

Branching profiles for characterising irregularly shaped objects

Thesis submitted to the College of Science and Engineering in partial fulfilment of the requirements for the degree of Master of Science (Mathematics) at Flinders University, Adelaide, Australia

Zayed Asiri, B.Sc.

11 October 2019

Contents

Abstract	2
Acknowledgement	3
1 Introduction	4
2 Background	6
2.1 Previous work on characterising irregular shape	6
2.2 K-means clustering	8
3 Branching profiles	10
3.1 Counting branches	10
3.2 Generating branching profiles	12
4 Applications	14
4.1 Regular patterns	14
4.1.1 Rectangles	14
4.1.2 Hexagons	17
4.2 Yeast colonies	19
4.2.1 Method	20
4.2.2 Results	20
4.2.3 Discussion	23
5 Conclusion	25
Bibliography	26

Abstract

The aim of this thesis is to characterise irregular shapes. Irregular shapes are those such as trees, clouds, land masses, yeast colonies, and cancellous bone. The method of branching profiles is introduced to characterise such shapes that cannot be characterised by using standard shape descriptors such as their height, width and length. The method of branching profiles is to count the number of intersections of the circle within the object at various scales. The number of intersections N are zero when $N < 2$ or $N - 2$ when $N > 2$. We applied the method of branching profiles on twenty images of the yeast colonies from two different strains. The method classified the yeast colonies perfectly. The method of branching profiles contributes to understanding irregular shapes.

Acknowledgement

I would like to express my deep and sincere gratitude to my supervisor A/Prof. Murk Bottema. He is an excellent mathematician and educator who recognises his students' individuality, catering to their strengths and helping them overcome their weaknesses. It's a rare combination of talents.

I am extremely grateful to my parents for their love and prayers. I am very much thankful to my wife, my son and my daughter for their love and prayers.

Chapter 1

Introduction

Topology is the science that studies shape characteristics [6]. Topological characteristics are those that the shape maintains if exposed to a tension without rupture or attachment. Topological characteristics are important for understanding shapes, particularly irregular shapes. Most man-made structures are regular in shape and may be described using a few parameters. For example, the shapes of many buildings are well characterised by their height, width and length. In nature, many objects such as trees, clouds and land masses are irregular in shape. Height, width and length may still have meaning with respect to the overall extent of these items, but do not characterise the local irregular and intricate patterns that define the structure or even function of such items. For example, Indonesia and Saudi Arabia have roughly the same surface area, but Indonesia comprises thousands of islands spread over a vast region while Saudi Arabia comprises a single connected region. The number of connected components is an example of a topological property. Other examples of topological properties are the number of holes and so is the number of branches. The analysis of objects using such topological properties leads to simple shape descriptors suitable for irregular objects and data. As an example, counting branches in irregular shapes is a topological measure. The analyses of shapes by using topology leads to

simplifying our understanding better of the shapes and extraction of data.

The objective of this thesis is to study irregular shapes by counting the number of branch events of a structure at various scales. The number of local branches as a function of the scale may be viewed as a local branching profile. We are going to count branch events as a function of scales to see how branch number changes. Examples of applications come from trabecular structure in cancellous bone, the shape of marbling in beef and the shapes of growing yeast colonies. These examples have appeared in the literature and will be reviewed below. This project will introduce the use of branching profiles to characterise irregular shape such as yeast colonies. Furthermore, although regular shapes have been studied before, we will also use branching profiles to study regular shapes including regular grids of squares and hexagons.

Chapter 2

Background

2.1 Previous work on characterising irregular shape

Digital topology is a strategy that arose in the late 1960s for the investigation of geometric and topological properties of objects in images [8]. Digital topology is a related field that focusses on digital implementations of topological ideas. The outcomes of digital topology provide the mathematical basis for techniques such as image thinning, border following, counter filling and object counting [6].

An example of a complex shape is marbling in beef that was studied [2]. The method that was used to characterise the complex shape of marbling was to find the quantity of branching [2]. In this example, five thousand points within the marbling area were chosen from each slice of marbling in beef [2]. A square box was centred at each chosen point P. This box was called the interrogation box [2]. Only the connected set within the interrogation box containing the point P was retained [2]. The intersection of this connected set and the boundary of the interrogation box include no less than two parts [2]. If a “rod” of marbling enters and exits the interrogation box, then this results in two connected components on the boundary of the interrogation box [2].

There were three branches inside the interrogation box, when the rod was branched only once inside the interrogation box and so forth [2]. Thus, if N is the number of connected components on the boundary of the interrogation box then the number of branch points is zero if $N = 1$ or $N = 2$, and the number of branch point is $N - 2$ if $N > 2$ [2].

Another example of a 3D object is cancellous bone, which was studied to characterise its structure, such as the irregular shape of trabeculae structure in rat tibiae [7]. Cancellous bone refers to the spongy bone at the ends of long bones. The amount of cancellous bone reduces due to ageing or disease but may increase after treatment. The method that was used in [7] to characterise irregular shapes of cancellous bone in rats is called oriented thickness textons, which is to measure the thickness of the bone. It comprises two important cases which are the thickness of the object at different points P , where P is the vector representing thickness at a different orientation [7]. Clustering was used to identify the common local patterns of cancellous bone [7].

A third example of characterising a complex shape is yeast colonies [5]. Yeasts are a set of cells on solid medium and we can study images of their formation. The datasets selected in the study were studied previously [1]. The aim of this study was to characterise the yeast colonies by quantifying the changes in morphology [5]. The method for classifying yeast colonies used was cluster shape primitives (CSPs) [5]. “A shape primitive is computed at an occupied pixel in a binary image by measuring the lengths of the longest line segments, oriented at equally spaced angles, that fit entirely inside the set of occupied pixels, which represent the local shape at the pixel. Common shape patterns may be identified by applying a clustering algorithm to the set of all shape primitives computed, and the centres of the resulting clusters, referred to as CSPs, represent shape patterns that occur commonly throughout a given set of images” [5]. In summary, by considering the changes in patterns of a binary image of yeast colonies we can characterise and identify different

strains or nutrient conditions.

As we have seen above in these examples the authors considered a single fixed radius to characterise such irregular shapes. But, in this thesis the main consideration is to characterise irregular and irregular shapes by looking at the branch profiles. Thus, in Chapter 3 we will demonstrate this technique to characterise different images of yeast colonies from two different strains; AWRI 796 and AWRI R2.

2.2 K-means clustering

K-means clustering is a method for automatically identifying clusters of points in high dimensional Euclidean space. Let S denote a set of points in Euclidean space

$$S = \{x_i = (x_{i_1}, x_{i_2}, \dots, x_{i_n}) : i = 1, 2, \dots, p\}$$

The objective is to group these points into K clusters so as to minimise the total sum distance between all the points and the cluster centres to which they are assigned.

The K-means clustering algorithm starts by assigning an initial set of K cluster centres. Each cluster centre is a point in the n -dimensional Euclidean space. These may be assigned randomly or otherwise. Next, each point in S is assigned to the nearest cluster centre leading to a vector C called the cluster assignment with $C(i)$ being the label (an integer in the range 1 to K) of cluster centre to x_i .

The following algorithm appears in [3]:

Algorithm for K-means clustering:

1. For a given cluster assignment C , the total cluster variance is minimised by

$$\min_{C, \{m_k\}_1^K} \sum_{k=1}^K \sum_{\{i: C(i)=k\}} \|x_i - m_k\|^2$$

yielding the means of the currently assigned clusters for the observations in the set S_k then this finds the new centre

$$\bar{x}_{S_k} = \operatorname{argmin}_m \sum_{i \in S_k} \|x_i - m\|^2$$

Here and throughout, $\|\cdot\|$ will denote the Euclidean norm.

2. Given a current set of means m_1, \dots, m_k

$$\min_{C, \{m_k\}_1^K} \sum_{k=1}^K \sum_{\{i: C(i)=k\}} \|x_i - m_k\|^2$$

is minimised by assigning each observation to the closest (current) cluster mean. That is,

$$C(i) = \operatorname{argmin}_{1 \leq k \leq K} \|x_i - m_k\|^2.$$

For more information see [3].

Chapter 3

Branching profiles

3.1 Counting branches

The theoretical properties of branching profiles will be developed in this thesis to characterise irregular shapes. Some idealised examples will be constructed to prove that the method is able to distinguish patterns that simple counting of branch events at one scale cannot accomplish. By looking at these examples we will illustrate that many different scales are important and prove the theoretical concept.

We will now demonstrate the idea of counting branches upon the branch events. Figure 3.1a shows an object (shaded) and four points for which the number of branches will be counted. To count the number of branches at a point, a circle of fixed radius is drawn centred at the point. This circle will intersect the object at a number of places. The number of branching events within the circle is nominally equal to number of intersections minus 2. Thus, for the left-most point in Figure 3.1a, there are two intersections of the circle and the object and there are no branching events within the circle. For the lower middle point in Figure 3.1a, there are four intersections between the circle and the object and so there are two branching events. The right-most point in Figure 3.1a illustrates that this rule is not quite correct. The

intersections minus 2 may be negative. For instance, in Figure 3.1b we can see there is a circle that is drawn at the top of vertical bar and centred by the point P , the circle intersects within the object at only one place and so is necessary to take the maximum of zero and the number of intersections minus 2.

3.2 Generating branching profiles

Now, in Figure 3.2 the number of branching events depends on the radius that is drawn. We can see there is an image, and a square is drawn around it. Inside the square, we will draw circles at different radii. Finally, a bar graph shows the branching profiles for each circle.

We consider different radii $r = 4, 8, 12, 16, 20, 24, 28$ inside the square. The circle with radius $r = 4$ has no branching events. The circles with $r = 8$ and $r = 12$ have the same number of branching events within the circle, that is two branching events. The circles with $r = 16, r = 20, r = 24$, and $r = 28$ all have three branching events.

We have demonstrated the concept of branching profiles. Thus, we can say that the branching profiles help to characterise different shapes. In the following chapter we will apply the method of branching profiles to characterise regular patterns as well as different images of yeast colonies from different strains.

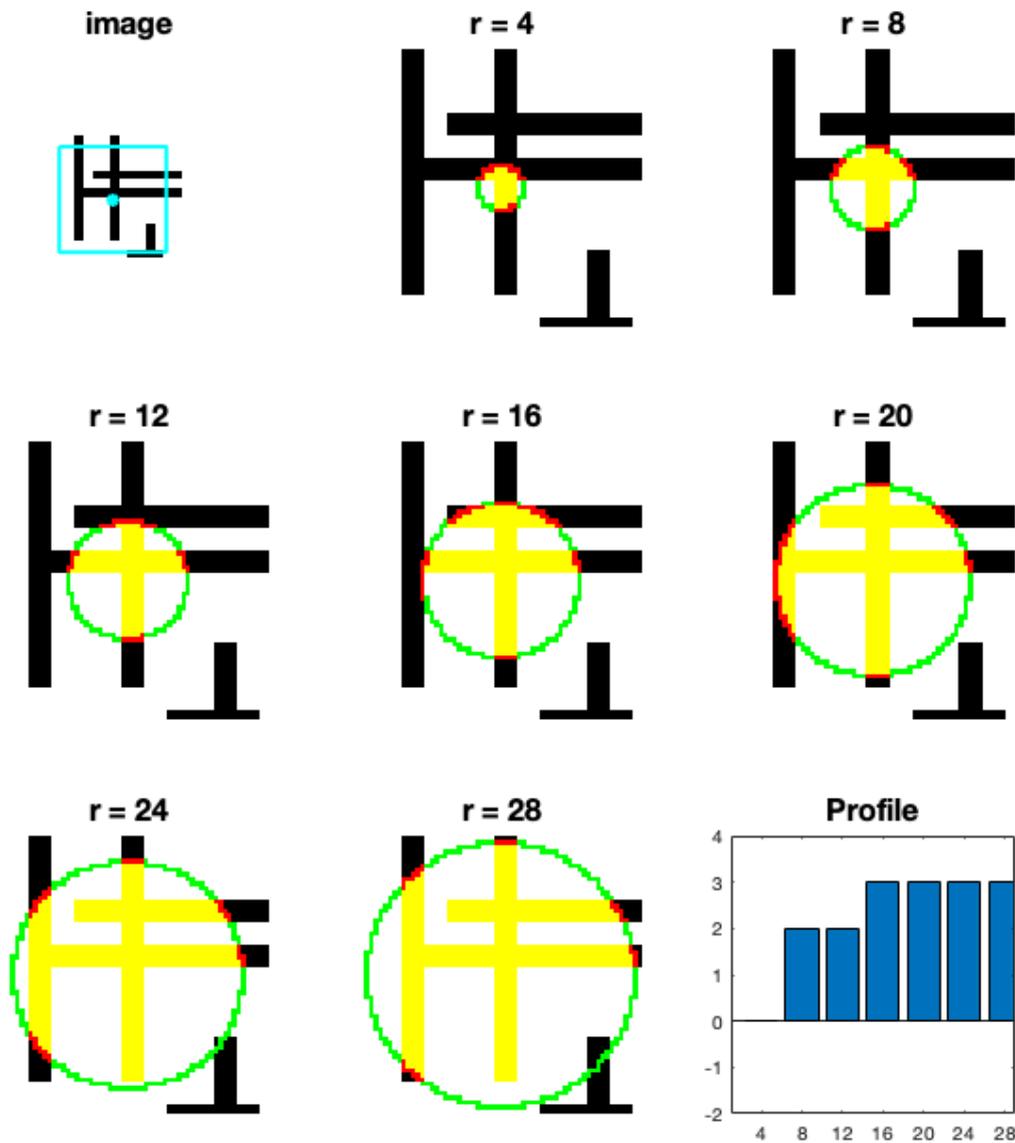


Figure 3.2: The top left panel shows a binary image and a selected point. The next seven panels show the region inside the rectangular box and circles centred at the point for increasing radii. The circle is green, the connected component of the object inside the circle and containing the point is yellow and the intersections of this component and the object are red. the bar graph in the lower right panel shows the number of branch points as a function of the radius. This number is called the branch profile for the selected point.

Chapter 4

Applications

4.1 Regular patterns

4.1.1 Rectangles

The main purpose of this method is to try to extract shape information on irregular features. Nonetheless, here we begin by exploring regular patterns. We apply the method of branching profiles to characterise a regular pattern. The object in Figure 4.1 consists of a grid of evenly spaced vertical and horizontal lines. To count the branch events, circles are drawn at different radii, centred at cross point. It is found that by starting from the smallest circle there are two branch events. Every time the circle expands, it reaches two more horizontal lines and two more vertical lines; for each of these, it gains an additional two intersections. For instance, the second smallest circle has twelve intersections and thus ten branch events. Thus, Table 4.1 describes for each circle the radius and the number of branching events. The following formula gives the number of branching events seen in this pattern

$$N = 2 + 8(\lceil r \rceil - 1),$$

where $\lceil r \rceil$ denotes ceiling or the smallest integer less than or equal to.

And so we can see the Table 4.1 as the radius increases then the number of branching events also increases.

The more interesting property of this rectangular grid is to take the ratio of the number of intersections to the radius and taking the limit. We obtain that the limit of that ratio at infinity is equal to 8:

$$\lim_{r \rightarrow \infty} \frac{N}{r} = \lim_{r \rightarrow \infty} \frac{2 + 8(\lceil r \rceil - 1)}{r} = 8.$$

In addition, using the same rectangular grid we can consider the branching profiles for the circle starting at a different centre as in Figure 4.2. Table 4.2 describes for each circle the radius and the number of branching events. The following formula will give the number of branching events:

$$N = 8\lfloor r \rfloor,$$

where $\lfloor r \rfloor$ denotes floor or the smallest integer less than or equal to.

Now, if we take the ratio for this case as the radius goes to infinity then we obtain the limit is equal to 8 which is same result that was for the same rectangular patterns at a cross centred point.

$$\lim_{r \rightarrow \infty} \frac{N}{r} = \lim_{r \rightarrow \infty} \frac{8\lfloor r \rfloor}{r} = 8.$$

Table 4.1: Number of intersections and branching events by radius of circle centred at cross point in a rectangular grid.

Radius	Intersections	Branching Events
0.4	4	2
1.2	12	10
1.8	12	10
2.2	20	18
2.8	20	18
3.2	28	26

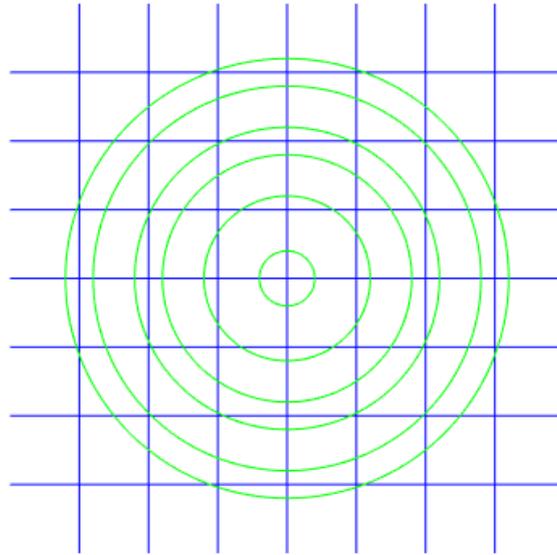


Figure 4.1: This figure shows a regular grid pattern centred at a cross points. The grid is blue. Six circles are drawn in green with some different radii. The circles intersect within the pattern in some places. Each time the circle expands, it reaches a two more lines in each direction.

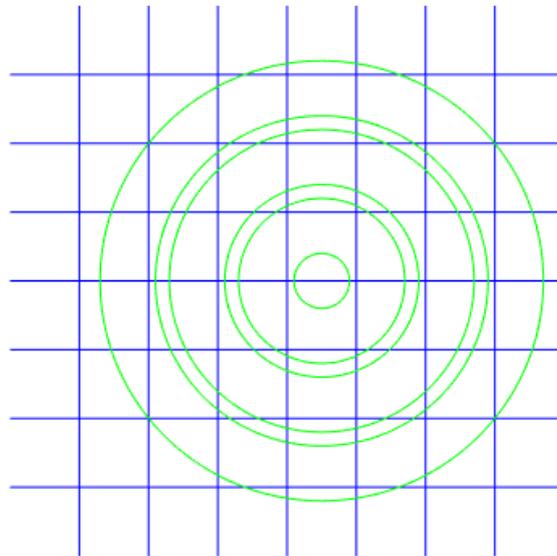


Figure 4.2: This figure shows a regular grid pattern centred at a point midway between to cross points. The grid is blue. Six circles are drawn in green with some different radii. The circles intersect within the pattern in some places. Each time the circle expands, it reaches two more lines in each direction.

By comparing between the two tables for the rectangular pattern, we can see the limit at infinity for two differently centred points is equal to 8.

Table 4.2: Number of intersections and branching events by radius of circle centred midway between two cross points in a rectangular grid.

Radius	Intersections	Branching Events
0.4	2	0
1.2	10	8
1.4	10	8
2.2	18	16
2.4	18	16
3.2	26	24

Thus, Table 4.3 is created to demonstrate that the limit at infinity for this rectangular pattern will always be the same. Because in Figure 4.1 the point was centred at the cross, where in Figure 4.2 the point was centred midway between two cross points. If we assume a new point, that is centred between the previous chosen points. Since every point on the grid is within 0.5 unit of across point, by symmetry, we can say that the limit in this rectangular patterns as the radius goes to infinity at any centre points will always equal 8.

Table 4.3: Ratios of the numbers of intersections to the radius for four different radii and two possible centre locations, and their limits as the radius goes to ∞ .

Radius	Ratio of intersections to radius	
	Cross-centred	Between crosses
0.4	10	5
1.2	10	8.3333
2.2	8.1818	7.2727
3.2	8.75	8.125
∞	8	8

4.1.2 Hexagons

Figure 4.3 shows a regular pattern hexagonal. We apply the method of branching profiles to characterise this pattern. The hexagons in Figure 4.3 have side lengths of one unit. Six circles are drawn in black colour at different

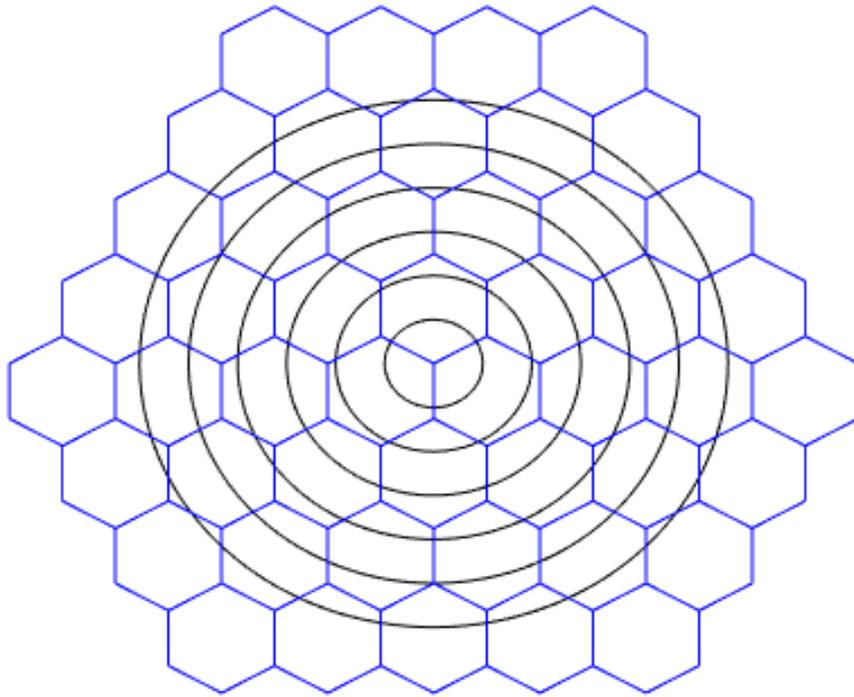


Figure 4.3: This figure shows a pattern of regular hexagons. Six circles are drawn in black with some different radii. The circles intersect within the object in some places. The number of intersections and the number of branching events are demonstrated in the Table 4.4.

radii. Table 4.4 describes for each circle the value of radii and the number of branching events.

Compare to the rectangular grid, we can see there are differences between the values of the branching events for each of these two regular patterns. Also, the formula to count the number of branching events was found for the rectangle patterns, whereas for the general formula for the hexagons was not found, but is clearly not the same as for the rectangular pattern. Hence, we can say that the method of branching profiles has distinguished between these two regular patterns.

Table 4.4: Number of intersections and branching events by radius of circle in a hexagons grid.

Radius	Intersections	Branching Events
0.8	3	1
1.6	6	4
2.4	9	7
3.2	15	13
4	15	13
4.8	24	22

4.2 Yeast colonies

Yeast is used in baking, brewing, wine making and many other industries. Baker’s yeast or *Saccharomyces cerevisiae* is well known as a dimorphic yeast which grows usually as single budding cells [5]. According to Gimeno et al., dimorphic yeasts are able to grow either by budding of single cells or as multicellular filaments called pseudohypha [4]. Baker’s yeast often develops by sprouting in the yeast stage; however, it tends to be incited to change to the pseudohyphae stage under certain pressure conditions [9]. There are thousands of strains of yeast and in industrial settings, identifying strains automatically and reliably is important. One way to distinguish yeast colony is by the patterns of growth at the edges.

Datasets

A set of 20 images of yeast colonies used in a previous study by Binder et al. [1] and Gontar et al. [5] was used to determine if branching profiles could be used to distinguish between yeast strains. The data consists of twenty images, ten images of colonies of strain AWRI 796 and ten images of colonies of strain AWRI R2. They were divided randomly into two sets of five each for training and five each for testing.

4.2.1 Method

Branching profiles were applied to identify the yeast colonies from different strains. In order to characterise the pattern of yeast colony filamentous that grows at the edge we consider the branching profiles of a binary image as in Figure 4.4. Black colour refers to the pixels of the yeast and the white colour refers to the pixels of the background.

We used the method of branching profile that was described in Chapter 3. Training was done by considering circles of seven different radii in units of pixel size lengths. After trying some different values, the following radii were considered to cover a range that is reasonable for problem. Circle radii were chosen to be 8, 16, 24, 32, 40, 48, 56. Five hundred points per image were chosen at random to compute branch profiles. For every point per image (20 images) we obtain a single element in seven-dimensional space because we were computing the branching profiles at seven different radii. The branch profiles were clustered using K -means with $K = 3$. Once the cluster centres were identified then every point was assigned to the closest cluster centre. We had 500 points per image, then there was a certain frequency for each cluster. The number of points per image assigned to each cluster formed a vector of length three and was used to represent the colony image. The Fisher classifier [3] was used to assign the strain of the yeast colony to each image based on the three dimensional cluster label vector. The Fisher classifier was used because the number of data points, five per class for training and five per class for testing, was too small to justify the use of a more complex classifier.

4.2.2 Results

The method classified the yeast colonies in the testing set hundred percent correct classification. The coordinates, in seven dimensions of the three cluster centres found are visualised in Figures 4.5a, 4.5b and 4.5c. Each

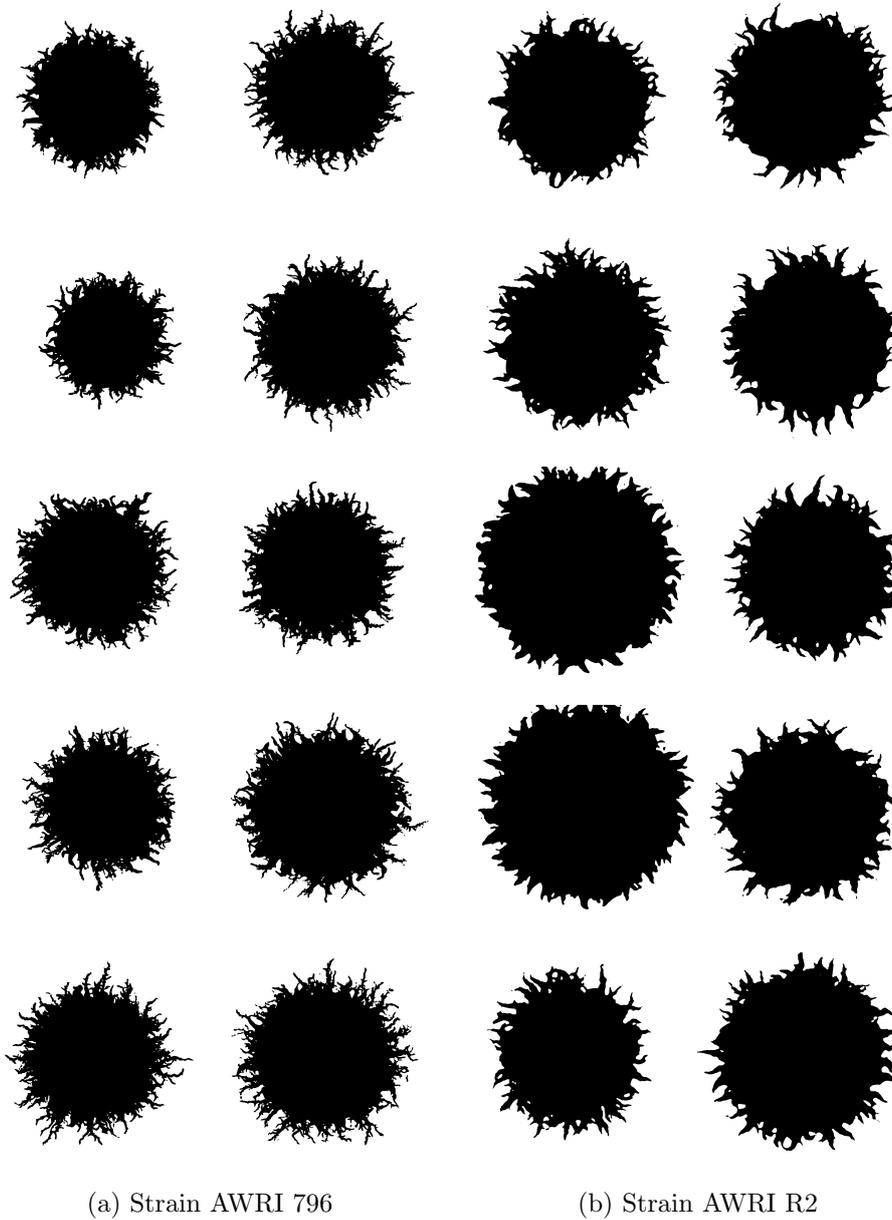


Figure 4.4: This shows four columns. Each has 5 images. From the left side; the first column is the training set colonies from strain AWRI 796, The second column is the testing set colonies from strain AWRI 796, the third column is the training set colonies from strain AWRI R2 and the last column is the testing set colonies from strain AWRI R2.

point per image was assigned to its closest cluster centre. Figure 4.5d shows the summaries of the two different strains of yeast colonies, 1 refers to the strain AWRI 796 and 2 refers to the strain AWRI R2. Each one of the bars demonstrate the number of points in each image which belong to one of these cluster centres. The blue bar refers to the points that belong to the first cluster, the red bar refers to the points that belong to the second cluster and the yellow bar refers to the points that belong to the third cluster. As we can see there is a significant difference between these two strains. The second group, that is the colonies from strain AWRI R2, has many more points in cluster three, whereas the first group, that is the colonies from strain AWRI 796, has proportionally more of cluster one and cluster two. To be more precise, Figure 4.6 shows a scatter plot of these 3-dimensional features in order to visualise the proportion of each colony of which cluster it belongs to. The red circles refer to the colonies from the strain AWRI 796 and blue crosses refer to the colonies from the strain AWRI R2. The result was that it takes five image from each group and train the Fisher classifier and then takes the other five that were not use for training. The result of testing set was ten out of ten.

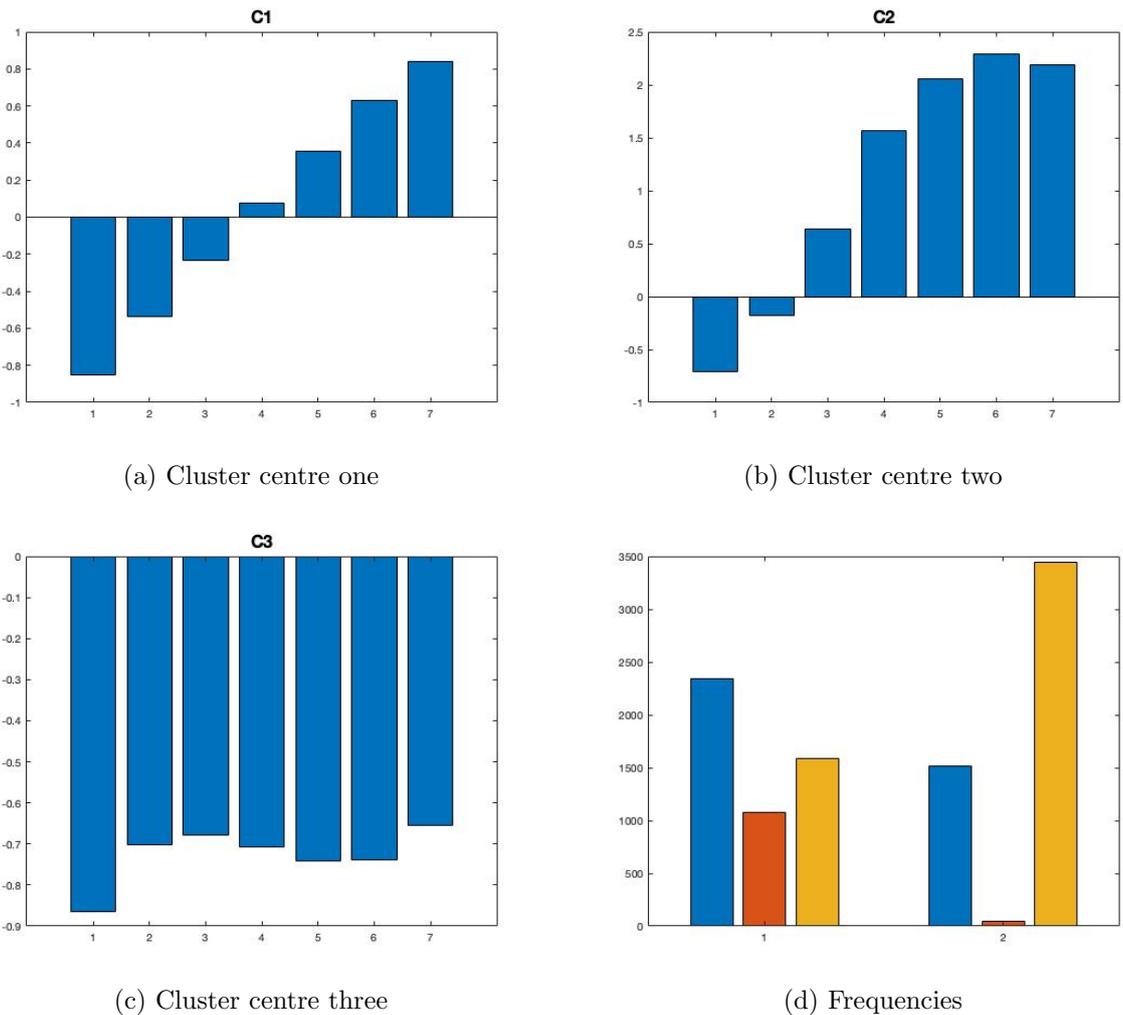


Figure 4.5: Figure shows four bar graphs. (a) The cluster one. (b) The cluster two. (c) The cluster three. (d) Frequencies of points classified as C1, C2, and C3 for (1)AWRI 796 and (2)AWRI R2. The blue bar refers to the points belong to C1, the red bar refers to the points belong to C2 and the yellow bar refers to the points belong to C3.

4.2.3 Discussion

The method of branching profiles has been used to classify two-dimensional binary images of yeast colonies. By eye it is difficult to know that is a such colony from this strain and the other from that strain as in Figure 4.4. Also, for the algorithm concerning they look very different. Hence, we can say that counting local branching is important to characterise irregular shapes such as yeast colonies.

On the other hand, we had only two strains and only ten images were

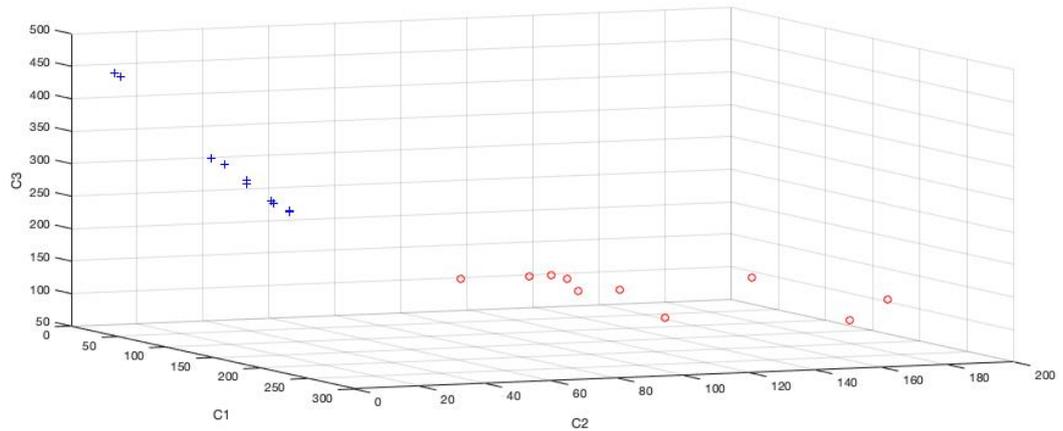


Figure 4.6: For each the 20 images, we consider the number of points in the three clusters as coordinates in 3 dimensional. These point are plotted here. The Red circles refer to the colonies from strain AWRI 796 and the blue circles refer to the colonies from strain AWRI R2.

considered in each strain. If we had 100 strains and each strain had that many images that may not have been so easy. This method may have a limitation due to some factors such as if we have that many strains. In previous work [2] authors found that branching and oriented shape feature were complementary in the sense that experiments showed that the two methods used together provided more information than either one on their own. In other words, the two methods do not find the same information. Thus, combining the methods may help in the case of many strains. A fruitful direction for future research would be to try adding more images from many different strains and seeing if the method will lead to improvements.

Also, instead of taking the $\max(N - 2, 0)$, we could allow the branching events to attain values of -1 and -2 . A score of -1 means that there is an end point as in Figure 3.1b and a score of -2 means that the connected component lies entirely within the circle and thus identifies this component as detached from the rest of the object. This does not make sense for yeast colonies, but might make sense for marbling in beef or other applications.

Chapter 5

Conclusion

After introducing the use of branching profiles to study shapes, it is found that by only counting the number of branching events with one radius is not sufficient to characterise shapes, whereas, counting the number of branching profiles lead to a useful description of the shape's structure, in particular, the irregular shapes. Irregular shapes are such as trees, clouds, land masses, yeast colonies, and cancellous bone that can not be characterised by using standard shape descriptors such as their height, width and length. Branching profiles contribute to our understanding of irregular shapes. Thus, the method of branching profiles leads to extracting and characterising shape characteristics. The method allows us to study shape characteristics at various scales, which may otherwise not be so easy to find.

Bibliography

- [1] B. J. Binder, J. F. Sundstrom, J. M. Gardner, V. Jiranek, and S. G. Oliver. Quantifying two-dimensional filamentous and invasive growth spatial patterns in yeast colonies. *PLoS computational biology*, 11(2): e1004070, 2015.
- [2] M. J. Bottema, Z. Kruk, A. Gontar, W. Pitchford, and C. Bottema. Evidence of marbling as a single connected entity. *Under review*.
- [3] J. Friedman, T. Hastie, and R. Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics New York, 2001.
- [4] C. J. Gimeno, P. O. Ljungdahl, C. A. Styles, and G. R. Fink. Unipolar cell divisions in the yeast *S. cerevisiae* lead to filamentous growth: regulation by starvation and RAS. *Cell*, 68(6):1077–1090, 1992.
- [5] A. Gontar, M. J. Bottema, B. J. Binder, and H. Tronolone. Characterizing the shape patterns of dimorphic yeast pseudohyphae. *Royal Society open science*, 5(10):180820, 2018.
- [6] T. Y. Kong and A. Rosenfeld. Digital topology: Introduction and survey. *Computer Vision, Graphics, and Image Processing*, 48(3):357–393, 1989.
- [7] B. L. Martin and M. J. Bottema. Textons for 3D binary data with applications to classifying cancellous bone. In *2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, pages 1–6. IEEE, 2015.

- [8] J. Šlapal. Closure operations for digital topology. *Theoretical Computer Science*, 305(1-3):457–471, 2003.
- [9] H. Tronnolone, J. M. Gardner, J. F. Sundstrom, V. Jiranek, S. G. Oliver, and B. J. Binder. TAMMiCol: Tool for analysis of the morphology of microbial colonies. *PLoS computational biology*, 14(12):e1006629, 2018.