

Quantitative measure of structural and geometric similarity of 3D morphologies

Maciej Komosinski Marek Kubiak

Poznan University of Technology
Institute of Computing Science
Piotrowo 2, 60-965 Poznan, Poland
fax +48 (61) 8771 525
Corresponding author:
maciej.komosinski@cs.put.poznan.pl

Abstract

This work describes a new heuristic algorithm that estimates structural and geometric similarity of three-dimensional morphologies. It is an extension to previously developed measure of similarity [25] that was only able to consider the structure of 3D constructs. Morphologies are modeled as graphs with vertices as points in a 3D space, and edges connecting these vertices. This model is very general, therefore the proposed algorithm can be applied in (and across) a number of disciplines including artificial life, evolutionary design, engineering, robotics, biology and chemistry. The primary areas of application of this fast numerical similarity measure are artificial life and evolutionary design, where great numbers of morphologies result from simulated evolutionary processes, and both structural and geometric aspects are significant. Geometry of 3D constructs (i.e., locations of body parts in space) is as important as the structure (i.e., connections of body parts), because both determine behavior of creatures or designs and their fitness in a particular environment. In this work both morphological aspects are incorporated in a single, highly discriminative measure of similarity.

Key Words: similarity; 3D morphology; structure; artificial life; evolution

1 Introduction

In the areas of artificial life and evolutionary robotics, just as in studies of real life, there is a need to compare individuals. They are called creatures, animats, agents or robots, and since they need to work in the world we live in, they are made three-dimensional constructs. They are also expected to handle all kinds of interaction with the environment – just as animals.

Researchers that design or evolve such constructs face the problem of classifying them. Humans encountered this issue earlier with natural life forms, and found it

The final version of this paper appeared in *Complexity* 16(6):40–52, 2011. <http://dx.doi.org/10.1002/cplx.20367>

desirable to introduce some order among organisms – primarily using morphological, and recently, molecular information [17]. In technology, continuous increase of computer power facilitated generating constructs that are increasingly complex to the point where humans are unable to understand them in depth. Just like there are huge amounts of data that cannot be handled manually, there are also algorithms that can generate complex systems too sophisticated to be analyzed manually. More and more often computer-generated solutions can be tested, proved successful and human-competitive, but their design cannot be understood, or it would cost a lot of human time.

This situation is common in artificial life and evolutionary design. Fortunately, generated or designed creatures exist in a digital form, so in order to avoid tedious and manual work on analyzing and systematizing them, one can employ various kinds of automated procedures (tools). These tools must be first created though. Once they exist, they can also be used for real creatures as long as real creatures can be represented in the model supported by the tools. Such analysis tools may be therefore useful both for computer science and biology. Measures of symmetry [3, 20] and similarity [25] are examples of such tools that deal with physical body structure – similarity measure being one of the simplest, yet quite powerful. Quantitative measure of similarity allows to

- analyze structure of populations of individuals (e.g. diversity, convergence, etc.), facilitating better interpretation of experimental results,
- discover clusters in groups of individuals,
- reduce large sets or populations of individuals to small subsets of diverse representatives, thus reducing complexity and size of experimental data and making it more comprehensible,
- infer dendrograms (and hopefully, phylogenetic trees) based on morphological distances between individuals,
- introduce artificial niches, or species, by modifying fitness values in evolutionary optimization [12, 33],
- restrict crossing over so that only similar parents are involved, which reduces the risk of impaired offspring,
- test correlation between similarity and quality of individuals, determine global convexity of the solution space [21] and develop efficient distance-preserving crossover operators [46, 31].

The graph-based model of morphology, considered here in the artificial life context, is popular in many disciplines and can represent any kind of structures where connected vertices are located in a 3D space as shown on Fig. 1, see also Sects. 1.2, 2.1 and 4.2. In particular, this model of morphology can represent skeletal and circulatory systems, molecules and chemical structures, trusses in architecture and structural engineering [16], constructs made of a finite number of elements (e.g. robots), computer graphics objects and meshes, and geolocated graphs or communication networks. Thus a compatible measure of similarity could also be used to compare constructs coming from different disciplines – e.g., to find a truss whose structure resembles most a given crystal structure, to direct simulated evolution to evolve a morphology that resembles (or differs from) a specific stick insect from the order of Phasmatodea, etc.

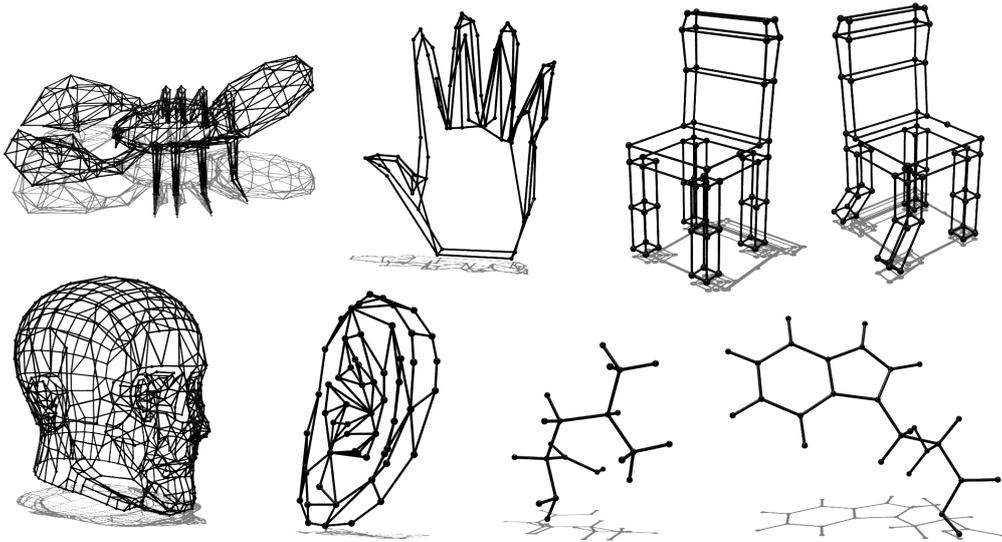


Figure 1: Sample disciplines where 3D objects are often represented as graphs: computer graphics, anatomy, design and engineering, chemistry.

1.1 Similarity and geometric properties of morphology

In [25] we introduced a heuristic, numerical measure of morphological similarity for 3D constructs. The measure, however, was only aware of structural properties of compared bodies, as illustrated in Fig. 2. Morphologies with identical structure and only differing in geometry were considered perfectly identical. However, since these constructs exist in a 3D world where body geometry plays an important role because of numerous physical interactions (e.g. touch) and physical properties of the environment (e.g. gravity, friction), it is necessary to consider body geometry as well.

In real and artificial life, geometric aspects determine various functions that an organism can perform (e.g. abilities of efficient locomotion [3], jumping, covering a large area or space, having a tall body, etc.). Body geometry is as important as body structure, and both properties influence efficiency – consider for example the ability to move fast in land and water environments – obviously, geometries of successful creatures differ in the two cases. Another example is a setting where successful creatures need tall bodies so that their top parts are high above the ground. Thus they need a stable base, and the geometry of the base part determines stability of individuals. Without considering geometry, we may be unable to capture any dissimilarity, as illustrated on Fig. 3.

For complex morphologies, it is more likely that any pair will be structurally different. The simpler the bodies and the smaller the number of their elements, the higher the probability of encountering identical structures. In this context, introducing geometry into the measure of similarity improves properties of the measure and increases *resolution* of the measure so that it is more sensitive to differences in morphologies and can discriminate constructs that were previously considered identical. This is especially important when the similarity estimation is applied to large numbers of individuals and no assumptions can be made regarding structural difference of their bodies.

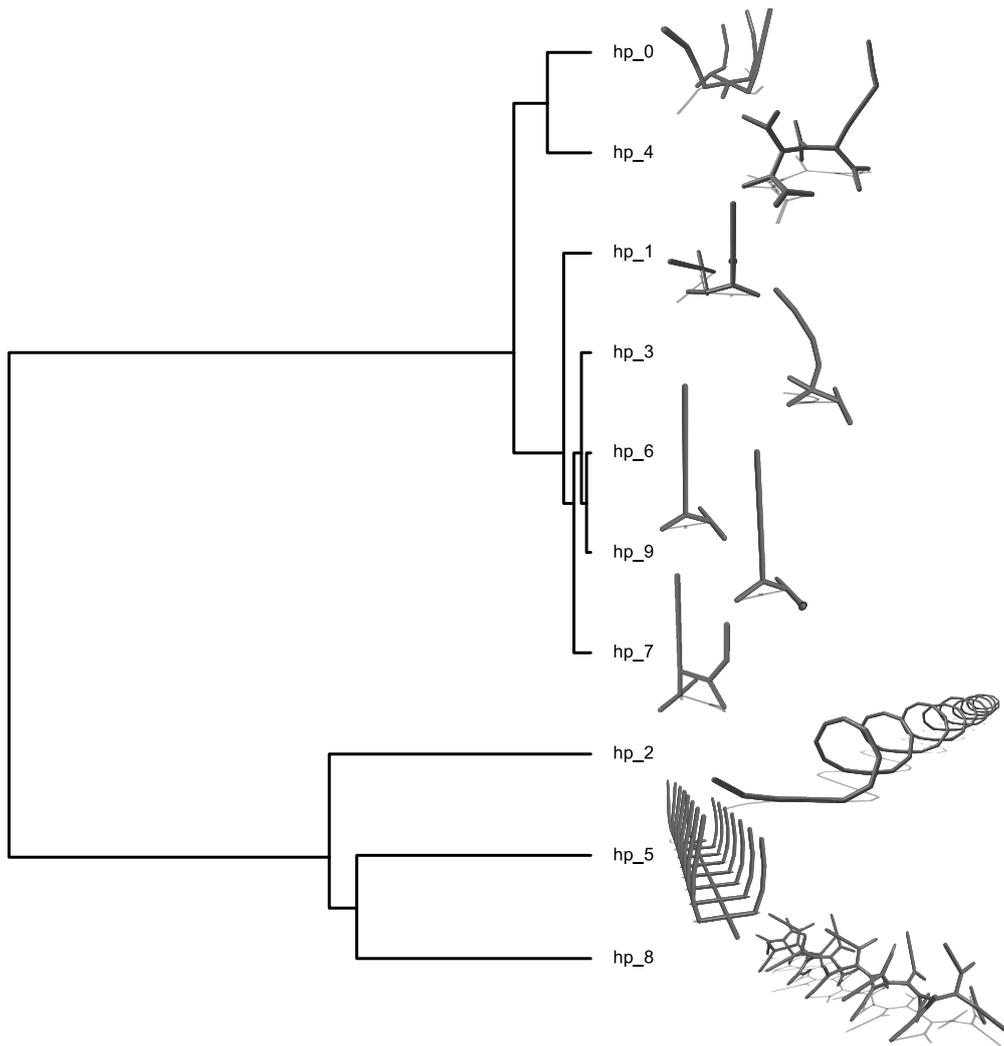


Figure 2: UPGMA clustering method [40] applied to 10 evolved morphologies, with similarity based on structural aspects only.

1.2 Related works

The problem of estimating similarity of 3D structures, primarily motivated by applications in computer vision and bioinformatics, is actively studied in many fields of science. However, existing solutions to this problem are either too specific or too inefficient to be directly applied to the model considered in this article (see also the list of requirements in Sect. 2.2). 3D objects can be represented in a number of ways – as solids, surface meshes, skeletal structures or descriptions of object properties. The latter representation has been employed in computer databases of 3D objects, where relational, quasi-symbolic methods can be used to find similar structures [39, 2, 34].

In computer vision, a number of approaches exist to find similar 3D objects [19, 45, 41, 14, 36, 11, 38]. These approaches are mainly based on extracting and then comparing some kind of characteristics (descriptions) of the objects, and are often used to recognize (identify) known 3D objects based on their 2D view(s). Many techniques

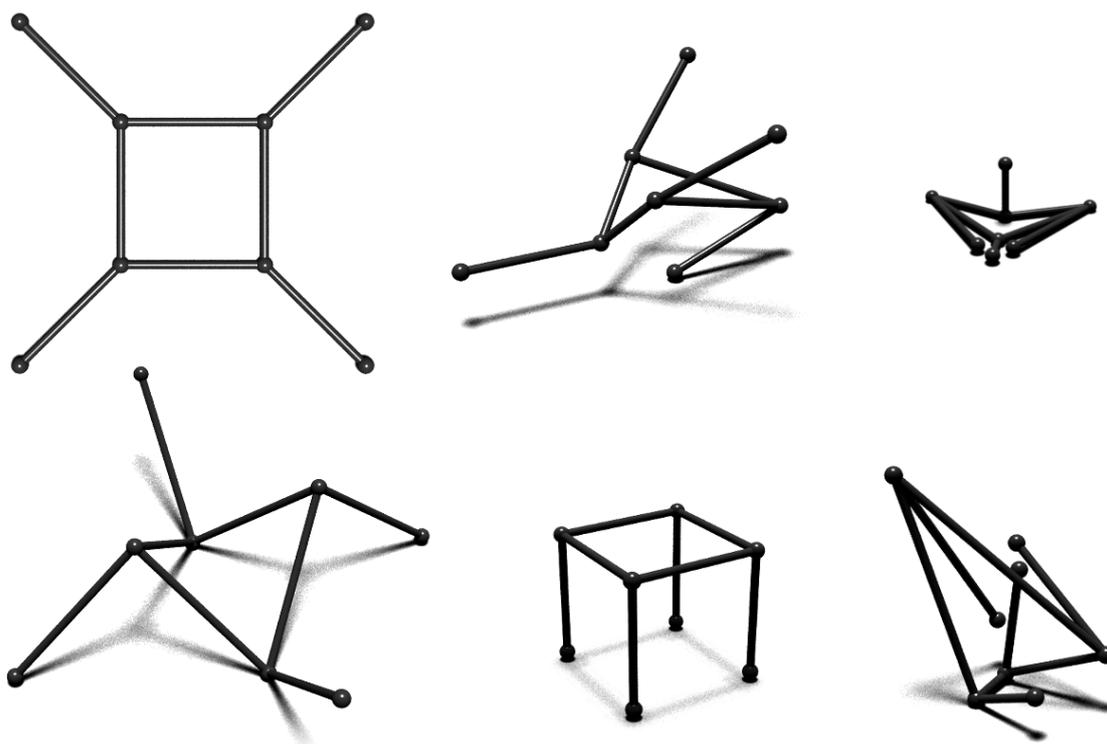


Figure 3: Human-designed morphologies that share the same structure. They are indistinguishable by the similarity measure that only considers structural aspects (dissimilarity is zero).

exist that only concern structural similarity and finding maximal common subgraphs, subgraph isomorphism or graph edit distance [22, 15, 28, 7].

In chemical informatics [27, 30, 6] and in bioinformatics, it is often sufficient to discover identity (isomorphism) of graphs or subgraphs [10], not the numerical estimate of dissimilarity. Moreover, in chemical and bioinformatic models, additional information is taken into account when comparing two structures, and these structures always satisfy a number of well-known constraints.

In chemical informatics, this additional knowledge may concern active reference structures. In bioinformatics, where protein structures are usually compared [4, 8], even more assumptions are made – e.g., that protein structures are linear and made of a number of smaller building blocks – amino acids. Then, the average distance between the backbones of superimposed proteins may be measured (using well-known RMSD [29] or some other measure [43, 1]). In chemical informatics, 2D fingerprints are usually compared [44] to identify common substructures, which suggests overall structural resemblance of the two molecules. Numerous 1D, 2D and 3D measures exist for this purpose, but methods used in cheminformatics and bioinformatics deal with chemical molecules made of atoms that can only bond in a finite number of ways. This specific property is exploited to identify smaller components and to compute chemical similarity.

This paper is organized as follows: Sect. 2 presents the model of morphology that

is considered throughout this work, describes in detail components of the dissimilarity estimation algorithm that handle geometry of constructs in addition to their structure, and finally summarizes properties of the measure. Sect. 3 illustrates the process of spatial alignment and matching of sample pairs of morphologies, while Sect. 4 discusses performance of the algorithm, gives an overview of possible generalizations, and outlines potential directions for future research.

2 Estimating structural and geometric similarity

2.1 Model of morphology

The morphological model considered here represents a construct as a graph where vertices correspond to body parts (e.g., material points), and arcs are connections between these parts. Let V denote a non-empty set of vertices, and let E be a set of edges. The function:

$$edge_connects : E \rightarrow V^2$$

represents connections (arcs) between pairs of body elements (vertices).

In our former works, the topology of the morphological graph only concerned the way vertices are connected [25]. This *structural* information was provided by the sets E , V , and the function *edge_connects*. To estimate *geometric* similarity, we additionally need information about coordinates of body parts in V as they are located in a 3D Euclidean space:

$$position : V \rightarrow \mathbb{R}^3$$

where \mathbb{R} denotes the set of real numbers.

In artificial life or evolutionary design applications, the morphologies belong to active or passive constructs or creatures that have a body and may have a brain (control system). Body is modeled here as an undirected, spatial graph, and may be composed of a finite number of elements – e.g., material points or balls connected with joints. Usually, the morphology graph is connected, which means that the construct does not have isolated parts.

The elements of body (parts, joints) and brain (neurons, receptors, effectors, neural connections) can be characterized by properties like mass, friction, stiffness, neural weights and other parameters – these constitute a variable number of additional attributes of vertices and edges in the model. For sample implementations following this model of morphology see chapters 3, 4, and 5 in [24], and [18]. Some variants of Artificial Chemistries (chapter 11 [24]) are also compatible with the model.

Compared to biochemical structures (e.g., proteins), artificial life constructs are more irregular and have less elements, since the number of vertices in the primary structure of a body is typically smaller than the number of atoms in complex molecules. Constructs in artificial life and evolutionary design are arbitrary, so no assumptions are made as to their structure, shape, geometry and occurrence of substructures.

It is possible to represent solid objects in the model described above by extracting their surface mesh (“wireframe”), or a graph that represents a complete volume of a solid [9]. This should be done carefully especially for smooth surfaces, as various mesh generation algorithms use different approaches to create edges that approximate a surface [35, 23]. Since the structure of the graph plays an important role in estimating dissimilarity, comparing complex surface meshes is primarily useful when their graph

structures are similar and only geometries differ. Otherwise it may be more appropriate to compare skeletal systems of 3D objects rather than their surface meshes.

2.2 Computing dissimilarity value

The model of morphology for artificial life and evolutionary design applications must be able to represent any possible 3D construct (i.e., phenotype), and there are also a number of requirements for the dissimilarity estimation algorithm that are imposed by these application domains:

- the algorithm will likely be applied to compare all pairs of constructs in large sets – therefore it should be fast and should have low complexity, so that execution time does not grow quickly when the complexity of morphologies increases,
- the dissimilarity measure should concern both structural and geometric aspects of compared morphologies at the same time, as they both determine mechanical properties and dynamics of constructs,
- the value of dissimilarity should be sensitive to small geometric and structural differences in compared morphologies,
- the algorithm should handle any number of properties associated with vertices and edges,
- the algorithm should be easily customizable so that importance of individual aspects of similarity can be adjusted depending on the experiment.

From a theoretical point of view, the problem of similarity estimation in the morphological model that is considered here is closely related to the problem of isomorphism of graphs. The task is to find the (best) matching between parts (V sets) of the two body graphs. However, an exact algorithm working on such sophisticated graph representations is not feasible. It would have an unacceptably high computational complexity (the problem of finding the maximal common subgraph is NP-hard). The typical size of V in our artificial life experiments ranges from 5 to 50, but constructs with thousands of elements exist, and there is no upper bound for the number of parts a creature or design may consist of.

Therefore, a fast heuristic method has been previously proposed to estimate dissimilarity of two individuals [25] that matches their body structures based primarily on degrees of vertices. The V sets are sorted by the degree of vertices. Vertices of the same degree are sorted by the number of control units (neurons, sensors and actuators) associated with these vertices. Then the algorithm matches vertices from both sets (thus constructing the matching function, *match*) starting from vertex groups with the highest degree and, subsequently, the highest number of control units. In case of ambiguity, additional properties of vertices are used as discriminating features (additional properties of edges are propagated to their incident vertices). Beginning with vertices that have the highest degrees helps preserve the most important parts of structures from being unmatched. When the two compared constructs have a different number of vertices, each missing vertex produces dissimilarity ingredient as big as would be produced by two maximally dissimilar vertices.

Note that although we usually speak about similarity, it is the dissimilarity (difference) between the two structures that is actually computed, i.e., the larger the value,

the more different the organisms are. The value of zero means that no differences could be determined. The general procedure of computing dissimilarity is as follows:

- create models corresponding to the two compared constructs – the V_1 , E_1 , V_2 and E_2 sets and the functions $edge_connects_1$, $edge_connects_2$, $position_1$, $position_2$, as described in Sect. 2.1,
- depending on the application area, consider a preprocessing step to balance the number of edges, vertices, edge lengths or the size of compared models – e.g., by adding or removing edges or vertices in a controlled way while preserving shapes,
- build the matching function, $match$, between V_1 and V_2 ,
- given $match$, compute individual components of dissimilarity,
- if necessary, aggregate individual components of dissimilarity to obtain a single value. The weighted sum formula can be used to compute the total value of dissimilarity:

$$dissim = d_V w_V + d_D w_D + d_N w_N + d_G w_G \quad (1)$$

where d_V is dissimilarity in the number of vertices (points), d_D estimates the difference in degrees of matched vertices, d_N reflects differences in neurons, sensors and actuators attached to body vertices, and d_G indicates geometric distance between matched points.

The weighted sum is often used as a simple way of aggregating individual aspects, but it is not required to use this particular form of aggregation, or to aggregate individual factors.

The most important role in this procedure is played by the $match$ function. Once this function is found, it further determines values of individual components of dissimilarity.

The algorithm that estimates structural and geometric dissimilarity builds on the former geometry-unaware algorithm. The four major enhancements that were introduced are described below.

2.2.1 The d_G component of the weighted sum

The d_G component reflects the distance between matched points from V_1 and V_2 . For each point $v_1 \in V_1$ matched with $v_2 \in V_2$, the d_G is increased by the Euclidean distance between v_1 and v_2 . If some point is not matched at all (it has no counterpart in the other organism), then d_G is increased by its distance to the origin of the coordinate system (see also Sect. 2.2.3).

If weighted sum aggregation is used as in (1), then the value of d_G is multiplied by the weight of the geometric aspect of similarity, w_G , and ultimately added to the total dissimilarity value, $dissim$.

2.2.2 Enhanced procedure that finds the $match$ function

The $match$ function matches points from both morphologies. In order to consider geometric aspects of bodies, coordinates of matched points must be taken into account while $match$ is constructed. The heuristic approach outlined earlier remains unchanged. It tries to match points that differ least, but apart from considering degrees of vertices

and their properties, it also considers the Euclidean distance between matched points (d_G) multiplied by the w_G weight. This way, the w_G weight reflects the importance of the geometric aspect while the matching is built. High values indicate high importance of geometry, while $w_G = 0$ ignores it both in the matching process and in the weighted sum of *dissim*.

2.2.3 The SVD transform

The matching that is constructed, as well as the value of d_G , depends on the coordinate systems of both models. Note that original coordinate systems of morphologies can be entirely unrelated: axes can be swapped or rotated, and origins can be relocated. Therefore it is crucial to first align both structures so that they overlap as much as possible.

To align the two structures, the SVD transform [32, 13, 42] is used. There is no information loss since a 3D space is transformed into another 3D space (dimensionality is not reduced). The SVD transform is performed for each of the two morphologies separately. As a result, the origin of the coordinate system is moved to the center of each construct (this is the location where the average distance to all vertices from V is minimal). Additionally, the three axes are aligned in such a way that the highest deviation of coordinates from the origin takes place along the first axis, the second highest deviation takes place along the second axis, and the remaining variance is left for the third axis.

The SVD transform is a simple and relatively fast method to align both constructs globally. It normalizes their spatial orientation based on the spatial distribution of vertex coordinates, but it does not handle *parts* of the constructs differently (e.g., it will not straighten a bent structure).

This transform aligns properly origins of coordinate systems of both morphologies. If the distribution of points in the 3D space is similar in both constructs, then the corresponding axes will be parallel in both models, but in some cases their direction may be opposite. This needs to be further addressed.

2.2.4 Symmetry operations

The SVD transform can, in some cases, yield opposite directions of axes for similar constructs. This is perfectly valid, yet not acceptable for our application, as coordinates resulting from two independent SVD transforms are directly compared. Therefore, additional configurations of the SVD-aligned models in the 3D space need to be tested. For three axes and two directions, there are $2^3 = 8$ possible configurations that correspond to the following operations:

- identity,
- reflection in planes: xy , xz , yz ,
- rotations by 180° about axes: x , y , z ,
- inversion through the origin, $(0, 0, 0)$.

One of the two models is subject to these transformations, and for each of them, the *match* function is constructed and the overall value of dissimilarity, *dissim*, is calculated.

Finally, the matching is chosen that minimizes *dissim*. This matching corresponds to the best alignment of coordinate systems of the two compared structures.

2.3 Properties of the proposed measure

The time complexity of the dissimilarity measure is $O(n^2)$, where n is the number of vertices in the more complex structure, $n = \max\{|V_1|, |V_2|\}$.

Some dissimilarity measures can act as metrics (distance functions). This is generally a desired property that allows to construct a metric space for a set S of compared structures. For the dissimilarity measure d to be a metric, the following conditions must be satisfied for all i, j and k in S :

$$d(i, j) \geq 0 \tag{2}$$

$$d(i, j) = 0 \text{ if and only if } i = j \tag{3}$$

$$d(i, j) = d(j, i) \tag{4}$$

$$d(i, j) \leq d(i, k) + d(k, j) \tag{5}$$

The proposed *dissim* measure (1), given non-negative weights, yields non-negative values for any pair of compared structures, as in (2). It is easy to demonstrate that it also satisfies (3) and (4). We proved formally that one of the components, d_D , satisfies triangle inequality (5), but were not able to prove this property for the complete measure.

To investigate if (5) is met, 30 sets have been created, each containing 30 to 50 diverse morphologies. These sets included 3D constructs that were either designed by a human, evolved, or first designed manually and then tuned using evolutionary optimization techniques. Among those evolved constructs, different fitness criteria were employed, and some morphologies only differed in geometry while sharing identical structure. Each of these sets formed a sample of experimental results, and for each set, a distance matrix was computed using *dissim*. From each distance matrix, 10^7 random triplets were selected and (5) was tested.

In matrices created with no geometric dissimilarity component ($w_G = 0$), all triplets satisfied (5). In matrices where $w_G > 0$, the percentage of triplets that satisfied (5) varied from 98.68% to 99.99% (99.54% on average). For those remaining 0.46% of triplets that did not satisfy triangle inequality, the average relative degree of violating (5) computed as

$$\frac{d(i, j)}{d(i, k) + d(k, j)} - 1$$

was 6%. This means that while the triangle inequality condition may be unsatisfied when the geometric component is employed, the proposed semimetric measure is “almost a metric”. In this testing sample, the triangle inequality condition was extremely rarely violated, and when it was, the amount of violation was small.

3 Spatial alignment case studies

To illustrate properties of the proposed dissimilarity measure and the way it deals with body geometry, three pairs of differing creatures have been selected. This section demonstrates the process of spatial alignment and matching for these pairs. Weights used for (1) are shown in Table 1.

weight	w_V	w_D	w_N	w_G
value	10	1	0.2	2

Table 1: Weights used to compute aggregated dissimilarity in the three spatial alignment case studies.

The weights were adjusted to focus on structural and geometrical aspects of morphologies. The highest value of w_V reflects the highest importance of differences in the number of vertices in compared designs (whether a body part exists or not). An intermediate value of w_D corresponds with the desired importance of differences in degrees of matched vertices. A larger value of w_G indicates significance of the geometric dissimilarity in the two compared morphologies. As neural aspects are not of interest here, a small value of w_N is set for only minor discrimination in the number of control units.

3.1 Highly similar evolved morphologies: *hp2* and *hp3*

First, two highly similar morphologies are compared. Both were evolved independently in two computer-simulated evolutionary processes, where fitness was proportional to the height of the center of mass of each creature. Neural networks were not simulated, so these were “passive” constructs. Their bodies are made of 9 vertices each and they are visually similar, as shown in Fig. 4. Both constructs resemble a long vertical shaft placed on a simple base.

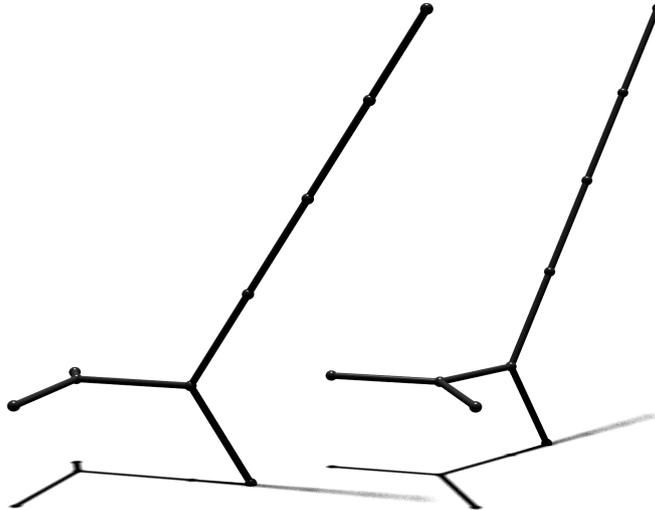


Figure 4: Morphologies of *hp2* and *hp3*.

Fig. 5 shows vertices projected onto the 2D plane (top: original xy coordinates, bottom: two most significant coordinates after the SVD transform). In this case, the SVD transform along with one symmetry operation (rotation by 180° about the z axis) allowed to nearly perfectly align both structures.

Table 2 shows indexes of the matched vertices. These indexes correspond to those shown in Fig. 5. It can be clearly seen that the structure-only matching is able to match just a few vertices properly – pairs $(1,0)$, $(3,1)$ and $(7,4)$, while others are matched

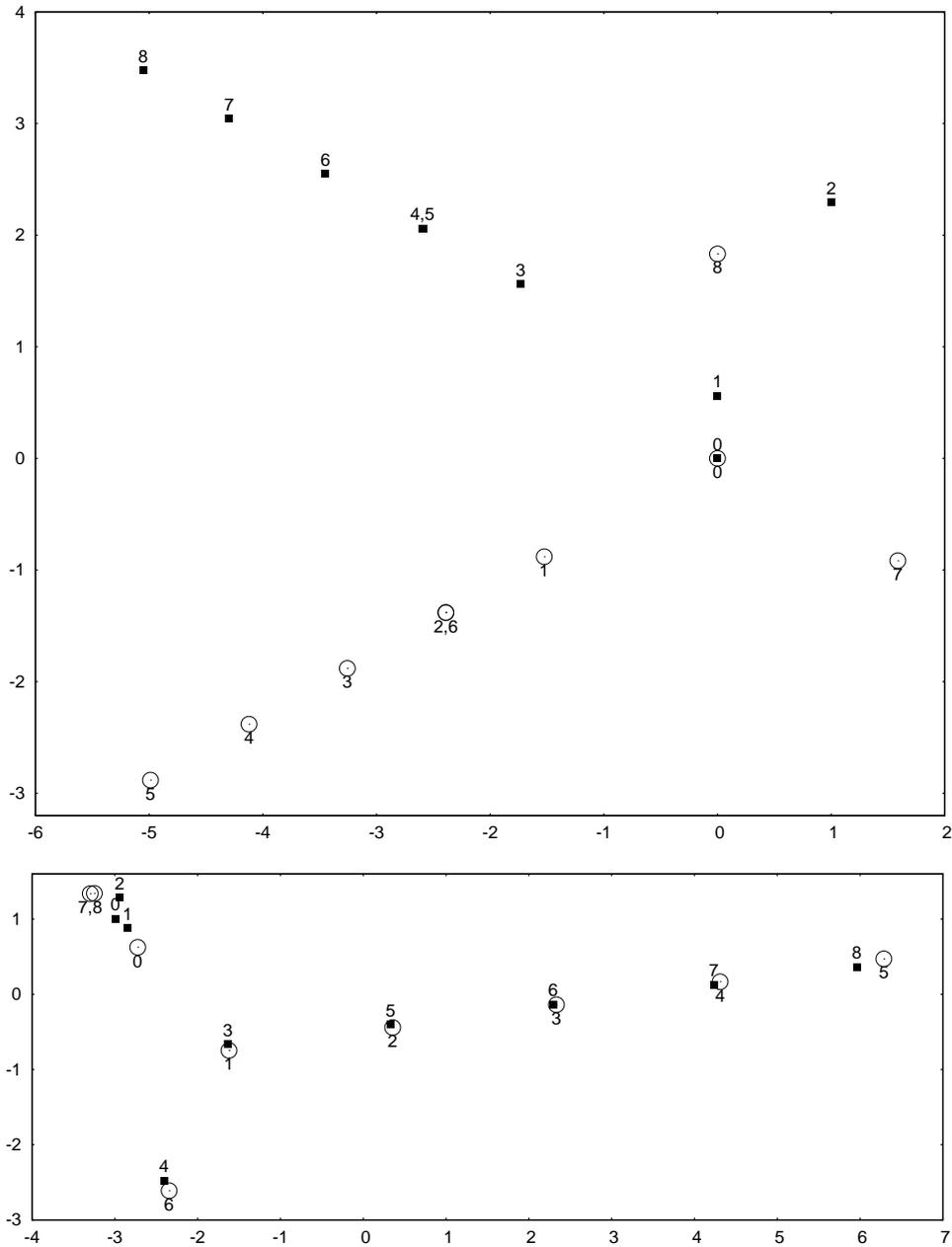


Figure 5: Original (top) and aligned (bottom) coordinates for creatures $hp2$ and $hp3$. Vertex index labels are placed above $hp2$ \blacksquare and below $hp3$ \circ .

inaccurately or completely incorrectly. The matching that uses geometrical information is able to associate all vertices perfectly, including the three mentioned pairs.

$hp2$	0	1	2	3	4	5	6	7	8
$hp3$ (structure only)	6	0	5	1	7	3	2	4	8
$hp3$ (geometry-aware)	8	0	7	1	6	2	3	4	5

Table 2: The influence of geometry on the matching process. Indexes of matched vertices



Figure 6: Morphologies of $s3$ and $s4$.

are shown.

Table 3 shows components of dissimilarity computed with and without geometrical information. It is important to note that even though both approaches managed to perfectly match structures, the structure-only approach matches vertices that are completely unrelated in terms of body geometry. For the geometry-aware approach, the small value of 2.658 corresponds to a minor geometrical difference of both spatially aligned morphologies.

The value of *dissim* for the structure-only approach is not given to avoid direct comparison with the geometry-aware approach. As the structure-only approach does not support body geometry, it would be unfair to compute geometric difference while the algorithm does not try to minimize it.

	d_V	d_D	d_N	d_G	<i>dissim</i>
structure only	0	0	12		
geometry-aware	0	0	12	2.658	7.717

Table 3: Components of dissimilarity as computed by the two approaches for $hp2$ and $hp3$.

3.2 Differing evolved morphologies: $s3$ and $s4$

The $s3$ and $s4$ creatures were evolved independently in two computer-simulated evolutionary processes, where fitness was proportional to the speed of movement. Neural networks, equipped with simple receptors like touch, processed environmental stimuli and controlled muscles which caused movement of bodies. The morphologies are made of 5 vertices each and are both linear, as shown in Fig. 6. Both of them are bent (one straight angle turn) and resemble a V-bar. Additionally, $s3$ has one arm bent in the z axis, which is geometrically different from $s4$. The bending location is also different: it is in the middle of the body of $s4$ while it is shifted in $s3$.

Fig. 7 shows vertices projected onto the 2D plane (left: original xy coordinates, right: two most significant coordinates after the SVD transform). In this case, the SVD transform was followed by one symmetry operation (rotation by 180° about the y axis). Even though aligned morphologies are not identical, both arms are turned in the same direction and geometric centers are close to each other.

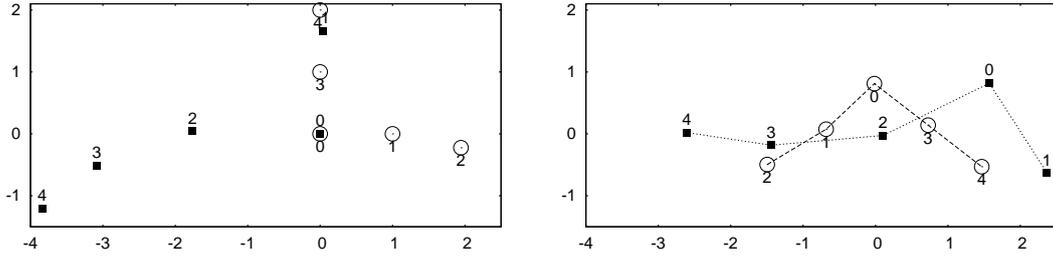


Figure 7: Original (left) and aligned (right) coordinates for creatures $s3$ and $s4$. Vertex index labels are placed above $s3$ \blacksquare and below $s4$ \circ .

Table 4 shows indexes of matched vertices. These indexes correspond to those shown in Fig. 7. In both approaches, bending vertices are matched – the (0,0) pair. Pairs (2,3) and (3,1) are also matched both by the structure-only and the geometry-aware algorithm. Note however vertices 1 and 4 of creature $s3$: the geometry-aware algorithm matched these outer vertices properly and consistently with the remaining parts, while the structure-only approach did just the opposite.

$s3$	0	1	2	3	4
$s4$ (structure only)	0	2	3	1	4
$s4$ (geometry-aware)	0	4	3	1	2

Table 4: The influence of geometry on the matching process. Indexes of matched vertices are shown.

Table 5 shows components of dissimilarity computed with and without geometrical information. As in the previous experiment, both approaches managed to perfectly match structures, but the structure-only approach matched vertices that are unrelated in terms of body geometry. Introducing geometric awareness not only allowed to evaluate geometric alignment, but also helped in matching corresponding parts of the body. For spatially aligned morphologies, the value of 5.339 corresponds to a small (yet higher than in the previous experiment with $hp2$ and $hp3$) geometrical difference.

	d_V	d_D	d_N	d_G	$dissim$
structure only	0	0	13		
geometry-aware	0	0	13	5.339	13.278

Table 5: Components of dissimilarity as computed by the two approaches for $s3$ and $s4$.

3.3 Diverse human-designed morphologies: *Centipede* and *Lizard*

Bodies of the *Centipede* and the *Lizard* creatures were manually designed. *Centipede*'s neural network was also designed by a human. *Lizard*'s neural network was pre-designed and then further evolved with fitness proportional to speed of movement on land. Both

creatures share a similar body concept, but they are highly different in size and complexity. *Centipede* consists of 52 vertices, while *Lizard* has only 15 vertices as shown in Fig. 8. Despite this divergence, both morphologies have a backbone, symmetrical legs and a head. Moreover, in the vertical axis, backbones of both creatures are elevated with respect to leg endings.

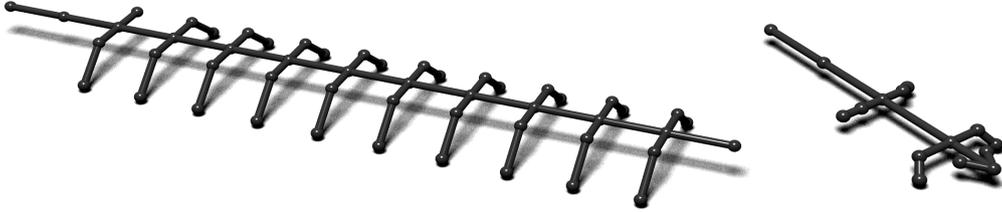


Figure 8: Morphologies of *Centipede* and *Lizard*.

Fig. 9 shows vertices projected onto the 2D plane (top: original xy coordinates, bottom: two most significant coordinates after the SVD transform). There was no need for any symmetry operations after the SVD transform has been performed. Despite high differences in size and complexity, the alignment is very good: geometric centers are close to each other, backbones are lined up, heads point in the same direction, and leg endings are on the same side of the xy plane.

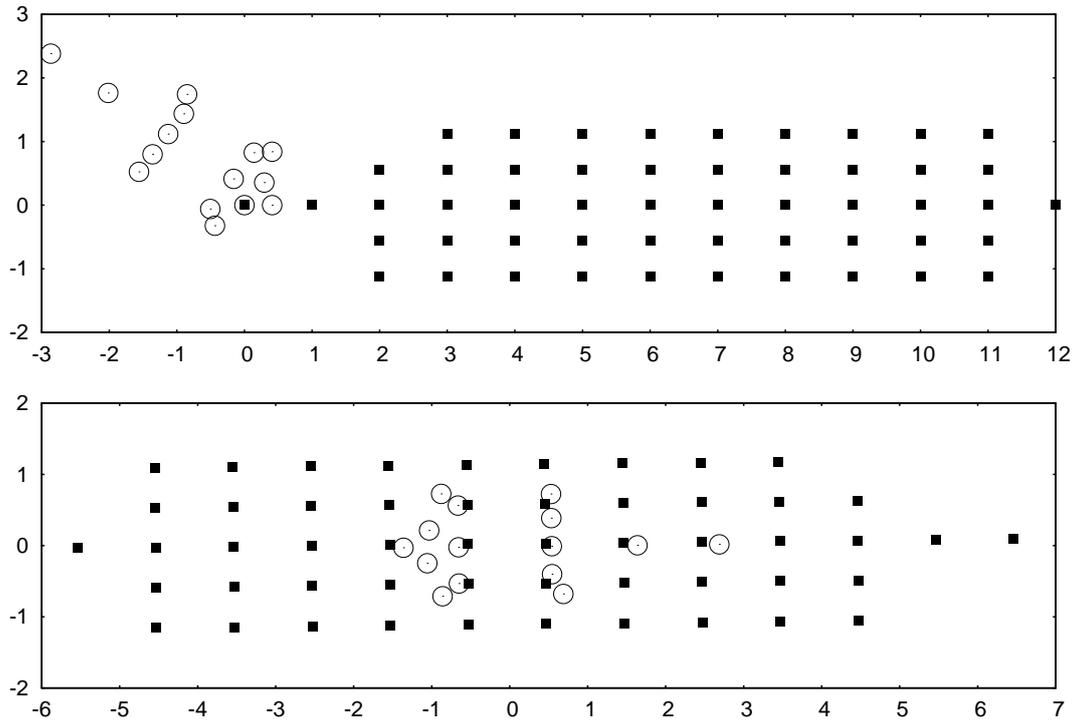


Figure 9: Original (top) and aligned (bottom) coordinates for creatures *Centipede*■ and *Lizard*○.

The structure-only algorithm was unable to match vertices of corresponding functional parts (backbones, legs, heads) except from two backbone vertices. The geometry-

aware approach successfully matched all parts of the *Lizard*'s backbone with the *Centipede*'s backbone, and most parts of *Lizard*'s legs with *Centipede*'s legs.

Table 6 shows components of dissimilarity computed with and without geometrical information. Both approaches matched structures of morphology graphs, but the geometry-aware approach took advantage of the additional geometrical information to match functionally corresponding body parts (backbones, legs), while – again – the structure-only algorithm paired vertices that are functionally unrelated.

	d_V	d_D	d_N	d_G	$dissim$
structure only	37	74	125		
geometry-aware	37	74	125	137.601	744.202

Table 6: Components of dissimilarity as computed by the two approaches for *Centipede* and *Lizard*.

4 Conclusions

4.1 Performance of the dissimilarity estimation algorithm

The experiments described in Sect. 3 illustrated the way spatial alignment is performed and the way the matching process works. Case studies that concerned selected pairs of diverse structures confirmed that the SVD transform allows to align and match geometries very accurately. An additional step of symmetry operations is sometimes necessary to improve alignment of coordinate systems. Introducing geometric aspects into the measure of similarity not only allowed to evaluate geometric alignment, but also helped in matching functionally corresponding parts of the body like bases, legs, backbones and heads.

The alignment and vertex matching were consistent with expectations of a human expert for structures of both similar and differing number of vertices. Experiments with structures that differ highly in complexity showed that it is hard for a human to identify the “best” matching of body parts (following common sense, not some numerical evaluation of the alignment). It seems that irregular structures of high complexity or highly differing structures may be the cases where the greedy algorithm employed here could lose most to a more global and elaborate approach.

The algorithm presented here is a heuristic developed with speed as an important requirement. Therefore the measure can be employed to compare large numbers of constructs (e.g. all pairs in a population) and efficiently produce reasonable matchings. It also yields a numeric estimate of dissimilarity that reflects intuitive, geometric difference of structures perceived by a human. Weights allow to easily adjust importance of individual aspects of dissimilarity.

Low complexity of the dissimilarity measure allows to compute a distance matrix not only for a single population, but for all (e.g. 10,000 or 100,000) constructs that appeared in an evolutionary process. The resulting distance matrix, after reducing to two or three dimensions, reveals the time-space phenetic trajectory of the evolving population of morphologies.

4.2 Possible generalizations

The proposed algorithm has been tested on 3D geometries from the domains of artificial life and evolutionary design. However, the model of morphology considered here is also compatible with chemical and biochemical molecules, biologic forms and technical constructs, or structures that are built from a finite number of connected simple elements. No specific assumptions are made regarding compared graphs.

The number of dimensions can be easily decreased to estimate dissimilarity of planar (2D) geometries, or increased to consider more than three dimensions. Some dimensions can be excluded from the SVD transform and symmetry operations (Sects. 2.2.3 and 2.2.4). The geometry could remain partially fixed in its original orientation which may be useful in some applications (e.g., the vertical direction can be considered fixed for plants).

There may be various properties assigned to vertices and edges. For artificial life forms, considered features may include the number of attached neurons, sensors and actuators, their parameters and connection weights, masses of body elements, their frictional coefficients or assimilation abilities. These properties are domain-dependent and can for example include bond strength, weight, material type, etc. The number and nature of these properties are arbitrary, the only basic requirement being the ability to estimate (dis)similarity of any pair of vertices.

To compute the total dissimilarity value, a simple weighted sum formula was employed here (Eq. 1 on page 8) to aggregate individual dissimilarity components. If necessary, a non-compensatory model of evaluation can be used to meet specific preferences [37, 5].

4.3 Further work

The dissimilarity measure described in this article has been employed in a number of applications, including reconstructions of sequences of random geometric mutations from phenetic dissimilarity matrices, building taxonomies, selecting representative constructs based on the distance matrices, and identifying optimization goals based on the geometry of optimized, structurally identical constructs. Due to space constraints, these experiments will be described separately. Despite very good results obtained with the heuristic algorithm presented here, it would be interesting to compare it with a more global (exact) method of matching vertices. For example, the Hungarian algorithm [26] may produce better matchings at the cost of higher time complexity, $O(n^3)$ – cf. Sect. 2.3. Both approaches could be compared using some objective quality measure like RMSD [29] to evaluate the tradeoff between time complexity and quality of the matching. Note however that in artificial life and evolutionary design, simple formulas like RMSD are not sufficient to evaluate the actual “quality of the matching”. Alignments suggested by experts usually consider functional elements of a body, and such alignments are suboptimal in terms of the total distances between matched vertices.

The adjustment of weights in the weighted sum formula, and value domains of individual components of dissimilarity need further investigation. For some applications, it may be advantageous to be able to normalize these components so that the meaning of weights depends less on the size and complexity of compared structures. However, a method to normalize ingredients of dissimilarity is not trivial. More theoretical work is required regarding mutual dependencies of the four components of dissimilarity (e.g., d_D and d_G likely correlate), and how these interactions should influence setting weights.

A related issue concerns size of compared structures. When chemical molecules are compared, their size (i.e., distance between vertices or atoms) is naturally constrained. As the algorithm described here does not assume any specific domain, scaled instances of structures may exist in some applications. However, the scale (size) of structures directly influences the value of the d_G component, so that the result of comparing A^* and B^* that are twice as big as A and B will proportionally increase d_G . Therefore, smaller structures may be considered similar more often than larger ones. This may or may not be a desired effect, so normalization of structure size might be appropriate.

Measuring geometrical dissimilarity opens up a way to evaluate similarity of moving constructs. Previously this was not possible, since in most cases the structure (connections of parts) does not change when a construct (a robot or an animal) moves. With the geometric aspect measured, it will be possible to estimate similarity of sequences of movement, and describe (evaluate) characteristics of the movement itself.

Finally, we would like to repeat experiments that concerned testing of global convexity of the space of solutions [21] in various evolutionary optimization tasks of evolving 3D constructs and agents. Previously, only structural similarity was computed and the correlation between quality of solutions and their similarity was relatively low. A more sensitive measure that takes into account geometry of morphologies is able to capture finer differences between these morphologies, allowing for better analysis of the convexity of the search space. This may further allow for development of efficient distance-preserving crossover operators [31, 46] for optimization of 3D constructs and spatial graphs.

Acknowledgement

This work has been supported by the Polish National Science Centre, grant no. N N519 441939.

References

- [1] Daniel Barthel, Jonathan Hirst, Jacek Blazewicz, Edmund Burke, and Natalio Krasnogor. ProCKSI: a decision support system for protein (structure) comparison, knowledge, similarity and information. *BMC Bioinformatics*, 8(1):416–438, 2007.
- [2] I. Blog. 3D shape matching and inspection using geometric features and relational learning. *Computer Vision and Image Understanding*, 72(3):340–350, 1998.
- [3] Josh C. Bongard and Chandana Paul. Investigating morphological symmetry and locomotive efficiency using virtual embodied evolution. In Jean-Arcady Meyer, editor, *Proceedings of the Sixth International Conference on Simulation of Adaptive Behaviour*, pages 420–429. MIT Press, 2000.
- [4] Philip Bourne and I. Shindyalov. Structure comparison and alignment. In P.E. Bourne and H. Weissig, editors, *Structural Bioinformatics*. Wiley-Liss, New York, 2003.
- [5] Denis Bouyssou, Thierry Marchant, Marc Pirlot, Patrice Perny, Alexis Tsoukiàs, and Philippe Vincke. *Evaluation and decision models: a critical perspective*. International series in operations research and management science. Kluwer Academic Publishers, 2000.

- [6] F.K. Brown. Chapter 35. Chemoinformatics: What is it and how does it impact drug discovery. *Annual Reports in Med. Chem.*, 33, 1998.
- [7] Fred W. DePiero and John K. Carlin. Structural matching via optimal basis graphs. *18th International Conference on Pattern Recognition (ICPR)*, 3:449–452, 2006.
- [8] K. Diederichs. Structural superposition of proteins with unknown alignment and detection of topological similarity using a six-dimensional search algorithm. *Proteins*, 23(2):187–95, 1995.
- [9] James Foley. *Computer Graphics: Principles and Practice*, chapter 11. Addison-Wesley, Boston, 1997.
- [10] H. H. Gan, S. Pasquali, and T. Schlick. Exploring the repertoire of RNA secondary motifs using graph theory; implications for RNA design. *Nucleic Acids Research*, 31(11):2926–2943, 2003.
- [11] H. El Ghawalby and E. R. Hancock. Measuring graph similarity using spectral geometry. *Image Analysis and Recognition*, 5112:517–526, 2008.
- [12] David E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Publishing Co., Reading, Mass., 1989.
- [13] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD, USA, third edition, 1996.
- [14] Lynne L. Grewe and Avinash C. Kak. Integration of geometric and nongeometric attributes for fast object-recognition. *Applications of Artificial Intelligence 1993: Machine Vision and Robotics*, 1964(1):13–28, 1993.
- [15] Douglas R. Heisterkamp and Prabir Bhattacharya. Matching of 3D polygonal arcs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(1):68–73, 1997.
- [16] Jacques Heyman. *The Science of Structural Engineering*. World Scientific Publishing Company, 1999.
- [17] Trevor R. Hodkinson, John A. N. Parnell, and Chen Wai-Kai. *Reconstructing the Tree of Life: Taxonomy and Systematics of Species Rich Taxa*. CRC Press, Boca Raton, 2007.
- [18] A. J. Ijspeert, A. Crespi, and J. M. Cabelguen. Simulation and robotics studies of salamander locomotion. Applying neurobiological principles to the control of locomotion in robots. *Neuroinformatics*, 3(3):171–196, 2005.
- [19] N. Iyer, S. Jayanti, K. Lou, Y. Kalyanaraman, and K. Ramani. Three-dimensional shape searching: state-of-the-art review and future trends. *Computer-Aided Design*, 37(5):509–530, 2005.
- [20] Wojciech Jaskowski and Maciej Komosinski. The numerical measure of symmetry for 3D stick creatures. *Artificial Life Journal*, 14(4):425–443, Fall 2008.

- [21] T. Jones and S. Forrest. Fitness distance correlation as a measure of problem difficulty for genetic algorithms. In L. J. Eshelman, editor, *Proceedings of the 6th International Conference on Genetic Algorithms*, pages 184–192. Morgan Kaufmann, 1995.
- [22] Y. Keselman, A. Shokoufandeh, M. F. Demirci, and S. Dickinson. Many-to-many graph matching via metric embedding. *Computer Vision and Pattern Recognition*, 1:850–857, 2003.
- [23] Bryan M. Klingner and Jonathan R. Shewchuk. Agressive tetrahedral mesh improvement. In *Proceedings of the 16th International Meshing Roundtable*, pages 3–23, October 2007.
- [24] Maciej Komosinski and Andrew Adamatzky, editors. *Artificial Life Models in Software*. Springer, London, second (revised and expanded) edition, 2009.
- [25] Maciej Komosinski, Grzegorz Koczyk, and Marek Kubiak. On estimating similarity of artificial and real organisms. *Theory in Biosciences*, 120(3-4):271–286, December 2001.
- [26] Harold W. Kuhn. The Hungarian method for the assignment problem. *Naval Res. Logist. Quart.*, 2:83–97, 1955.
- [27] Andrew Leach. *An Introduction to Chemoinformatics*. Kluwer Academic Publishers, Boston, 2003.
- [28] B. Luo and E. R. Hancock. Structural graph matching using the EM algorithm and singular value decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:1120–1136, 2001.
- [29] V. N. Maiorov and G. M. Crippen. Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. *Journal of Molecular Biology*, 235(2):625–634, 1994.
- [30] Ana Maldonado, J. Doucet, Michel Petitjean, and Bo-Tao Fan. Molecular similarity and diversity in chemoinformatics: From theory to applications. *Molecular Diversity*, 10:39–79, February 2006.
- [31] Peter Merz. Advanced fitness landscape analysis and the performance of memetic algorithms. *Evolutionary Computation*, 12(3):303–325, 2004.
- [32] Carl D. Meyer. *Matrix Analysis and Applied Linear Algebra*. Society for Industrial and Applied Mathematics (SIAM), 2000.
- [33] Zbigniew Michalewicz. *Genetic Algorithms + Data Structures = Evolution Programs*. Springer-Verlag, 1996.
- [34] P. J. Neal, L. G. Shapiro, and C. Rosse. The digital anatomist structural abstraction: a scheme for the spatial description of anatomical entities. *Journal of American Medical Informatics Association AMIA*, 98:423–427, 1998.
- [35] S. J. Owen. A survey of unstructured mesh generation technology. *7th International Meshing Roundtable*, 3(6), 1998. Online at <http://ima.udg.edu/~sellares/ComGeo/OwenSurv.pdf>.

- [36] X. Pennec and N. Ayache. A geometric algorithm to find small but highly similar 3D substructures in proteins. *Bioinformatics*, 14(6):516–522, 1998.
- [37] Vladislav V. Podinovskii. Criteria importance theory. *Mathematical Social Sciences*, 27(3):237–252, June 1994.
- [38] H. Qiu and E. R. Hancock. Graph matching and clustering using spectral partitions. *Pattern Recognition*, 39(1):22–34, 2006.
- [39] L. G. Shapiro, J. D. Moriarty, R. M. Haralick, and P. G. Mulgaonkar. Matching three-dimensional objects using a relational paradigm. *Pattern Recognition*, 17(4):385–405, 1984.
- [40] P. H. A. Sneath and R. R. Sokal. *Numerical Taxonomy: The Principles and Practice of Numerical Classification*. W. H. Freeman and Company, San Francisco, 1973.
- [41] F. Stein and G. Medioni. Structural indexing: Efficient 3-D object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):125–145, 1992.
- [42] Gilbert Strang. *Introduction to Linear Algebra*. Wellesley Cambridge Pr, third edition, March 2003.
- [43] Stefan Wallin, Jochen Farwer, and Ugo Bastolla. Testing similarity measures with continuous and discrete protein models. *Proteins: Structure, Function, and Bioinformatics*, 50(1):144–157, 2003.
- [44] P. Willett. Similarity-based virtual screening using 2D fingerprints. *Drug Discov. Today*, 11:1046–1053, 2006.
- [45] H. J. Wolfson and I. Rigoutsos. Geometric hashing: An overview. *IEEE Computational Science & Engineering*, pages 10–21, 1997.
- [46] D. H. Wolpert and W. G. Macready. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1:67–82, 1997.