

Topological Saliency

Harish Doraiswamy^a, Nithin Shivashankar^b, Vijay Natarajan^{b,c}, Yusu Wang^d

^aDepartment of Computer Science and Engineering, Polytechnic Institute of New York University, USA

^bDepartment of Computer Science and Automation, Indian Institute of Science, Bangalore, India

^cSupercomputer Education and Research Centre, Indian Institute of Science, Bangalore, India

^dDepartment of Computer Science and Engineering, The Ohio State University, USA

Abstract

Topological methods have been successfully used to identify features in scalar fields and to measure their importance. In this paper, we define a notion of topological saliency that captures the relative importance of a topological feature with respect to other features in its local neighborhood. Features are identified by extreme points of an input scalar field, and their importance measured by the so-called topological persistence. Computing the topological saliency of all features for varying neighborhood sizes results in a saliency plot that serves as a summary of relative importance of all topological features. We develop a convenient tool for users to interactively select and inspect features using the saliency plot. We demonstrate the use of topological saliency together with the rich information encoded in the saliency plot in several applications, including key feature identification, scalar field simplification, and feature clustering.

Keywords: saliency, persistence, computational topology

1. Introduction

The use of topological methods is becoming popular for analyzing, visualizing, and exploring scalar fields. It is being used for a wide range of applications including topological simplification and cleaning [1, 2, 3, 4], surface segmentation and parametrization [5, 6, 7], topology based shape matching [8, 9], and designing transfer functions for volume rendering [10, 11, 12, 13]. Distinguishing between significant and unimportant features of the input forms an integral part of the methodologies used in these applications. In this paper we focus on this theme to define a notion of saliency for topological features and explore its application to visual analysis of features.

1.1. Related work

Features in scalar fields are represented by critical points of the field. A common approach to identifying features in 3D geometric models is to first design a descriptor function that captures important information about the input and use critical points of such a descriptor function as representatives of features. Various methods have been proposed to produce a meaningful descriptor function. Once a descriptor function is given, one can measure the topological importance of the features (critical points) associated with it based on the so-called persistence homology, originally proposed by Edelsbrunner et al. [1]. Indeed, topological persistence has been demonstrated

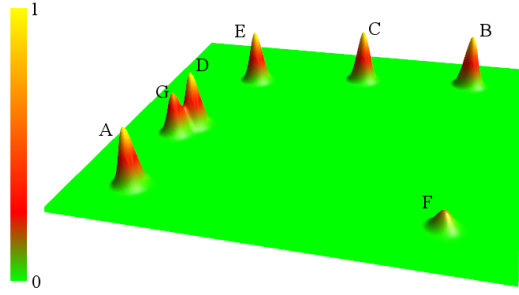
to be an effective importance measure and has been used for many applications, including scalar field simplification [14, 15] and shape matching [8]. While the notion of persistence can effectively describe the importance of a feature with respect to an input scalar function or filtration, it is somewhat oblivious to other geometric information not encoded in the input function. In particular, it does not reflect how important a feature is relative to other features in its neighborhood.

In this paper, we aim to initiate a study in this direction by defining a notion of topological saliency for features in the input that captures the relative importance of a feature within a spatial neighborhood. We propose to do this following an approach similar to saliency models used in image and geometric mesh analysis. Many models for obtaining salient locations in images have been proposed [16, 17, 18, 19, 20]. In particular, Itti et al. [19] propose a model that computes the saliency of a pixel in an image based on the properties of pixels in its neighborhood. Lee et al. [21] extend this model to geometric features and propose a notion of mesh saliency that captures the saliency of a point in a surface or volume mesh. It is computed as the curvature at a point weighted by the average curvature within a small neighborhood.

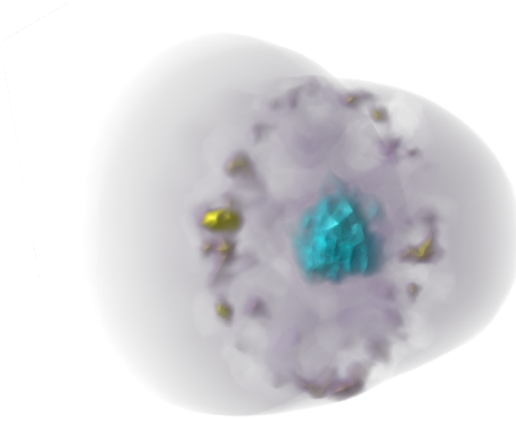
Our topological saliency framework can be viewed as a way to combine geometry information with topological methods. Note that this measure is not intended to replace existing global importance measures. Rather, we expect it to complement existing measures when applied to the visual analysis of features.

We remark that the theme of combining geometry and topology is not new. For example, Carr et al. [22] employed geometric measures computed on contour trees to find and simplify less significant features. Weber et al. [12] used Reeb graphs to

Email addresses: hdoraisw@poly.edu (Harish Doraiswamy),
nithin@csa.iisc.ernet.in (Nithin Shivashankar),
vijayn@csa.iisc.ernet.in (Vijay Natarajan),
yusu@cse.ohio-state.edu (Yusu Wang)



(a)



(b)

Figure 1: **(a)** A sample terrain with seven peaks. Traditional topological methods identify peaks *A*, *B*, *C*, *D* and *E* as important. Even though peak *F* remains a lone peak for a significantly large part of the domain, it is not considered to be important. **(b)** Volume rendering of a breast data set. The tumor, highlighted in cyan, corresponds to a low persistent feature in the input.

identify significant features in volumetric data in order to design transfer functions for rendering the volume. The concept of topological landscapes [23, 24] provides an intuitive view of the data by displaying its topological features, abstracted by the contour tree, as a terrain. In Agarwal et al. [25], the authors aim to use a descriptor function to encode certain geometric information of interest, and use topological persistence to identify geometric features from this function. The concept of “localized homology” was later proposed by computing the homology from local pieces to global pieces, so that the generators of homology classes are localized in local pieces [26]. Reininghaus et al. [27] proposed an importance measure for critical points in two dimensional scalar fields called the scale space persistence, which combines the notion of deep structure of the scale space with topological persistence. The scale space persistence is computed by accumulating the persistence values of a critical point through its evolution in the scale space [28].

1.2. Motivating example

Consider the terrain shown in Figure 1(a) with seven peaks. Existing topology-based methods would ignore peak *F* even though it dominates a large area of the domain in the sense that it remains an important feature within a large neighborhood size. Similarly, based on persistence, peaks *B* and *D* would have been declared as equally important even though *D* is surrounded by other peaks of similar height making it not as dominant as *B*. In general, spatial distribution of topological features has not been considered while measuring the size of a feature and its significance.

Such scenarios are common when studying medical data obtained using diffuse optimal tomography. For example, Figure 1(b) shows a volume rendering of a breast dataset having a tumor. The high persistent features in the input corresponds to the fibre bundles on the periphery of the volume, shown in yellow. However, the tumor corresponds to a low persistent feature that is isolated within the volume. Therefore, using persistence

alone, it may not be possible to identify the tumor as a significant feature of the input.

1.3. Results

We address the problem raised in the above example by defining a notion of *topological saliency* that considers the presence or absence of other features within the neighborhood while measuring the importance of a topological feature. A feature in this paper is always represented by an extremum (minimum or maximum) of the input scalar field. The topological saliency of a feature is computed as a weighted average of the topological persistence of features within its local neighborhood. We also introduce a *saliency plot* that is generated by computing the topological saliency of all features for varying neighborhood sizes. This plot can be considered as an augmentation or refinement of persistence, obtained by injecting certain spatial geometry information into it. We also propose the use of topological saliency for simplifying the input in order to remove noise.

Further, we develop visualization software to facilitate the use of the saliency plot. In particular, the software can compute the saliency plots of minima and maxima for a given input field as well as a decomposition of the input domain around these features. It also allows the users to interactively select and inspect features of the input using the saliency plot. Finally, we demonstrate the use of this software and the concept of topological saliency in the following applications.

- Identify key features that may be missed by standard persistence. In particular, we use topological saliency to identify breast tumors.
- Identify craters on Mars. We use topological saliency in conjunction with standard persistence to identify significant craters on the surface of Mars.
- Extract similar features. We use the topological saliency plot to identify and group similar features of the input.

2. Background

We briefly introduce some necessary notations and refer the readers to appropriate textbooks [29, 30, 31] for more precise definitions and comprehensive discussions of these concepts.

2.1. Morse function

Let $\mathbb{M}$ denote a d -manifold with or without boundary. Given a smooth, real-valued function $f : \mathbb{M} \rightarrow \mathbb{R}$ defined on $\mathbb{M}$, the *critical points* of f are exactly where the gradient becomes zero. The function f is called a *Morse function* if it satisfies the following conditions [32]:

1. All critical points of f are non-degenerate and lie in the interior of $\mathbb{M}$.
2. All critical points of the restriction of f to the boundary of $\mathbb{M}$ are non-degenerate.
3. All critical values are distinct *i.e.*, $f(p) \neq f(q)$ for all critical points $p \neq q$.

In the remaining discussion, we assume that the scalar function defined on our input is a Morse function. In case the above conditions do not hold, simulated perturbation of the function [33, Section 1.4] ensures that no two critical values are equal. For a Morse function f defined on a d -manifold $\mathbb{M}$, there are $d + 1$ types of critical points indexed from 0 to d . The two most familiar types are *minimum* (with index 0) and *maximum* (with index d), corresponding to a point p whose function value is smaller, or larger, than all other points within a sufficiently small neighborhood of p , respectively.

2.2. Topological persistence

Given a Morse function $f : \mathbb{M} \rightarrow \mathbb{R}$, the topological persistence algorithm sweeps the manifold $\mathbb{M}$ in increasing order of function value¹, and inspects the changes in the homology groups of the sublevel set $\mathbb{M}^{(-\infty, \alpha]} := \{x \in \mathbb{M} \mid f(x) \leq \alpha\}$, which only happens when the sweep passes a critical point of f . In particular, at a critical point, either new topology is generated or some topology is destroyed, where topology is quantified by a class of ‘cycles’. For example, a 0-dimensional cycle represents a connected component, a 1-dimensional cycle is a loop that represents a tunnel, and a 2-dimensional cycle bounds a void. A critical point is a creator if new topology appears and a destroyer otherwise. It turns out that one can pair up each creator v_1 uniquely with a destroyer v_2 that destroys the topology created at v_1 . The persistence value of v_1 and v_2 is defined as $f(v_2) - f(v_1)$, which intuitively indicates the lifetime of the feature created at v_1 , and thus the importance of v_1 and v_2 .

In this paper, we only consider extreme points of the input function as features. Given an input domain of size n , the persistence of such features can be computed efficiently in $O(n \log n + n\alpha(n))$ time using the union-find data structure, versus the usual cubic-time algorithm to compute general topological persistence [1, 34].

¹The persistence algorithm works for more general topological spaces than manifolds. We only describe the case when it is induced by a function defined on a manifold.

3. Topological Saliency

In this section we describe the notion of topological saliency, which combines the topological persistence of a feature with its neighborhood information, and discuss applications.

3.1. Definition

Let the set $C = \{c_1, c_2, \dots, c_t\}$ be the set of minima of the input function $f : \mathbb{M} \rightarrow \mathbb{R}$. Let $P(i)$ denote the persistence of the topological feature created at c_i . Let $d_g(p, q)$ denote the geodesic distance between two points $p, q \in \mathbb{M}$. Consider a r -neighborhood $N_r(i) = \{x \in \mathbb{M} \mid d_g(x, c_i) \leq r\}$, which is the geodesic ball of radius r centered at critical point c_i . We define the *topological saliency* $T_r(i)$ of the feature created at c_i as

$$T_r(i) = \frac{\omega_i^i P(i)}{\sum_{c_j \in C} \omega_j^i P(j)},$$

where ω_j^i is a weighting function for the feature j with respect to i . The topological saliency at a maximum is defined in a symmetric manner. Two common choices of the weighting function are (a) uniform weight:

$$\omega_j^i = \begin{cases} 1 & \text{if } c_j \in N_r(i) \\ 0 & \text{otherwise} \end{cases}$$

and (b) Gaussian weight:

$$\omega_j^i = e^{-\frac{d_g(c_i, c_j)^2}{r^2}}$$

The topological saliency of a topological feature essentially normalizes the persistence of that feature based on the features that are present in its neighborhood. A Gaussian weighting function reduces the influence of farther features, while a uniform weighting function treats all features within the neighborhood equally.

Due to the discontinuous nature of the definition of topological saliency when using uniform weights, a small change in the location of a critical point (due to noise) could notably effect the saliency of itself, as well as that of other features. However, such changes do not have a significant impact when using Gaussian weights. This is because topological saliency changes continuously with varying neighborhood sizes. We have implemented both weighting schemes in our software. Unless otherwise mentioned, we use Gaussian weights for all experiments reported in this paper.

Note that when computing the topological saliency of a minimum (resp. a maximum), we only consider features of the same type, *i.e.*, other minima (resp. maxima) in its neighborhood. Intuitively, critical points of different indices capture different types of features: a minimum captures a valley while a maximum captures a mountain peak. We also remark that one could extend the topological saliency to critical points of other indices (*i.e.*, various saddle points). An index- k saddle point indicates the formation of a k -cycle. However, the meaning of a neighborhood of such features become less clear, and we leave the definition of salient index- k features for future work.

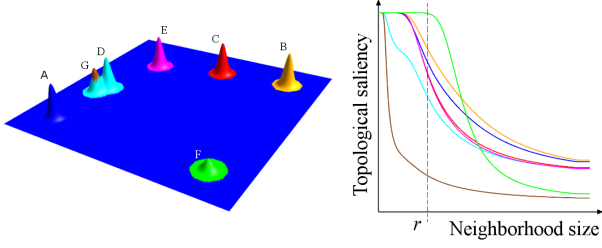


Figure 2: The topological saliency plot of the terrain data. Features in the input and the corresponding curve plots are highlighted using a common color. Note that the green colored peak F , which has low persistence, maintains a topological saliency close to 1 for up to a large value of neighborhood size.

3.2. Topological saliency plot

Consider the neighborhood of a feature c_i when $r = 0$. It consists of just the critical point c_i . Its topological saliency $T_0(i)$ equals 1. As we increase the neighborhood size r , $T_r(i)$ remains close to 1 until $N_r(i)$ includes another feature represented by, say c_j . At this point, the value of $T_r(i)$ reduces depending on the value of $P(j)$. Note that, at this value of r , $T_r(j)$ also decreases simultaneously. We can continue increasing r until it equals the diameter D of the input domain $\mathbb{M}$, at which point $N_D(i)$ covers $\mathbb{M}$. Plotting the values of $T_r(i)$ for all features from $r = 0$ to D , we obtain a *topological saliency plot*. Figure 2 shows the topological saliency plot for maxima in the terrain data from Figure 1(a).

Under the uniform weighting scheme, when r equals the diameter D of the input domain $\mathbb{M}$, the topological saliency $T_D(i)$ of c_i is equal to the standard persistence $P(i)$ scaled down by the total persistence $\sum_i P(i)$. Hence, by varying the parameter r from 0 to D , we move from a local perspective of the feature to its global perspective. One can recover the traditional persistence of a feature by looking at the corresponding value of T_D .

3.3. Significant features

The topological saliency plot can be used to identify significant features in multiple ways. In this paper, we use topological saliency defined for a fixed neighborhood size r as a measure to order features. Applying this alternative notion of importance to the terrain dataset for the value of r shown in Figure 2, the features are ordered as follows: F , B , A , C , E , D and G . This notion helps resolve our problem of identifying the green peak F in this input as being most significant. Note that the brown peak G is not considered to be significant because of its neighborhood even though its persistence is similar to F .

Ordering features based on its topological saliency requires choosing an appropriate value of r , which is application dependent. The user can compute the order at different perspectives, from local to global, by suitably specifying the neighborhood size r .

3.4. Saliency based simplification

Practical datasets usually contain noise, which may decrease topological saliency of a feature for small neighborhood sizes.

This is particularly true if that feature is present within a noisy region of the input. One way to prevent this artifact is to simplify the input, and thus remove noise. Simplification based on persistence could possibly remove salient features. For example, if we were to simplify the terrain dataset using persistence, then the peak F may be simplified away at a small threshold.

To address this issue, we propose a saliency based simplification method, which uses the topological saliency at a fixed neighborhood size r in order to simplify features. Removal of a feature during simplification affects the saliency of the remaining features, and hence the saliency of these features is recalculated. As a side effect of this simplification process, we obtain a good segmentation from the resulting set of features. This is attributed to the fact that features that are close to each other merge early, as opposed to persistence based simplification, where no spatial information is used when merging an existing feature.

3.5. Feature similarity

The saliency plot of a single feature can be considered as its descriptor and used to find similarity between features. The behavior of the plots of various features also aids in studying the relationship between features. Consider the plots corresponding to similar peaks A and B , colored blue and orange, respectively, in Figure 2. We observe that the two plots corresponding to them have a similar behavior. Another observation is that, though peaks C , D , and E have persistence similar to peaks A and B , they differ in terms of the behavior of their corresponding plots. The neighborhood of A and B are similar, in the sense that they contain other peaks whose relative sizes are similar. In fact, the neighborhood is similar for different sizes of r . The same is not true for the peaks C , D , and E , and hence their plots are different. In order to automatically capture this similarity between features, we define the distance between two features as the area between their corresponding plots. Two features are said to be similar if the area between the corresponding plots is close to zero. Note that using persistence, it is not possible to distinguish between the features A , B , C , D , and E . In Section 5, we show that these observations indeed hold for many datasets.

4. Computing and Representing Features

The input to our experiments consists of piecewise linear functions defined over simplicial meshes and piecewise trilinear functions defined over structured grids. Two common topological abstractions used for representing scalar fields are the contour tree [35], and the Morse-Smale complex [36]. They identify the set of critical points in the input, and associate regions corresponding to these critical points [37]. We use both these data structures to identify and represent the set of features in the input.

4.1. Contour trees

The contour tree of a scalar function tracks the evolution of the topology of its level sets. It is computed by merging the

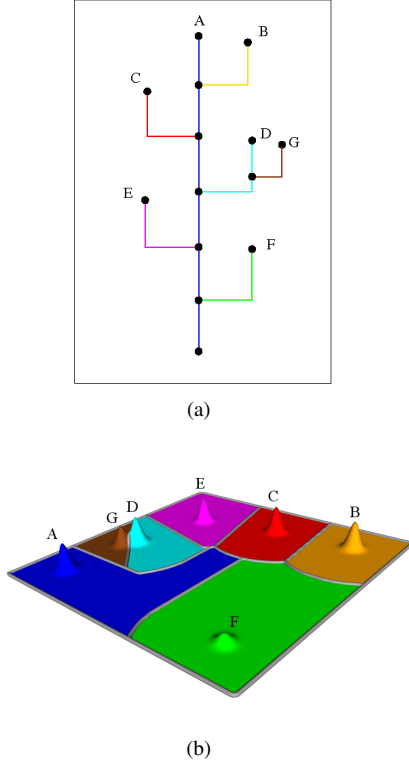


Figure 3: (a) The join tree of the terrain shown in Figure 1(a). Notice that the seven branches of the branch decomposition correspond to the seven peaks in the input. (b) Descending manifolds of the maxima partition the terrain.

join tree and the split tree of the input [35] and represented as a branch decomposition [38]. A join tree tracks the topology of the super-level sets of the input function, while the split tree tracks the topology of the sub-level sets. Since we are interested in features corresponding to only the set of maxima or the minima, it is sufficient to compute only the join tree or the split tree respectively. The join / split tree is computed using the union-find data structure to keep track of the connected components of the super-level set (or the sub-level set). This procedure also returns the set of saddle-extremum pairs that represent the topological features. A saddle-extremum pair corresponds to a branch in the join (or split) tree. The set of points from the input domain corresponding to each branch form the influence region of the corresponding extremum. Hence the branch decomposition of the join and split tree also induces a decomposition (segmentation) of the input domain into influence regions of features. Figure 3(a) shows the join tree for the terrain input from Figure 1(a). Each branch of this join tree is colored using the color of the corresponding region in Figure 2.

A simplification procedure is defined on the contour tree to remove noise from the input. Each step of the simplification removes the least significant leaf of the contour tree until a simplification threshold is reached. This segment in the input domain that corresponds to the removed branch merges with the segment corresponding to the parent branch. In all our experiments, unless otherwise mentioned, we use topological saliency as the significance measure for the simplification.

4.2. Morse-Smale complex

The Morse-Smale (MS) complex of a scalar function f partitions the domain of the function based on the gradient of f . The *gradient curves* of f are maximal curves on the domain whose tangent at each point aligns with the gradient of the function at that point. Gradient curves begin and end at critical points of f , referred to as their source and destination respectively. The MS complex of f partitions the domain based on the source and destination critical point of its gradient curves. The *combinatorial structure* of the MS complex is a graph whose nodes correspond to the critical points and an arc exists between two nodes if there is a gradient curve between the corresponding critical points and the indices of the critical points differ by one. The *descending manifold* of a critical point is the set of gradient curves that originate from it. The *ascending manifold* of a critical point is the set of gradient curves that terminate at it. The ascending manifold of a minimum defines its influence region, while the descending manifold of a maximum defines the influence region of the maximum. Figure 3(b) shows the partition of the terrain from Figure 1(a) based on the descending manifolds of the set of maxima.

Many algorithms to compute the MS complex have been proposed in the literature [14, 36, 39, 40, 41, 42, 43]. We use an implementation based on discrete Morse theory. Simplification of a pair of critical points p, q of index $i+1$ and i respectively, connected by a unique gradient curve is realized by locally modifying the function in the neighborhood of the gradient curve [44]. After simplification, p and q are no longer critical. The change in the combinatorial structure is realized by removing all arcs in the graph that are incident on either p or q and inserting an arc from every index- i critical point connected to p to every index- $i+1$ critical point connected to q . The ascending manifold of q merges with the ascending manifolds of all index- i critical points connected to p , and the descending manifold of p merges with the descending manifolds of all index- $i+1$ critical points connected to q . Arcs are scheduled for simplification in increasing order of the topological saliency of the extremum end point node. Arcs whose end points do not correspond to any feature have least priority and are simplified as soon as they appear.

4.3. Analysis

Computing the topological saliency for a given neighborhood size r requires the computation of persistence of all features, as well as the distances between all pairs of features. As mentioned in Section 2.2, computing the persistence of all features can be accomplished in $O(n \log n)$ time, where n is the number of triangles in the input. In case of a surface mesh, computing the geodesic distance between all pairs of features takes $O(tn \log n)$ time in the worst case, where t is the number of features. When the input is a structured grid, this can be accomplished in $O(t^2)$ time. The influence regions of various features are obtained using either the contour tree or the MS-complex, which also takes time polynomial in the size of the input.

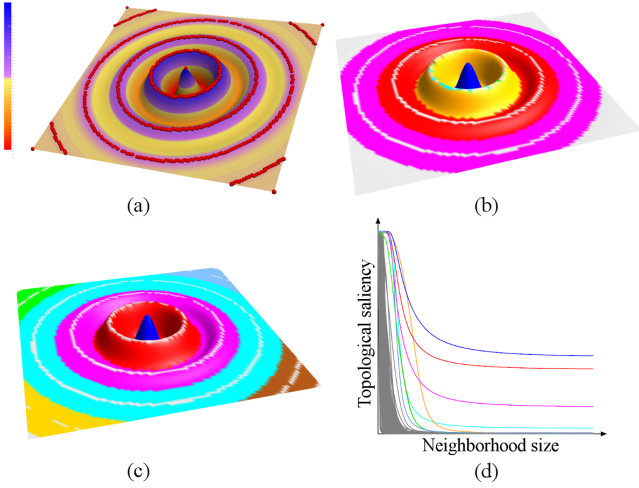


Figure 4: The *wgauss* dataset. (a) Input scalar field. The red spheres correspond to the set of maxima of the input. The large number of maxima cause the entire crest to be filled with red spheres. (b) The top 10 persistent regions. Note that small regions (cyan, brown) on the rim of the rings closer to the center are considered more important than the parts of the ring farther away. (c) The top 10 salient regions. Note that the regions corresponding to the ring farther away from the center is salient, even though it has low persistence. (d) The topological saliency plot highlighting the plots corresponding to the top 10 salient features.

4.4. Software

We first identify the set of all features in the input and compute the corresponding regions in the domain. We then simplify the input based on the simplification threshold provided as input. As mentioned in Section 3.4, removal of a feature during the simplification process required the recalculation of the saliency of the remaining features. In order to efficiently perform the simplification, we approximate the saliency of a feature by considering only the weights contributed by features within a $4r$ neighborhood. When a feature is removed, we only need to update the saliency of the features that remain within this neighborhood. Here, r is the neighborhood size at which saliency value is calculated. Note that in practice, the contribution of the features beyond the $4r$ neighborhood is negligible. Using this cut-off therefore gives a good approximation of the saliency value while also improving the efficiency of the computation.

We plot the topological saliency of the features that remain after simplification at varying neighborhood sizes ranging from 0 to the diameter D of the input to obtain the topological saliency plot. Each curve in the topological saliency plot maps to the influence region of the corresponding extremum. Our software displays the topological saliency plot along with the input. By selecting a curve in this plot, the user can view the corresponding region of the input. The user can order features based on topological saliency, by specifying a neighborhood size r . Our software also allows the user to view the various clusters that are identified by grouping similar features based on the topological saliency plot.

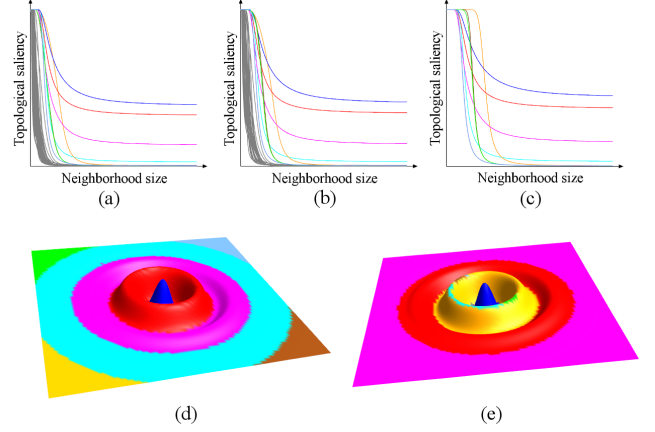


Figure 5: Effect of simplification on the saliency plot and the segmentation. (a) 100 features remaining. (b) 50 features remaining. (c) 8 features remaining. (d) Resulting segmentation when 8 features are remaining. (e) Segmentation after persistence-driven simplification. The outer rings merge whereas the inner ring is over-segmented. Note that the saliency plots of the top salient regions do not change significantly with different levels of simplification.

5. Applications

In this section, we first describe experiments on a synthetic dataset that demonstrates key properties of the topological saliency measure. Next, we describe three applications of topological saliency – identification of significant craters on the surface of Mars, locating breast tumors from optical tomography data, and detecting similar features in polygonal models. We use the height function as the scalar field for the experiment on synthetic data in Section 5.1, and the average geodesic distance (AGD) function [9] for surface meshes in Section 5.4. The scalar field is provided as input for all volumetric data and for the mars dataset used in Section 5.2.

5.1. Salient features of the *wgauss* dataset

The *wgauss* dataset is similar to an iso-surface of the popular Marschner-Lobb dataset [45], which is widely used to study reconstruction / interpolation filters for volume rendering. The dataset consists of concentric rings of crests and troughs, see Figure 4(a). The analytic form of *wgauss* is given by a 2D Gaussian distribution centered at the origin and weighted by the cosine of the distance from the center of the grid. The cosine term causes concentric crests and troughs whose number is controlled by the frequency. The amplitude of the crests and troughs are modulated by the 2D Gaussian. Sampling effects results in considerable noise in terms of critical points that are located on crests and troughs. Note that this effect exists in high resolution grids also due to the high curvature at the crest/trough.

We demonstrate the difference between topological saliency and topological persistence with the aim of extracting and ordering the rings corresponding to crests of the *wgauss* dataset. They are extracted as the influence regions of maxima, using the branch decomposition of the contour tree. The unsimplified input has a total of 211 maxima, most of which corresponds to

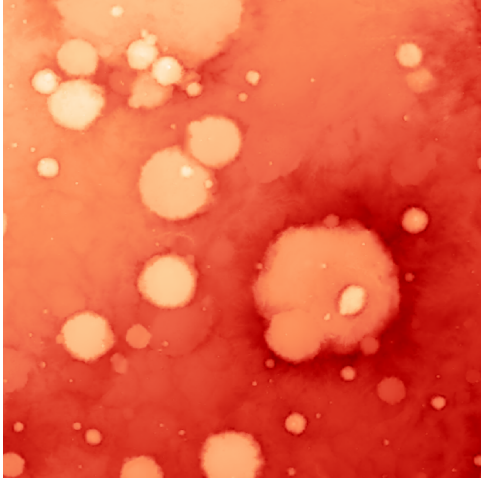


Figure 6: The region of Mars selected as input for our experiment.

noise in the input. Therefore, there exists more than one maxima on each ring, see Figure 4(a). In the ideal case, at least one maxima from a ring has to be considered important. Figure 4(b) shows the top 8 persistent features, and Figure 4(c) shows the top 8 salient features. Notice that the ring that is farthest away from the center is not considered important when using persistence. Instead, persistence identifies multiple maxima on rings closer to the center as important. These maxima are similar to the peak G in the motivating example shown in Figure 1.

Figure 5 shows the topological saliency plots at different levels of simplification. The plots corresponding to the top 8 salient features are highlighted in Figures 5(a), 5(b) and 5(c). Due to the use of the saliency metric as a simplification measure, the various rings are preserved during the simplification process, see Figure 5(d). This is not true when using persistence for simplification, since the regions corresponding to the outermost rings were simplified into the inner rings, see Figure 5(e). Also, note that the saliency plots of the remaining features are similar to the corresponding plots before simplification, indicating that in practice, the saliency measure of a salient feature is stable during simplification. This observation is true even for the real world datasets used in the following sections.

5.2. Identifying craters on the surface of Mars

We use topological saliency together with traditional persistence to identify significant craters on the surface of Mars. The input is an elevation map of a region on Mars. This data set was acquired by Mars Orbiter Laser Altimeter (MOLA) and is provided by NASA [46]. In such data, the domain scientists are interested in identifying craters, since many properties of celestial bodies can be inferred using this information. The mars data set has a resolution of 128 pixels per degree, and we chose a 1024×1024 region that was rich with craters, see Figure 6.

Craters correspond to valleys with low elevation. We use the set of minima in the input to represent features. The set of minima and their influence regions are identified by computing the MS complex of the input and extracting the ascending manifolds. We set the neighborhood size $r = 50$ and simplify the

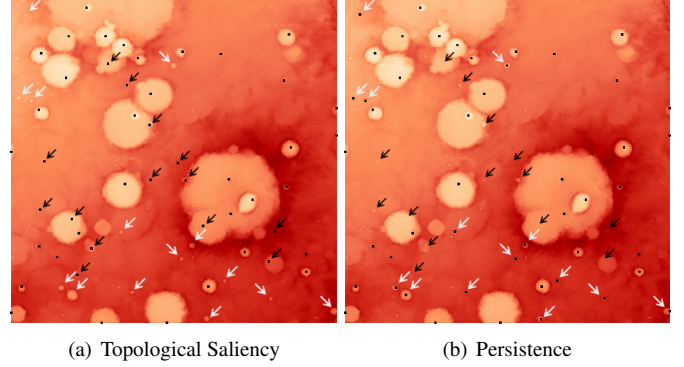


Figure 7: Top 50 craters as identified by topological saliency and persistence indicated by the black circles. Salient craters that are missed when using standard persistence are highlighted using black arrows, while persistent craters that are missed using saliency are highlighted using white arrows.

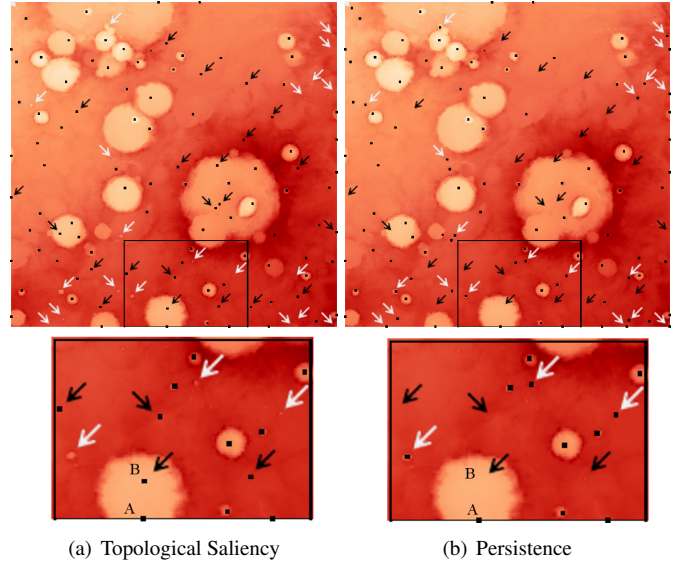


Figure 8: Top 100 craters as identified by topological saliency and persistence. The bottom figure focuses on a small rectangular region of the input.

input until 1000 features remain. This operation removes noise from the input. Figures 7(a) and 8(a) show the top 50 and top 100 craters, respectively, that are identified. Figures 7(b) and 8(b) show the top 50 and top 100 craters, respectively, as identified when using standard persistence. For the experiment using persistence, the simplification of the input to remove noise was also directed by persistence. Salient features that are missed by persistence are highlighted using black arrows, while features that have high persistence but are not salient are highlighted using white arrows. Note that, using topological saliency we are not only able to identify large and shallow craters, but can also identify small craters that are isolated. For example, consider the craters marked as A and B in Figure 8. Figure 9 shows the terrain corresponding to the two craters. Notice that crater B, which corresponds to a significant crater on mars, is not considered important when using persistence. Craters missed by our method correspond to those that are close to other high persistence craters, and therefore not considered salient.

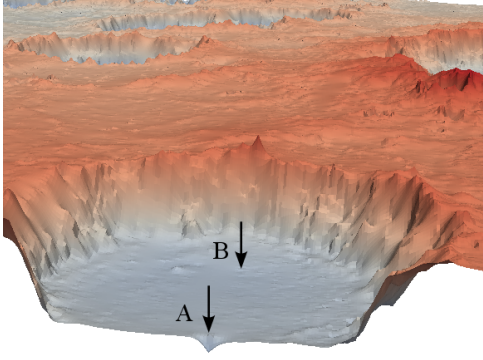


Figure 9: Even though crater *B* corresponds to a significant crater on the surface of Mars, it is not identified as significant when using traditional persistence.

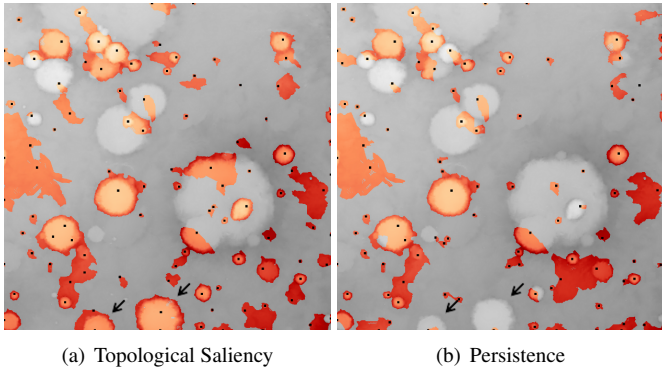


Figure 10: Segmented regions corresponding to the identified craters. Using saliency we obtain segments that covers entire craters. The arrows point to two such craters.

The regions corresponding to various craters are given by the influence regions corresponding to the minima. We use the segmentation obtained after simplification (either saliency based or persistence based), and refine the boundary of each crater using the snakes active contour model [47]. The result of this operation is shown in Figure 10, which compares the influence regions of the top salient and the top persistent craters. Note that many craters are completely covered by the influence regions of the set of minima when using topological saliency. As mentioned earlier in Section 3.4, this is attributed to the fact that small and close-by regions get simplified and merged earlier, thus providing a good segmentation. We observe that setting $r = 50$ provides good results for this data set, while the results may vary when changing this neighborhood size. We are currently exploring methods to automatically identify a good value for r .

As mentioned in Section 5.1, the saliency plots and hence the saliency values of the most salient regions do not drastically change with simplification. To verify the same with real world data, we repeated the above experiment by simplifying the input containing around 19000 minima until 5000, 4000, 3000 and 2000 features, respectively, were remaining. While we noticed a minor change in the order of the top 50 salient features, the set of top 50 features did not significantly change. We computed the difference between the sets of top 50 salient features com-

puted using simplification thresholds of 1000 and 5000. The two sets differed at only three features. We observed that the ranks of the features that were replaced, as well as that of the new features were greater than 40. Also, their saliency values were within 0.05 of each other, indicating that the saliency of a salient region is stable during simplification.

5.3. Identifying breast tumors

Diffuse optical tomography is used as an adjunct imaging modality for breast and brain imaging to provide functional images. Non-ionizing near infrared (NIR) light with wavelength in the range of 600-1000 nm is the interrogating medium of choice [48, 49]. Typically, the NIR light is delivered and collected using fibre bundles at the boundary of tissue. These boundary measurements are used to reconstruct the internal distributions of optical absorption and scattering coefficients. The data is available as a tetrahedral mesh where the scattering coefficient at each vertex defines the input scalar function.

Figure 11(a) shows the volume rendering of two breast data sets that have a tumor. Features are represented by the set of maxima in the input. The branch decomposition of the join tree of the input is used to segment the volume. The topological saliency plots for the two data sets are shown in Figure 11(b). The persistence of the fibre bundles at the periphery of the volume, that are used to collect data, is higher than that of the tumor itself. Therefore, a persistence based ordering would identify one such fibre bundle as the most significant feature. Using an appropriate value for neighborhood size r (10% of the diagonal of the volume's bounding box for this experiment), shown in the saliency plot, we are able to identify and isolate the region corresponding to the tumor as the most salient feature, see Figure 11(c).

In order to test the stability of the topological saliency measure with respect to noise, we artificially induced noise in both the input scalar field and the location of the extrema of the breast data set. The scalar field was perturbed using a Gaussian with mean equal to zero and standard deviation equal to 1% of the function range. The locations of extrema were perturbed using a Gaussian with mean equal to zero and standard deviation equal to 1% of the diagonal of the volume's bounding box. We repeated the above experiment on the resulting noisy data sets. Figure 12 shows the saliency plots corresponding to data obtained by inducing different kinds of noise. Even though the tumor has low persistence, it still remains the most salient feature. Note that, these plots are visually similar to the ones shown in Figure 11. The area between the curves corresponding to the retained extrema before and after the addition of noise is equal to 0.01 on an average. Note that curves in the saliency plot are contained within a unit square (both topological saliency (y-axis) and neighborhood size (x-axis) are between 0 and 1). This value is small and essentially implies that topological saliency is indeed stable in the presence of noise.

5.4. Extracting similar features

We group similar features by analyzing the topological saliency plot. As mentioned in Section 3.5, we use the area

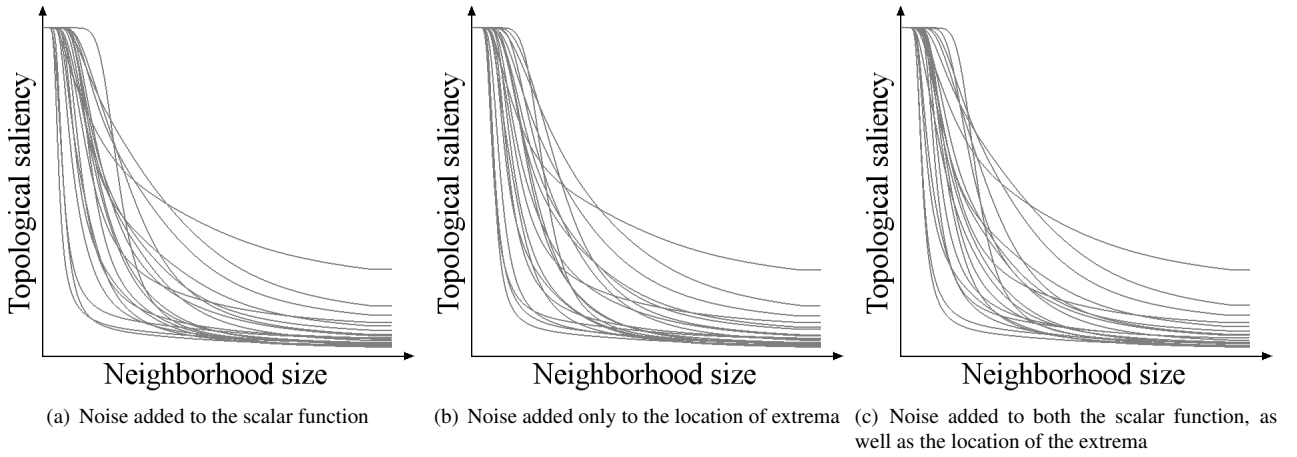


Figure 12: Stability of topological saliency in the presence of noise. Noise was added to the input scalar function, as well as the location of the extrema of the `breast` dataset. Note that the saliency plots obtained for the noisy data sets are similar, and the significant feature remains the same even in the presence of noise.

between two plots as a measure of similarity between the corresponding features. We then group similar features using single linkage clustering.

Figure 13 shows three surface meshes used as input in our experiments – `horse`, `human`, and `memento`. The AGD function defined on the mesh is used as its shape descriptor. We compute and plot the topological saliency for varying r . Figure 14 shows the topological saliency plot for these models. Similar features in these models are highlighted in the figure. The similar plots and the corresponding regions of the mesh are represented by the same color. Segments in the model are computed using the branch decomposition of the join tree. Notice that the legs of the `horse` model, shown in Figure 14(a), are grouped together in a single cluster. Ears of the horse are also grouped together. For the `human` model shown in Figure 14(b), the legs and hands form groups. An interesting point to note is that the plot corresponding to the head of the `human` does not cluster together with any of the other plots signifying that it is different from the other features. The torso and the hands of the three humanoid figures in the `memento` model form two groups, while the base of the model and the lone leg form separate groups.

As discussed in Section 3, the topological saliency plot can be used to distinguish between features that have similar persistence. For example, consider the features corresponding to the hands and the base of the `memento` model. Even though they have similar persistence, the fact that their saliency plots differ helps distinguish between them. Figure 15, which highlights regions that are grouped together when using traditional persistence, demonstrates that persistence alone is not sufficient to distinguish between features.

Even though we perform an initial simplification in order to remove noise, repeating the above experiment without this simplification step does not change the set of clusters obtained. This is because simplification only removes the features that are not significant, and does not affect the salient features. Also, since the behavior of the saliency plots of the salient regions

does not change, the clustering obtained remains the same. The saliency plots corresponding to the surface meshes obtained before simplification is shown in Figure 16. Note that the clustering of the main features are preserved even after simplification.

We repeated the above experiment using the heat kernel signature (HKS) function [50] as the scalar field instead of the AGD function. The HKS function provides a better segmentation of the features in the input. While we could again group similar features using the saliency plot, it was still not possible to distinguish between features using persistence.

Figure 17(a) shows a volume rendering of the `silicium` and `hydrogen atom` datasets. The topological saliency plot and the features that are grouped in the `silicium` dataset is shown in Figure 17(b). Notice that all the curves corresponding to individual atoms have a similar behavior, and are clustered together. The plots corresponding to the two spherical lobes form a group in the `hydrogen atom` dataset, see Figure 17(c). Other dissimilar branches correspond to the toroidal region and the outer envelope.

6. Conclusion

The problem of identifying important features forms an essential component in many of the topological methods used for geometry processing and scientific visualization. We address this problem by defining a notion of topological saliency that extends existing concepts by considering the spatial proximity of a feature to other features. Computing the topological saliency of all features for varying neighborhood sizes results in a topological saliency plot. We use this plot to define an order on features and show its utility in identifying important features that are possibly ignored by traditional approaches. We also analyze this plot to group similar features and demonstrate its application to various surface and volumetric inputs.

The measure of importance defined using topological saliency may not be effective when the absolute size of features is important. It also considers an isotropic neighborhood. It will

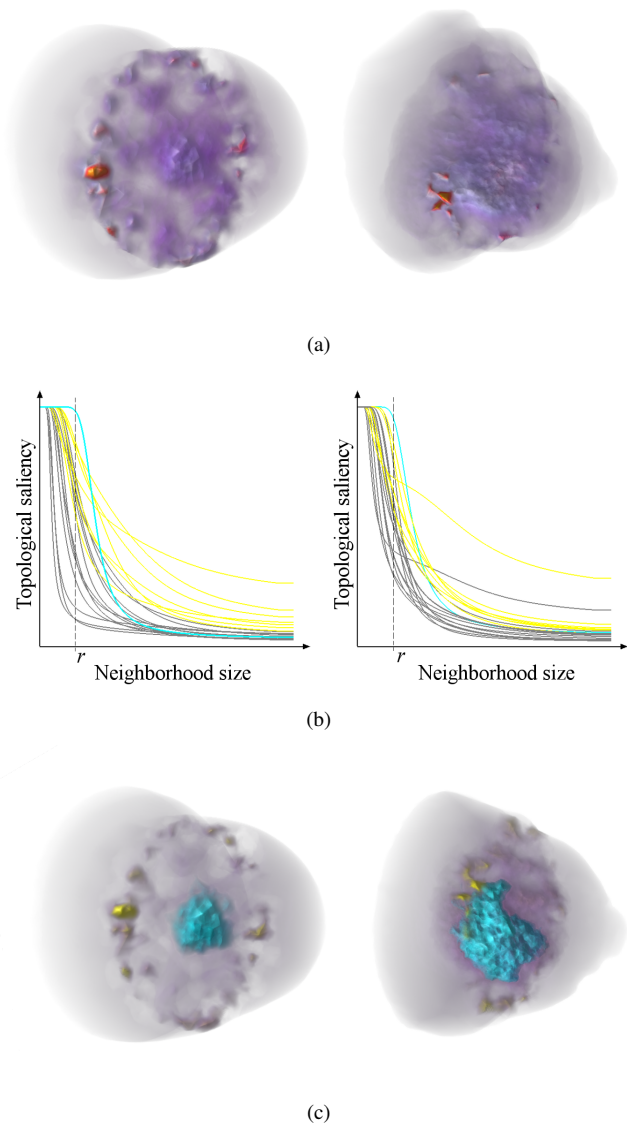


Figure 11: Identifying tumors using topological saliency. (a) Volume rendering of the two breast datasets. (b) The most salient feature of the dataset is highlighted in cyan in the topological saliency plot. The plots corresponding to the sensors are colored yellow. (c) The most salient feature corresponds to the region containing the tumor. The volume rendering highlights the most salient feature.

therefore be interesting to consider alternate methods of defining the importance of a feature based on its topological saliency that gives preference to certain directions. We could also potentially improve the sensitivity of the proposed measure by representing features using regions instead of critical points. While we observe that topological saliency is stable with respect to both simplification and noise in practice, it will be interesting to either theoretically prove the same, or to identify conditions under which this measure is stable.

We envision the use of topological saliency together with established measures like topological persistence for various feature detection applications. We believe that the notion of topological saliency will be useful for higher dimensional data, where explicit visualization of the data becomes difficult. The

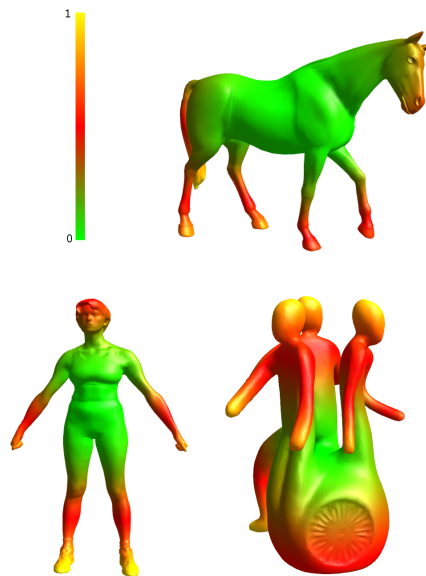


Figure 13: The horse, human and memento models used in the experiments with the shape descriptor function mapped to color.

topological saliency plot enables the representation of relative importance of features. This could potentially lead to a good user interface for exploring feature rich data.

Acknowledgements Harish Doraiswamy was supported by Microsoft Corporation and Microsoft Research India under the Microsoft Research India PhD Fellowship Award. This work was partially supported by the Department of Science and Technology, India, under Grant SR/S3/EECE/0086/2012 and the DST Center for Mathematical Biology, IISc, under Grant SR/S4/MS:419/07, and partially supported by the National Science Foundation (NSF) of the United States under grants CCF-0747082, DBI-0750891, and CCF-1048983.

References

- [1] Edelsbrunner H, Letscher D, Zomorodian, A. Topological persistence and simplification. *Disc Comput Geom* 2002;28(4):511–33.
- [2] Guskov I, Wood Z. Topological noise removal. In: *Proc. Graphics Interface*. 2001, p. 19–26.
- [3] Takahashi S, Nielson GM, Takeshima Y, Fujishiro I. Topological volume skeletonization using adaptive tetrahedralization. In: *GMP '04: Proceedings of the Geometric Modeling and Processing 2004*. Washington, DC, USA: IEEE Computer Society. ISBN 0-7695-2078-2; 2004, p. 227.
- [4] Wood Z, Hoppe H, Desbrun M, Schröder P. Removing excess topology from isosurfaces. *ACM Trans Graph* 2004;23(2):190–208.
- [5] Hétry F, Attali D. Topological quadrangulations of closed triangulated surfaces using the Reeb graph. *Graph Models* 2003;65(1-3):131–48.
- [6] Mortara M, Patané G. Affine-invariant skeleton of 3d shapes. In: *SMI '02: Proceedings of the Shape Modeling International 2002 (SMI'02)*. Washington, DC, USA: IEEE Computer Society. ISBN 0-7695-1546-0; 2002, p. 245.
- [7] Zhang E, Mischaikow K, Turk G. Feature-based surface parameterization and texture mapping. *ACM Trans Graph* 2005;24(1):1–27.
- [8] Dey T, Li K, Luo C, Ranjan P, Safa I, Wang Y. Persistent heat signature for pose-oblivious matching of incomplete models. *Computer Graphics Forum* 2010;25:1545–54.
- [9] Hilaga M, Shinagawa Y, Kohmura T, Kunii TL. Topology matching for fully automatic similarity estimation of 3d shapes. In: *Proc. SIGGRAPH*. 2001, p. 203–12.

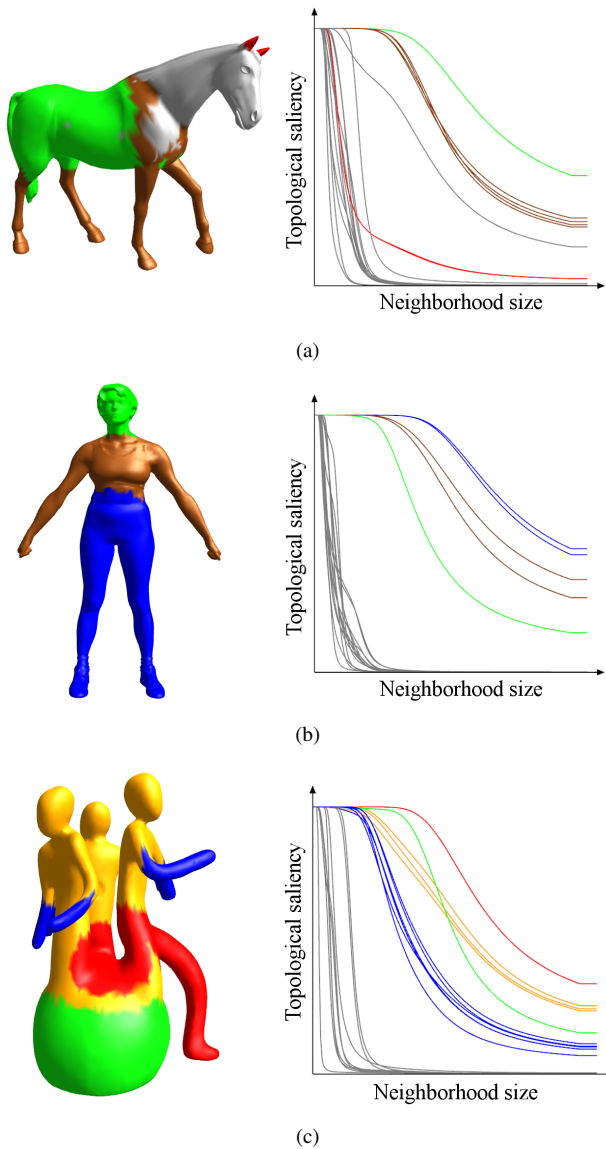


Figure 14: Topological saliency plots for three surface meshes. Different features of the input surface like arms and legs are grouped together based on the similarity between their saliency plot.

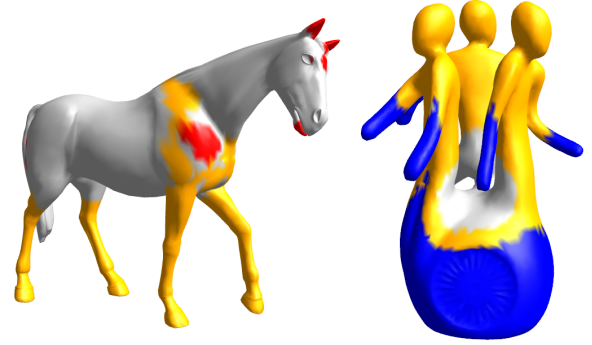


Figure 15: Using persistence to group features. Note that features having similar persistence need not correspond to similar features. Notice that the ears the horse are grouped with its jaw and a few small patches on its face and body. Similarly, the hands and the base of the memento model are grouped together.

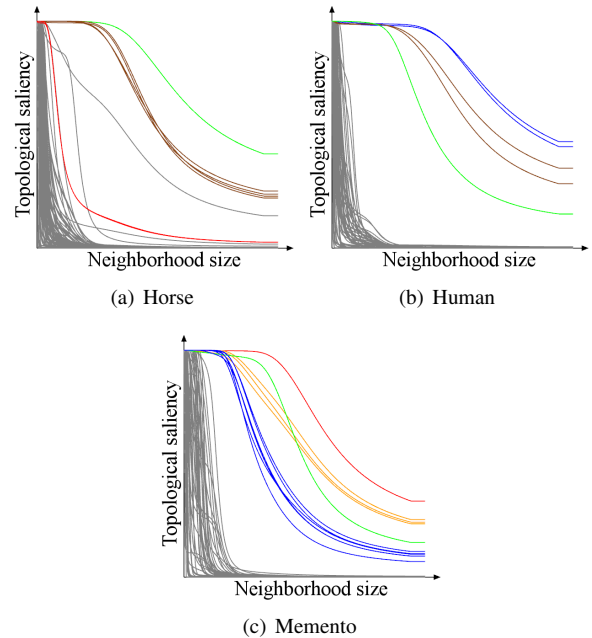


Figure 16: Clustering of similar features obtained using the saliency plot of the unsimplified surface mesh. Note that the main clusters obtained does not change.

- [10] Fujishiro I, Takeshima Y, Azuma T, Takahashi S. Volume data mining using 3d field topology analysis. *IEEE Computer Graphics and Applications* 2000;20:46–51.
- [11] Takahashi S, Takeshima Y, Fujishiro I. Topological volume skeletonization and its application to transfer function design. *Graphical Models* 2004;66(1):24–49.
- [12] Weber GH, Dillard SE, Carr H, Pascucci V, Hamann B. Topology-controlled volume rendering. *IEEE Trans Vis Comput Graph* 2007;13(2):330–41.
- [13] Zhou J, Takatsuka M. Automatic transfer function generation using contour tree controlled residue flow model and color harmonics. *IEEE Transactions on Visualization and Computer Graphics* 2009;15(6):1481–8.
- [14] Bremer PT, Edelsbrunner H, Hamann B, Pascucci V. A topological hierarchy for functions on triangulated surfaces. *IEEE Transactions on Visualization and Computer Graphics* 2004;10(4):385–96.
- [15] Gyulassy A, Natarajan V, Pascucci V, Bremer PT, Hamann B. A topological approach to simplification of three-dimensional scalar fields. *IEEE Transactions on Visualization and Computer Graphics (special issue IEEE Visualization 2005)* 2006;:474–84.

- [16] Koch C, Ullman S. Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology* 1985;4:219–27.
- [17] Milanese R, Wechsler H, Gil S, Bost J, Pun T. Integration of bottom-up and top-down cues for visual attention using non-linear relaxation. In: *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*. 1994, p. 781–5.
- [18] Tsotsos JK, Culhane SM, Wai WYK, Lai Y, Davis N, Nuflo F. Modeling visual attention via selective tuning. *Artif Intell* 1995;78:507–45.
- [19] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans Pattern Anal Mach Intell* 1998;20:1254–9.
- [20] Rosenholtz R. A simple saliency model predicts a number of motion popout phenomena. *Vision Research* 1999;39:3157–63.
- [21] Lee CH, Varshney A, Jacobs DW. Mesh saliency. *ACM Trans Graph* 2005;24:659–66.
- [22] Carr H, Snoeyink J, van de Panne M. Simplifying flexible isosurfaces using local geometric measures. In: *Proc. IEEE Conf. Visualization*. 2004, p. 497–504.

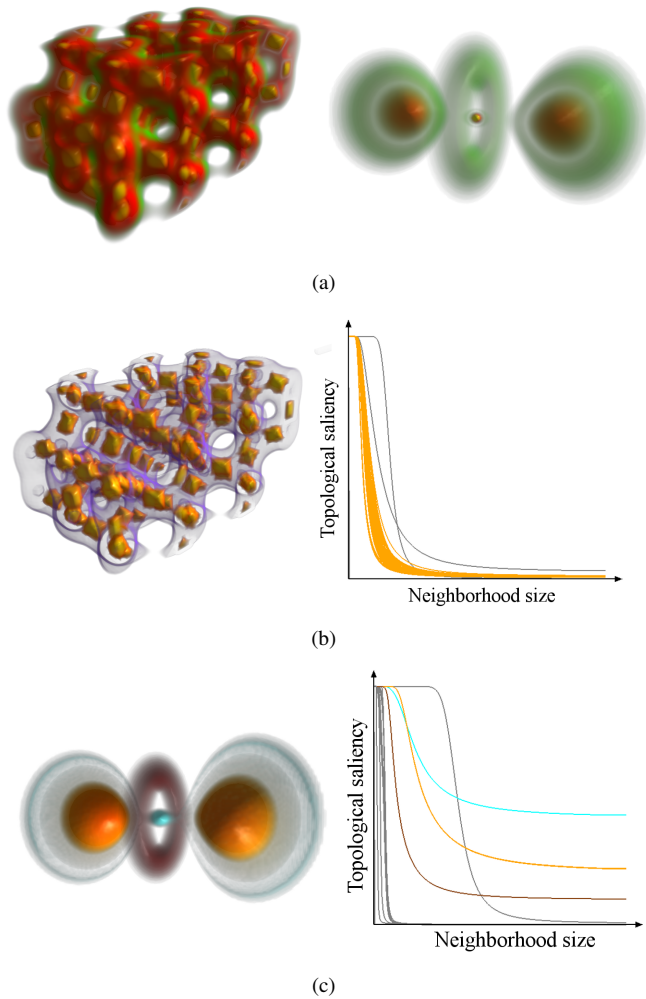


Figure 17: The topological saliency plots for the the `silicium` and the `hydrogen atom` datasets. All the atoms of `silicium`, and the two spherical lobes of the `hydrogen` data are grouped together using the saliency plot.

[23] Harvey W, Wang Y. Topological landscape ensembles for visualization of scalar-valued functions. *Computer Graphics Forum* 2010;29:993–1002.

[24] Weber G, Bremer PT, Pascucci V. Topological landscapes: A terrain metaphor for scientific data. *IEEE Transactions on Visualization and Computer Graphics* 2007;13:1416–23.

[25] Agarwal PK, Edelsbrunner H, Harer J, Wang Y. Extreme elevation on a 2-manifold. *Disc Comput Geom* 2006;36(4):553–72.

[26] Zomorodian A, Carlsson G. Localized homology. *Comput Geom Theory Appl* 2008;41:126–48.

[27] Reininghaus J, Kotava N, Guenther D, Kasten J, Hagen H, Hotz I. A scale space based persistence measure for critical points in 2d scalar fields. *IEEE Transactions on Visualization and Computer Graphics* 2011;17:2045–52.

[28] Chen C, Edelsbrunner H. Diffusion runs low on persistence fast. In: *Proceedings of the 2011 International Conference on Computer Vision*. 2011, p. 423–30.

[29] Edelsbrunner H, Harer J. *Computational Topology: An Introduction*. Amer. Math. Soc., Providence, Rhode Island; 2009.

[30] Hatcher A. *Algebraic Topology*. New York: Cambridge U. Press; 2002.

[31] Milnor J. *Morse Theory*. New Jersey: Princeton Univ. Press; 1963.

[32] Cole-McLaughlin K, Edelsbrunner H, Harer J, Natarajan V, Pascucci V. Loops in Reeb graphs of 2-manifolds. *Disc Comput Geom* 2004;32(2):231–44.

[33] Edelsbrunner H. *Geometry and Topology for Mesh Generation*. England: Cambridge Univ. Press; 2001.

[34] Edelsbrunner H, Harer J. Persistent homology — a survey. In: Goodman JE, Pach J, Pollack R, editors. *Surveys on Discrete and Computational Geometry. Twenty Years Later*. Amer. Math. Soc., Providence, Rhode Island; 2008, p. 257–82. *Contemporary Mathematics* 453.

[35] Carr H, Snoeyink J, Axen U. Computing contour trees in all dimensions. *Comput Geom Theory Appl* 2003;24(2):75–94.

[36] Edelsbrunner H, Harer J, Zomorodian A. Hierarchical Morse-Smale complexes for piecewise linear 2-manifolds. *Disc Comput Geom* 2003;30(1):87–107.

[37] Biasotti S, De Floriani L, Falcidieno B, Frosini P, Giorgi D, Landi C, et al. Describing shapes by geometrical-topological properties of real functions. *ACM Comput Surv* 2008;40(4):12:1–12:87.

[38] Pascucci V, Cole-McLaughlin K, Scorzelli G. The TOPORRERY: computation and presentation of multi-resolution topology. In: *Mathematical Foundations of Scientific Visualization, Computer Graphics, and Massive Data Exploration*. Mathematics and Visualization; Springer; 2009, p. 19–40.

[39] Edelsbrunner H, Harer J, Natarajan V, Pascucci V. Morse-Smale complexes for piecewise linear 3-manifolds. In: *Proc. Symp. Comput. Geom.* 2003, p. 361–70.

[40] Cazals F, Chazal F, Lewiner T. Molecular shape analysis based upon the Morse-Smale complex and the Connolly function. In: *Proc. 19th Ann. ACM Sympos. Comput. Geom.* 2003, p. 351–60.

[41] Gyulassy A, Bremer PT, Pascucci V, Hamann B. A practical approach to Morse-Smale complex computation: scalability and generality. *IEEE Transactions on Visualization and Computer Graphics* 2008;14(6):1619–26.

[42] Robins V, Wood PJ, Sheppard AP. Theory and algorithms for constructing discrete morse complexes from grayscale digital images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2011;99(PrePrints).

[43] Shivashankar N, Maadaswamy S, Natarajan V. Parallel computation of 2D Morse-Smale complexes. *IEEE Transactions on Visualization and Computer Graphics* 2011;99(PrePrints).

[44] Matsumoto Y. *An Introduction to Morse Theory*. Amer. Math. Soc.; 2002. Translated from Japanese by K. Hudson and M. Saito.

[45] Marschner SR, Lobb RJ. An evaluation of reconstruction filters for volume rendering. In: *Proc. IEEE Conf. Visualization*. 1994, p. 100–7.

[46] The Geosciences Node of NASA's Planetary Data System. ??? URL <http://pds-geosciences.wustl.edu/>.

[47] Kass M, Witkin A, Terzopoulos D. Snakes: Active contour models. *INTERNATIONAL JOURNAL OF COMPUTER VISION* 1988;1(4):321–31.

[48] Boas DA, Brooks DH, Miller EL, DiMarzio CA, Kilmer M, Gaudette RJ, et al. Imaging the body with diffuse optical tomography. *IEEE Signal Processing Magazine* 2001;18(6):57–75.

[49] Gibson A, Hebden JC, Arridge SR. Recent advances in diffuse optical tomography. *Physics in Medicine and Biology* 2005;50:R1–.

[50] Sun J, Ovsjanikov M, Guibas L. A concise and provably informative multi-scale signature based on heat diffusion. In: *Proceedings of the Symposium on Geometry Processing. SGP '09; Aire-la-Ville, Switzerland, Switzerland: Eurographics Association; 2009, p. 1383–92.*