Persistence modules, shape description, and completeness

F. Cagliari[1], M. Ferri[1,3], L. Gualandri[1], and C. Landi[2,3]

[1] Dip. di Matematica, Univ. di Bologna, Italy
{francesca.cagliari,massimo.ferri,luciano.gualandri}@unibo.it,
[2] Dip. di Scienze e Metodi dell'Ingegneria, Univ. di Modena e Reggio Emilia, Italy
claudia.landi@unimore.it,
[3] ARCES, Univ. di Bologna, Italy

Abstract. Persistence modules are algebraic constructs that can be used to describe the shape of an object starting from a geometric representation of it. As shape descriptors, persistence modules are not complete, that is they may not distinguish non-equivalent shapes. In this paper we show that one reason for this is that homomorphisms between persistence modules forget the geometric nature of the problem. Therefore we introduce geometric homomorphisms between persistence modules, and show that in some cases they perform better. A combinatorial structure, the H_0-tree, is shown to be an invariant for geometric isomorphism classes in the case of persistence modules obtained through the 0th persistent homology functor.

Keywords: geometric homomorphism; rank invariant; H_0-tree

1 Introduction

The shape description problem is at the core of many shape recognition methods used in computer vision and computer graphics. It is based on the fundamental idea of using compact representations of shapes, namely, shape descriptors, to analyze, understand, and compare objects [17]. Roughly speaking, a shape descriptor is complete when any two different shapes have two different descriptions.

In this paper we consider persistence modules as shape descriptors, focusing on the completeness problem, that is the problem of deciding whether persistence modules are able to discriminate between different shapes. We highlight some differences that arise in this respect depending on whether we work in a purely algebraic setting or we also keep memory of the underlying geometric setting to define the set of homomorphisms between persistence modules.

Approaching the shape description problem by persistence, one usually models the shape of an object as a geometric pair (X, f), where X is a geometric representation of the object under study (e.g., a manifold or a triangular mesh), and f is a function on X measuring some shape property of the object (e.g., curvature, height, distance from a fixed point). Two objects are considered to

have the same shape whenever the pairs (X, f) and (X', f') modeling them are isomorphic in a suitable category $\mathcal{C}$ [12]. However, since answering the isomorphism question in $\mathcal{C}$ is not an easy problem, one is usually satisfied with moving to persistence modules via the persistent homology functor, and studying the isomorphism question for persistence modules [18, 4]. In this way the shape descriptors that are actually used for comparison are persistence modules.

The rationale behind this approach is that the persistent homology functor does not change the isomorphism classes. Moreover, the isomorphism question for persistence modules is easier than for objects in $\mathcal{C}$. Sometimes one simplifies further the problem by only considering as shape descriptors invariants of isomorphism classes of persistence modules, such as barcodes [5], or size functions [7].

The completeness problem for persistence modules is studied in [11] for the case of curves. The authors show that two different shapes can have isomorphic persistence modules (i.e. the persistent homology functor forgets some relevant geometric features of the original shape), and prove that completeness can be achieved by increasing the number of components of the measuring function f.

In this paper, after reporting the basic definitions about persistence in Sect. 2, in Sect. 3 we present an example showing that, in some cases, persistence modules cannot discriminate non-equivalent shapes because in the category of persistence modules there are homomorphisms that are purely algebraic. More precisely, they do not reflect geometric transformations between the original shapes. For this reason, in Sect. 4, we introduce the notion of geometric homomorphisms between persistence modules and show that in some case they perform better than algebraic homomorphisms. Finally, in Sect. 5, we consider the problem of completeness for invariants. As for invariants with respect to algebraic isomorphisms, we review some results about the rank invariant from [2]. In the case of invariants with respect to geometric isomorphisms, we present an invariant in the case of 0th homology, the H_0-tree. We end the paper with a brief list of open questions in Sect. 6.

2   Background on persistence

2.1   The geometric approach to persistence

According to the spirit of the original persistence papers [10, 1], one has some category $\mathcal{C}$ of interest of geometric nature, and a functor F from that category to the category n-filt of n-filtrations. One then studies and works with the functor $H_i \circ F$, H_i being the ordinary ith homology functor. This composite functor is generally called an *ith persistent homology functor*.

In the most simple case, fixed $n \in \mathbb{N}$, $\mathcal{C} = \mathcal{C}(n)$ is the category defined as follows [15]:

1. Objects of $\mathcal{C}$ are pairs (X, f), where X is a topological space and $f = (f_1, \ldots, f_n) \colon X \to \mathbb{R}^n$ is a continuous function.

2. If $(X, f), (X', f') \in obj(\mathcal{C})$, then the set of morphisms of $\mathcal{C}$ from (X, f) to (X', f') is the set of continuous functions $\gamma: X \to X'$ such that $f(x) \geq f'(\gamma(x))$ for all $x \in X$ (with the convention that $u = (u_i) \leq v = (v_i)$ in $\mathbb{R}^n$ means $u_i \leq v_i$ for all i).

We observe that an isomorphism in $\mathcal{C}$ is a homeomorphism mapping level sets into level sets. Therefore we obtain the following easy remark.

Proposition 1. *If two pairs (X, f) and (X', f') are isomorphic in $\mathcal{C}$ then their natural pseudo-distance vanishes:*

$$\delta((X, f), (X', f')) := \inf_{\gamma} \sup_{x \in X} \|f(x) - f(\gamma(x))\|_\infty = 0,$$

where γ varies in all possible homeomorphisms between X and X'.

More details on the natural pseudo-distance and its relationship with persistence can be found, e.g., in [6, 8]. We remark that the converse of Proposition 1 in general is false, although it can be true in certain cases (cf., e.g., [3]).

For $(X, f) \in obj(\mathcal{C})$, and for $u = (u_i) \in \mathbb{R}^n$, let $X_u = \cap_{i=1}^n f_i^{-1}((-\infty, u_i])$. If $u \leq v \in \mathbb{R}^n$, then there is an inclusion $i_X(u, v): X_u \hookrightarrow X_v$. Thus the collection $\{X_r\}_{r \in \mathbb{R}^n}$ is an n-filtration of X. If $(X, f), (X', f') \in obj(\mathcal{C})$, and $\gamma: X \to X'$ is a morphism from (X, f) to (X', f'), then the restriction of γ to X_u, denoted by γ_u, maps X_u to X'_u, for all $u \in \mathbb{R}^n$. Moreover, for all $u \leq v \in \mathbb{R}^n$, $\gamma_v \circ i_X(u, v) = i_{X'}(u, v) \circ \gamma_u$. Thus the collection $\{\gamma_r\}_{r \in \mathbb{R}^n}$ is a morphism of n-filtrations. The functor $F: \mathcal{C} \to n$-filt which maps (X, f) to $\{X_r\}_{r \in \mathbb{R}^n}$ and γ to $\{\gamma_r\}_{r \in \mathbb{R}^n}$ is called the *sublevelset filtration functor*.

2.2 The algebraic approach to persistence

In [2, 18], the authors showed that persistence can be defined at algebraic level directly, without the need for an underlying topological setting. More precisely, they introduced the concept of a *persistence module* $\mathbf{M}$ as the one of a family $\{M_u\}_{u \in \mathbb{R}^n}$ of vector spaces (or modules over the same commutative ring) together with a family of homomorphisms $\{\iota_M(u, v): M_u \to M_v\}_{u \leq v \in \mathbb{R}^n}$ such that $\iota_M(u, w) = \iota_M(v, w) \circ \iota_M(u, v)$ and $\iota_M(u, u) = \mathrm{id}_{M_u}$ for all $u \leq v \leq w \in \mathbb{R}^n$.

Given two persistence modules $\mathbf{M}$ and $\mathbf{N}$, the set of homomorphisms from $\mathbf{M}$ to $\mathbf{N}$ consists of collections of homomorphisms of vector spaces $\mathbf{h} = \{h_u: M_u \to N_u\}_{u \in \mathbb{R}^n}$ such that $\iota_N(u, v) \circ h_u = h_v \circ \iota_M(u, v)$ for all $u \leq v \in \mathbb{R}^n$. Therefore, in a purely algebraic setting two persistence modules $\mathbf{M}$ and $\mathbf{N}$ are isomorphic if there exists a collection $\mathbf{h} = \{h_u: M_u \to N_u\}_{u \in \mathbb{R}^n}$ of isomorphisms of vector spaces such that $\iota_N(u, v) \circ h_u = h_v \circ \iota_M(u, v)$ for all $u \leq v \in \mathbb{R}^n$. We call $\mathbf{h}$ an *algebraic isomorphism of persistence modules*.

The category of persistence modules will be denoted by $\mathcal{M}$. Clearly, objects and homomorphisms of $\mathcal{M}$ can be constructed by applying the persistent homology functor to objects and morphisms of $\mathcal{C}$. It is known that, in the case of objects, the converse is also true, at least for finite persistence modules [2, Th. 2]. For this reason, in this paper we focus on morphisms rather than on objects of $\mathcal{M}$.

3 A preliminary example

Let us begin considering the following example. It shows that, even in the very basic case of curves endowed with simple Morse functions, we can find non-isomorphic pairs (X, f) and (X', f') in $\mathcal{C}$ taken by the persistent homology functor into algebraically isomorphic modules.

Example 1. Let (X, f) and (X', f'), with $X = X' = S^1$, be the two curves displayed in Figure 1. (X, f) and (X', f') are not isomorphic in $\mathcal{C}$. Indeed, an isomorphism between (X, f) and (X', f') necessarily takes a critical point of f to the critical point of f' at the same height, which is clearly impossible in this case. On the other hand, the persistence modules $\mathbf{M}$ and $\mathbf{N}$ obtained by applying the 0th persistent homology functor to the pairs (X, f) and (X', f'), respectively, are isomorphic in $\mathcal{M}$. To see this, it is sufficient to consider the diagram

$$
\begin{array}{ccc}
M_5 = <z_1, z_2, z_3 | z_1 = z_2 = z_3> & \xrightarrow{h_5} & N_5 = <z'_1, z'_2, z'_3 | z'_1 = z'_2 = z'_3> \\
\uparrow & & \uparrow \\
M_4 = <z_1, z_2, z_3 | z_1 = z_3> & \xrightarrow{h_4} & N_4 = <z'_1, z'_2, z'_3 | z'_2 = z'_3> \\
\uparrow & & \uparrow \\
M_3 = <z_1, z_2, z_3> & \xrightarrow{h_3} & N_3 = <z'_1, z'_2, z'_3> \\
\uparrow & & \uparrow \\
M_2 = <z_1, z_2> & \xrightarrow{h_2} & N_2 = <z'_1, z'_2> \\
\uparrow & & \uparrow \\
M_1 = <z_1> & \xrightarrow{h_1} & N_1 = <z'_1>,
\end{array}
\tag{1}
$$

where the vertical maps are induced by inclusions, and the horizontal maps are defined by setting $h_1(z_1) = z'_1$, $h_2(z_1) = z'_1$, $h_2(z_2) = z'_2$, $h_3(z_1) = z'_1$, $h_3(z_2) = z'_2$, $h_3(z_3) = z'_1 + z'_2 - z'_3$, $h_4(z_1) = z'_1$, $h_4(z_2) = z'_2$, $h_5(z_1) = z'_1$.
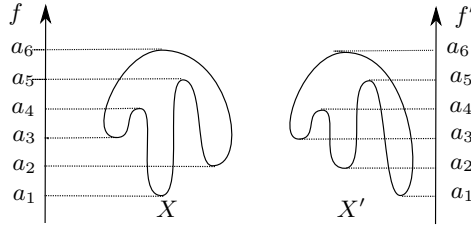

Fig. 1. Two curves not distinguishable by persistence modules in $\mathcal{M}$.

The previous example proves the following result, saying that algebraic isomorphism of persistence modules may not distinguish the shapes of two pairs.

Proposition 2. *The persistent homology functor $H_i \circ F \colon \mathcal{C} \to \mathcal{M}$ does not reflect isomorphisms (i.e., $H_i \circ F(X, f)$ isomorphic to $H_i \circ F(X', f')$ in $\mathcal{M}$ does not imply (X, f) isomorphic to (X', f') in $\mathcal{C}$).*

This prompts a new definition that will be given in the next section.

4 Geometric homomorphisms

In this section we focus on particular homomorphisms between persistence modules that we call *geometric* because they are the image of a morphism in $\mathcal{C}$. In Proposition 3 we prove that, restricting to geometric homomorphisms, we can distinguish the curves of Example 1.

Definition 1. *Let (X, f) and (X', f') be two objects in $\mathcal{C}$. A homomorphism (resp., isomorphism) $\mathbf{h}$ between the persistence modules $H_i \circ F(X, f)$ and $H_i \circ F(X', f')$ is called a* geometric homomorphism *(resp.* geometric isomorphism*) if it belongs to the image of the persistent homology functor.*

We now prove that no isomorphism between the persistence modules $\bigoplus_{i \in \mathbb{Z}} H_i \circ F(X, f)$ and $\bigoplus_{i \in \mathbb{Z}} H_i \circ F(X', f')$ of Example 1 belongs to the image of the persistent homology functor.

Proposition 3. *Let (X, f) and (X', f') be as in Example 1. No morphism γ between (X, f) and (X', f') is taken by $\bigoplus_{i \in \mathbb{Z}} H_i \circ F$ into an isomorphism of persistence modules.*

Proof. Let $\gamma \colon X \to X'$ be a continuous function such that $f(x) \geq f'(\gamma(x))$ for every $x \in X$ (i.e., γ is a morphism in $\mathcal{C}$). Let us assume that $H_i \circ F(\gamma)$ is an isomorphism for every $i \in \mathbb{Z}$. Hence, in particular, $H_1(\gamma) \colon H_1(X) \to H_1(X')$ is an isomorphism, and therefore the degree of γ is non-zero (recall that $X = X' = S^1$). It follows that $\gamma(X) = X'$.

Let a_i with $i = 1, \ldots, 6$ be as in Figure 1 and let p_i (resp. q_i) be the only critical point of f in $f^{-1}(a_i)$ (resp. of f' in $(f')^{-1}(a_i)$). Since $f(p_1) \geq f'(\gamma(p_1))$, necessarily $\gamma(p_1) = q_1$. Moreover, it must hold $\gamma(p_6) = q_6$. Indeed, since $\gamma(X) = X'$, there is some $p \in X$ such that $\gamma(p) = q_6$. Thus we get $f(p) \geq f(\gamma(p)) = a_6$, implying $p = p_6$. Using again $f(x) \geq f'(\gamma(x))$, we deduce that $f'(\gamma(p_2)) \leq a_2$. Hence, either $\gamma(p_2) = q_2$ or $\gamma(p_2)$ belongs to the arc of X' containing q_1 and staying under a_2.

Let us assume $\gamma(p_2) = q_2$. By considering the arcs in which p_1, p_6 (resp. q_1, q_6) split the curve X (resp. X'), since p_3 does not belong to the arc containing p_2, by continuity we get that $\gamma(p_3)$ does not belong to the arc containing $\gamma(p_2)$. Moreover, $\gamma(p_3)$ stays under a_3. Thus, the classes of $\gamma(p_1)$ and $\gamma(p_3)$ are homologous in $N_3 = H_0(X'_{a_3})$. Since the classes of p_1 and p_3 are not homologous in $M_3 = H_0(X_{a_3})$, we conclude that $H_0 \circ F(\gamma)$ is not an isomorphism.

Otherwise, if $\gamma(p_2)$ belongs to the arc of X' containing q_1 and staying under a_2, then $\gamma(p_1)$ and $\gamma(p_2)$ are homologous in $N_2 = H_0(X'_{a_2})$. Since the classes of p_1 and p_2 are not homologous in $M_2 = H_0(X_{a_2})$, we conclude that $H_0 \circ F(\gamma)$ is not an isomorphism in this case either, yielding the claim. $\qquad \square$

Thus, if we consider the subset of geometric isomorphisms we can distinguish the curves of Example 1. This seems to suggest that the image of the persistent homology functor is better suited for the aims of shape comparison than persistence modules.

5   Invariants

In this section we study invariants for isomorphism classes of persistence modules in the algebraic as well as in the geometric case.

5.1   Algebraic setting

Invariants for classes of persistence modules up to algebraic isomorphism have been thoroughly studied for cases $n = 1$ in [18] and $n > 1$ in [2]. The main invariant proposed is the *rank invariant*.

Definition 2. *Given a persistence module* $\mathbf{M}$ *consisting of vector spaces* $\{M_u\}_{u \in \mathbb{R}^n}$ *and homomorphisms* $\{\iota_M(u,v) \colon M_u \to M_v\}_{u \leq v \in \mathbb{R}^n}$, *its rank invariant is an integer-valued function* $\rho_{\mathbf{M}}$ *of two variables* $u \leq v \in \mathbb{R}^n$, *defined by* $\rho_{\mathbf{M}}(u,v) = \mathrm{rk}\left(\iota_M(u,v)\right)$.

In [18] the authors show that, for $n = 1$, the rank invariant is a complete invariant for algebraic isomorphism of persistence modules admitting a finite presentation (in terms of generators and relators). This means that any two such persistence modules are algebraically isomorphic if and only if their rank invariants coincide.

The analogous property for $n > 1$ is false, as the following example shows (see also [2]).

Example 2. Consider the bi-dimensional persistence modules $\mathbf{M}$ and $\mathbf{N}$, given by

where the row matrix displayed on each arrow represents the homomorphism between the modules connected by the arrow, with respect to the bases enclosed by angle brackets.
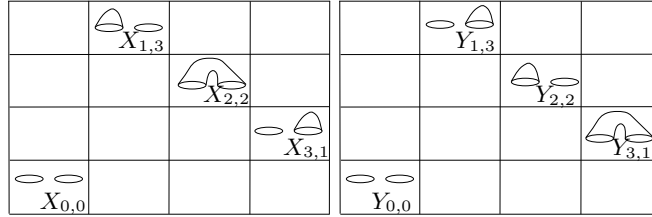
Fig. 2. Filtrations taken by the 1st homology functor into **M** (left) and **N** (right).

For instance, **M** and **N** can be obtained by applying the 1st homology functor with coefficients in $\mathbb{Z}_2$ to the 2-filtrations displayed in Fig. 2.

M and **N**, as persistence modules over $\mathbb{Z}_2$, are not isomorphic, although their rank invariants coincide.

5.2 Geometric setting

We now consider invariants for classes of persistence modules up to geometric isomorphism. We confine our treatment to the case $n = 1$, that is to scalar functions $f: X \to \mathbb{R}$. In this setting we can consider the so-called H_0-tree of f introduced in [1].

Trees are of widespread use in topology, either for invariant computations in a discrete setting (e.g., spanning trees for the fundamental group [13], and spanning forests for homology [16]) or as signatures in a continuous setting (e.g., contour trees [14] for domains of the plane, and merge trees [9] for arbitrary manifolds). The use we make of trees in this section is in the latter spirit.

H_0-trees can be defined on any topological space endowed with a (suitable) scalar function used to filter the space. Intuitively, the connected components of the sub-level sets of a function f, thanks to the inclusion relation, can be organized in a directed tree structure where node parenthood maps component inclusion. We shall prove that H_0-trees are invariant for geometric isomorphism classes of 0th homology.

Definition 3. *For a closed (i.e. compact and without boundary) connected manifold X and a simple Morse function $f: X \to \mathbb{R}$, the H_0-tree of f is a rooted binary tree labeled on the nodes defined as follows:*

- *the set of nodes is equal to the set of points of X such that for every sufficiently small real value $\varepsilon > 0$, the homomorphism induced by the inclusion $\iota(f(p) - \varepsilon, f(p) + \varepsilon): H_0(X_{f(p)-\varepsilon}) \to H_0(X_{f(p)+\varepsilon})$ is not an isomorphism;*
- *the label of a node p is equal to $f(p)$;*
- *p is a child of q if q has the lowest label among the nodes for which $f(p) < f(q)$ and $\iota(f(p), f(q)): H_0(X_{f(p)}) \to H_0(X_{f(q)})$ takes the class of p to that of q.*

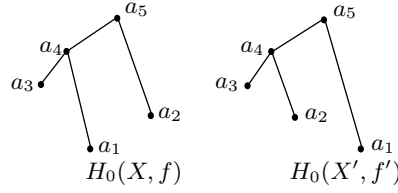The H_0-trees corresponding to the curves of Fig. 1 are displayed in Fig. 3.

Fig. 3. The H_0-trees associated with the two curves of Fig. 1.

Proposition 4. *Let X, X' be closed connected manifolds, and $f\colon X \to \mathbb{R}$, $f'\colon X' \to \mathbb{R}$ be simple Morse functions. If $\mathbf{h}$ is a geometric isomorphism between $\mathbf{M} = H_0 \circ F(X, f)$ and $\mathbf{N} = H_0 \circ F(X', f')$, then the H_0-trees of f and f', say T and T', are isomorphic as labeled trees.*

Proof. Since $\mathbf{h}$ is an isomorphism, there is a label preserving bijection σ between the set of nodes of T and that of T'. Let us see that σ also preserves the edges. Let p be a child of q in T. Then $\iota(f(p), f(q))$ sends the class of p into that of q. Since $\mathbf{h}$ is induced by a morphism $\gamma\colon X \to X'$, $\iota(f'(\gamma(p)), f'(\gamma(q)))$ sends the class of $\gamma(p)$ into that of $\gamma(q)$, and the class of $\gamma(p)$ (resp. $\gamma(q)$) coincides with that of $\sigma(p)$ (resp. $\sigma(q)$). Therefore $\sigma(p)$ is a child of $\sigma(q)$. $\qquad\square$

We remark that H_0-trees are not invariant for algebraic isomorphism classes of 0th persistence modules. Indeed, the curves of Example 1 have non-isomorphic H_0-trees whereas their persistence modules are algebraically isomorphic.

We can see that H_0-trees are not complete invariants for geometric isomorphism classes by taking the examples in Fig. 1 with $-f$ and $-f'$ instead of f and f'. However, in the case $X = S^1$, we can prove that a curve (X, f), $f\colon X \to \mathbb{R}$ being a simple Morse function, can be completely reconstructed up to f-preserving homeomorphisms, from the H_0-trees of f and $-f$.

6 Open questions

We think that further investigations on the subject presented here could tackle the following problems:

1. Is it true that the image of the persistent homology functor is a category?
2. Does a simple characterization of geometric homomorphisms exists?
3. Can H_0-trees be generalized so to obtain combinatorial structures providing invariants for geometric isomorphism classes for other homology degrees?

References

1. Cagliari, F., Ferri, M., Pozzi, P.: Size functions from the categorical viewpoint. Acta Appl. Math. 67, 225–235 (2001)

2. Carlsson, G., Zomorodian, A.: The theory of multidimensional persistence. Discrete Comput. Geom. 42(1), 71–93 (2009)
3. Cerri, A., Di Fabio, B.: On certain optimal diffeomorphisms between closed curves. Tech. rep., Università di Bologna (2010), http://amsacta.cib.unibo.it/3010/
4. Chazal, F., Cohen-Steiner, D., Glisse, M., Guibas, L.J., Oudot, S.Y.: Proximity of persistence modules and their diagrams. In: Proc. SCG '09. pp. 237–246 (2009)
5. Collins, A., Zomorodian, A., Carlsson, G., Guibas, L.: A barcode shape descriptor for curve point cloud data. Computers and Graphics 28, 881–894 (2004)
6. d'Amico, M., Frosini, P., Landi, C.: Natural pseudo-distance and optimal matching between reduced size functions. Acta Appl. Math. 109, 527–554 (2010)
7. Dibos, F., Frosini, P., Pasquignon, D.: The use of size functions for comparison of shapes through differential invariants. J. Math. Imag. Vis. 21, 107–118 (2004)
8. Donatini, P., Frosini, P.: Natural pseudodistances between closed manifolds. Forum Math. 16(5), 695–715 (2004)
9. Edelsbrunner, H., Harer, J.: Computational Topology: An Introduction. American Mathematical Society (2009)
10. Edelsbrunner, H., Letscher, D., Zomorodian, A.: Topological persistence and simplification. Discrete Comput. Geom. 28(4), 511–533 (2002)
11. Frosini, P., Landi, C.: Uniqueness of models in persistent homology: the case of curves. Inverse Problems (to appear)
12. Frosini, P., Landi, C.: Size theory as a topological tool for computer vision. Patt. Recog. Image Anal. 9, 596–603 (1999)
13. Hatcher, A.: Algebraic topology. Cambridge University Press (2002)
14. van Kreveld, M., van Oostrum, R., Bajaj, C., Pascucci, V., Schikore, D.: Contour trees and small seed sets for isosurface traversal. In: Proc. 13th Annu. ACM Sympos. Comput. Geom. pp. 212–220 (1997)
15. Lesnick, M.: The optimality of the interleaving distance on multidimensional persistence modules (2011), http://arxiv.org/abs/1106.5305
16. Molina-Abril, H., Real, P.: Homological computation using spanning trees. In: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. Proceedings of the 14th Iberoamerican Conference on Pattern Recognition (CIARP '09). LNCS, vol. 5856, pp. 272–278. Springer-Verlag (2009)
17. Veltkamp, R.C., Hagendoorn, M.: Principles of Visual Information Retrieval, chap. State-of-the-art in shape matching, pp. 87–119. Springer-Verlag (2001)
18. Zomorodian, A., Carlsson, G.: Computing persistent homology. Discrete Comput. Geom. 33(2), 249–274 (2005)