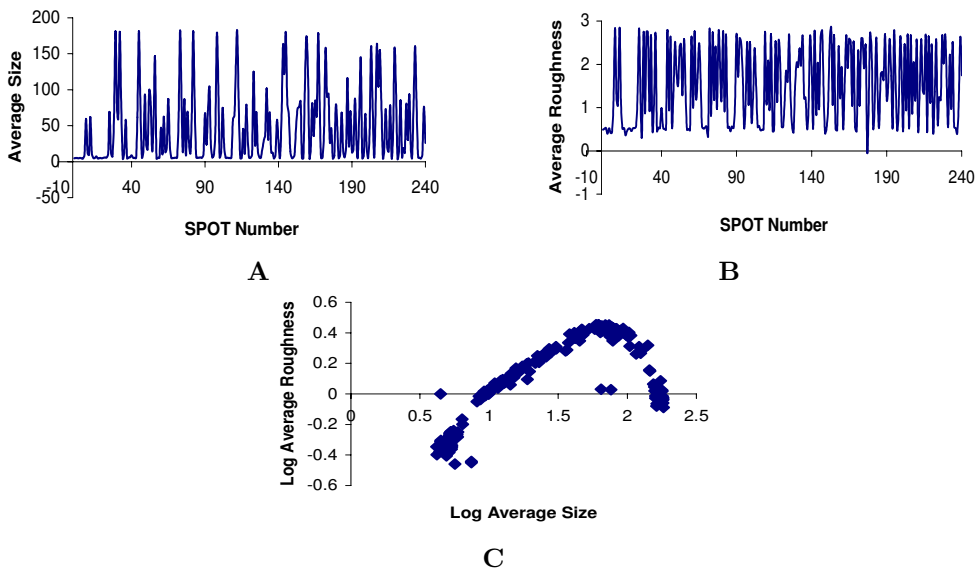


Fig. 5 A. Graph showing spot-wise average sizes. **B.** Graph showing spot-wise average roughness values. **C.** Log-log graph between average size and average roughness.

range possess significantly different roughness values. The larger the average size, the larger the size of octagonal element in a spot. From this fact, we infer that these spots consist of larger size of octagonal element. A graph is plotted with spot numbers on X-axis and average roughness on Y-axis (Figure 5B). Similarly, higher average size values of more than 145 were observed for the bright spots with 29, 32, 39, 45, 49, 56, 69, 73, 82, 94, 99, 110, 111, 112, 117, 172, 179, 196, 203, and 233. The corresponding roughness values were observed in the range of 0.8 and 1.1. It is inferred from this unsupervised classification that the spots with higher average size values, which are also bright spots, possess the roughness values between 0.8 and 1.1. This study, which was carried out in unsupervised way, needs further supporting evidence that is possible if the results derived from core microbiological analysis are available.

Conclusion

This study provides a simple framework to classify DNA microarray spots based upon their gene expression levels. This work demonstrates possibility to unravel topological characteristics of microarrays using powerful computers and established tools like mathematical morphology. These characteristics provide new insights in understanding microbiological phenomena in a quantitative manner.

A brief idea about the gist of this work is to identify the presence of gene expression level in the desired DNA sequence. The DNA sequence is converted into a microarray image, which is a gray level image depicting gene expression levels in terms of intensities. Analysis of this microarray image is performed by using concepts from mathematical morphology and computational tools. Then, the spots in the image are categorized based on the average roughness values into bright and dull categories. The motivation for this work is to show a relation between the roughness value and the brightness of a spot. From this study, we can infer that the average roughness is proportional to the brightness or luminescent property of a spot, which is proportional to its gene expression level or hybridisation level. From the sample microarray spot image considered, based on this approach, the salient features observed include:

1. Larger values of average size estimated by the pattern spectrum procedure indicate spots with higher luminescence characters.

2. Roughness values would provide a basis to understand the topology-based classification.

3. Bright spots that are obvious from Figure 1A evidently possess average roughness values in the range of 0.8 and 1.1, while spots beyond this range possess significantly different roughness values.

4. The higher the average size, the higher the size of octagonal element that a spot consists of.

5. Higher average size values of more than 145 are observed for the bright spots with 29, 32, 39, 45, 49, 56, 69, 73, 82, 94, 99, 110, 111, 112, 117, 172, 179, 196, 203, and 233.

Using morphology, the rich topological properties extracted from DNA chips are immensely valuable to cell biologists and scientists who study the roots of cancer and other complex diseases. The application areas include the research on gene function, gene pathways, disease classifications, and disease origin. The parameters (shape-size complexity measures) that we computed may change as we change the characteristic information of the structuring element. By changing the structuring element or template, we plan to characterise the spots in our future studies to generate a spectrum of these topological parameters.

Materials and Methods

Array fabrication

Array fabrication involved preparing a glass slide, obtaining the DNA sequence representing genes of a genome of interest, and depositing (printing) the cDNA sequences on the glass slide. A cDNA sequence, called a cDNA probe, was selected to make the arrays. In order to get this cDNA sequence, one requires cDNA clones and cDNA library. Each cDNA clone was amplified to get many copies using the PCR (polymerase chain reaction) technique. After amplification, the PCR product (the liquid containing the amplified cDNA probes) was deposited on the polylysine-coated glass slide using a set of microspotting pins. A typical example was the spotting of more than 6,000 different PCR generated DNA samples on a polylysine-coated slide measuring 18×18 mm^2. The drops of solution containing cDNA probes formed the spots on the array.

Sample preparation

The experimental material was a tissue sample from a patient. From the cell, total RNA was extracted and

out of which mRNA was isolated. This mRNA was converted into a more stable cDNA by the process of reverse transcription. Thus, one pool of cDNA was prepared from the experimental sample. In the same way, another pool of cDNA was prepared from the normal cell. Hence, for the preparation of microarrays, two pools of cDNA were synthesized.

Experimental and reference target cDNAs' labeling

The cDNAs obtained from experimental mRNA were labeled with red fluorescent dye Cy5 and those from reference sample were labeled with green fluorescent dye Cy3. These are called target cDNAs.

Hybridization

Hybridization refers to the binding of two complementary DNA strands by base pairing. The mixed solution of experimental and reference target cDNAs was applied to the array, which contained the probe cDNA in each spot. A specific spot on the array contained cDNA probes for gene A. The target cDNA in the mixed solution, which was complementary to the probe cDNA of gene A, bound together by base-pairing which was nothing but hybridization.

Microarray quantification

The expression levels of a gene in the experimental and reference cells were measured by the spot intensities of the Cy5 (red) dye and Cy3 (green) dye respectively. Dyes or fluorescent intensities were obtained by scanning the array using a confocal laser microscope. The products resulting from the array scanning process were two 16-bit tagged image file format (TIFF) images. The scanned array area was divided into equally sized pixels and the resulting image contained fluorescence intensities for corresponding pixels.

References

1. Weinzierl, R.O.J. 1999. *Mechanisms of Gene Expression.* Imperial College Press, London, United Kingdom.
2. Solomon, E.P., *et al.* 2002. *Biology.* Thomson Learning, Florence, United States.
3. Balagurunathan, Y., *et al.* 2002. Simulation of cDNA microarrays via a parameterized random signal model. *J. Biomed. Opt.* 7: 507-523.
4. Eisen, M.B. and Brown, P.O. 1999. DNA arrays for analysis of gene expression. *Methods Enzymol.* 303: 179-205.
5. Alon, U., *et al.* 1999. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. Natl. Acad. Sci. USA* 96: 6745-6750.
6. Amaratunga, D. and Cabrera, J. 2001. Analysis of data from viral DNA microchips. *J. Am. Stat. Assoc.* 96: 1161-1170.
7. Angulo, J. and Serra, C. 2003. Automatic analysis of DNA microarray images using mathematical morphology. *Bioinformatics* 19: 553-562.
8. Jornsten, R., *et al.* 2003. Microarray image compression: SLOCO and the effect of information loss. *Signal Processing* 83: 859-869.
9. Deutsch, J.M. 2003. Evolutionary algorithms for finding optimal gene sets in microarray prediction. *Bioinformatics* 19: 45-52.
10. Wernisch, L., *et al.* 2003. Analysis of whole-genome microarray replicates using mixed models. *Bioinformatics* 19: 53-61.
11. Maragos, P. 1989. Pattern spectrum and multiscale shape representation. *IEEE Trans. Pattern Anal. Machine Intell.* 11: 701-716.
12. Serra, J. 1982. *Image Analysis and Mathematical Morphology.* Academic Press, London, United Kingdom.