

Please cite the Published Version

Arslan, Mazlum Ferhat, Haridis, Alexandros, Rosin, Paul L, Tari, Sibel, Brassey, Charlotte , Gardiner, James D, Genctav, Asli and Genctav, Murat (2022) SHREC'21: Quantifying shape complexity. *Computers & Graphics*, 102. pp. 144-153. ISSN 0097-8493

DOI: <https://doi.org/10.1016/j.cag.2021.09.005>

Publisher: Elsevier BV

Version: Accepted Version

Downloaded from: <https://e-space.mmu.ac.uk/628468/>

Usage rights:  [Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Additional Information: This is an Author Accepted Manuscript of a paper accepted for publication in *Computers and Graphics*, published by and copyright Elsevier.

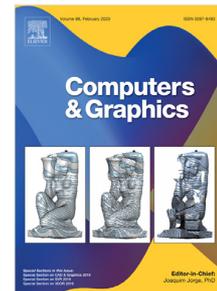
Enquiries:

If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)

Journal Pre-proof

SHREC'21: Quantifying shape complexity

Mazlum Ferhat Arslan, Alexandros Haridis, Paul L. Rosin, Sibel Tari,
Charlotte Brassey, James D. Gardiner, Asli Genctav, Murat Genctav



PII: S0097-8493(21)00191-6

DOI: <https://doi.org/10.1016/j.cag.2021.09.005>

Reference: CAG 3421

To appear in: *Computers & Graphics*

Received date: 2 April 2021

Revised date: 22 August 2021

Accepted date: 16 September 2021

Please cite this article as: M.F. Arslan, A. Haridis, P.L. Rosin et al., SHREC'21: Quantifying shape complexity. *Computers & Graphics* (2021), doi: <https://doi.org/10.1016/j.cag.2021.09.005>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 Published by Elsevier Ltd.



SHREC'21: Quantifying shape complexity

Mazlum Ferhat Arslan^{a,1,*}, Alexandros Haridis^{b,1}, Paul L. Rosin^{c,1}, Sibel Tari^{a,1}, Charlotte Brassey^d, James D. Gardiner^e, Asli Genctav^f, Murat Genctav^g

^aDepartment of Computer Engineering, Middle East Technical University, Turkey

^bDepartment of Architecture, Massachusetts Institute of Technology, USA

^cSchool of Computer Science & Informatics, Cardiff University, UK

^dSchool of Science and The Environment, Manchester Metropolitan University, UK

^eDepartment of Musculoskeletal & Ageing Science, University of Liverpool, UK

^fDepartment of Computer Engineering, TED University, Turkey

^gMicroelectronics, Guidance and Electro-Optics Division, ASELSAN A.S., Turkey

ARTICLE INFO

Article history:

Received June 30, 2021

Keywords: 3D Shape Complexity

ABSTRACT

This paper presents the results of SHREC'21 track: Quantifying Shape Complexity. Our goal is to investigate how good the submitted shape complexity measures are (*i.e.* with respect to ground truth) and investigate the relationships between these complexity measures (*i.e.* with respect to correlations). The dataset consists of three collections: 1800 perturbed cube and sphere models classified into 4 categories, 50 shapes inspired from the fields of architecture and design classified into 2 categories, and the data from the Princeton Segmentation Benchmark, which consists of 19 natural object categories. We evaluate the performances of the methods by computing Kendall rank correlation coefficients both between the orders produced by each complexity measure and the ground truth and between the pair of orders produced by each pair of complexity measures. Our work, being a quantitative and reproducible analysis with justified ground truths, presents an improved means and methodology for the evaluation of shape complexity.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

Shape complexity is studied across several fields such as psychology [1], design [2, 3], computer vision [4]. In the context of 3D shapes, it has the potential to be useful in shape retrieval [5, 6], measuring neurological development and disorders [7, 8], in determining the processes and costs involved for manufacturing products [2, 9], etc. Early work on shape

complexity appears in the literature of experimental psychology as well as in literature related to design and aesthetics. The classical aesthetic notions of “unity” and “variety” [10], or comparably, “order” and “complexity” [11] are directly connected to the complexity of spatial objects. One of the first measures of complexity for polygonal shapes can be found in [11]. Attneave [1] conducted human experiments to seek correlations of shape complexity with scale, curvedness, symmetry and number of turns. On the basis of the variety in the responses from human subjects, Attneave states that shape complexity is ill-

*Corresponding author: Tel.: +90-539-478-4937

e-mail: mferhata@gmail.com (Mazlum Ferhat Arslan)

¹Track organizers

defined. With the premise of circles being the simplest shapes, a natural candidate for the quantification of shape complexity is P^2/A . In several works ([1, 3]) it is used as a measure of the complexity along with other indicators. In most other works [4, 12, 13, 14, 15, 16], tools from information theory, on top of various geometric features are used to quantify complexity. Work that relates complexity to algorithmic information theory and is applied to objects of art and design can also be found in Stiny and Gips [17]. Rossignac [18] provides a classification of shape complexity that focuses on measuring different aspects of computer representations for 3D shapes. The variety of approaches taken in the quantification of shape complexity further supports the claim that complexity can obtain a variety of meanings based on the approach that one chooses to take in a particular research area and for the particular task at hand.

There is a lack of benchmark datasets for shape complexity. Even the methodologies in the literature need improvements. For example, in many cases just visual results are reported without quantitative analysis [19, 13]. The methods are neither compared to other methods nor evaluated in terms of statistical consistency. In this track paper, we aim to account for and investigate different aspects of complexity that can help other researchers to develop and test their methods. In particular, we investigate how good the submitted shape complexity measures are (*i.e.* with respect to ground truth) and investigate the relationships between these complexity measures (*i.e.* with respect to pairwise correlations). Due to the ill-defined nature of complexity, a linear order may not make sense. Hence, we propose to explore complexity using multiple tasks and multiple shape collections.

The first collection is composed of subgroups obtained by introducing additive or subtractive noise to two basic shapes: sphere and cube. The purpose is to investigate the relation of complexity to noise level. The second collection is composed of artificial 3D shapes constructed by transforming and combining multiple elements, and evaluated by experts to provide ground truth. The purpose is to investigate the complexity methods in relation to perceptual categories. The final collection is an

already existing 3D shape dataset which was originally developed as a segmentation benchmark. We repurpose this data and use the segmentation ground truth as a means to investigate 3D shape complexity via a proxy (secondary) task. The main contributions of this work are as follows:

- Generation of two novel shape collections with associated ground truth, and repurposing of a previous segmentation benchmark for assessing complexity measures.
- Systematic evaluation of the performance of a selection of both 2D and 3D classical and recent shape complexity measures.
- Assessment of similarities and differences between different measures by using pairwise correlations and clustering based on their performance with respect to multiple ground truths.

Note that due to the ill-defined nature of shape complexity our dataset contains three shape collections with different characteristics. Each collection in the dataset contains a different type of 3D object with ground truth defined and obtained in a different manner.

The paper is organized as follows: In Section 2 the dataset is introduced. In Section 3 the ground truths and the evaluation strategy are explained. In Section 4 the short descriptions of the participating methods are included. In Section 5 the the evaluation results for each collection in the dataset are presented. Finally, Section 6 is Discussion and Section 7 is Conclusion.

2. Dataset

The used dataset consists of three collections each aiming to account for a different aspect of shape complexity. The first two collections are created synthetically, and the third is an existing collection consisting of natural shapes. The ground truth for the first collection is based on the parameters used in creating the collection. For the second collection, the ground truth is provided by two design experts on the final design object. The purpose of the third collection is to investigate how esti-

1 mated complexity is related to the number of parts perceived by
 2 humans, which we hypothesise is related to shape complexity.

3 2.1. Collection 1 – Perturbed basic shapes

4 In this collection we aim to explore the correlation between
 5 shape complexity and magnitude of perturbations of a cube and
 6 a sphere.

7 A cube of side length 199 voxels and a sphere of radius
 8 100 voxels are stochastically perturbed additively and subtractively,
 9 separately. This forms four families (additively perturbed
 10 cubes, subtractively perturbed cubes, and so on). The algorithm
 11 used in perturbing a shape introduces a perturbation at a random
 12 location on the shape’s boundary in each application. The
 13 algorithm has two parameters: *i*) width (w) determining the area
 14 of effect of the perturbations and *ii*) number of times of appli-
 15 cation (c) determining how many times a local perturbation is
 16 introduced. Both parameters are set to three different values,
 17 $w \in \{3, 4, 5\}$ and $c \in \{25, 50, 75\}$. This results in a group of
 18 nine shapes. A sample group for an additively perturbed cube
 19 is displayed in Fig. 1. Fifty such groups form a family.

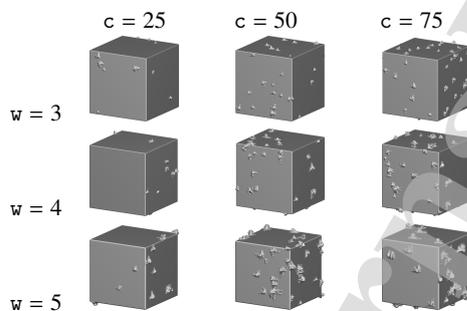


Fig. 1. A sample group of additively perturbed cubes.

20 The shapes in this collection were distributed to the participants
 21 as both volumetric data and triangular meshes.

22 2.2. Collection 2 – Parametric shape families

23 The second collection is made up of two distinct families of
 24 shapes and each family contains twenty five shapes. The shapes
 25 in this collection are inspired from 3D models and designs that
 26 are commonly found in the fields of architecture and urban de-
 27 sign. The shapes were created with two primary objectives in

28 mind. The first objective was to have shapes that vary para-
 29 metrically in terms of a few spatial features. The set of spatial
 30 features is different for each family (see below), but in both fam-
 31 ilies, these features guide the generation of the shapes in a sys-
 32 tematic way via algorithms. The second objective was to have
 33 shapes that on the one hand are spatially “rich”, in the sense
 34 that they can be deployed in a variety of realistic design scenar-
 35 ios and problems, and on the other hand, are abstract enough
 36 to not suggest fixed typological interpretations or the shapes of
 37 everyday objects. While such shapes make the task of measur-
 38 ing complexity significantly more challenging, they present an
 39 opportunity for a broader exploration of what constitutes com-
 40 plexity of spatial objects.

41 The shapes in both families are generated with the built-
 42 in scripting language of the Rhinoceros 3D software package
 43 (Robert McNeel & Associates, USA). All shapes are repre-
 44 sented as watertight triangle meshes and were distributed to the
 45 workshop’s participants in this format.

46 In the first family, the shapes are generated by stacking
 47 cuboids. The main spatial features that control the generation
 48 are the number of cuboids and the length of the side faces of
 49 each cuboid. An additional rotational factor is used for eleven
 50 out of twenty five shapes in the family. In more detail, for each
 51 shape in the family the opposite faces of each cuboid are equal
 52 and all cuboids have constant height (heights are adjusted au-
 53 tomatically based on the number of cuboids, for a fixed total
 54 height). The sizes of the cuboids are controlled by varying the
 55 length of the side faces according to the function of a prede-
 56 fined curve (either a sine or a Bezier curve). For the shapes that
 57 are controlled by a rotational factor, variation is also achieved
 58 by varying the angle of rotation of the cuboids around a central
 59 vertical axis. The resulting shapes can be understood as design
 60 objects ranging anywhere between pedestals and columns, to
 61 high-rise buildings and towers.

62 In the second family, the shapes are generated by aggregating
 63 three, four or five cuboids within a predefined rectangular area
 64 in the plane. All the resulting shapes form connected config-
 65 urations (that is to say, there are no gaps between neighboring

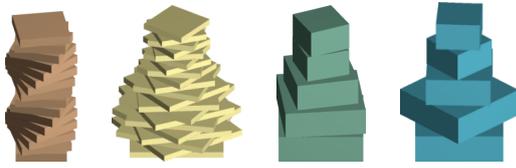


Fig. 2. Sample shapes from the first family of Collection 2.

cuboids) and the cuboids are merged in a single solid. The main spatial features that control the generation are the number of cuboids, the locations of the cuboids within the rectangular area and their individual heights. The lengths of the side faces of all cuboids are equal. Variation in the way cuboids are aggregated in the plane is achieved mainly by varying the locations and the heights of the cuboids. This is done by randomly sampling values from the allowable ranges specified for the locations and the heights. The resulting shapes can be understood as designs ranging anywhere between furniture and stairs, to volumes of buildings that form city blocks.

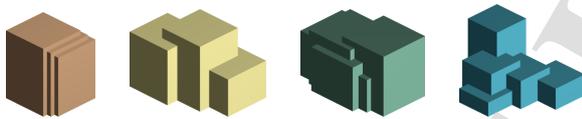


Fig. 3. Sample shapes from the second family of Collection 2.

2.3. Collection 3 – *Manually segmented shapes*

Collection 3 is the dataset for Princeton mesh segmentation benchmark [20]. We use this set with the primary objective of exploring how shape complexity measures correlate with the uniformity of the number of segments of the segmentations of the shape. The benchmark consists of 380 shapes across 19 categories and their human-generated segmentations. As opposed to the synthetic shapes in the first two collections, the shapes in the benchmark are natural. As such, they have a particular semantic content, which may affect the perception of complexity. The availability of manual segmentations for this collection makes it an ideal candidate to be used in exploring complexity by using segmentation as a proxy task.

The shapes in this collection were distributed to the participants as triangular meshes.

2.4. 2D Collections

Most of the shape quantifying methods in the literature work exclusively in 2D. To include such methods into this study we have created the 2D analogues of the shapes. We create twelve 2D silhouettes of each shape in the above collections from the views determined by the azimuthal angles ($\{0^\circ, 30^\circ, 60^\circ, \dots, 330^\circ\}$) and the elevation angle (30°). The resulting silhouettes of a shape are similar in size, thus, the collections do not pose a challenge in terms of scale-invariance. The contributing 2D methods report the averaged score over the twelve silhouettes as the measure of complexity for the corresponding 3D shape.

The families consisting of subtractively perturbed spheres and cubes are excluded because the resulting silhouettes highly depend on whether the perturbations appear on the 2D boundary of a given view or not, rather than the controlling parameters.

3. Ground truths and evaluation

3.1. Collection 1

For the first collection, the two parameters w and c used in creating the shapes constitute the ground truth. We expect the complexity scores to increase as either of the parameters increase. The performance of the methods are measured in a controlled experiment manner: we keep one of the parameters fixed and let the other vary. The performance of a method is then measured by averaging the Kendall rank correlation coefficient over the groups. This results in six measures of performance (one for each value of the parameters) for a family.

3.2. Collection 2

In Collection 2 the ground truth is provided manually. Two experts on the topics of Computer-Aided Design and 3D Shape Modeling determined the ground truth complexities of the shapes in the two families. The shapes in each family were presented to an expert in a random presentation order.

The evaluation of the experts was based on a qualitative comparison of the shapes in each collection that aimed to determine how simple or difficult it would be to model or execute a shape in three-dimensional space in a finite number of steps.

1 The shapes in a family were divided into groups of shapes of
2 comparable executional difficulty. Shapes in different groups
3 were considered incommensurate from this standpoint.

4 For the first family, the evaluation produced the following
5 five groups in which shapes are listed in increasing order of
6 complexity:

7 Group 1: (16, 14, 12, 17, 18, 19, 20)

8 Group 2: (15, 13, 24, 22)

9 Group 3: (23, 21, 25)

10 Group 4: (11, 10, 7, 1, 2)

11 Group 5: (9, 8, 4, 5, 6, 3)

12 For the second family, the evaluation produced the following
13 six groups in which shapes are considered to be of equal com-
14 plexity:

15 Group 1: (18, 24, 17, 20)

16 Group 2: (25, 21, 22, 19)

17 Group 3: (5, 3, 2)

18 Group 4: (23, 16, 15)

19 Group 5: (6, 4, 14, 12, 13)

20 Group 6: (7, 9, 10, 1, 11, 8)

21 As the shapes from different groups are incommensurate, we
22 provide ground truths only for the shapes in the same group.
23 While many qualitative notions of complexity could equally ap-
24 ply to shapes of the kind we use in this collection, we consider
25 the aforementioned notion of executional complexity one of the
26 best determinants of complexity for 3D models representing de-
27 sign objects (i.e. objects that can be used for design purposes).
28 It is also an approach to the characterization of complexity that
29 has not been investigated in the literature.

30 Since we have a total order on the groups of the first family,
31 we measure the performance of the methods using the Kendall
32 rank correlation coefficient between the complexity order indi-
33 cated by the assigned complexity scores and the ground truth.
34 For the second collection, we measure the uniformity of in-
35 group scores. The scores are first normalized to the range $[0, 1]$.
36 The pairwise absolute differences of the normalized scores are
37 summed to yield the performance measure of a group. Note that

38 the lower score indicates a better performance, in contrast to the
39 rest of the performance measures.

3.3. Collection 3

40 The shapes in the third collection are segmented by both hu-
41 mans and computer algorithms in [20]. We consider the data
42 collected from humans to be an indicator of a shape's complex-
43 ity. The fact that these human annotations differ is consistent
44 with the ill-posed nature of specifying both segmentation and
45 complexity. For each shape, there are 11 human-generated seg-
46 mentations and 7.90 segments, on average. **We use two ground**
47 **truths: one is the order induced on the shapes by the mean num-**
48 **ber of segmentations (μ) and the other is the order acquired by**
49 **the standard deviation (σ) of the number of segments.** For each
50 ground truth, we calculate *i*) Kendall rank correlation coeffi-
51 cient over all the shapes in the collection which we refer to
52 as $\tau_{\mu_{\text{all}}}$ and $\tau_{\sigma_{\text{all}}}$ in Table 4 *ii*) the averaged coefficients $\frac{1}{N} \sum_i \tau_i$
53 where τ_i is the correlation coefficient for the *i*th category, re-
54 ferred to as $\tau_{\mu_{\text{cat}}}$ and $\tau_{\sigma_{\text{cat}}}$.

4. Methods

55 We present the examined methods in this section. A total of 19
56 methods are presented in 6 groups:

- 57 1. A multi-scale measure of complexity for arbitrary dimen-
58 sional discrete shapes [16] by M. F. Arslan, § 4.1,
59
- 60 2. Alpha-shape complexity [21] by J. Gardiner and C.
61 Brassey, § 4.2
62
- 63 3. Discrepancy [15] by A. Genctav, § 4.3,
64
- 65 4. PARCELLIN distance [14] by M. Genctav, § 4.4,
66
- 67 5. 2D multi-view based shape convexity measures C_1 , C_2 ,
68 [22], [23] by P. L. Rosin § 4.5
69
- 69 6. 2D multi-view based shape complexity measures [24], [4],
70 C_{CRE} ([25]), [26] and C_{σ} , [27], [28], [29], C_{PC} ([19]) from
71 the literature, § 4.6.

4.1. A Multi-scale Measure of Complexity for Arbitrary Dimensional Discrete Shapes (M.F. Arslan)

Assuming the space (of any dimensions) in which the shape S is embedded has uniform grid, we solve the following partial

differential equation (PDE) inside S

$$\left(\Delta_\infty - \frac{1}{\rho^2}\right)f_S = -1 \text{ subject to } f_S|_{\partial S} = 0 \quad (1)$$

where Δ_∞ is the Laplace operator in L^∞ . The term $\Delta_\infty f$ is the minimizer of $\int |\nabla f|^p$ as $p \rightarrow \infty$. The parameter ρ is chosen as the maximum of the L^∞ distance transform of S (the field is referred to as t from now on). This choice ensures the robustness of solutions under changes in scale.

We construct f_S using the iterative scheme given in [16]. However, as the 3D shapes in the collections contain high number of voxels, instead of applying the convergence conditions used there, we start with a guided initial assumption and solve for a fixed number of steps. For the shapes in Collection 1, we solve for 200 steps, whereas for the shapes in Collection 2 and 3 (sampled to fit into a rectangle of total volume $300 \times 300 \times 300$ voxels) we solve for 600 steps. The guided initial assumption is the analytical solution of (1) for an axis-aligned origin-centered rectangle whose value at the point $(x, y) \in S$ is given as:

$$f_S^{(0)}(x, y) = \rho^2 - \rho^2 \frac{e}{e^2 + 1} \times \left(\exp\left\{\frac{\max\{|x|, |y|\}}{\rho}\right\} + \exp\left\{\frac{-\max\{|x|, |y|\}}{\rho}\right\} \right).$$

In vague terms, f_S can be regarded as a well-behaving distance transform. The discrepancy between f_S and t is due to the smoothed propagation of the level sets of f_S in comparison to those of t . We use the entropy of the values of \hat{f}_S (f_S normalized to $[0, 1]$) collected from a level set $t = t_0$ to measure the discrepancy at the scale t_0 . We construct a pseudo-probability distribution acquired from $\hat{f}_S|_{t=t_0}$ by partitioning it into 1024 bins and normalizing it to have a total sum of 1. The entropy of this distribution gives the complexity of the shape at the scale $t = t_0$.

The submitted scores are the summation of the complexities at scales $t \leq 0.1$.

4.2. Alpha-shape Complexity (J. Gardiner and C. Brassey)

Model pre-processing. All data collections were pre-processed to produce the prerequisite 3D point clouds for subsequent alpha-shape complexity analysis. Each model's original volume and a reference length (to be used as a metric of the

model's scale) were also calculated during pre-processing. Models in Collection 1 were stored in a 3D voxel format, and the points clouds were produced by taking the row, column and depth of each voxel's location as the x , y and z coordinates respectively and randomly down-sampling to 100,000 points. Previous analyses [30, 21, 31] have found 100,000 points to be a good compromise between retaining sufficient detail of the original model and minimising calculation times. The volume of each model in Collection 1 was calculated as the sum of the number of voxels in the original model. The models in Collections 2 and 3 were stored as watertight surface meshes. Point clouds were produced by generating a spatially random distribution of 100,000 points inside each mesh (Fig. 4). The volume of each model in Collections 2 and 3 was calculated as the volume of the original watertight mesh.

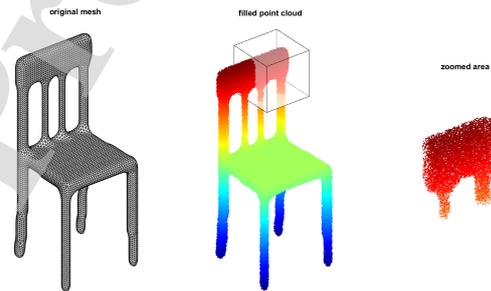


Fig. 4. Example generation of a point cloud from an original watertight mesh by filling the internal volume of the mesh with a random distribution of 100,000 points.

Across all models, a reference length was calculated to be used for model scaling within the complexity algorithm. Reference length was calculated as the mean of the distance from 10,000 random points (10% of the point cloud) to their nearest 100 neighbors.

Alpha-shape complexity algorithm. Once model point clouds, and their associated volumes and reference lengths, had been calculated, the alpha-shape complexity algorithm (originally developed for analysing biological datasets lacking homologous landmarks [30, 21, 31]) was run. Alpha-shapes [32] are a suite of shapes fitted to underlying point clouds, with the 'tightness' of their fit being determined by the value of the radius α .

1 For large values of α , the fit is coarse and tends to a convex
 2 hull as α approaches infinity (Fig. 5). For smaller values of α
 3 the fit conforms tightly to the underlying ‘shape’ of the object,
 4 until the single alpha shape fit breaks down and begins to form
 5 multiple separate objects as α approaches the smallest distance
 6 between any two points within the point cloud, where no fit will
 7 be achieved.

To calculate the shape complexity of each model, ten separate
 alpha-shapes were fitted to each point cloud across a range of
 α values (Fig. 5), from highly refined (corresponding to fine
 scale complexity) to very coarse (corresponding to gross scale
 complexity). To account for differences in the absolute size of
 models, the α used for each model was scaled by the reference
 length such that

$$\alpha_m = k \times I_{\text{ref}}$$

8 where α_m is the model-specific alpha radius, k the refinement
 9 coefficient and I_{ref} the point cloud reference length calculated
 10 in pre-processing. For the ten alpha-shape fits calculated,
 11 the same ten values of k (equally spaced on a logarithmic
 12 scale) were used, ensuring fits are equivalent across each
 13 model despite differences in absolute scale. Alpha-shapes
 14 were fitted using the ‘alphavol’ function of Jonas Lundgren
 15 ([www.mathworks.co.uk/matlabcentral/fileexchange/28851-](http://www.mathworks.co.uk/matlabcentral/fileexchange/28851-alpha-shapes)
 16 alpha-shapes).

17 Following shape fitting, ‘volume ratios’ were calculated for
 18 each of the fits as the ratio of alpha-shape volume to the origi-
 19 nal model’s volume (calculated in pre-processing). Relatively
 20 larger volume ratios therefore correspond to greater complex-
 21 ity in the model at any given scale. To further boil down the
 22 results of the alpha-shape analysis, the ten volume ratios pro-
 23 duced for each model were subject to principal component anal-
 24 ysis (PCA) as a dimension reduction technique. PCAs were run
 25 for each collection separately and the first two principal com-
 26 ponent scores were taken as the complexity metrics for each
 27 model. All data pre-processing and analysis were performed in
 28 Matlab R2020b (Mathworks Inc., Natick, USA).



Fig. 5. Example chair model illustrating the ten alpha-shape fits used in the complexity analysis, ranging from highly refined (top left) to very coarse (bottom right). Particular underlying details of the model’s point cloud are resolved at the different scales of fit, from only gross shape at the coarsest fits to details of the chair’s legs and back at finer scales.

4.3. Discrepancy (A. Genctav)

Discrepancy [15] is a field $D : \Omega \rightarrow \mathbb{R}$ defined over shape domain Ω . At each shape point p , it measures local deviation from a reference disk shape. Radius of the reference disk is defined using a global shape property A that is radius of maximally inscribed circle of Ω . The deviation is measured indirectly using an auxiliary field, which is selected as solution of the following screened Poisson equation:

$$\left(\Delta - \frac{1}{A^2}\right) v = 0 \text{ subject to } v|_{\partial\Omega} = 1 \quad (2)$$

The auxiliary field is obtained numerically for Ω and analytically for the reference disk. For each shape point p , discrepancy $D(p)$ is computed as the difference between value of the auxiliary field at p and value of the auxiliary field at the corresponding point of p in the reference disk. For each shape point p , the corresponding point in the reference disk is specified using its minimal distance to the shape boundary $\partial\Omega$. Due to uniform inhomogeneous boundary condition in (2), the auxiliary field is circularly symmetric for the reference disk, so it takes the same value at each point with the same minimal distance to the boundary. For more information including the implementation details, the reader is referred to [15].

Discrepancy is uniformly zero for a perfect disk and, hence, the entropy is 0. As shape deviated from a disk, discrepancy takes its highest positive values on central regions and lowest negative values on appendages, protrusions and boundary detail, and the entropy increases.

In this work, we used discrepancy entropy for measuring shape complexity. As discrepancy is presented in [15] for 2D shapes, our method works on the 2D views of 3D shapes.

4.4. PARCELLIN Distance (M. Genctav)

In this method, following the idea presented in [14], (3) is solved simultaneously with different source functions, which are designed for exploration of the shape volume, subject to homogeneous Dirichlet boundary conditions.

$$(\Delta - \alpha) \Phi_i = -f_i \quad (3)$$

where Δ denotes the Laplace operator, and α is a small damping parameter introduced for numerical conditioning and $i = 1, 2, \dots, n$.

By design, each source function represents an initial hypothesis for a decomposition of the shape volume into central and outer regions which correspond to positive and negative sets in the steady state distribution, respectively. Specifically, the source function selected for the i th solution is

$$f(x)_i = \text{sign}(d(x) - i \times s)$$

where $s = 1/n$ is the step size and $d(x)$ is the normalized signed distance between the location x and the boundary point nearest to x . The normalization is performed by dividing the raw distances to their maximum value.

Once a set of $n = 70$ equations are solved, the shape information contained in the solutions Φ_i are aggregated by assigning each shape location the number of solutions in which the location falls into the outer region, i.e. attains a negative value.

Finally, to obtain a measure of complexity, the entropy is computed on the 70 bin histogram of function values near shape boundary.

4.5. 2D Multi-view Based Shape Convexity (P.L. Rosin)

Several methods for measuring convexity were tested as it is hypothesised that a convexity measure can act as a shape complexity measure. It is likely that oscillations and irregularities in a shape's boundary which lead to scores indicating lower convexity will also indicate high complexity.

The method [22] measured convexity by applying a polygonal convexification process which applies a *flip* operation that reflects a polygon's concavities about their corresponding edges (termed *lids*) in the convex hull. The process is guaranteed to converge to a convex polygon in a finite number of flips. To ensure repeatability for similar shapes, the order of flipping is standardised. At each iteration the maximum deviation between each pocket and its lid is determined, and the pocket with the largest deviation is selected for flipping. Convexity is measured as the ratio of the areas of the original and convexified polygon. An alternative version is also considered, in which the pocket is flipped and also has the order of its vertices reversed (a *flipturn*). To improve computational efficiency and also reduce sensitivity to digitisation effects, polygonal approximation is first applied to the shape boundaries [33] using a small error tolerance (0.5).

The convexity measure [23] of shape S is given by

$$C(S) = \min_{\theta \in [0, 2\pi]} \frac{\mathcal{P}_2(\mathcal{R}(S, \theta))}{\mathcal{P}_1(S, \theta)}$$

where $\mathcal{P}_1(S, \theta)$ denotes the l_1 perimeter of S after rotation by angle θ , and $\mathcal{P}_2(\mathcal{R}(S, \theta))$ is the l_2 (Euclidean) perimeter of the minimum area bounding rectangle of S . Polygonal approximation is first applied to the shape boundaries, using an error tolerance of 2.

The two standard convexity measures in the literature are included: If we denote the convex hull of polygon S by $\mathbf{CH}(S)$ then the measures are defined as

$$C_1(S) = \frac{\text{area}(S)}{\text{area}(\mathbf{CH}(S))}$$

and

$$C_2 = \frac{\mathcal{P}_2(\mathbf{CH}(S))}{\mathcal{P}_2(S)}.$$

4.6. 2D Multi-view Based Shape Complexity

The method [24] attempts to capture global and local aspects of a shape in order to measure its complexity. It uses a linear combination of three quantitative terms: the number of notches (non-convex vertices) normalized to be in the range $[0, 1]$, and two terms similar to C_1 and C_2 .

The method [4] uses the entropy of boundary turning angles (i.e. the subtended angle at each point); this is used as a discrete

1 alternative to curvature. We use Sturges' rule to select the bin
2 size for estimating the probability distribution. An alternative
3 version was tested, where cumulative residual entropy [25] was
4 used instead of Shannon entropy, and is denoted as C_{CRE} .

5 Another approach that considers curvature is [26], who use
6 the sum of absolute Gaussian curvature to calculate the com-
7 plexity of curved surface shapes. We apply a version to two-
8 dimensional shapes, and also calculate the standard deviation
9 of signed curvature as another measure of complexity, denoted
10 as C_{σ} . In order to make these measures scale invariant, the
11 shapes are first scaled to a fixed area (100,000) and uniformly
12 sampled along the interpolated boundary at a fixed resolution
13 (single unit steps).

The method [27] measures k-regularity (i.e. wiggleness or
fractal dimension) of curves based on a ratio between lines
lengths (distances between points on the curve) at different
scales. The local value at point p_i of S is given as

$$r_{s,k}(S)(i) = \frac{\|p_{i+ks} - p_i\|}{\sum_{j=1}^k \|p_{i+js} - p_{i+(j-1)s}\|}$$

14 and the k-regularity of the shape is the mean value of $r_{s,k}(S)(i)$
15 over S . Our experiments used the values $s = 2$ and $k = 3$.

16 The method [28] computes the fractal dimension of a curve
17 by estimating a shape's perimeter using a series of ruler lengths.
18 We use the hybrid (Clark) method which is a combination of
19 two other methods, the fast and exact algorithms. Fractal di-
20 mension is then estimated as the slope of the regression line
21 computed for log versus log plots of ruler length versus perime-
22 ter.

23 Another fractal approach (the averaged mass dimension
24 method) is given by [29] who use a version of box counting
25 but replace the box with a circular neighborhood. To obtain a
26 more robust line fit, we find the line with least median absolute
27 error rather than least mean squared error.

28 We have implemented a method, denoted as C_{PC} , that is a
29 simplified version of [19]. Their insight was that simple shapes
30 lead to similar views whereas complex ones result in dissimilar
31 views. In our version this is measured by performing a pairwise
32 comparison of the boundaries of all the views for a given model,
33 and returning the mean score across all comparisons. Arkin *et*

34 *al.*'s [34] method for comparing polygons is used since it is
35 invariant under translation, rotation, and scaling.

5. Results

36 Since ground truths provide only the order information, we are
37 interested in the order relations rather than linear relationship
38 between actual values that could be measured by Pearson cor-
39 relation coefficient or any other parametric relation. Even for
40 pairwise comparison of measures in Section 6, order correlation
41 seems as a more meaningful measure rather than some preas-
42 sumed parametric relation which may or may not exist. Hence,
43 we use only Kendall rank correlation as a robust rank correla-
44 tion measure. We report Kendall rank correlation coefficients
45 between the participating methods and the ground truths in Ta-
46 bles 1-4. In the tables, we mark the scores of the best perform-
47 ing methods with red, the second best performers with green,
48 and the third best performers with blue.

5.1. Collection 1

50 The Kendall rank correlation coefficient (τ) for the additively
51 perturbed cubes and spheres are given in Table 1. For the
52 cubes [16], and for the spheres [15] induce the correct order on
53 all considerations. In both cases, [14] and [21]-1 follow very
54 closely. Some of the methods ([4], [27], C_{CRE} , [23], [22]-1,
55 [22]-2 and C_2) achieve strong correlations when the parameter
56 c is varied, yet the correlations weaken when the parameter w is
57 varied. This suggests that it is easier to account for the number
58 of perturbations than it is for the magnitude of perturbations.
59 Likewise, performances of some of the methods are sensitive to
60 the base shape. Notable are [15], [28], [29] and C_1 .

61 Comparing 3D methods with 2D ones, we see that 3D meth-
62 ods [16], [14] and [21]-1 consistently score highly whereas the
63 best performing 2D methods performs only partially well. For
64 example, [15] only achieves high scores for the sphere-related
65 tasks and [24], [4], C_{CRE} , C_{PC} , and so on, score high only when
66 the width parameter w is kept fixed.

67 Four submissions have been run on the subtractively per-
68 turbed cubes and spheres. The performances of these are re-
69

Table 1. The averaged Kendall τ for the additively perturbed cubes (the first value) and spheres (the second value).

Method	w = 3	w = 4	w = 5	c = 25	c = 50	c = 75
[16]	1.00 / 1.00	1.00 / 1.00	1.00 / 0.99	1.00 / 1.00	1.00 / 1.00	1.00 / 1.00
[14]	0.87 / 0.93	0.95 / 0.97	0.93 / 1.00	0.89 / 0.93	0.92 / 1.00	0.96 / 1.00
[21]-1	0.99 / 0.97	0.99 / 1.00	0.99 / 1.00	1.00 / 0.99	1.00 / 1.00	1.00 / 1.00
[21]-2	0.17 / 0.57	0.24 / 0.80	0.52 / 0.88	-0.09 / 0.35	0.03 / 0.69	0.39 / 0.69
[15]	0.25 / 1.00	0.31 / 1.00	0.68 / 1.00	0.13 / 1.00	0.39 / 1.00	0.52 / 1.00
[29]	0.67 / 0.91	0.84 / 0.99	0.80 / 0.97	0.84 / 0.68	0.95 / 0.71	0.93 / 0.68
[28]	0.45 / 0.97	0.65 / 0.99	-0.11 / 0.96	0.44 / 0.83	0.43 / 0.89	0.04 / 0.84
[27]	-0.97 / -1.00	-0.96 / -1.00	-0.97 / -1.00	-0.29 / -0.08	-0.29 / 0.05	-0.24 / 0.28
C_{CRE}	0.97 / 1.00	0.95 / 1.00	0.96 / 1.00	0.33 / 0.12	0.47 / -0.07	0.47 / -0.35
[4]	0.93 / 1.00	0.96 / 1.00	0.96 / 1.00	0.20 / 0.28	0.39 / 0.17	0.32 / -0.21
[26]	0.79 / 0.87	0.87 / 0.93	0.87 / 0.93	0.83 / 0.83	0.88 / 0.83	0.88 / 0.79
C_r	0.51 / 0.72	0.60 / 0.80	0.64 / 0.87	0.68 / 0.60	0.68 / 0.63	0.69 / 0.64
C_1	-0.73 / -0.97	-0.83 / -0.97	-0.76 / -0.95	-0.77 / -0.89	-0.81 / -0.95	-0.93 / -0.96
C_{PC}	0.93 / 0.96	0.91 / 0.93	0.88 / 0.97	0.61 / 0.53	0.71 / 0.51	0.64 / 0.36
[24]	0.93 / 1.00	0.95 / 0.99	0.91 / 0.99	0.73 / 0.69	0.83 / 0.77	0.79 / 0.64
[23]	-1.00 / -0.99	-0.96 / -0.99	-0.91 / -0.96	-0.68 / -0.59	-0.77 / -0.68	-0.76 / -0.61
C_2	-0.95 / -0.99	-0.96 / -0.99	-0.93 / -0.99	-0.67 / -0.53	-0.80 / -0.63	-0.77 / -0.51
[22]-1	-0.96 / -0.99	-0.93 / -0.99	-0.92 / -0.99	-0.59 / -0.68	-0.76 / -0.76	-0.76 / -0.64
[22]-2	-0.96 / -0.99	-0.95 / -0.99	-0.92 / -0.99	-0.65 / -0.68	-0.76 / -0.76	-0.75 / -0.63
MA	0.79 / 0.94	0.83 / 0.96	0.82 / 0.97	0.60 / 0.65	0.68 / 0.69	0.68 / 0.68

ported in Table 2. For the cubes, [14] ranks the first in all measurements, and for the spheres there is no clear winner.

In the last rows of Tables 1 & 2 we provide the mean of the absolute scores, denoted as MA . The mean absolute scores show that the most challenging case is the cubes with $c = 25$ for both the additive and subtractive cases. We also note that for the additive perturbations it is significantly harder for the considered methods to correlate with the ground truth when the parameter w is varied.

Table 2. The averaged Kendall τ for the subtractively perturbed cubes (the first value) and spheres (the second value).

Method	w = 3	w = 4	w = 5	c = 25	c = 50	c = 75
[16]	1.00 / 0.99	1.00 / 0.99	1.00 / 1.00	0.67 / 0.92	0.89 / 1.00	0.91 / 0.97
[14]	1.00 / 0.89	1.00 / 0.88	1.00 / 0.93	0.97 / 0.97	0.97 / 1.00	0.99 / 1.00
[21]-1	0.68 / 0.81	0.81 / 0.95	1.00 / 1.00	0.77 / 0.83	0.91 / 1.00	0.97 / 1.00
[21]-2	0.21 / 0.48	0.48 / 0.83	0.87 / 0.96	0.37 / 0.61	0.69 / 0.80	0.81 / 0.95
MA	0.72 / 0.79	0.82 / 0.91	0.97 / 0.97	0.70 / 0.83	0.87 / 0.95	0.92 / 0.98

5.2. Collection 2

For the first family of Collection 2, the top five methods are C_{CRE} , C_{PC} , [27], [4] and C_2 according to the summed scores given in Table 3. Note that all of these are 2D methods, three of which are convexity measures. The best performing 3D method is [14] and places the sixth. [26] and [21]-2 perform the poorest on this family both having almost no correlations with the ground truth considering all of the groups.

For each group except the fifth, there is at least one method that completely agrees (or disagrees) with the ground truth. It seems that none of the considered methods is able to capture the

notion of complexity that induces the order given by the ground truth for Group 5. The MA scores are in alignment with this, indicating that Group 5 is the most challenging group. We also note that the highest MA is attained by Group 3 that consists only of three elements which is the minimum number required to attain a non-trivial Kendall rank correlation coefficient (τ).

Strangely, some of the methods ([28], [29], [15] and [21]-2) have both strongly positive and strongly negative correlations with the ground truth.

Table 3. The Kendall τ for the first family (the first value) and the non-uniformity measurements for the second family (the second value) of Collection 2.

Method	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Sum
[16]	0.33 / 2.91	0.67 / 2.08	0.00 / 0.16	1.00 / 0.66	0.47 / 2.53	- / 2.75	2.47 / 11.09
[14]	0.81 / 1.82	1.00 / 1.31	0.00 / 0.02	0.80 / 0.28	0.33 / 2.78	- / 0.52	2.94 / 6.73
[21]-1	-0.52 / 0.88	-0.33 / 1.65	-1.00 / 0.41	0.80 / 0.06	0.60 / 0.75	- / 1.05	-0.46 / 4.79
[21]-2	0.62 / 0.87	0.33 / 1.83	-1.00 / 0.53	-0.60 / 0.09	0.73 / 0.49	- / 1.46	0.09 / 5.27
[15]	-0.81 / 0.86	0.33 / 1.60	1.00 / 0.38	-0.20 / 0.63	-0.60 / 0.67	- / 3.33	-0.28 / 7.47
[29]	-0.81 / 1.37	0.67 / 1.91	1.00 / 0.31	0.40 / 0.35	0.47 / 1.00	- / 1.62	1.72 / 6.56
[28]	-0.43 / 1.57	-0.67 / 0.71	1.00 / 1.93	0.00 / 0.08	0.33 / 0.12	- / 0.24	0.24 / 4.65
[27]	-1.00 / 1.75	-1.00 / 2.15	-1.00 / 1.08	-0.40 / 0.34	-0.47 / 1.43	- / 5.55	-3.87 / 12.31
C_{CRE}	0.90 / 1.72	1.00 / 2.10	1.00 / 0.99	0.60 / 0.26	0.47 / 0.86	- / 4.28	3.97 / 10.21
[4]	0.81 / 2.01	0.67 / 2.98	1.00 / 1.76	0.80 / 1.63	0.47 / 0.80	- / 6.04	3.74 / 15.23
[26]	-0.71 / 1.06	-0.33 / 0.84	0.33 / 0.24	0.40 / 0.21	0.33 / 0.80	- / 1.33	0.02 / 4.47
C_r	-0.81 / 1.44	0.00 / 1.29	0.33 / 0.17	0.40 / 0.17	0.33 / 0.61	- / 1.19	0.26 / 4.87
C_1	0.05 / 1.31	-0.33 / 1.88	-0.33 / 0.10	-0.80 / 0.09	-0.73 / 0.45	- / 0.59	-2.15 / 4.41
C_{PC}	0.62 / 1.19	0.67 / 1.49	1.00 / 0.20	1.00 / 0.14	0.60 / 1.13	- / 0.62	3.89 / 4.78
[24]	0.43 / 1.17	0.33 / 1.66	1.00 / 0.16	0.60 / 0.10	0.47 / 0.33	- / 0.92	2.83 / 4.53
[23]	-0.52 / 0.98	-0.33 / 1.37	-1.00 / 0.10	-0.60 / 0.25	-0.47 / 0.75	- / 1.42	-2.92 / 4.88
C_2	-0.43 / 1.28	-0.33 / 1.41	-1.00 / 0.19	-0.80 / 0.18	-0.47 / 0.29	- / 1.02	-3.03 / 4.37
[22]-1	0.14 / 0.93	-0.33 / 1.59	-1.00 / 0.11	-0.80 / 0.20	-0.47 / 0.43	- / 1.00	-2.46 / 4.27
[22]-2	-0.05 / 1.12	-0.33 / 1.85	-1.00 / 0.13	-0.80 / 0.16	-0.60 / 0.43	- / 0.82	-2.78 / 4.51
MA	0.57 / 1.38	0.51 / 1.67	0.79 / 0.47	0.62 / 0.31	0.49 / 0.88	- / 1.88	2.11 / 6.59

For the second family of Collection 2, we start by remarking that the reported scores indicate better performances when they are close to 0, in contrast with the other reported scores. Similar to the case in the first family, 2D methods take the lead (listed from best to worst: [22]-1, [24], C_2 , C_1 , [26]), with the best performing 3D method ([21]-1) placing the 8th. The worst performing method is [4]. This is interesting because it is also the third best performing method in the first family. In a similar manner, we note that there is no overlap between the top five performers of the two families except for C_2 .

The top three performers for both families are 2D methods based on the summed scores. The highest scoring 3D method for the first family is [14] and [21]-1 for the second family.

5.3. Collection 3

The Kendall rank correlation coefficients computed for Collection 3 are reported in Table 4. The best performers are [21]-2, [26], [4] and [4] for τ_{cut} , τ_{hall} , τ_{cut} and τ_{all} , respectively.

1 All of the methods, except [21]-1, perform better when the
2 correlations are computed over the whole collection, regardless
3 of the ground truth.

4 For the tasks of this collection we observe that [21]-2 out-
5 performs [21]-1. This is interesting because [21]-1 is a better
6 performer for the majority of tasks involving the other two col-
7 lections. Since these two are the first two principal components
8 of the method of [21], this suggests that the segmentation ac-
9 counts for an aspect of shape complexity different than those of
10 the other collections.

11 We note that all of the top three performers are 2D methods,
12 except for $\tau_{\mu_{\text{cut}}}$.

Table 4. Kendall τ when the ground truth is the mean and the standard deviation of the number of segments of the human segmentations.

Method	$\tau_{\mu_{\text{cut}}}$	$\tau_{\mu_{\text{all}}}$	$\tau_{\sigma_{\text{cut}}}$	$\tau_{\sigma_{\text{all}}}$
[16]	0.148	0.346	0.072	0.234
[14]	0.041	0.354	0.018	0.203
[21]-1	0.110	0.105	0.055	-0.006
[21]-2	0.151	0.417	0.065	0.262
[15]	-0.022	0.251	0.013	0.138
[29]	0.061	0.401	0.027	0.202
[28]	0.140	0.375	0.087	0.167
[27]	-0.082	-0.458	-0.089	-0.282
C_{CRE}	0.110	0.585	0.083	0.331
[4]	0.132	0.600	0.117	0.350
[26]	0.131	0.671	0.066	0.283
C_{σ}	0.075	0.540	0.077	0.244
C_1	-0.037	-0.255	-0.037	-0.078
C_{PC}	0.099	0.464	0.078	0.217
[24]	0.041	0.326	0.029	0.123
[23]	-0.105	-0.486	-0.056	-0.219
C_2	-0.112	-0.501	-0.084	-0.227
[22]-1	-0.069	-0.372	-0.085	-0.168
[22]-2	-0.079	-0.395	-0.090	-0.188
MA	0.092	0.416	0.065	0.206

13 6. Discussion

14 Despite the lack of full shape information, 2D methods are ob-
15 served to perform unexpectedly well when compared with 3D
16 methods, especially for Collection 3. However, it should be
17 noted that for the results presented in this paper, the participat-
18 ing 3D methods also do not make use of the full shape informa-
19 tion because some of them ([14, 21]) down-sample the shapes

20 in Collection 1 and they all need to voxelize the shapes in Col-
21 lection 2 and 3. Also, [16] solves for only a limited number of
22 steps for all of the collections due to the size and the number of
23 the shapes and relatively costly computation.

24 One of the interesting observations is that [4], [27] and [26]
25 perform poorly under the changes of the parameter w despite
26 their high scores under the changes of the parameter c . This
27 might be explained by the fact that changing the width parame-
28 ter w has a greater impact on the local changes in curvature than
29 the parameter c and that these methods are inherently curvature-
30 dependent. Similar results for the methods that measure convexity
31 can be explained in a similar manner since convexity
32 can be related to curvature for the examples in our datasets.
33 Note also that this observation highlights the non-triviality of
34 the noisy collections.

35 Assessing the results for Collection 1 suggests the use of dif-
36 ferent methods for different use cases. For example, [16], [21]-1
37 can be used in applications involving additive perturbations and
38 [15] can be used in applications involving noisy spheres. Pro-
39 vided one has information about the type of the noise present in
40 their use cases, one can settle for [4], [27], C_2 , [22], or [23]. For
41 overall robustness [14] can be preferred. The results for Collec-
42 tion 2 suggest that classical measures supported by psychology
43 experiments are still better alternatives for quantifying percep-
44 tual complexity as judged from the final product of the design
45 process (*i.e.* ignoring the generation level complexity). For
46 Collection 3, we observe that the performances of the methods
47 improve significantly when the entire collection is considered.
48 In this sense, we can say that the task of correlating complexity
49 with the segmentation is harder when the shapes are from the
50 same category.

51 In Fig. 6, two 2D embeddings of the evaluated complexity
52 measures using Stochastic Neighborhood Embedding (t-SNE)
53 [35] are depicted. For each measure a high dimensional feature
54 vector is formed using the Kendall rank correlation coefficients
55 reported in Tables 1-4. For the plot on the left, 17-dimensional
56 feature vectors (whose components are the twelve τ scores from
57 Table 1 and five τ scores from Table 3) are used. For the plot on

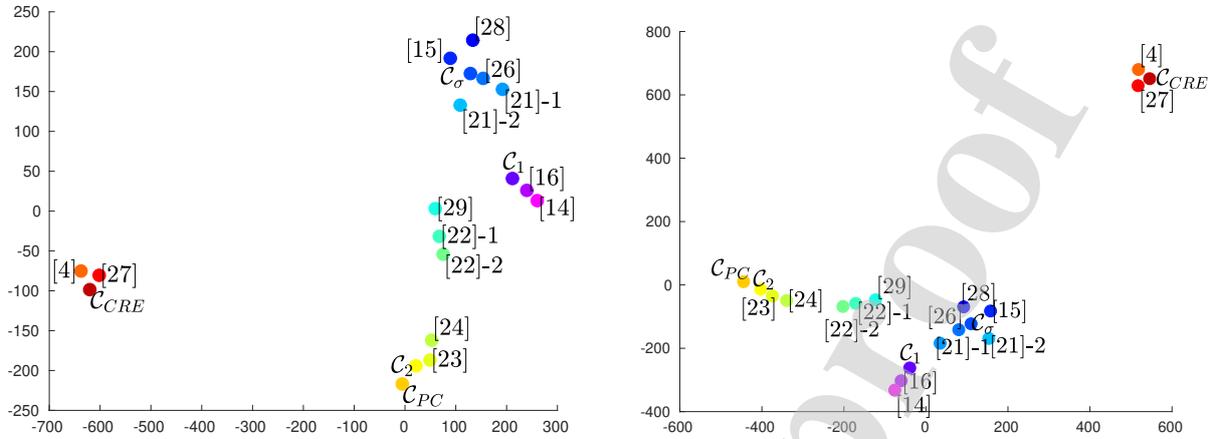


Fig. 6. Clustering of the methods in τ -based feature space: 2D embedding applied to 17 (left) and 21 (right) τ -values

the right 21-dimensional feature vectors are used by augmenting the 17-dimensional vectors with four additional τ scores from Table 4. We considered the scores from Collection 3 as optional because we feel that the nature of this collection is different from those of the first two collections. Note that we negate the τ scores of the methods, [27], C_1 , [23], C_2 , [22]-1 and [22]-2 as they serve as measures of simplicity rather than complexity. In both plots the perplexity parameter is set to 2. Nevertheless, we have observed that doubling or even quadrupling the perplexity parameter does not make a significant qualitative change except that the spread gets larger. Notice that [4], [27] and [26] form a distinct cluster. Another interesting observation from Fig. 6 is that the two methods [15] and [16], both employing real valued fields computed using a common Partial Differential Equation, are not close in the τ -based feature space. This is because these methods use different metrics. The choice of the metric makes [16] an ideal method for noisy cubes whereas the other is better suited for noisy spheres.

In addition to correlations between the ground truths and the order induced by the measures, we believe that the correlations among the orders induced by the measures convey insight into the ill-defined concept of shape complexity. Hence, we report in Fig. 7 Kendall rank correlation coefficients (τ) for each pair of methods over the dataset. Specifically, for Collection 1 we compute the mean of τ_i ($i \in \{1, 2, \dots, 50\}$) for the groups of each family, for Collection 2, we compute τ over the families

(i.e. disregarding the groups), and for Collection 3 we compute both the mean of τ_{cat} over the categories and τ_{all} over the whole collection. Here also we negate the scores of the methods, [27], C_1 , [23], C_2 , [22]-1 and [22]-2. The results show that the methods correlate the most to each other over the additively perturbed spheres. This could be explained by noting that the different approaches of the methods towards complexity, such as uniformity of curvature, convexity, or the agreement of the shape with the underlying grid, more or less agree for this family. Similar clusters to the ones seen to emerge in Fig. 6 can be identified, such as [4], [27] and C_{CRE} or the cluster consisting of convexity measures. Yet, for example, the second family of Collection 2 provides a means of distinguishing [4] from C_{CRE} and [27]. The same family also allows us to observe the differences between the behaviors of the 3D methods. Similarly, the results acquired for Collection 3 by comparing shapes from the same categories show that [23] and C_2 are more close to each other than they are to [22]-1 and [22]-2, and vice versa. Together, these provide support for our claim that the three collections account for different aspects of shape complexity.

The correlations for Collection 3 are generally lower than those for Collections 1 and 2. This could be a consequence of either the data being more challenging, or else that the proxy task does not map strongly to complexity. This needs further study, and can be explored in future work.

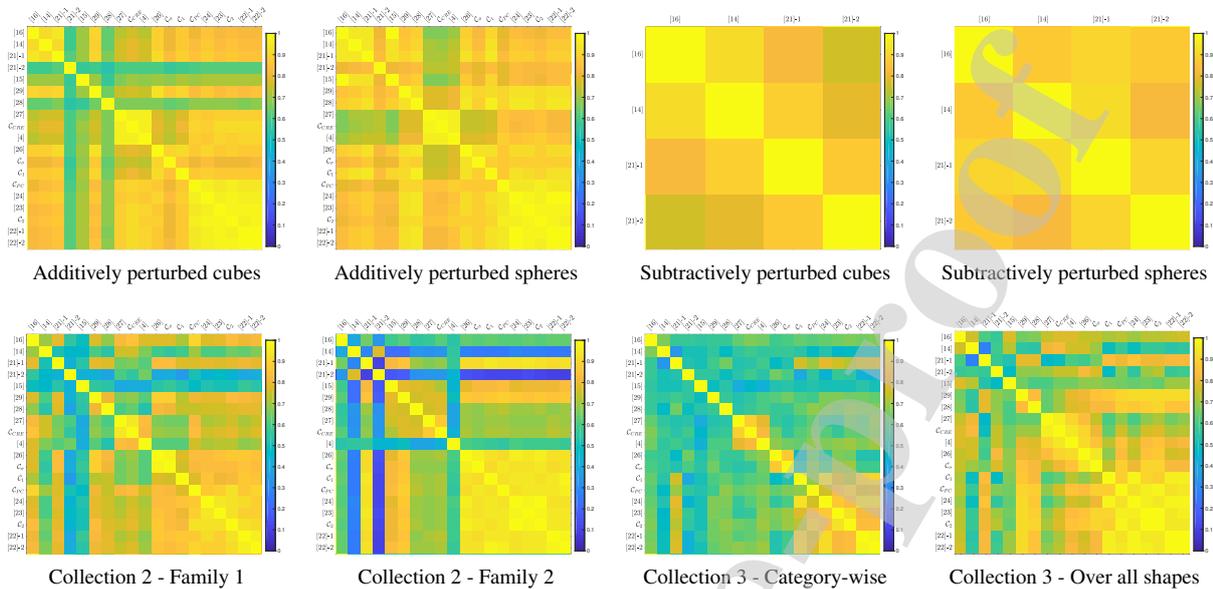


Fig. 7. Correlations between pairs of methods

7. Conclusion

We have introduced a novel 3D dataset to evaluate shape complexity measures. Using this dataset we not only evaluated the methods with respect to ground truth but also with respect to each other under a rich variety of ordering tasks in order to see how they are related in the context of shape complexity. To evaluate methods with respect to each other, we clustered measures in the τ -based feature space, and displayed pairwise rank correlations between orders induced by all pair of methods.

We conclude the paper by noting that the evaluation methodology of the paper is a significant improvement on the current literature in the sense that the reported scores are quantitative with justified ground truths and the analysis is reproducible. Since the research in 3D shape complexity is still in its infancy, we believe that this work will encourage further explorations of the field.

References

- [1] Attneave, F. Physical determinants of the judged complexity of shapes. *Journal of experimental Psychology* 1957;53(4):221.
- [2] Joshi, D, Ravi, B. Quantifying the shape complexity of cast parts. *Computer-Aided Design and Applications* 2010;7(5):685–700.
- [3] Wing, CK. On the issue of plan shape complexity: plan shape indices revisited. *Construction Management & Economics* 1999;17(4):473–482.

- [4] Page, DL, Koschan, AF, Sukumar, SR, Roui-Abidi, B, Abidi, MA. Shape analysis algorithm based on information theory. In: *International Conference on Image Processing*; vol. 1. 2003, p. 229–232.
- [5] Backes, AR, Eler, DM, Minghim, R, Bruno, OM. Characterizing 3d shapes using fractal dimension. In: Bloch, I, Cesar, RM, editors. *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*. 2010, p. 14–21.
- [6] Arai, K. Visualization of 3d object shape complexity with wavelet descriptor and its application to image retrievals. *Journal of Visualization* 2012;15(2):155–166.
- [7] Kim, SH, Lyu, I, Fonov, VS, Vachet, C, Hazlett, HC, Smith, RG, et al. Development of cortical shape in the human brain from 6 to 24 months of age via a novel measure of shape complexity. *NeuroImage* 2016;135:163–176.
- [8] Nitzken, M, Casanova, M, Gimel'farb, G, Elnakib, A, Khalifa, F, Switala, A, et al. 3d shape analysis of the brain cortex with application to dyslexia. In: *IEEE International Conference on Image Processing*. 2011, p. 2657–2660.
- [9] Volarevic, N, Cosic, P. Shape complexity measure study. *Annals of DAAAM & Proceedings* 2005;:375–377.
- [10] Hutcheson, F. An inquiry into the original of our ideas of beauty and virtue. *Natural Law and Enlightenment Classics*; revised edition ed.; Carmel: IN: Liberty Fund, Inc.; 2008.
- [11] Birkhoff, GD. *Aesthetic measure*. Harvard University Press; 1933.
- [12] Maragos, P. Pattern spectrum and multiscale shape representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1989;11(7):701–716.
- [13] Rigau, J, Feixas, M, Sbert, M. Shape complexity based on mutual information. In: *International Conference on Shape Modeling and Applications 2005 (SMI'05)*. IEEE; 2005, p. 355–360.
- [14] Genctav, M, Genctav, A, Tari, S. Nonlocal via local—nonlinear via linear: A new distance field via screened Poisson equation. *J Math Imaging Vis* 2016;55(2):242–252.
- [15] Genctav, A, Tari, S. Discrepancy: Local/global shape characterization with a roundness bias. *J Math Imaging Vis* 2019;61:160–171.
- [16] Arslan, MF, Tari, S. Complexity of shapes embedded in \mathbb{Z}^n with a bias towards squares. *IEEE Transactions on Image Processing* 2020;29:8870–8879.
- [17] Stiny, G, Gips, J. *Algorithmic aesthetics: computer models for criticism and design in the arts*. Univ of California Press; 1978.
- [18] Rossignac, J. Shape complexity. *The Visual Computer* 2005;21(12):985–996.

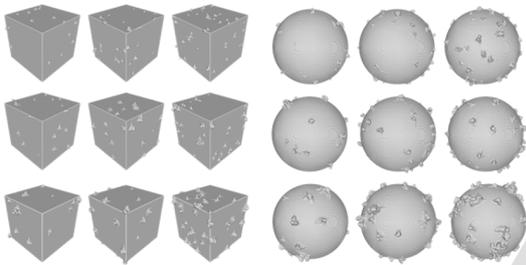
- 1 [19] Saleem, W, Belyaev, A, Wang, D, Seidel, HP. On visual complexity of
2 3d shapes. *Computers & Graphics* 2011;35(3):580–585.
- 3 [20] Chen, X, Golovinskiy, A, Funkhouser, T. A benchmark for 3D mesh
4 segmentation. *ACM Trans Graph* 2009;28(3):73:1–73:12.
- 5 [21] Gardiner, JD, Behnsen, J, Brassey, CA. Alpha shapes: Determining 3D
6 shape complexity across morphologically diverse structures. *BMC Evo-*
7 *lutionary Biology* 2018;18(184):1–16. URL: <https://bmcecol1.evol.biomedcentral.com/articles/10.1186/s12862-018-1305-z>.
8 doi:10.1186/s12862-018-1305-z.
- 9 [22] Rosin, PL. Classification of pathological shapes using convexity mea-
10 sures. *Pattern Recognition Letters* 2009;30(5):570–578.
- 11 [23] Žunić, J, Rosin, PL. A new convexity measurement for poly-
12 gons. *IEEE Transactions on Pattern Analysis and Machine Intelligence*
13 2004;26(7):923–934.
- 14 [24] Brinkhoff, T, Kriegel, HP, Schneider, R, Braun, A. Measuring the com-
15 plexity of polygonal objects. In: *International Workshop on Advances in*
16 *GIS*; vol. 109–117. 1995.
- 17 [25] Wang, F, Vemuri, BC, Rao, M, Chen, Y. Cumulative residual entropy,
18 a new measure of information & its application to image alignment. In:
19 *International Conference on Computer Vision*; vol. 2. 2003, p. 548–548.
- 20 [26] Matsumoto, T, Sato, K, Matsuoka, Y, Kato, T. Quantification of “com-
21 plexity” in curved surface shape using total absolute curvature. *Comput-*
22 *ers & Graphics* 2019;78:108–115.
- 23 [27] Vasselle, B, Giraudon, G. 2-d digital curve analysis: A regularity mea-
24 sure. In: *International Conference on Computer Vision*. 1993, p. 556–561.
- 25 [28] Hayward, J, Orford, JD, Whalley, WB. Three implementations of fractal
26 analysis of particle outlines. *Computers & Geosciences* 1989;15(2):199–
27 207.
- 28 [29] Borkowski, W. Fractal dimension based features are useful descrip-
29 tors of leaf complexity and shape. *Canadian Journal of Forest Research*
30 1999;29(9):1301–1310.
- 31 [30] Orbach, DN, Brassey, CA, Gardiner, JD, Brennan, PLR. 3D genital
32 shape complexity in female marine mammals. *Ecology and Evolution*
33 2021;(December 2020):1–9. URL: <https://onlinelibrary.wiley.com/doi/full/10.1002/ece3.7269>. doi:10.1002/ece3.7269.
- 34 [31] Brassey, CA, Behnsen, J, Gardiner, JD. Postcopulatory sexual
35 selection and the evolution of shape complexity in the car-
36 nivoran baculum. *Proceedings of the Royal Society B: Bi-*
37 *ological Sciences* 2020;287(1936):20201883. URL: <https://royalsocietypublishing.org/doi/10.1098/rspb.2020.1883>.
38 doi:10.1098/rspb.2020.1883.
- 39 [32] Edelsbrunner, H, Mücke, EP. Three-Dimensional Alpha Shapes. *ACM*
40 *Transactions on Graphics* 1994;13(1):43–72. doi:10.1145/174462.156635. arXiv:9410208.
- 41 [33] Ramer, U. An iterative procedure for the polygonal approximation of
42 plane curves. *Computer Graphics and Image Processing* 1972;1:244–256.
- 43 [34] Arkin, E, Chew, L, Huttenlocher, D, Kedem, K, Mitchell, J. An effi-
44 ciently computable metric for comparing polygonal shapes. *IEEE Trans-*
45 *actions on Pattern Analysis and Machine Intelligence* 1991;13(3):209–
46 216.
- 47 [35] van der Maaten, L, Hinton, G. Visualizing high-dimensional data using
48 t-SNE. *J Mach Learn Res* 2008;9:2579–2605.
- 49
50
51
52

Research Highlights

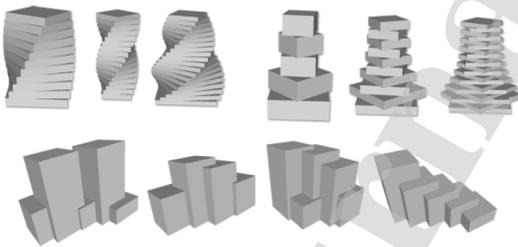
- ❖ Two novel datasets with associated ground truth are introduced.
- ❖ Segmentation is used as a proxy task for measuring complexity.
- ❖ The performance of 2D and 3D complexity measures are systematically compared and evaluated.

Data Contribution

Collection 1



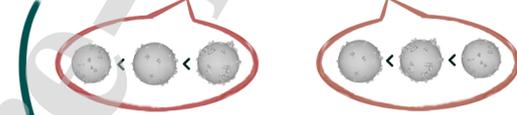
Collection 2



Methodology

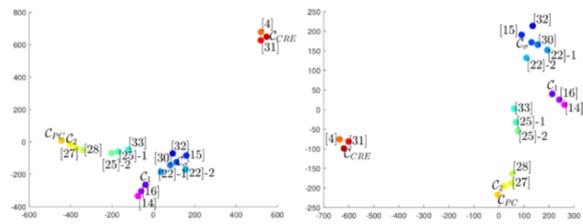
Analysis with Ground Truth

Kendall_tau (Ground Truth, Method Result)



Method	w = 3	w = 4	w = 5	c = 25	c = 50	c = 75
[16]	1.00 / 1.00	1.00 / 1.00	1.00 / 0.99	1.00 / 1.00	1.00 / 1.00	1.00 / 1.00
[14]	0.87 / 0.93	0.95 / 0.97	0.93 / 1.00	0.89 / 0.93	0.92 / 1.00	0.96 / 1.00
[22]-1	0.99 / 0.97	0.99 / 1.00	0.99 / 1.00	1.00 / 0.99	1.00 / 1.00	1.00 / 1.00
[22]-2	0.17 / 0.57	0.24 / 0.80	0.52 / 0.88	-0.09 / 0.35	0.03 / 0.69	0.39 / 0.69
[15]	0.25 / 1.00	0.31 / 1.00	0.68 / 1.00	0.13 / 1.00	0.39 / 1.00	0.52 / 1.00
[33]	0.67 / 0.91	0.84 / 0.99	0.80 / 0.97	0.84 / 0.68	0.95 / 0.71	0.93 / 0.68
[24]	0.45 / 0.97	0.65 / 0.99	-0.11 / 0.96	0.44 / 0.83	0.43 / 0.89	0.04 / 0.84
[31]	-0.97 / -1.00	-0.96 / -1.00	-0.97 / -1.00	-0.29 / -0.08	-0.29 / 0.05	-0.24 / 0.28

Cross-method Analysis



Title page for SHREC'21: Quantifying Shape Complexity

M. Ferhat Arslan, Department of Computer Engineering, Middle East Technical University, Turkey, email: mferhata@gmail.com

Alexandros Haridis, Department of Architecture, Massachusetts Institute of Technology, USA, email: charidis@mit.edu

Paul L. Rosin, School of Computer Science & Informatics, Cardiff University, UK, email: Paul.Rosin@cs.cf.ac.uk

Sibel Tari, Department of Computer Engineering, Middle East Technical University, Turkey, email: stari@metu.edu.tr

Charlotte Brassey, School of Science and The Environment, Manchester Metropolitan University, UK, email: c.brassey@mmu.ac.uk

James D. Gardiner, Institute of Ageing and Chronic Disease, University of Liverpool, UK, email: jjdg@liverpool.ac.uk

Aslı Gençtav, Department of Computer Engineering, TED University, Turkey, email: asli.genctav@tedu.edu.tr

Murat Gençtav, Microelectronics, Guidance and Electro-Optics Division, ASELSAN A.S., Turkey, email: email: mgenctav@aselsan.com.tr

Mazlum Ferhat Arslan: Conceptualization, Methodology, Software, Formal analysis, Data Curation, Writing - Original Draft, Review & Editing, Visualization.

Alexandros Haridis: Conceptualization, Methodology, Data Curation, Writing - Original Draft, Review & Editing.

Paul L. Rosin: Conceptualization, Methodology, Formal analysis, Writing - Original Draft, Review & Editing, Supervision.

Sibel Tari: Conceptualization, Methodology, Formal analysis, Writing - Original Draft, Review & Editing, Supervision.

Charlotte Brassey: Writing - Original Draft.

James D. Gardiner: Writing - Original Draft.

Asli Genctav: Writing - Original Draft, Visualization.

Murat Genctav: Writing - Original Draft.

The authors declare no conflicts

Journal Pre-proof