Branch Profiles for Shape Analysis

Zayed M Asiri
College of Community n Muhayil
Department of General Course
King Khalid University
Abha, Saudi Arabia
Email: aalmfrari@kku.edu.sa

Brianna L Martin
School of Natural Sciences
University of Tasmania
Hobar, TAS, 7001, Australia
Email: brianna.martin@utas.edu.au

Murk J Bottema
College of Science and Engineering
Flinders University, Tonsley Campus
Clovelly Park, SA, 5042, Australia
Email: murk.bottema@flinders.edu.au

***Abstract*—The necessity for characterising highly irregularly shaped objects appears in many circumstances, most prominently in biology and medicine, but also in physical sciences and elsewhere. Here, a multi-scale method for quantifying the level of branching in irregular structures is presented to extend the repertoire descriptors of shape. The method was used to classify strains of yeast colonies and to demonstrate differences in structure of newly formed cancellous bone in rats under various experimental conditions. Yeast colonies were classified with an accuracy of 1.000 (n = 10) and classification of newly formed cancellous bone into three classes achieved mean accuracy of $0.853 \pm .088$ over 10 runs with data randomly sampled from the same 15 rats each run.**

I. Introduction

Irregular and contorted shapes and structures abound in nature. Commonly encountered examples include vines, clouds, roots, caves, as well as a plethora of internal bodily structures such as the vascular and pulmonary systems. In studying such entities, the need to quantify irregular shapes arises naturally and, indeed, a variety of methods for quantifying shape have appeared in the literature. Many methods are designed for studying the shape of objects with relatively simple gross geometry typical of man-made entities such as buildings, vehicles, and household goods, etc. These objects are of primary interest in many problems in defence, surveillance, robotics, remote sensing and so forth. Here the focus is on methods for analysing shapes that are not well quantised in terms of standard geometrical measures. To do so, a method called branch profiles is proposed for determining the density of branching events at different scales. The motivation for this approach is that branching patterns are not efficiently captured by existing methods but constitute a physical feature intuitively associated with descriptions of irregular structure. Two examples, psuedohyphae of yeast colones and trabecular structure in cancellous bone (Fig. 1) will recur throughout this work and are briefly described below.

Strains of common brewer's yeast, *Sacchromyces cerevisiae*, underpin both the baking and brewing industry and so contribute substantially to the production of basic foods. Yeast are single celled fungi and typically reproduce by budding. When grown on solid substrates, yeast form round colonies. Under certain conditions such as low nutrient levels, cells elongate and form irregular, branching structures called pseudohyphae [1] as seen in Figure 1 a and b. In industrial settings, automatically identifying specific strains of *S. cereviciae* based on images of colonies expedites processing and quality control.

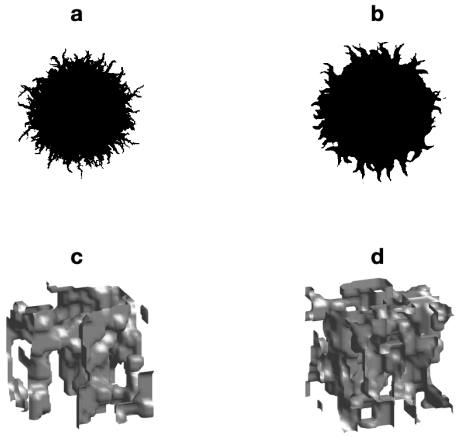

Fig. 1. Yeast colonies and cancellous bone examples. **a**; An example image of a yeast colony from strain AWRI 796. **b**; Strain AWRI R2. **c**; A portion of cancellous bone from the sham group (Section III-B) **d**; A portion of cancellous bone from the ovx+zol group. The bone blocks in **c** and **d** are cubes of side length approximately 261 μm.

Cancellous bone refers to the spongy bone present in the ends of long bones such as the tibia. The individual structures comprising cancellous bone are called trabeculae (Fig. 1 c and d). Due to age or disease, the volume of cancellous bone may reduce. Treatment with drugs such as bisphophonates helps to arrest the loss of bone. If bones are still growing, drug treatment may also cause an increase in volume per unit time of the newly emerging cancellous bone. During bone loss, the number and thickness of individual trabeculae are known to reduce. However, very little is known about changes, if any, regarding other aspects of structure as a result of bone loss or treatment. Here, branch profiles are used to investigate whether or not patterns of branching of trabeculae differ between three classes of bone in rats; normal bone, bone after loss due to oestrogen deprivation, and bone after treatment with the bisphosphonate zoledronic acid.

A. General shape analysis

A wide variety of methods have been used to study shape in biological settings, although many of these methods have applications in other contexts. Elliptic Fourier analysis was used to study otolith shape in whiting [2], identify plant species based on leaf shape [3] and study variation in shell shape [4]. Skeletonisation has been used to detect pseudopods on blood cells [5] and early versions of landmark methods were have to characterise human [6] and mole skulls [7]. Cootes et al. extended landmarks to point distribution models [8] which, in turn, were extended to statistical shape models including active shape models and active appearance models [9]. These are versatile models for classifying classes of shapes when the overall shape is common to the examples in the class but local variations are of interest. Applications have included recognising hands in various poses [10], segmenting ventricles in echocardiograms [11] and estimating risk of fracture from duel energy x-ray images of femurs [12]. Extensions to 3D applications include segmenting MR magnetic resonance data of heads [13]. While these methods have enjoyed substantial success, they require the identification of landmark points which is labour intensive if done manually and sometimes unreliable if done automatically. Shape context uses histograms of local shape feature but also requires finding analogous points of similarity between objects studied [14].

There are two key attributes of the classes of irregularly shaped objects that motivated this study. First, two objects within a class are not necessarily similar enough to identify consistent landmarks or to match control points consistently. One could argue that the outer tips of pseudohyphae are landmarks or that the points where the pseudohyphae meet the round central part of the colony could play this role, but these will not be consistent in number and are not always clearly defined (Fig. 1). Certainly in the case of cancellous bone, the individual trabeculae cannot be identified and are not even well defined. There is no real sense as to where one trabecula starts or stops. The lack of natural landmarks eliminates many of the methods reviewed above for studying pseudohyphae or trabecular shapes. Second, the overall shape of the object is not of interest. Of interest is the local shape or structure. The yeast colonies may be described as essentially round with a fuzzy border, but the shapes of interest are the patterns of the fuzz comprising the border. The samples of cancellous bone available from micro-computed tomography (μ-CT) scans form rectangular blocks but this is certainly not the shape of interest. The shapes of interest are the local structure patterns that must be quantised and then combined to describe the block of bone.

B. Shape descriptors for filamentous growth

For these reasons, shape descriptions have been invented specific to individual problems. In the case of yeast colonies, the area covered by the colony has been compared to square of the perimeter. The ratio of the perimeter squared to the area is 4π for a perfect disc but increases as the boundary gets fuzzier. This simple metric was useful for improving understanding of relative fitness of strains [15]. A natural extension is to consider the fractal dimension of the boundaries although this has been done for bacteria instead of yeast [16]. The fractal dimension and ratio of perimeter squared to area are single metrics over the entire colony and do not address shapes of individual pseudohyphae. Spatial distributions were taken into account by Binder et al. [17] [18]. A clear focus on shapes of individual psuedohypha was presented in [19] and [20] where local orientation of pseudohyphae were determined and the length of line segments internal to psuedohyphae were used to understand the sizes of the filaments.

C. Shape descriptors for cancellous bone

Historically, the structure of cancellous bone has been described in terms of general topological features such as bone surface area (BS), bone volume per unit tissue volume (BV/TV), bone surface area per unit tissue volume (BS/TV), Euler's characteristic (EUL) and connective density (CONND) as well as descriptors specific to cancellous bone such as trabecular bone pattern factor (TBPF), structural model index (SMI), trabecular thickness (TB.Th), trabecular spacing (TB.Sp) and trabecular number (TB.N) [21] [22] [23] [24]. Initially these measures were extracted using 2D histomorphometry, meaning that animals had to sacrificed and cross sections of bones were examined manually [25]. The measures TBPF, SMI, TB.Th, TB.Sp and TB.N are based on the rod and plate model for the structure of cancellous bone. As such they are limited by the extent to which this model reflects the structure. The rod and plate model suits healthy human bone well, but is less apparent in juvenile rats and mice. With the advent of μ-CT, full three dimensional structure became available (without the need to sacrifice animals). New methods appeared, for example, based on finding the lengths of line segments within the structure at various orientations [26] [20].

D. Focus of the paper

Here the density of branching at various scales, called a branch profile, is proposed as a possible shape feature to be used on its own or as part of a suite of shape features for a particular task. Branch profiles are explained in Section II. Details on the application of branch profiles for classifying psuedohyphae in yeast colonies is presented in Section III-A and work on analysing trabecular structure in cancellous bone appears in Section III-B. All computations, including the generation of figures, were performed in Matlab.

II. Branch Profiles

Let Ω denote the object of interest for which branching patterns are to be described, where Ω may be viewed as a set in $\Re^n$. Most often in practice, $n = 2$ as in the case of pictures of yeast colonies or $n = 3$ as in the case of trabecular structure in cancellous bone. Let p be a point in Ω and set $B_p^r = \{x \in \Re^n : \|x - p\| \leq r\}$, the ball of radius r centred on p. Loosely speaking, the number of branches at p at scale r is given by the number of places where the structure Ω 'sticks out' of the ball B_p^r. However, this is not quite right

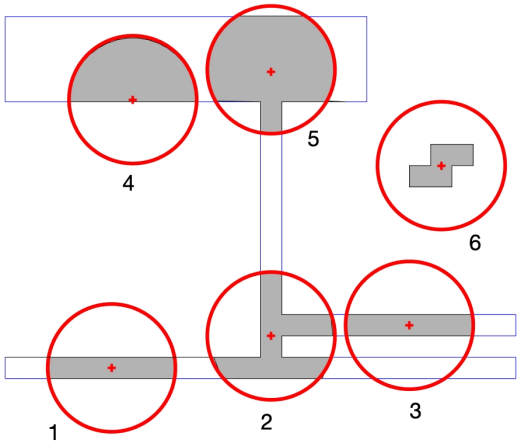

Fig. 2. A toy example in 2D. The thin line outlines the structure in question. There are two components, one large component and one small component lying entirely within the circle labelled 6. Six examples of points p are indicated and for each, the boundaries of the balls B_p^r for a fixed radius r are shown. The branching parameter T_p^r has values 2, 4, 2, 1, 3, 0 for points 1, 2, 3, 4, 5, 6 respectively as given by the number of connected components of the intersections of the structure and the red circles. The number of branching events within each circle is $b = T_p^r - 2$ when this number is non-negative. In the case of point 3, there are four intersections of the structure and the circle but there are no branching events within the circle. This illustrates the necessity of using C_p^r, the connected component of the structure restricted to B_p^r containing p, for counting the number of intersection with the circle. The dependence on r is illustrated by point 5, for example. If a slightly smaller radius is used, T_p^r changes from 3 to 2. A further reduction in r leads to $T_p^r = 1$.

since if, for example, the restriction of Ω to B_p^r consists of two non-intersecting beams, then the structure will stick out of the ball in four places but there are no branching events within the ball (Fig. 2, circle 3). In order to count the number of branching events correctly, attention should be restricted to a single connected component of $\Omega \cap B_p^r$. Since $p \in \Omega$, p is in one of the connected components of $\Omega \cap B_p^r$ and this is the natural component to choose. Thus a careful definition of the branching parameter, T_p^r at p for fixed r is as follows. Let C_p^r denote the connected component of $\Omega \cap B_p^r$ that contains p. Then T_p^r is the number of connected components of $C_p^r \cap \partial B_p^r$. Here ∂B_p^r denotes the boundary of B_p^r which is the same as the sphere $\{x \in \Re^n : \|x - p\| = r\}$.

Depending on the shape of Ω, the location of the point p and the radius r, $C_p^r \cap \partial B_p^r$ may be the empty set, in which case $T_p^r = 0$. This means that the p lies in an isolated component of Ω that lies entirely within the ball B_p^r (Circle 6 in Fig. 2). Thus the branch profile T_p^r is able to identify isolated components of Ω. If B_p^r lies entirely within Ω or if B_p^r lies partially outside Ω as one component, then $T_p^r = 1$ (Circle 4 in Fig. 2). If, near p, Ω comprises a single beam that enters and exits B_p^r at disjoint locations, then $T_p^r = 2$ (Circle 1 in Fig. 2). Only when $T_p^r > 2$ does it live up to its name as a reasonable branching parameter since the number of branching events of Ω inside B_p^r (restricted to the connected component C_p^r) is then $b = T_p^r - 2$. Figure 2 illustrates these notions.

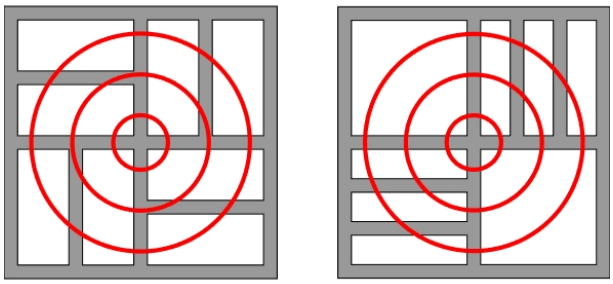

Fig. 3. The two structures each comprise 12 T-intersections and one cross intersection and so simply counting branching events does not distinguish these figures. The profiles evaluated at the radii of the indicated circles distinguish the structures. Taking p to be the central point the branch profile for the structure on the left is $T_p = (4, 4, 8)$ while the branch profile at the analogous point for the structure on the right is $T_p = (4, 6, 8)$.

Note that the terminology 'number of branches' has been deliberately avoided since there seems to be no consensus in common speech whether a single branching point yields two branches or three branches.

The discussion above has already hinted at the obvious observation that the branching parameter at a point p depends heavily on the radius r. Thus a fuller picture of the branching pattern is obtained by viewing T_p^r as a function of r for a fixed point p. This function is what is meant by the branch profile of Ω at p. With apologies for the banality of the notation, the branch profile at p is denoted by T_p and is defined by $T_p(r) = T_p^r$. In practice, everything is usually discrete and the branch profile becomes the vector $T_p = (T_p^{r_1}, T_p^{r_2}, \ldots, T_p^{r_M})$, where M denotes the number of radii at which the branch profile is computed.

A pair of toy examples (Fig. 3) demonstrates the benefit of branch profiles over just counting the number of branches. In these examples, the number of T-intersections is the same and the number of cross intersections is the same. Hence the number and types of branch events do not distinguish the two structures. The branch profiles computed only at the centre points of these two examples do distinguish the structures. In particular, the branches at the scale of the middle circle are different (Fig. 3, legend).

For a given structure Ω, the collection of branch profiles $\{T_p : p \in \Omega\}$, or a sampled subset, encode the branching information of Ω. Exactly how to make use of this information is not part of the method presented here, per se, but several standard strategies are available. For example, statistical summaries of the T_p are used in Section III-B to obtain feature vectors for each rat. For classifying yeast colonies in Section III-A, an approach is taken similar to 'bag of words' [27]. In the same way that textons may be viewed as 'bag of textures', the vectors T_p may be used to implement 'bag of branch profiles'. Thus for a given class, clustering may applied to the collection of branch profiles from that class to obtain K clusters for some pre-assigned number K. The centres of these clusters A_k, $k = 1, 2, \ldots, K$ may be called branch primitives.

Each point $p \in \Omega$ is now mapped to the branch primitive closest to its profile T_p according to the Euclidean norm in the space of branch profiles $\Re^M$, where M is the number of radii used to construct the vectors T_p. Thus every point $p \in \Omega$ is represented by a cluster index in $\{1, 2, \ldots, K\}$. The set Ω is represented by the normalised histogram (distribution) of cluster indices associated with the points $p \in \Omega$. This normalised histogram becomes the feature vector of length K that represents Ω for classification or regression.

Since branch profiles are computed for every in $p \in \Omega$ or for a subsample of these points, the same branching event will contribute to T_p^r for many points p and fixed r. Thus the quantity $\sum_{p\in\Omega} T_p^r$ for fixed r is not equal to the number of branching events in Ω and does not seem to correspond to a natural physical quantity. However, if T_p^r is computed at S points $p_1, p_2, \ldots, p_S \in \Omega$, then $\frac{1}{S}\sum_{i=1}^{S} T_{p_i}^r$ may be interpreted as a measure of branching density.

III. Implementations of branch profiles

A. Classifying yeast colonies

The data for classifying yeast colonies comprised 10 images each from two strains of *S. cereviciae*; AWRI 796 and AWRI R2. This set of images is a subset of the data described in [18] and were kindly provided for this study by the authors of that work. The images are binarised top views of yeast colonies at resolution 1.52 - 1.53 μm per pixel side length.

The branch profile was computed at 500 random points in each image. Each profile was computed using seven radii; 8, 16, 24, 32, 40, 48, 56 voxels (12.2, 24.4, 36.6, 48.8 61.0, 73.2, 85.4 μm). Five colonies from each strain were randomly assigned to the training set and five were assigned to the testing set. The branch profiles from the images in the training set were pooled. K-means clustering with $K = 3$ was used to find three clusters in the 7D space of pooled branch profiles from the training set.

Each point, p, in each image, in both the training set and testing set was mapped to the index of the cluster closest to T_p according to the Euclidean norm in the 7D space of branch profiles. Next, each image in the training and testing set was represented by the normalised distribution of cluster indices associated with the points in the image. Since there were three clusters, the feature vector representing each image was of length three. The feature vectors from the training set were used to train a simple Fisher classifier and the resulting classifier was used to assign the feature vectors in the testing set as stemming from yeast strain AWRI 796 or AWRI R2. All images in the testing set were assigned correctly.

Because this was a small study in terms of the number of images from each class, there is no claim that the perfect classification results are expected to extend to images of these two strains of yeast generally. This was not the objective of the study. The objective of the study was to ascertain if branch profiles capture sufficient information regarding the shape to be of use as a shape feature on its own or to complement existing shape features.

B. Branch patterns in cancellous bone

Data for studying changes in structure of cancellous bone were described in [28]. Briefly, 30 female rats were randomly assigned to three groups of 10 animals each. At time 0, rats in groups 2 and 3 underwent ovariectomy to induce oestrogen deficiency resulting in bone loss over time. Rats in group 1 underwent sham surgery, meaning that they were opened up the same as the rats in groups 2 and 3, but ovaries were not removed. Group 1 is called the sham group. Treatment by zoledronic acid commenced for rats in group 3 after 14 days. This group, having undergone ovariectomy followed by treatment is called the ovx+zol group and group 2 is called the ovx group since these rats underwent ovariectomy but were not treated.

Micro-CT scans of the left tibias of all rats were taken at days 0, 14, 28, 56 and 84. CT reconstructions yielded cubic voxels of side length 8.702 μm. Otsu's method was used to produce binary 3D arrays of size $121 \times 121 \times 400$ voxels (approximately $(1 \times 1 \times 3.5$ mm). Three rats died between the μ-CT scans at day 56 and 84 and so, for the present study, only data from scans at day 56 was used. The rats were young and still growing so new bone was emerging at the growth plates of the long bones. Since interest here was focused on bone generated under the influence of the three experimental conditions, only the the first 100 slices closest to the growth plate were processed as these would have emerged since the start of treatment.

Each treatment group was represented by 10 blocks of cancellous bone of size $121\times121\times100$ voxels ($\approx 1.0\times1.0\times0.87$ mm. For each block, 5000 points were selected randomly with replacement. For each of these points, the branch profile $T_p = (T_p^{r_1}, T_p^{r_2}, T_p^{r_3}, T_p^{r_4})$ was computed for radii $r_1 = 16$, $r_2 = 24$, $r_3 = 32$, $r_4 = 40$ voxels (139, 209, 278, 348 μm). These radii were determined by a few preliminary computations of T_p^r for several combinations of p and r and visual inspection of the variation in T_p. The 10 blocks in each experimental group were pooled so that each experimental group, sham, ovx and ovx+zol, was represented by 50,000 branch profiles. One-way ANOVA for the three groups was applied to each of the four components of the vectors T_p separately. For each component, the result indicated that null hypothesis of the three group means being equal should be rejected ($\mathrm{p} << 0.001$).

Next, an experiment was conducted to classify the rats into the three experimental groups based on branch profiles only. To do so, five rats from each group were randomly assigned to the training set and five rats from each group were assigned to a testing set. Thus the training and testing sets comprised 15 rats each. This time the data were not pooled. Instead, the medians and standard deviations (SD) of the branch profiles were computed separately for each rat. Thus rat j was represented by the feature vector of length eight given by

$$F_j = (\mathrm{med}_1^j, \mathrm{SD}_1^j, \mathrm{med}_2^j, \mathrm{SD}_2^j, \mathrm{med}_3^j, \mathrm{SD}_3^j, \mathrm{med}_4^j, \mathrm{SD}_4^j),$$

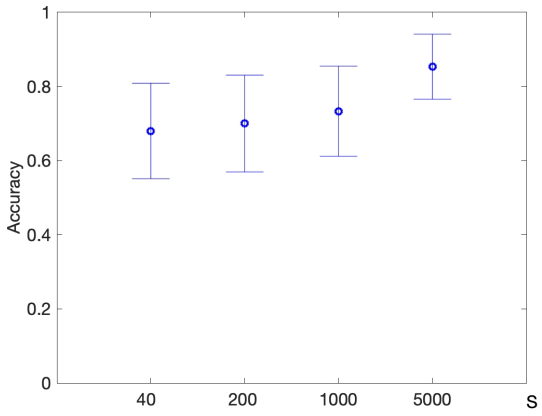

Fig. 4. Mean accuracy and SD of the three-way classification of cancellous bone as sham, ovx or ovx+zol as a function of the number sample points per data block. The means and standard deviations are $S = 40$: 0.68 ± 0.13, $S = 200$: 0.70 ± 0.13, $S = 1000$: 0.73 ± 0.12, $S = 5000$: 0.85 ± 0.09.

where med_i^j and SD_i^j are the median and standard deviation of the set

$$\{T_p^{r_i} : p \in \Omega_j\}$$

and Ω_j is the data block for rat j. In fact, branch profiles were not computed at all $p \in \Omega_j$. As in the ANOVA experiment above, each Ω_j was sampled S times with replacement. The size of the sample, S, needed to achieve a sensible balance between computational load and accuracy was not known in advance and so trials were run with sampling sizes $S = 40, 200, 1000, 5000$. In order to compare results, the entire process of computing branch profiles at S random points for each of the 30 rats and training and testing the classifier was performed 10 times for each value of S. Classification into three classes was based on linear discriminant analysis using the 'classify' command in Matlab.

Each run produced a confusion matrix listing for each group the number of assignments correctly to the same group and the number of assignments made incorrectly to each of the other groups. The number of correct assignments divided by 15 was reported as the accuracy. For a each value of S, the 10 repeat runs were averaged to form a mean accuracy (Fig. 4) and the 10 confusion matrices were averaged to obtain mean confusion matrices (Table I).

IV. Discussion and conclusion

The sample sizes in both experiments were small and so one could speculate that the classifier found fortuitous hyperplanes to separate the classes that could not be expected to generalise. Classification of the yeast colonies relied on feature vectors of length three derived from branch profiles and so may be visualised as points in 3D (Fig. 5). The vectors from the two strains form well separated clusters indicating that these feature vectors capture actual differences between the two groups. Although the gap between the two clusters in Fig. 5 may well reduce if more yeast colonies are included, the gap does show that separating hyperplanes exist that are well bounded away from both clusters.

TABLE I

Mean confusion matrix $\pm$ SD for classifying cancellous bone into three classes using branch profiles for the four sampling rates S. Rows are true class memberships, columns are class memberships assigned by the classifier. Thus the entry in row i and column j is the proportion of rats in testing set in group i assigned by the classifier to to group j.

$S = 40$	sham	ovx	zol
sham	0.62 ± 0.26	0.16 ± 0.18	0.22 ± 0.18
ovx	0.34 ± 0.21	0.48 ± 0.22	0.18 ± 0.18
zol	0.06 ± 0.14	0	0.94 ± 0.14
$S = 200$	sham	ovx	zol
sham	0.60 ± 0.27	0.14 ± 0.19	0.26 ± 0.28
ovx	0.22 ± 0.18	0.58 ± 0.29	0.20 ± 0.23
zol	0.08 ± 0.14	0	0.92 ± 0.14
$S = 1000$	sham	ovx	zol
sham	0.66 ± 0.27	0.04 ± 0.08	0.30 ± 0.24
ovx	0.10 ± 0.11	0.62 ± 0.18	0.28 ± 0.17
zol	0.08 ± 0.14	0	0.92 ± 0.17
$S = 5000$	sham	ovx	zol
sham	0.82 ± 0.22	0	0.18 ± 0.22
ovx	0.04 ± 0.08	0.74 ± 0.10	0.22 ± 0.06
zol	0	0	1.00 ± 0.00

The classification of cancellous bone to experimental groups sham, ovx and ovx+zol using branching alone is comparable to results in [26] where a search was made over combinations of traditional shape features for cancellous bone based on histomorphometry and oriented thickness measured on 3D μ-CT data. There the best combination of features achieved an accuracy of 0.866 for the three classes of bone considered here. Comparisons must be made carefully since the experiments in [26] classified cancellous bone into nine groups instead of three. Even so, the results obtained here show branch profiles performing not quite, but close to the level of the best combination of existing features.

On its own, branching density based on branch profiles at several scales cannot expected to provide a full description of local shape patterns. Intuitively branching density does not say anything about the number of holes or connectivity, for example. Branching density was only anticipated to provide another measure of shape to complement existing measures. In this light, the results from the experiments reported here on classifying yeast colonies and cancellous bone using branch profiles alone were better than expected. Unsurprisingly, the rate of classification improved as the number sample points increased (Fig. 4). However, the mean accuracy achieved with only 40 samples per rat was not anticipated. To the eye, the patterns of trabeculae seem quite complex. To put this in perspective, there was no previous knowledge on whether branching patterns were different between cancellous bone from normal rats, oestrogen deprived rat and treated rats. First, this result says that there are differences between the three groups of cancellous bone and second, the result shows that measuring branching density at four scales (four values of r)

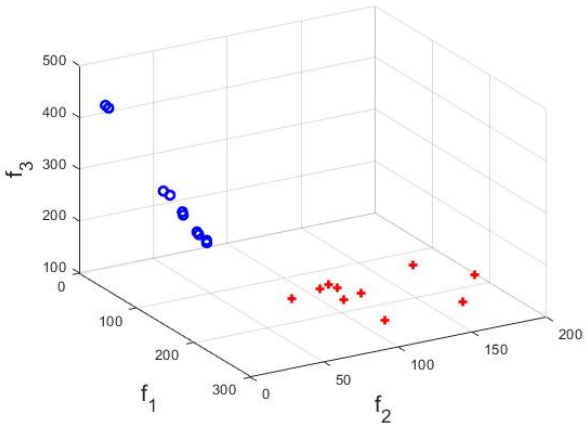

Fig. 5. A scatter plot of the feature vectors for the yeast colonies. The symbols o (in blue) denote feature vectors for strain AWRI R2 and + (in red) denote strain AWRI 796.

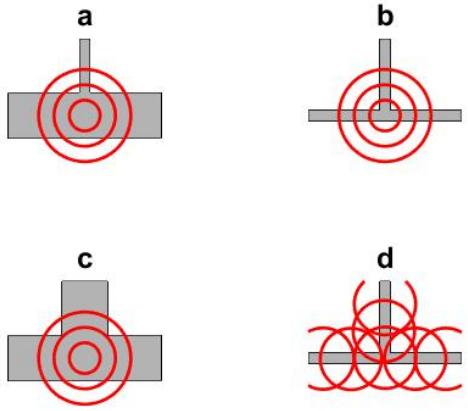

Fig. 6. In figures a, b and c, the branch profile T_p^r is computed at the single point of intersection of the beams forming the structure for three values of r. In each case, the structure is a T intersection, but the profiles at these points are $T_p = (1, 3, 3)$ in figure a, $T_p = (3, 3, 3)$ in figure b and $T_p = (1, 1, 3)$ in figure c. The profiles differ because of the combinations of thicknesses of the beams. In part d, for a single fixed value of r, T_p^r takes on the values 1,2 or 3 depending on the juxtaposition of the p and the point of intersection of the beams. If the beams were thicker, the T_p^r would still take on the same three values but the distribution of these values would be different. A greater proportion of points within a fixed distance of the intersection would have $T_p^r = 1$, for example.

at only 40 randomly selected points suffices to distinguish the three groups to some reasonable extent.

As it turns out, branch profiles capture more than just branching. Already, the isolated component of Ω within the circle labelled 6 in Figure 2 demonstrates that disconnected components may be identified if they are small compared to the radius r. Probably, this aspect of branch profiles does not contribute significantly to the performance in the examples considered here. A property of branch profiles that is likely to be of greater importance in these examples is their ability to quantify thickness to some extent (Fig. 6, a, b and c). The figure shows that branch profiles computed at analogous points of the same structure differ due to relative thicknesses of the components of the structure. For a fixed radius, the proportion of values of T_p^r within a given distance of a point of intersection also differ according to the thicknesses of the branches (Fig. 6 d).

The accuracy improved only slightly with the five fold increase in the number of points sampled per rat from $S = 40$ to $S = 200$ and again only a slight improvement with another five fold increase to $S = 1000$ (Fig. 4). Only with another five fold increase to $S = 5000$ samples per rat did the accuracy improve noticeably, albeit only a 25 percent increase in accuracy at the expense of a 625 fold increase in computing time compared to $S = 40$. This may not seem worthwhile, but accuracy rates, especially for multi-class classification can be misleading. The mean confusion matrix for $S = 5000$ (Table I) demonstrates important characteristics of the method at this level of sampling. For the zol group, the classifier has perfect sensitivity but poor sensitivity while for the ovx group, the classifier has perfect sensitivity but poor specificity. The sham group has good sensitivity (0.82) and very good specificity (0.95). These attributes of the method will likely be important for interpreting results from models for changes of bone over time in the three experimental groups of rats. From this perspective, $S = 5000$ may be worthwhile.

For the experiments on quantifying structure of cancellous bone, computing a single branch profile T_p required approximately 0.11 seconds (MacBook Pro, 2.4 GHz Quad-Core IntelCor i5) running Malab code. The slow steps of the calculation are the labelling of connected sets of voxels twice for each calculation of T_p^r; once to identify C_p^r, the connected set containing p, and once to count the connected components of $C_p^r \cap \partial B_p^r$ (Section II). Since there were four radii in the experiment, labelling of connected sets was required eight times to find the vector T_p. The time needed for the labelling is proportional to r^3 since the sets in question are in 3D. The preliminary runs used to determine suitable values of r for this experiment indicated the need to use radii up to $r = 40$ voxels. This resulted in the long run times for computing T_p. Run times for computing branch profiles for the yeast colonies were much shorter even though the radii were larger and the branch profiles were vectors or length seven compared to length four for cancellous bone. The reason is that, for yeast, the data is 2D and so labelling time is proportional to r^2.

The method requires a choice of the number of radii, M, and a range $[a, b]$ such that $a = r_1 < r_2 < \cdots < r_M = b$. There is no general theory for choosing a, b and M. Suitable values depend on the problem at hand and must be determined based on domain knowledge and possibly some experimentation balanced by practical considerations such as the resolution of the data and computational load. There is no *a priori* reason for the radii to be equally spaced. In cases where large intervals are desirable but the number of radii must be kept low due to computational load, the spacing could increase exponentially.

In the two examples presented here, the first choices for these parameters, in both cases equally spaced, provided satisfactory results and brief experiments with different choices (results not shown) indicated that the method did not depend critically upon these parameters.

Similarly, the number of sample points must be determined by practical considerations and perhaps some experimentation. Figure 4 shows that, as expected, performance tends to increase with the number of points sampled but this increase is small over several orders of magnitude for the cancellous bone data. Whether this trend extends to other problems is not known. The issue did not arise with the yeast colony because perfect classification was achieved on the initial run.

Statistics on classification performance similar to that reported in Table I for cancellous bone, was not performed for the yeast colony data. This was because the scatter plot of all the data for yeast colonies in terms of the shape features (Fig. 5) indicated that the classes were extremely well separated. Clearly, the implementation of any reasonable classifier will report perfect classification whether used as stand alone or as part of a cross validation scheme.

As noted in Section II, the focus of this paper is on developing branch profiles as features of potential use for characterising or classifying irregular shapes. How these features are used is not part of the method. For cancellous bone, simple statistics of these features, mean and standard deviation, provided useful classification. In the case of yeast colonies, the features were used as input into a "bag of words" scheme to determine ubiquitous patterns and classification was based on the distributions of these patterns. In doing so, these two examples illustrate diverse approaches for gleaning information from branch profiles. As always, the best advice is to start simple and consider higher complexity as needed.

REFERENCES

[1] P. J. Cullen and G. F. Sprague, "The regulation of filamentous growth in yeast," *Genetics*, vol. 190, no. 1, pp. 23–49, 2012.

[2] J. Mapp, E. Hunter, J. Van Der Kooij, S. Songer, and M. Fisher, "Otolith shape and size: the importance of age when determining indices for fish-stock separation," *Fisheries Research*, vol. 190, pp. 43–52, 2017.

[3] J. C. Neto, G. E. Meyer, D. D. Jones, and A. K. Samal, "Plant species identification using elliptic fourier leaf shape analysis," *Computers and Electronics in Agriculture*, vol. 50, no. 2, pp. 121 – 134, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168169905001560

[4] D. Innes and B. JA, "Morphological variation of Mytilus edulis and Mytilust trossulus in eastern Newfoundland," *Marine Biology*, vol. 133, pp. 691–699, 1999.

[5] A. R. Dill, M. D. Levine, and P. B. Noble, "Multiple resolution skeletons," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, no. 4, pp. 495–504, 1987.

[6] F. Bookstein, K. Schäfer, H. Prossinger, H. Seidler, M. Fieder, C. Stringer, G. W. Weber, J.-L. Arsuaga, D. E. Slice, F. J. Rohlf, W. Recheis, A. J. Mariam, and L. F. Marcus, "Comparing frontal cranial profiles in archaic and modern homo by morphometric analysis," *The Anatomical Record*, vol. 257, no. 6, pp. 217–224, 1999.

[7] F. J. Rohlf, A. Loy, and M. Corti, "Morphometric analysis of old world talpidae (mammalia, insectivora) using partial-warp scores," *Systematic Biology*, vol. 45, no. 3, pp. 344–362, 09 1996.

[8] T. Cootes, C. Taylor, D. Cooper, and J. Graham, "Active shape models - their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[9] T. Heimann and H.-P. Meinzer, "Statistical shape models for 3d medical image segmentation: A review," *Medical Image Analysis*, vol. 13, no. 4, pp. 543 – 563, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1361841509000425

[10] D. Hogg and R. Boyle, Eds., *Active shape models - 'smart snakes'*. Springer, 1992.

[11] T. F. Cootes, A. Hill, C. J. Taylor, and J. Haslam, "The use of active shape models for locating structures in medical images," in *Information Processing in Medical Imaging*, H. H. Barrett and A. F. Gmitro, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1993, pp. 33–47.

[12] R.-S. Lu, E. Dennison, H. Denison, C. Cooper, M. Taylor, and M. J. Bottema, "Texture analysis based on gabor filters improves the estimate of bone fracture risk from dxa images," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 6, no. 4, pp. 453–464, 2018.

[13] A. Hill, A. Thornham, and C. J. Taylor, "Model-based interpretation of 3d medical images." in *BMVC*. Citeseer, 1993, pp. 1–10.

[14] G. Mori and J. Malik, "Estimating human body configurations using shape context matching," in *European conference on computer vision*. Springer, 2002, pp. 666–680.

[15] L. Chen, J. Noorbakhsh, R. M. Adams, J. Samaniego-Evans, G. Agollah, D. Nevozhay, J. Kuzdzal-Fick, P. Mehta, and G. Balázsi, "Two-dimensionality of yeast colony expansion accompanied by pattern formation," *PLoS Comput Biol*, vol. 10, no. 12, p. e1003979, 2014.

[16] M. Matsushita and H. Fujikawa, "Diffusion-limited growth in bacterial colony formation," *Physica A: Statistical Mechanics and its Applications*, vol. 168, no. 1, pp. 498–506, 1990.

[17] 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 Comput Biol*, vol. 11, no. 2, p. e1004070, 2015.

[18] H. Tronnolone, J. M. Gardner, J. F. Sundstrom, V. Jiranek, S. G. Oliver, and B. J. Binder, "Quantifying the dominant growth mechanisms of dimorphic yeast using a lattice-based model," *Journal of The Royal Society Interface*, vol. 14, no. 134, p. 20170314, 2017.

[19] A. Gontar, M. J. Bottema, B. J. Binder, and H. Tronnolone, "Characterizing the shape patterns of dimorphic yeast pseudohyphae," *Royal Society open science*, vol. 5, no. 10, p. 180820, 2018.

[20] A. Gontar, H. Tronnolone, B. J. Binder, and M. J. Bottema, "Characterising shape patterns using features derived from best-fitting ellipsoids," *Pattern Recognition*, vol. 83, pp. 365–374, 2018.

[21] M. Hahn, M. Vogel, M. Pompesius-Kempa, and G. Delling, "Trabecular bone pattern factor—a new parameter for simple quantification of bone microarchitecture," *Bone*, vol. 13, no. 4, pp. 327–330, 1992.

[22] T. Hildebrand and P. Rüegsegger, "A new method for the model-independent assessment of thickness in three-dimensional images," *Journal of microscopy*, vol. 185, no. 1, pp. 67–75, 1997.

[23] T. Hildebrand, A. Laib, R. Müller, J. Dequeker, and P. Rüegsegger, "Direct three-dimensional morphometric analysis of human cancellous bone: microstructural data from spine, femur, iliac crest, and calcaneus," *Journal of bone and mineral research*, vol. 14, no. 7, pp. 1167–1174, 1999.

[24] C. T. Dodson, P. E. Parker, and P. Parker, *A user's guide to algebraic topology*. Springer Science & Business Media, 1997, vol. 387.

[25] A. M. Parfitt, M. K. Drezner, F. H. Glorieux, J. A. Kanis, H. Malluche, P. J. Meunier, S. M. Ott, and R. R. Recker, "Bone histomorphometry: Standardisation of nomenclature, symbols, and units," *Journal of Bone and Mineral Research*, vol. 2, no. 6, pp. 595–610, 1987.

[26] 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)*. IEEE, 2015, pp. 1–6.

[27] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in *European conference on machine learning*. Springer, 1998, pp. 137–142.

[28] E. Perilli, V. Le, B. Ma, P. Salmon, K. Reynolds, and N. Fazzalari, "Detecting early bone changes using in vivo micro-ct in ovariectomized, zoledronic acid-treated, and sham-operated rats," *Osteoporosis international*, vol. 21, no. 8, pp. 1371–1382, 2010.