Deformation-Driven Shape Correspondence via Shape Recognition

CHENYANG ZHU, Simon Fraser University and National University of Defense Technology
RENJIAO YI, Simon Fraser University and National University of Defense Technology
WALLACE LIRA, Simon Fraser University
IBRAHEEM ALHASHIM, Simon Fraser University
KAI XU, National University of Defense Technology
HAO ZHANG, Simon Fraser University

Many approaches to shape comparison and recognition start by establishing a shape correspondence. We "turn the table" and show that quality shape correspondences can be obtained by performing many shape recognition tasks. What is more, the method we develop computes a *fine-grained*, *topology-varying* part correspondence between two 3D shapes where the core evaluation mechanism only recognizes shapes *globally*. This is made possible by casting the part correspondence problem in a deformation-driven framework and relying on a *data-driven* "deformation energy" which rates visual similarity between deformed shapes and models from a shape repository. Our basic premise is that if a correspondence between two chairs (or airplanes, bicycles, etc.) is correct, then a reasonable deformation between the two chairs anchored on the correspondence ought to produce *plausible*, "chair-like" in-between shapes.

Given two 3D shapes belonging to the same category, we perform a top-down, hierarchical search for part correspondences. For a candidate correspondence at each level of the search hierarchy, we deform one input shape into the other, while respecting the correspondence, and rate the correspondence based on how well the resulting deformed shapes resemble other shapes from ShapeNet belonging to the same category as the inputs. The resemblance, i.e., plausibility, is measured by comparing multi-view depth images over category-specific features learned for the various shape categories. We demonstrate clear improvements over state-of-the-art approaches through tests covering extensive sets of man-made models with rich geometric and topological variations.

CCS Concepts: • **Computing methodologies** → **Computer graphics**; *Shape analysis*;

General Terms: shape correspondence, topology variation, data-driven shape plausibility

ACM Reference format:
Chenyang Zhu, Renjiao Yi, Wallace Lira, Ibraheem Alhashim, Kai Xu, and Hao Zhang. 2017. Deformation-Driven Shape Correspondence via Shape Recognition. *ACM Trans. Graph.* 36, 4, Article 51 (July 2017), 12 pages.
DOI: http://dx.doi.org/10.1145/3072959.3073613

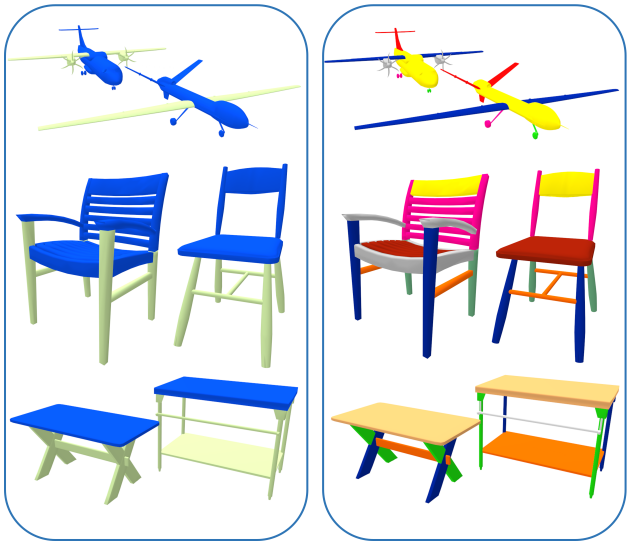


Fig. 1. Shape correspondences, derived by our hierarchical search, between 3D shapes exhibiting significant geometry and topology variations. Results at two granularity levels are shown. The topology-varying part matchings are computed where the core evaluation mechanism only performs shape recognition with respect to a shape repository. Matched parts share the same color; unmatched parts are in gray.

1 INTRODUCTION

Shape correspondence is one of the most fundamental and challenging problems in shape analysis [van Kaick et al. 2010]. The main difficulty arises when the goal is to match shapes that are dissimilar, geometrically, topologically, or organizationally. At the forefront of this pursuit, there have been recent attempts to compute *fine-grained, topology-varying* shape correspondences, resulting in part matchings at a fine granularity level [Alhashim et al. 2015; Fish et al. 2016; Kleiman et al. 2015; Tevs et al. 2014].

As the shape dissimilarities become significant, even the most advanced correspondence methods to date are far from satisfactory. Yet, humans always seem unfettered by these dissimilarities, owing mainly to their rich prior knowledge of the shapes and how to match them. It is then natural to consider a knowledge-driven approach, e.g., to turn correspondence into a supervised part recognition problem [Laga et al. 2013; van Kaick et al. 2011]. However, to recognize fine-grained parts amid significant variations, supervised learning still faces much challenge from feature space design, data collection, and label explosion. For example, labeling for the chair and bicycle models would have to reach the level of individual slats and frames, which can be tedious. Training directly on user-provided correspondences is an alternative, but collecting fine-grained correspondence data in the large scale is already a daunting task.

DOI: http://dx.doi.org/10.1145/3072959.3073613

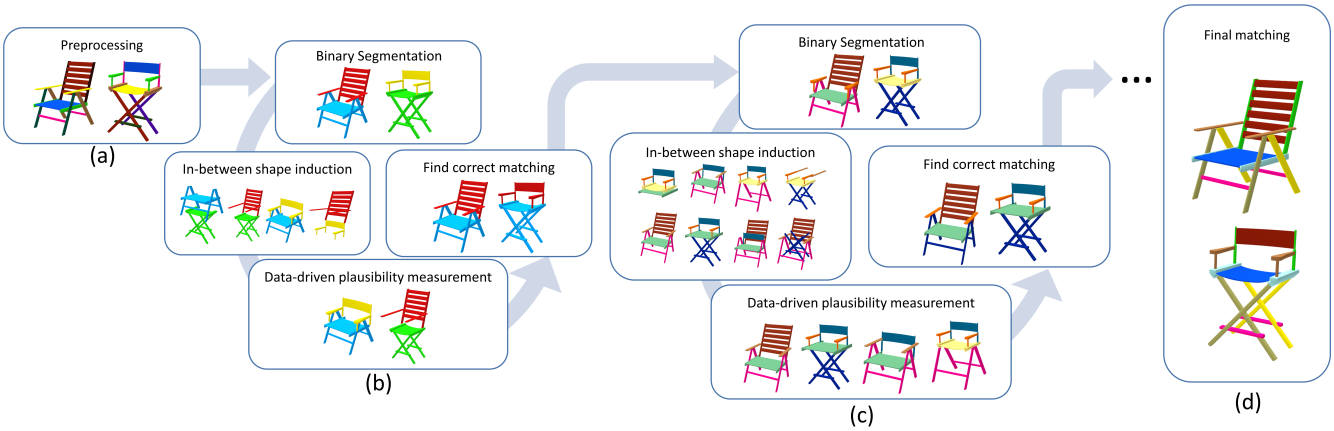


Fig. 2. An overview of our deformation-driven correspondence algorithm. The input consists of two pre-segmented or over-segmented 3D shapes (a). We recursively split and match substructures according to a data-driven plausibility criterion that relies exclusively on shape recognition. The first iteration (b) splits the input shapes into two sub-shapes. Given a pair of matched components, the algorithm recursively splits and matches them (c). Finally, after termination conditions are met for each substructure, we obtain a final part correspondence (d).

In this paper, we develop a *data-driven* approach to compute fine-grained, topology-varying part correspondences between a pair of 3D shapes, which bypasses the above difficulties. Our approach never solicits or relies on direct human knowledge on shape correspondence. Instead, we evaluate correspondences by performing *global shape recognition* tasks, which are predicated on available data. The required data is only a collection of 3D shapes belonging to the same category as the two input shapes, for shape recognition. Such shape collections spanning many object categories are already widely available, e.g., ShapeNet [Su et al. 2015]. No additional human annotations beyond object categorization are needed. All of the above are made possible by posing the correspondence problem under a *deformation-driven* framework [Alhashim et al. 2015; Sederberg and Greenwood 1992; Zhang et al. 2008].

The input to our algorithm consists of two man-made 3D objects belonging to the same category. The shapes can be pre-segmented or un-segmented and in the latter case, we first perform an over-segmentation. Our method treats the shape segments as the lowest-level primitives and follows a *coarse-to-fine, recursive beam* search. To start the search, each input shape is partitioned into two sub-shapes in multiple ways and for each pair of binary partitions, we evaluate the two (only) possible correspondences. Conceptually, to evaluate a part correspondence, we deform one input shape (source) into the other (target), based on the correspondence, and rate the correspondence by a "deformation energy". A small set of candidate correspondences, up to the beam width, are chosen based on their ratings. The recursive binary partition continues for each sub-shape in a candidate correspondence.

The main premise of deformation-driven correspondence is that if a part correspondence is correct and its implied deformation is properly defined, then the effort (or energy) exerted for the deformation should be low [Sederberg and Greenwood 1992]. For instance, it should take less energy to deform various parts of a chair into their counterparts in another chair, but more effort to deform a chair back into a set of legs or armrests. Our key observation is that the notion of "deformation energy" can be interpreted by how far the *non-transient* in-between shapes over the deformations deviate from other shapes belonging to the same object category as the input shapes. In other words, low deformation energy is synonymous with high *plausibility* of the in-betweens according to the semantics and functionalities of the object category. For example, if the inputs are both chairs, then the in-betweens for a correct correspondence should be "chair-like".

To model "chair-likeness", we utilize an existing, preferably large, collection of chair models. The plausibility of a shape is characterized only by a global visual similarity measured against these known chairs. Specifically, we compare multi-view depth images of two 3D shapes over a set of category-specific features learned during pre-processing. During feature learning, e.g., for the chair category, we mine representative and discriminative patches over the depth images of the chair collection.

In general, defining a proper deformation model between 3D shapes which may differ significantly in geometry and topology is challenging [Alhashim et al. 2015]. Without a clear ground truth to adhere to, it is conceivable that there are multiple "reasonable" deformations between two shapes for a given correspondence [Alhashim et al. 2014]. Computationally, keeping track of and evaluating continuous sequences of part deformations and doing so over a large search space of part correspondences is prohibitively expensive. In our work, we address these challenges by making simplifications to a full-fledged deformation-driven correspondence search.

First, instead of examining all the in-between shapes over possibly multiple deformation paths between the source and target shapes, we only evaluate in-betweens resulting from part *replacement*. Specifically, we replace parts in the source by matched parts from the target, based on the candidate correspondence, and deform the replacement parts or part groups in a structure-preserving manner to better fit them with their new surroundings in the source shape. Note that this way, we can also avoid determining what makes an in-between transient or not. Second, we do not perform

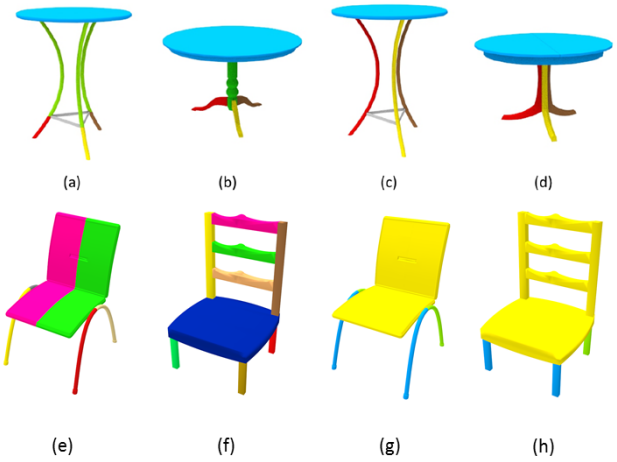


Fig. 3. Shape segmentations adapting to correspondences. Top: when the same table is matched to structurally different shapes (b, d), our method returns different table segmentations (a, c). Bottom: when fine-level segmentations (e, f) of two shapes are not matchable, our method returns a correct correspondence at a coarser level (g, h).

all possible part exchanges for a given candidate correspondence. Instead, we enumerate and evaluate candidate correspondences from coarse to fine and only sample a few part exchanges for each candidate, those which were not already evaluated higher up in the search hierarchy.

The plausibility measure defined above allows us to terminate a search path when the deformation energy increases significantly with the next recursion, which typically occurs when further segmentations of the two shapes would not yield a good correspondence. As a result, our method may return part correspondences at different levels of granularities, depending on the matchability between the input shapes and their fine-level segmentations. Figure 3 (top row) shows that our method returns different final segmentations based on the target shape to be matched, while in the bottom row, we show that when the fine-level segmentations of two input shapes are not exactly matchable, our method would return a correct correspondence at a coarse level.

Compared to *GeoTopo* [Alhashim et al. 2015], the most advanced solution for fine-grained, topology-varying shape correspondence to date, the key novelty of our work lies in defining the deformation energy via a data-driven plausibility. On the other hand, our deformation is also controlled by an energy to ensure structure preservation [Mitra et al. 2013] and this energy is *decoupled* from the other, the plausibility, which serves to rate part correspondences. In GeoTopo however, the energy which controls the deformation is also employed to rate correspondences. As a result, an inaccurate correspondence may be rated high due to the manner in which structure preservation is implemented in the deformation model, as shown in Figure 4. Our method explicitly compares the deformed in-betweens with models from the right category and leads to more

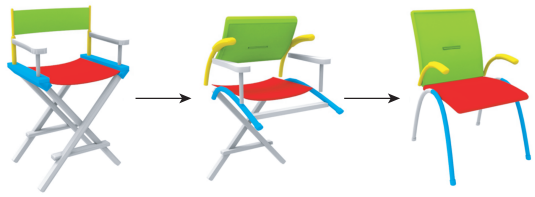


Fig. 4. An in-between shape obtained via a structure-preserving deformation [Alhashim et al. 2015] is not plausible. Yet, its associated energy is relatively low due to the strong effort in the structure preservation.

accurate correspondences, as we demonstrate over extensive sets of shapes with rich variations.

In retrospect, our work also reveals a refreshing connection between shape correspondence and (global) shape recognition. While many approaches to the latter would start by establishing a shape correspondence, we show that their roles can be exchanged. We can effectively solve difficult instances of the shape correspondence problem by performing many shape recognition tasks and only utilizing *weak* human knowledge (objective categorization) and readily available data (e.g., ShapeNet) for that purpose.

2 RELATED WORK

Shape correspondence [van Kaick et al. 2010] and data-driven shape processing [Xu et al. 2016] are well-studied in graphics. In this section, we only cover the most closely related works.

Knowledge-driven shape correspondence. Recent research trends on shape correspondence have shifted from low-level geometric analysis to higher-level semantic reasoning. The latter often involves recognizing shape parts and correlating them based on their semantics, contexts, and functionalities [Hu et al. 2016]. Shape semantics can be either inferred as prior knowledge from a set of exemplar shapes, through supervised learning [Pitiot et al. 2007; van Kaick et al. 2011], unsupervised co-analysis [Huang et al. 2011; Kim et al. 2013; Xu et al. 2010, 2012; Zheng et al. 2014] or implied by contexts [Laga et al. 2013] or symmetry priors [Tevs et al. 2014; Wang et al. 2011]. Among these methods, supervised learning is comparatively successful in handling shape variations. However, there are two major challenges limiting its performance when dealing with significant variations. Firstly, obtaining training data is an involved process, where the task of fine-grained part labeling is especially complex when topological changes are taken into consideration. Secondly, in order to model the prior knowledge of part correspondence to facilitate knowledge transfer, a proper feature space needs to be crafted, which demands intensive human knowledge.

Deformation-driven shape correspondence. This approach has been shown to be effective in matching shapes exhibiting significant topological differences [Alhashim et al. 2015; Blanz and Vetter 1999; Sederberg and Greenwood 1992; Zhang et al. 2008]. The key for better performance is the definition of a deformation energy for penalizing wrong correspondences. Up to now, the deformation energies employed have been based on either purely geometric criteria or hand-crafted rules. In particular, in the closest recent related

work, Alhashim et al. [2015] devise a deformation energy that simultaneously penalizes geometric distortion and encourages structure preservation, while allowing topological changes. Although this approach can lead to fine-grained part correspondences, both the deformation operations and the distortion measures need a careful design. To avoid hand-crafting the deformation energy, we take a data-driven approach to learn from and utilize implicit knowledge implied by a large collection of plausible shapes.

Fine-grained shape classification or similarity. Fine-grained classification is typically coupled with fine-grained correspondence since the latter would enable detailed comparison between highly similar shapes. Huang et al. [2013] proposes to learn a distance metric for fine-grained shape classification by globally aligning a collection of shapes for localized shape comparison. The recent work of Kleiman et al. [2015] is closely related to ours. They compute fine-grained shape *similarity* by introducing a shape edit distance (SHED) which is defined as the amount of effort needed to transform one shape into another. The edit distance is indirectly modeled as part correspondence, where topological changes are possible. However, their goal of correspondence-assisted shape similarity is almost the opposite of ours, which is shape correspondence aided by shape similarity. As a side product, SHED does produce a shape matching. But for shape pairs belonging to the same category exhibiting large variations, their matching results are quite far from natural. This is acceptable in their problem setting as these shapes are then deemed to be distant from each other. Technically, their deformation cost is measured as a weighted sum of the changes on segment geometry and context, where the weights are learned from examples. However, the specific terms of the transformation cost are still hand-crafted.

Data-driven shape space. The idea of learning plausible shape spaces from example data has been exploited in the modeling of morphable human faces [Blanz and Vetter 1999] and bodies [Allen et al. 2003]. Recently, Gao et al. [2013] propose to model the plausible shape deformation space for guiding shape morphing, using a set of models belonging to the same category as the shapes being blended. These works focus on the construction of plausible shape spaces spanned by example data points, which itself is based on existing correspondence computed independently. Our work, on the other hand, aims to tackle the difficult problem of shape correspondence via data-driven plausibility assessment.

Co-analysis. Co-analysis [Golovinskiy and Funkhouser 2009; Hu et al. 2016; Huang et al. 2014; Kim et al. 2013; Sidi et al. 2011; van Kaick et al. 2013; Xu et al. 2010, 2012; Zheng et al. 2014] also utilizes weak knowledge from shape collections, but doing so in a different manner from our work. Co-analysis works directly on the shape collection, with the goal of inferring commonality and diversity among the shapes and their part structures. In our work, the input consists of a pair of shapes to be matched, while the large shape collection only provides *background* knowledge to assist in plausibility evaluation for different object categories.

In a recent work on structure-oriented networks of shape collections, Fish et al. [2016] developed a projective matching scheme that can handle topological variations between two 3D shapes. However, their method was designed to work on large shape collections and assumes sufficient density in terms of shape and structure. If a shape pair is deemed too dissimilar, the correspondence computed for it can be highly partial and unreliable. Part of the reason for the partiality of the matching is that their method only computes one-to-one matchings. In contrast, our method can produce one-to-many and many-to-one matchings between shapes which differ significantly in geometry and structure.

3 OVERVIEW

Figure 2 shows an overview of our correspondence method. The input of our algorithm is either a pair of pre-segmented or unsegmented 3D shapes. If the input 3D shapes are unsegmented, it is necessary to perform an over-segmentation first. The output of our algorithm is a coarse-to-fine part correspondence hierarchy. Aside from the hierarchical correspondences, the key improvements of our approach over GeoTopo [Alhashim et al. 2015] include the utilization of only weak human knowledge in the form of objective categorization and less sensitivity to the initial segmentation of the shapes, as demonstrated by examples in Figure 3.

Per-shape preprocessing. The preprocessing step starts with a convexity-based over-segmentation [van Kaick et al. 2014] of shapes with many interconnected components if an input shape S is not segmented. The aim of this step is to ensure that all connected components c_i have a relatively similar segmentation granularity, which is important to improve finer level correspondences. Additionally, we detect and group symmetric components using a co-hierarchical analysis of shapes [van Kaick et al. 2013]:

$$S = \{g_0, g_1, ..., g_n\}$$
$$\text{where } g_i = \{c_{i_0}, c_{i_1}, ..., c_{i_k}\}, i \in \{0, 1, .., n\}.$$

Finally, for both presegmented and not presegmented input shapes, a unique undirected connected graph G for each shape is needed, since our correspondence search is based on graph partitioning. Each node in G is a symmetry group g_i, and each edge e_{ij} represents a physical connection between g_i and g_j.

Data-driven correspondence evaluation. The fundamental premise of our work is that in-between shapes induced from correct part correspondences have greater plausibility than others induced from incorrect part correspondences. The plausibility of an in-between shape is estimated using shape recognition with respect to known 3D shapes. We utilize a set of shapes from ShapeNet [2015] that belongs to the same class of objects as the two input shapes. During an offline preprocessing step, we project each shape in ShapeNet to form multi-view depth images and then sample patches in each image. Our approach mines the most discriminative and representative patches based on the shape's category. These patches are used to encode the in-between shapes and compute shape similarity based on visual features. Section 4.1 provides more details on this process.

Graph partition based correspondence search. Since the aim of our approach is to find fine-level correspondences based on a segmentation, which may not be perfectly accurate, the search space would be too large for a comprehensive search. Therefore, a coarse-to-fine, recursive beam search scheme is used to find near optimal shape correspondences in a short amount of time.

In detail, given a source and a target input 3D shape, the initial step contains the matching hierarchy at the coarsest level. In other words, a set containing all segmented components of the source shape corresponds to a set containing all segmented components of

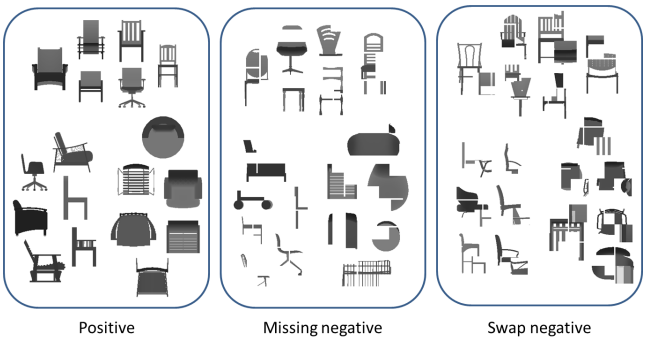


Fig. 5. Samples of our plausibility measure training data. The "Missing negative" column shows examples with patches that are not present in the shape, while the "swap negative" column shows structures that have some undesirable combination of patches. At least one of the negative examples directly affects the visual perception of shape implausibility.

the entire target shape. Any given matched set in each input shape is then recursively partitioned into two subsets each. Two in-between shapes are then induced by the deformation of the matched subsets from the source shape to the target shape. This recursive process continues until a termination condition is met. At each step, these partitions are optimally matched by evaluating the generated in-between shapes using the data-driven correspondence evaluation method previously mentioned. This approach allows us to significantly reduce the search space by halting further recursions if a specific termination condition is met during the part correspondence step, while also providing us a hierarchical part correspondence, as shown in Figure 12. Furthermore, the proposed approach is robust in finding reasonable coarser level part correspondences even if a given initial segmentation may be less than perfect. More details can be found in Section 4.2.

4 CORRESPONDENCE ALGORITHM

In this section, we provide more details on our coarse-to-fine, recursive beam search and data-driven plausibility measurement from the preprocessed ShapeNet data.

4.1 Data-driven correspondence evaluation

With a data-driven approach to define shape plausibility, the premise is that if a shape belongs to a specific shape class, then its structure should look similar to some other instances in that class. Furthermore, an in-between shape that is generated using incorrect correspondences should either lack some relevant sub-structures or have a messy global structure. It is important to point out that all shapes in the reference dataset and the input pair of shapes to be matched must have a consistent upright orientation.

4.1.1 Data preparation. To evaluate whether a shape is plausible or not in a data-driven approach, both positive and negative examples are needed. The shapes in ShapeNet [Su et al. 2015] (Figure 5) with class labels are used as our positive examples. This dataset is also used to artificially generate negative examples.

Positive examples. Measuring structural features over 3D shapes is a difficult task in itself. To tackle this challenge, we employ a projective approach, which has been shown to be an effective way to compare 3D shapes [Chen et al. 2003; Qi et al. 2016]. Working with view-based projections also allows the use of a large amount of prior successes on distance/feature learning for 2D images.

Man-made objects typically exhibit major global symmetries aligning with their perceptually principal directions [Kazhdan et al. 2004; Podolak et al. 2006]. This directs us to a key observation for a visual evaluation of shape plausibility: the geometric and structural plausibility of man-made shapes should be inspected using their canonical views as references. In light of this, we project 2.5D depth images of 3D shapes into three canonical views (front, left and top).

Negative examples. Shapes generated by an incorrect correspondence either (a) have an undesirable sub-structure combination or (b) miss some critical sub-structures. The generated negative examples, shown in Figure 5, should reflect these characteristics. Therefore, two types of data are needed: shapes with random parts swapped and shapes with random parts removed. The swap and remove operations are applied on projected 2.5D depth images of positive data, providing a reliable dataset of negative examples.

4.1.2 Plausibility measure training. When measuring the similarity of two depth images, instead of relying on local features such as HOG [Dalal and Triggs 2005], we learn characteristic mid-level elements [Singh et al. 2012] from all training depth images and use them to perform feature encoding. A good feature of such mid-level representation is one that is neither too local, which allows the capture of relevant structures of 3D shapes, nor too global, aiming to remain flexible when modeling intra-class shape variation.

Furthermore, we extend the well-known convolution operation to perform feature encoding of a given depth image. The convolution operation is the fundamental building block of Convolutional Neural Networks (CNNs); however, instead of training a very large number of convolution filters as is required in CNNs, we take a light-weight approach by using the pre-analyzed mid-level elements as convolution filters. This approach was inspired by the work of Bansal et al. [2015]. It is worth noting that the mining of pre-analyzed mid-level elements in our work is *unsupervised* as opposed to supervised mining performed when using CNNs.

Mid-level element extraction. Mid-level element extraction is performed on each view separately. For all the database shapes belonging to the same class, we sample from each canonical view a set of mid-level patches using the selective search method [Uijlings et al. 2013]. More specifically, the size of each projected depth image is 224×224 pixels, where the brightness encodes depth information. We use a 56×56 sliding window moving from the top left of each depth map. If in a given window we find more than 80% non-void pixels, we keep this patch as a potential source of mid-level elements.

Let P be the number of patches collected for each view on all shapes in a certain class. The value of P is usually over 100K. We apply self-tuning spectral clustering [Zelnik-Manor and Perona 2004], with their HOG feature as a similarity measure to extract mid-level elements. To make the spectral clustering feasible and faster, we first use k-means to perform a pre-clustering of the P patches into $P/4$ clusters. The spectral clustering is then performed over these $P/4$ cluster centers. At last, the self-tuning clustering

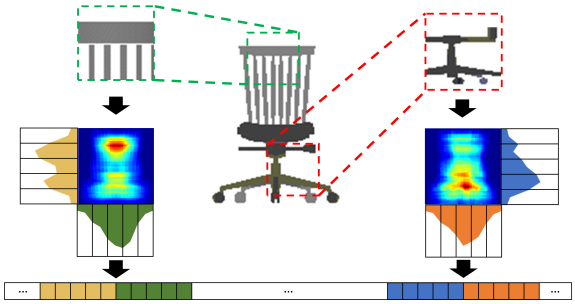


Fig. 6. Extracting a feature vector from the HOG representation of the depth image and the mid-level patches. Feature vectors are constructed by generating a response map using a convolution operation between patches and depth maps. We obtain a 10-dimensional feature vector for each depth patch by averaging five segments of the response map in the vertical and horizontal directions. The feature vectors of all the patches of a given depth image are then combined into one single feature vector that represents the depth image.

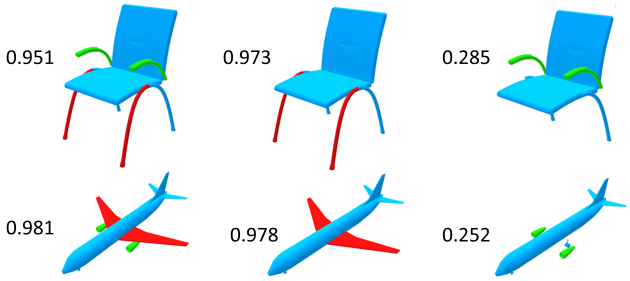


Fig. 7. In-between shapes and plausibility scores by our plausibility measure. Removing key substructures from a plausible shape significantly decreases our plausibility score. In the examples at the top row, the plausibility score would not be affected by removing an unnecessary part (armrests), but it would decrease significantly after removing a critical part (front legs). Same for the bottom examples, engines under wings are not necessary since there are other airplanes with no engines in the training set. The score would significantly decrease when removing the wings, since all airplanes in the training set have wings. This shows that our plausibility measure is reasonable and accurate.

returns from 200 to 400 clusters. We then apply an IF-IDF similar approach to select out the most discriminative 60 clusters.

A discriminative patch cluster should not cover all or most of the given images dataset. Instead, patch clusters should appear mainly in a specific subset of the whole dataset. Moreover, the frequency in which patch clusters appear in a single image is indicative of the relevance of this cluster. Accordingly, we determine the term frequency of a given cluster i for image j as below:

$$tf_{i,j} = n_{i,j} / \sum_k n_{i,k},$$

where $n_{i,j}$ is the number of patches in cluster i appear in image j. The sum of highest 20 $tf_{i,j}$ is accumulated as the term frequency tf_i of cluster i. Let n_i be the number of patches in cluster i appear in all images, the inverse document frequency idf_i for a cluster i determined as below:

$$idf_i = n_i / |k : n_{i,k} \neq 0|.$$

Finally, we take the center of the 60 clusters with highest $tf_i \times idf_i$ as the mid-level elements for this class.

Feature encoding. Only some specific combinations of sub-structures could construct plausible structures for a given shape category. Suppose that a sub-structure appears at a wrong place or incorrect relative position. This would cause structure implausibility, as shown in Figure 9. Therefore, the feature should be able to describe the position properties of each discriminative sub-structure.

We perform view-based feature encoding, and each extracted mid-level elements could be seen as a projected sub-structure. Given a depth image of a 3D shape, we perform sliding-window convolution operation using the mid-level patches extracted for the corresponding view and shape class. The convolution is performed in HOG space, based on the HOG representation of the depth image and the mid-level patches. The convolution returns an activation value for each window position, see Figure 6. This activation map would contain the position properties we need for a mid-level element based feature encoding.

We need to extract a 1D feature vector from each 2D activation map while, at the same time, we store the position of the patch that generated this feature. We encode the 2D information into 1D by defining 5 slices of accumulation histograms along both the horizontal and vertical directions, resulting in a feature vector of length 10. By placing all the normalized histograms activated by the extracted mid-level elements in a row, we obtain an element-based feature which encodes position distribution properties.

SVM train and test. Finally, we obtain an approximately even number of instances classified as: positive, negative by missing some parts, and negative by random correspondences. We train a three-class SVM with these instances using the feature vector we introduced previously. During the testing of the trained SVM, the plausibility score of a shape is given by its probability of belonging to the positive class.

Note that there are three SVM classifiers, each trained for depth images from each canonical view: front, left, and top. We choose the worst score from these classifiers as the plausibility score for a shape, since some structure errors usually appear in just one view.

Part necessity evaluation. To determine whether a part is necessary or not in a plausible shape is key when deciding whether to consider it for a one-to-null correspondence. The reasoning is that if a part is very important for a plausible structure, it should not be a one-to-null candidate in the correspondence matching. For instance, if the legs of a chair are removed, the plausibility should significantly decrease, which means this part is critical for a chair structural soundness. On the other hand, removing optional parts from a plausible shape, e.g. removing the arms of a chair, should not significantly impact the plausibility score since such a part is not critical for a chair structural soundness. Based on this idea, we decide whether removing a part is acceptable during the shape binary partition operation. Figure 7 shows how the plausibility measure proposed in this work can describe whether a part is critical for a given shape. Note that removing key parts from a plausible shape leads to a significant decrease of its plausibility score. Further details are presented in Subsection 4.2.

4.2 Correspondence search via graph partition

Since there is a large number of possible matching hierarchies representing a correspondence of two input shapes, we propose a coarse-to-fine beam search scheme to find an optimal hierarchy correspondence among two input shapes S_0 and S_1.

Binary shape partition and in-betweens. Let G_i represent a connected graph representing structural connections between symmetry groups extracted from shape S_i, $i = 0, 1$, as shown in Figure 8(a). In our implementation, we only consider reflectional and translational symmetries. A binary graph partition is obtained by splitting G into two vertex-disjoint subgraphs G_i^a and G_i^b. For each binary partition pair over G_0 and G_1, there are two possible matchings: $\{G_0^a \leftrightarrow G_1^a, G_0^b \leftrightarrow G_1^b\}$ and $\{G_0^a \leftrightarrow G_1^b, G_0^b \leftrightarrow G_1^a\}$. Under a deformation-driven framework, the best matching would maximize the plausibility of in-between shapes resulting from a proper deformation from the source shape S_0 to the target shape S_1.

Deforming S_0 into S_1 is a process which would transform the two part groups (i.e., subgraphs) in G_0 into their matched part groups in G_1, both position-wise and shape-wise. A deformation could execute the part positioning and shape-wise deformation in an interleaving manner or decouple them. To simplify matters, we choose the latter and let shape-wise deformation precede part positioning. With two part groups in a binary partitioning of the source, there are two possible deformation paths from the source depending on which subgraph (G_0^a or G_0^b) is deformed first. Then to simplify matters further, we only evaluate plausibility on *one* in-between shape along each deformation path. This in-between is obtained at the moment when a source part is fully deformed into its counterpart in the target — it is simply the result of a part (group) replacement. Figure 9 shows the four in-betweens we evaluate for a pair of binary partitions of two chair models, along the four deformation paths.

However, mere part replacement may break some of the connectivity and symmetry relations originally found in the input shapes. Hence, it is necessary to perform a structure-preserving local part deformation to arrive at a more reasonable in-between. With respect to proximity constraints, we aim to preserve a displacement vector between every two connected parts in each G_i. This vector is determined by the closest pair of points between the bounding boxes of every two connected parts. Simply put, if we can make every displacement vectors remain similar after the deformation, in-between shape structure would also be preserved. Symmetry preservation is enforced separately on each symmetry group as done in [Alhashim et al. 2015].

For the matching $\{G_0^a \leftrightarrow G_1^a, G_0^b \leftrightarrow G_1^b\}$, we evaluate plausibility, denoted by ϕ, on two in-betweens: $G_1^a \bigcup G_0^b$ and $G_0^a \bigcup G_1^b$, where $\bigcup$ denotes part replacement followed by structure-preserving part (group) deformation described above. We rate the matching as

$$min(\phi_{\{G_0^a \cup G_1^b\}}, \phi_{\{G_1^a \cup G_0^b\}}).$$

Matching hierarchy. There are several possibilities of hierarchically matching the components of S_0 and S_1. Each matching hierarchy is represented as a binary tree, where each node is a single correspondence pair (G_0^a, G_1^a) and child nodes correspond to a unifying subset of the components represented by its parent. Since there is a prohibitively large number of hierarchy combinations at each level, exhaustively searching for all possible matching hierarchies would be impractical. To deal with this, it is fundamental to observe that all matching hierarchies have the same root node, i.e., the whole shape S_0 matches to the whole shape S_1. Therefore, we apply a recursive coarse-to-fine beam search scheme to efficiently find a reasonable matching hierarchy. The recursion terminates when one of the three conditions occurs regarding a correspondence pair:

- contains a one-to-one or one-to-many correspondence, since it is not possible to further decompose it;
- the decrease of both in-between shapes plausibility measure is below a threshold β. This condition happens when we only have bad fine-level correspondences;
- both in-between shapes' plausibility measures are approximately equal within a tolerance margin of γ. This condition happens when we are starting to see diminishing returns of in-between shape variance on each iteration.

It is important to point that either G_i^a or G_i^b could be void, since our search scheme allows one-to-null correspondences. This approach does not aim to necessarily map every component of the two input shapes to a corresponding component. Moreover, considering all possible one-to-null correspondences is a challenging task, since it would significantly increase the number of possible cuts in a graph G_i. To address this challenge, we apply a heuristic to select only meaningful one-to-null correspondences — only noncritical groups may match a void component. Here we define a noncritical group as a group whose removal would not lead to a decrease of its plausibility score by more than 20%. In other words, we only consider the one-to-null option when G^a or G^b contains a noncritical group.

Searching for best matching hierarchy. All possible binary partitions of G_i into the connected graphs G_i^a and G_i^b are considered as the first child nodes connected directly to the root of a matching hierarchy. Since the graph partition requires that each subgraph is connected, the number of possible binary partitions is limited by an enumeration test. For every single correspondence pair $(p_{i_k}^0, p_{j_k}^1)$, all possible pairs of these binary partitions are evaluated with the plausibility measure described in the Subsection 4.1. Finally, the best k_{node} correspondence pairs are selected, resulting in k_{node} candidate hierarchies with two levels. In our experiments, we set $k = 5$. This node split processing would be applied recursively on every child node of each candidate hierarchy.

In the following iterations, the score of a candidate hierarchy is given by a combination of correspondence pair scores, since the candidate hierarchy usually has more than one leaf node. It is worth mentioning that by keeping only the top k_{tree} candidate hierarchies in each iteration are kept, the search space is significantly reduced.

At the end of this process, we select the matching hierarchy candidate with the best score among all the computed hierarchies. The output of this step is the information encoded in the leaf nodes of the optimal hierarchy, which is the fine-level correspondences between substructures from S_0 and S_1.

This optimal hierarchy can also be used to discover in-between shapes from S_0 to S_1. The closer to the leaf nodes we get, the closer the generated in-between shape would be to either input shape. As we examine closer to the root, we will find in-between shapes more balanced, with about half substructures originating from each input shape.

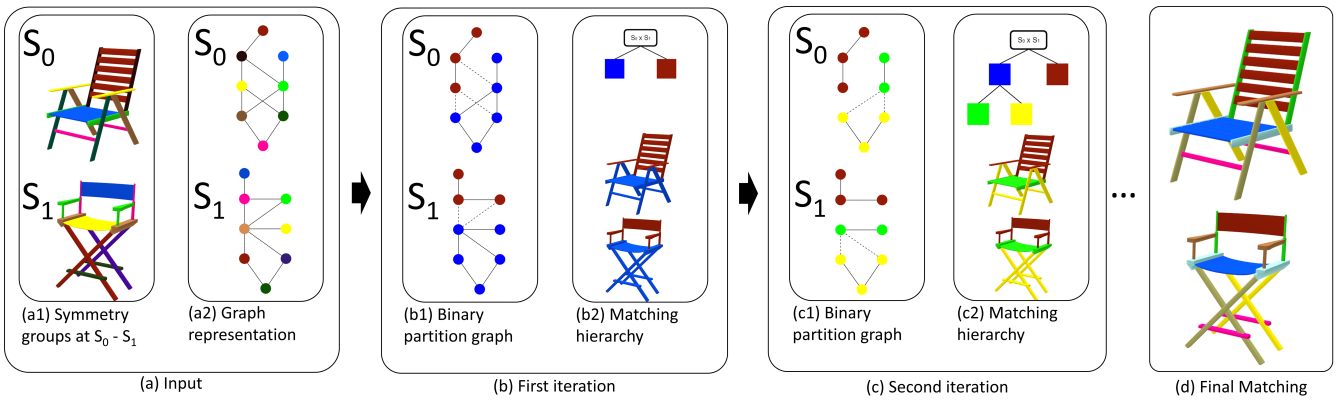


Fig. 8. Correspondence search, starting from two shapes whose symmetry groups are given (a1). The connectivity of the symmetric components is then represented as a connected graph (a2). The best cut at a graph is selected in a way to maximize the plausibility measure (b1), resulting in a matching hierarchy (b2). The algorithm runs recursively, decomposing each paired subgraph into two other subgraphs (c). Termination is achieved when every leaf node of the matching hierarchy cannot be further decomposed (d).

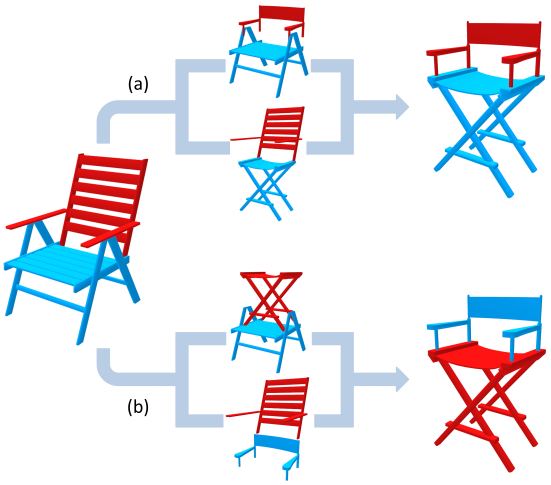


Fig. 9. In-between shapes obtained by deforming a source shape (left) to a target shape (right) based on two part (group) matchings, (a) and (b), signified by matching colors. The bifurcations in (a) and (b) represent two distinct deformation paths for each matching, resulting in four in-between shapes evaluated in total to rate two matchings. Matching (a) is better rated because of the higher plausibility, or chair-likeness, of its associated in-betweens.

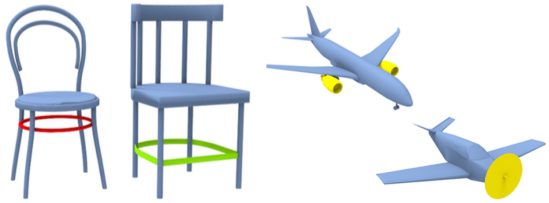


Fig. 10. Issues with GT setup via part labeling in GeoTopo. For the chair pair, the red part was labeled as "seat bar" and the green part as "leg bar", hence they were not matched in the GT. In the airplane pair, both yellow parts were labeled as engine in GeoTopo GT, yet no human users matched them when working on the pair.

5 RESULTS

In this section, we show extensive results and evaluation of our correspondence method including quantitative and visual comparisons to GeoTopo [Alhashim et al. 2015], the current best solution for fine-grained and topology-varying shape correspondence. For this particular problem, comparative studies by Alhashim et al. [2015] already demonstrated that GeoTopo performs better than state-of-the-art co-analysis methods including [Kim et al. 2013; Xu et al. 2012; Zheng et al. 2014]. Hence we only focus on comparisons to GeoTopo. Indeed, co-analysis methods are designed to analyze a set and amid diversity in geometry or structure among the set of the shapes, a set correspondence would necessarily return a coarse-level, as opposed to a fine-grained, matching across the set.

Dataset. Most of the 3D shapes used in our experiments are of man-made objects and gathered from GeoTopo. The total number of models is 185, spanning 5 object categories: chairs, tables, airplanes, beds, and velocipedes. Their part counts range between 4 and 35. Since GeoTopo requires initial segmentations of the input shapes, during our comparison, both GeoTopo and our method are provided with the same segmented input shapes from the GeoTopo data.

For data-driven plausibility estimation, the number of 3D shapes from ShapeNet [2015] considered for the five object categories for our quantitative experiments are: 6,778 chairs, 8,509 tables, 254 beds, 4,045 airplanes, and 59 velocipedes. For each category, it requires between 1 and 2 hours of off-line processing. Three views were selected to generate depth images of size 224×224 pixels, where each patch is of size 56×56 pixels.

Ground-truth (GT) correspondences. To evaluate the performance of our correspondence method and compare to GeoTopo, we asked human users to specify *pairwise* part correspondences to produce ground-truth (GT) data. Note that the way GT is set up in our work is different from GeoTopo, which matched fine-grained parts for

Fig. 11. A gallery of part correspondences computed by our algorithm (bottom pair of each two shape pairs compared) with comparisons to GeoTopo [Alhashim et al. 2015]. Matched parts share the same color; unmatched parts are in gray.

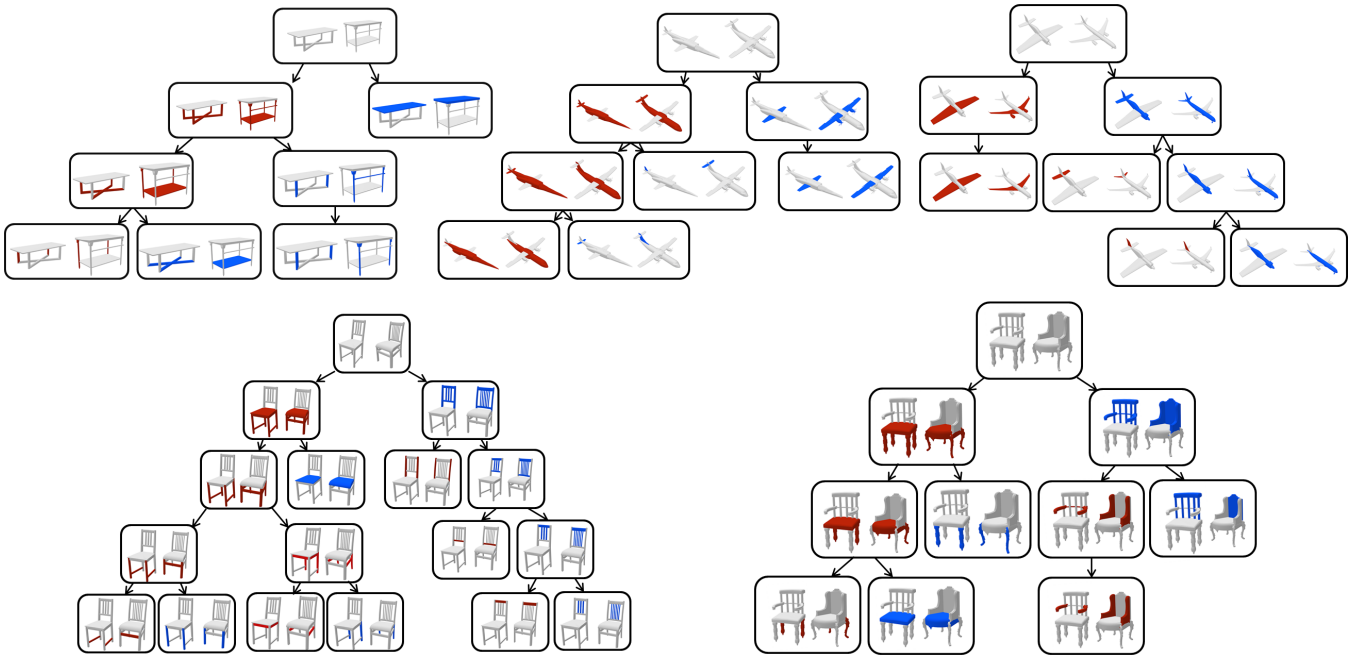
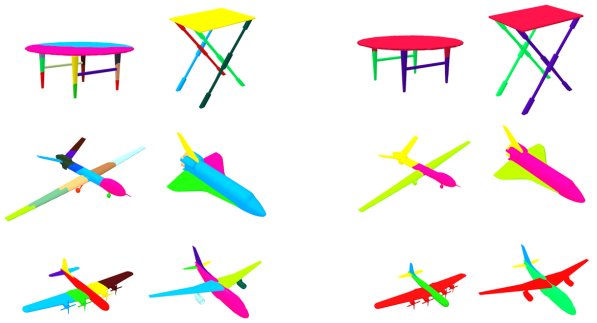


Fig. 12. Some hierarchical correspondence of input shape segments obtained by our method.

shapes within a given category *implicitly* by *independently* labeling the parts. The resulting GT may not be as reliable, as shown in Figure 10. We asked five users to manually produce fine-grained, topology-varying part correspondences on top of the given initial shape segmentations from GeoTopo. The resulting GT correspondences are used in two tests comparing our method to GeoTopo. Each user was asked to produce hundreds of correspondences between pairs of shapes and multiple users worked on each shape pair to obtain GT. At the end, we arrive at a single best GT correspondence for each shape pair by majority votes.



Initial segmentations Final correspondences

Fig. 13. Correspondence results computed by our method when the input segmentations are imperfect and not well-matched. With a hierarchical search, our method finds the correct correspondences at a coarser level.

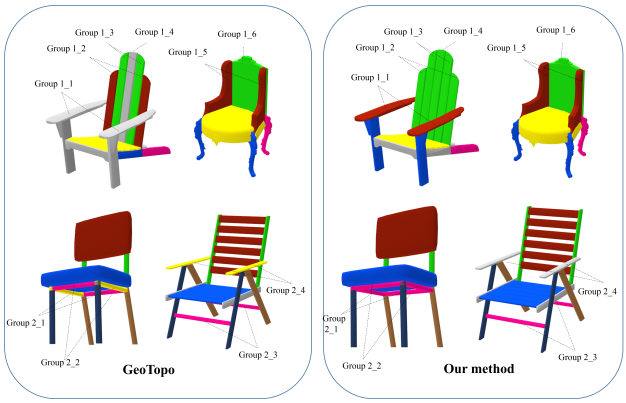


Fig. 14. GeoTopo (left) cannot produce many-to-one matchings due to its bottom-up search paradigm. As the two examples show, it cannot find the correct result when it involves many-to-one matchings (group 1_2, group 1_3 and group 1_4 cannot correspond to their corresponding group 1_6 in the top-left pairs). Sometimes their bottom-up search scheme would lead to further inaccuracies (in the bottom-left pair, group 2_3 corresponds to only one group, and miss another correspondence to group 2_1. This in turn leads to another wrong correspondence, namely group 2_1 and group 2_4). With a hierarchical search, our method performs well in such cases.

Visual correspondence results. Figure 11 shows a gallery of correspondence results comparing our method to GeoTopo. For a meaningful comparison, the results from our method in Figure 11 were all produced by first generating part groups from the initial segmentations used by GeoTopo on the input shapes. It is visually evident that our correspondence results exhibit higher levels of accuracy, especially when the structural variations between the two matched shapes become greater. This is also confirmed by numerical results on precision and recall reported in Table 1.

With our hierarchical correspondence search, by the time a final part correspondence is reached for two input shapes, a hierarchical correspondence between the shapes, as well as a hierarchical segmentation for each shape, can be obtained by backtracking the search process. Figure 12 shows some results.

Figure 13 shows how our method is able to deal with input segmentations that are imperfect and not well-matched. With the hierarchical search scheme, correct correspondences can be found at coarser levels of the search hierarchies and the plausibility measure would allow the search to terminate since unmatched segmentations at the finer level could not have led to a better correspondence.

Many-to-many matching. Figure 14 shows how GeoTopo's limitation to only searching for one-to-one matchings between parts or part groups can lead to sub-optimal correspondence results. On the other hand, our method's ability to produce many-to-many matchings is attributed to the hierarchical correspondence search paradigm; see figure caption for detailed explanations.

Quantitative comparison to GeoTopo. In this comparison, we conduct tests entirely on the benchmark dataset in GeoTopo, which consists of 75 models covering five object categories (chair, table, airplane, velocipedes, and bed). There are 15 models per category, resulting in $15 \times 14/2 = 105$ shape pairs to be matched. To assess the quality of the correspondences produced, we employ the precision+recall measures developed by GeoTopo. These measures are computed with respect to GT correspondences described above. Given a pair of shapes with a GT correspondence, i.e., a set of part pairs, G, let M be a set of part pairs computed by an algorithm

Category	GeoTopo		Ours		
	Precision	Recall	Precision	Recall	β
Chair	0.69	0.67	**0.83**	**0.83**	0.2
Table	0.63	0.61	**0.81**	**0.86**	0.2
Bed	0.60	0.62	**0.78**	**0.81**	0.3
Airplane	0.60	0.68	**0.80**	**0.85**	0.25
Velocipedes	**0.47**	0.44	0.43	**0.49**	0.35

Table 1. Numerical comparisons between our method (with parameters γ fixed at 0.1 and β set category-specifically) and GeoTopo on precision and recall rates measured against user-specified ground truth. The best rates are shown in bold. The table shows results run on input pairs and object categories tested in the original GeoTopo paper.

and $R = M \bigcap G$. Then, the precision rate is defined as $|R|/|M|$ and recall as $|R|/|G|$. When making comparisons, we report both rates averaged over the shape pairs tested.

It is interesting to observe that among the five object categories, velocipedes were the only one for which GeoTopo performs comparably to our method. In other words, there are essentially no obvious gains in using the data-driven plausibility. This situation could be explained by two reasons. First, the knowledge base for velocipedes (only 59 models) is rather limited, compared to other categories. A possible consequence is that the learned feature encoding for velocipedes is ineffective. Second, many of the parts in the velocipede objects are not visually prominent in depth images since they are either quite thin or small. As a result, the global visual similarity we employ as the plausibility measure may not be as discriminative for this category as for the others.

Statistics and parameters. Table 2 provides timing and other informative statistics obtained by running our algorithm on the model pairs shown in Figure 1. Aside from the plausibility measure, which

Input pair	$\#P_s$	$\#P_t$	$\#G_s$	$\#G_t$	t_C
Airplanes	11	8	7	5	5.32
Chairs	27	11	7	7	9.45
Tables	6	7	4	5	2.04

Table 2. Statistics and timing for our correspondence method, when applied to model pairs in Figure 1. We report various statistics which influence the search cost: the total number of fine-level parts in the source ($\#P_s$) and target ($\#P_t$) shapes, the total number of part groups ($\#G_s$ and $\#G_t$) to be matched. Execution times, t_C, for the correspondence search are in seconds and measured on an Intel(R) Core(TM) i7, 3.4GHz with 16 GB RAM.

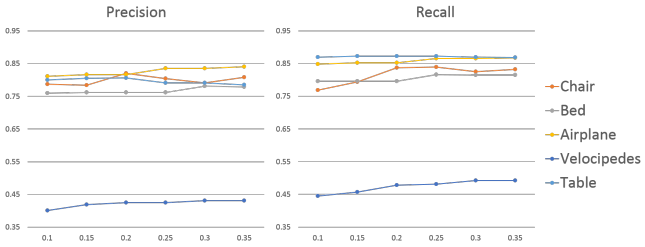


Fig. 15. Impact of tunable parameters: varying parameter β while keeping $\gamma = 0.1$ constant does not produce dramatic changes on the precision and recall rates resulting from our method.

is based on category-specific feature patches learned during preprocessing, there are two tunable parameters, β and γ. Our experiments indicate that changing γ had minimal impact on the precision and recall rates when keeping β constant. In Figure 15, we show how these rates are influenced as β changes while keeping a constant $\gamma = 0.1$, an empirically chosen value leading to the best results. As can be observed, varying β only led to small fluctuations in the results over all object categories tested.

Failure cases. Figure 16 reveals several failure cases generated by our method. While our data-driven plausibility measure generally works quite well, the structure restoration component in our deformation mechanism may produce in-betweens that are visually plausible for an imperfect correspondence. As discussed previously, when the input shapes are composed of many small parts, e.g., for the bicycles shown in the figure, our current plausibility measure can become inadequate. Other possible causes for failures may include discrepancies between symmetry characteristics (the table pair in the figure) and search bounds imposed by the beam width.

6 DISCUSSION, LIMITATIONS, AND FUTURE WORK

We position our work as an advance in solving the difficult problem of computing fine-grained, topology-varying correspondences between 3D shapes with geometric and topological variations. Such an advance is clearly demonstrated when comparing our results to those obtained by GeoTopo, the current state of the art.

The main novelty of our work is showing that the difficult shape correspondence problem can be solved by performing (many) shape recognition tasks. This is made possible within a data-driven and deformation-driven solution framework. Moreover, as a side product

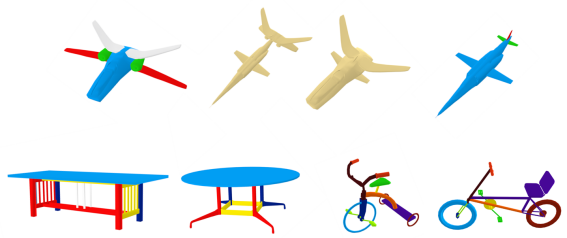


Fig. 16. Some failure cases obtained by our correspondence method. Top row: plausible in-betweens resulting from a less-than-perfect part correspondence. Bottom row: inability to infer proper fine-grained correspondences when the input shapes are composed of many small parts.

of our search paradigm, our method not only yields a fine-level part correspondence, but also a hierarchical one. It is worth noting however that there is no provision in our method to ensure that the hierarchical segmentation for each input shape, implied from the correspondence, is the most meaningful one.

All three key components of our problem, segmentation, correspondence search, and setting up objectives or evaluation criteria for the search, remain challenging. Our current solution is still limited in several aspects, leaving much room for future work.

Initial segmentation. A pre-requisite for GeoTopo to obtain a correct part correspondence is that the input segmentations must be perfect. In our work, this requirement is relaxed, since even if the initial, fine-level input segmentations are imperfect and not well-matched, our hierarchical search may still find a proper correspondence at a coarser level of the search hierarchy. However, it still has limitations, since the search cannot extrapolate the given input segmentations. An engaging future work would be to combine segmentation and correspondence search in a unified framework, so that as the search proceeds, the initial segmentation can be updated to achieve a better correspondence.

Man-made vs. organic shapes. Our correspondence scheme is designed to handle man-made shapes, not organic shapes like humans or creatures. This limitation is attributed to our choices of the deformation model and the use of multi-view depth images for shape similarity. Neither of these is well suited to handle organic shapes possessing free-form articulations. For these shapes, there have been recent efforts attempting to learn *intrinsic* invariants to compute correspondences [Masci et al. 2015; Rodolà et al. 2014].

Plausibility measure. Our current plausibility measure is based on visual similarity, while a better way to determine whether an in-between shape is "chair-like" should be based on judging its *functionality*. It will be interesting to employ the latest tools for functionality analysis, e.g., [Hu et al. 2016], for this task. Another issue to examine is how to select non-transient in-betweens to evaluate plausibility. In general, it may be more robust to estimate plausibility over multiple, well-selected in-betweens and combine the measures to rate a candidate correspondence. Our current method selects one and only one in-between that is deemed to be the most

plausible; this is an obvious choice, but may be too sensitive to any imperfections in the global similarity measure.

Future work. Aside from addressing the above limitations, we would also like to further speed up the data-driven correspondence evaluation, perhaps with a more selective subset of 3D shapes from the knowledge base. This would be possible if a fine-grained shape categorization [Huang et al. 2013] can be obtained and utilized. Partial matching is another area to explore, where a deformation-driven approach is still applicable, but the data-driven energy needs to be re-designed. Extending our approach to the co-analysis setting is also an interesting avenue to explore, especially when the problem is posed under a hierarchical shape co-analysis setting [van Kaick et al. 2013]. For a start, we shall take a closer look at the correspondence-driven hierarchical shape segmentations obtained by our current search scheme. Finally, an obvious question is whether deep learning techniques can be developed for fine-grained, topology-varying shape correspondence.

ACKNOWLEDGEMENTS

We would like to thank the reviewers for their valuable comments and feedback. We are grateful to Min Liu for his early help in the project. Thanks also go to Noa Fish and Oliver van Kaick for the fruitful discussions. This work is supported in part by grants from NSERC Canada (611770), China Scholarship Council, and NSF China (61572507, 61532003, 61622212).

REFERENCES

Ibraheem Alhashim, Honghua Li, Kai Xu, Junjie Cao, Rui Ma, and Hao Zhang. 2014. Topology-Varying 3D Shape Creation via Structural Blending. *ACM Trans. on Graphics* 33, 4 (2014), 158:1–158:10.

Ibraheem Alhashim, Kai Xu, Yixin Zhuang, Junjie Cao, Patricio Simari, and Hao Zhang. 2015. Deformation-driven Topology-varying 3D Shape Correspondence. *ACM Trans. on Graphics* 34, 6 (2015), 236:1–236:13.

Brett Allen, Brian Curless, and Zoran Popović. 2003. The Space of Human Body Shapes: Reconstruction and Parameterization from Range Scans. *ACM Trans. on Graphics* 22, 3 (2003), 587–594.

Aayush Bansal, Abhinav Shrivastava, Carl Doersch, and Abhinav Gupta. 2015. Mid-level elements for object detection. *arXiv preprint arXiv:1504.07284* (2015).

Volker Blanz and Thomas Vetter. 1999. A morphable model for the synthesis of 3D faces. In *Proc. SIGGRAPH*. 187–194.

Ding-Yun Chen, Xiao-Pei Tian, Yu-Te Shen, and Ming Ouhyoung. 2003. On visual similarity based 3D model retrieval. *Computer Graphics Forum (Proc. EUROGRAPHICS)* 22, 3 (2003), 223–232.

Navneet Dalal and Bill Triggs. 2005. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, Vol. 1. IEEE, 886–893.

Noa Fish, van Kaick, Amit Bermano, and Daniel Cohen-Or. 2016. Structure-oriented Networks of Shape Collections. *ACM Trans. on Graphics* 35, 6 (2016), 171:1–171:14.

Lin Gao, Yu-Kun Lai, Qixing Huang, and Shi-Min Hu. 2013. A Data-Driven Approach to Realistic Shape Morphing. *Computer Graphics Forum* 32, 2 (2013), 449–457.

Aleksey Golovinskiy and Thomas Funkhouser. 2009. Consistent segmentation of 3D models. *Computers & Graphics (Proc. SMI)* 33, 3 (2009), 262–269.

Ruizhen Hu, Oliver van Kaick, Bojian Wu, Hui Huang, Ariel Shamir, and Hao Zhang. 2016. Learning How Objects Function via Co-Analysis of Interactions. *ACM Trans. on Graphics* 35, 4 (2016), 47:1–47:13.

Qixing Huang, Vladlen Koltun, and Leonidas Guibas. 2011. Joint shape segmentation with linear programming. *ACM Trans. on Graphics* 30, 6 (2011), 125:1–125:12.

Qixing Huang, Fan Wang, and Leonidas Guibas. 2014. Functional map networks for analyzing and exploring large shape collections. *ACM Trans. on Graphics* 33, 4 (2014), 36.

Qi-Xing Huang, Hao Su, and Leonidas Guibas. 2013. Fine-grained Semi-supervised Labeling of Large Shape Collections. *ACM Trans. on Graphics* 32, 6 (2013), 190:1–190:10.

Michael Kazhdan, Thomas Funkhouser, and Szymon Rusinkiewicz. 2004. Symmetry descriptors and 3D shape matching. In *Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing*. 115–123.

Vladimir G. Kim, Wilmot Li, Niloy J. Mitra, Siddhartha Chaudhuri, Stephen DiVerdi, and Thomas Funkhouser. 2013. Learning Part-based Templates from Large Collections of 3D Shapes. *ACM Trans. on Graphics* 32, 4 (2013), 70:1–70:12.

Yanir Kleiman, Oliver van Kaick, Olga Sorkine-Hornung, and Daniel Cohen-Or. 2015. SHED: Shape Edit Distance for Fine-grained Shape Similarity. *ACM Trans. on Graphics* 34, 6 (2015), 235:1–235:11.

Hamid Laga, Michela Mortara, and Michela Spagnuolo. 2013. Geometry and Context for Semantic Correspondences and Functionality Recognition in Man-made 3D Shapes. *ACM Trans. on Graphics* 32, 5 (2013), 150:1–150:16.

Jonathan Masci, Davide Boscaini, Michael Bronstein, and Pierre Vandergheynst. 2015. Geodesic convolutional neural networks on Riemannian manifolds. In *Proc. of ICCV Workshops*. 37–45.

Niloy Mitra, Michael Wand, Hao (Richard) Zhang, Daniel Cohen-Or, Vladimir Kim, and Qi-Xing Huang. 2013. Structure-aware Shape Processing. In *SIGGRAPH Asia 2013 Courses*. 1:1–1:20.

Alain Pitiot, Herve Delingette, and Paul Thompson. 2007. Learning Shape Correspondence for n-D curves. *Int. J. Computer Vision*. 71, 1 (2007), 71–88.

Joshua Podolak, Philip Shilane, Aleksey Golovinskiy, Szymon Rusinkiewicz, and Thomas Funkhouser. 2006. A planar-reflective symmetry transform for 3D shapes. *ACM Trans. on Graphics* 25, 3 (2006), 549–559.

Charles Qi, Hao Su, Matthias Niessner, Angela Dai, Mengyuan Yan, and Leonidas Guibas. 2016. Volumetric and Multi-View CNNs for Object Classification on 3D Data. In *Proc. IEEE Conf. on CVPR*. 5648–5656.

Emanuele Rodolà, Samuel Rota Bulo, Thomas Windheuser, Matthias Vestner, and Daniel Cremers. 2014. Dense non-rigid shape correspondence using random forests. In *Proc. IEEE Conf. on CVPR*. 4177–4184.

T. W. Sederberg and E. Greenwood. 1992. A physically based approach to 2-D shape blending. In *Proc. SIGGRAPH*. 25–34.

Oana Sidi, Oliver van Kaick, Yanir Kleiman, Hao Zhang, and Daniel Cohen-Or. 2011. Unsupervised Co-Segmentation of a Set of Shapes via Descriptor-Space Spectral Clustering. *ACM Trans. on Graphics* 30, 6 (2011), 126:1–126:9.

Saurabh Singh, Abhinav Gupta, and Alexei Efros. 2012. Unsupervised discovery of mid-level discriminative patches. In *Proc. Euro. Conf. on Comp. Vis. (ECCV)*. 73–86.

Hao Su, Manolis Savva, Li Yi, Angel X. Chang, Shuran Song, Fisher Yu, Zimo Li, Jianxiong Xiao, Qixing Huang, Silvio Savarese, Thomas Funkhouser, Patrick Hanrahan, and Leonidas J. Guibas. 2015. ShapeNet: An Information-Rich 3D Model Repository.

Art Tevs, Qixing Huang, Michael Wand, Hans-Peter Seidel, and Leonidas Guibas. 2014. Relating Shapes via Geometric Symmetries and Regularities. *ACM Trans. on Graphics* 33, 4 (2014), 119:1–119:12.

Jasper RR Uijlings, Koen EA van de Sande, Theo Gevers, and Arnold WM Smeulders. 2013. Selective search for object recognition. *Int. J. Computer Vision*. 104, 2 (2013), 154–171.

Oliver van Kaick, Noa Fish, Yanir Kleiman, Shmuel Asafi, and Daniel Cohen-Or. 2014. Shape Segmentation by Approximate Convexity Analysis. *ACM Trans. on Graphics* 34, 1 (2014), 4:1–4:11.

Oliver van Kaick, Andrea Tagliasacchi, Oana Sidi, Hao Zhang, Daniel Cohen-Or, Lior Wolf, and Ghassan Hamarneh. 2011. Prior Knowledge for Shape Correspondence. *Computer Graphics Forum* 30, 2 (2011), 553–562.

Oliver van Kaick, Kai Xu, Hao Zhang, Yanzhen Wang, Shuyang Sun, Ariel Shamir, and Daniel Cohen-Or. 2013. Co-Hierarchical Analysis of Shape Structures. *ACM Trans. on Graphics* 32, 4 (2013), 69:1–69:10.

Oliver van Kaick, Hao Zhang, Ghassan Hamarneh, and Daniel Cohen-Or. 2010. A Survey on Shape Correspondence. *Computer Graphics Forum* 30, 6 (2010), 1681–1707.

Yanzhen Wang, Kai Xu, Jun Li, Hao Zhang, Ariel Shamir, Ligang Liu, Zhiquan Cheng, and Yueshan Xiong. 2011. Symmetry Hierarchy of Man-Made Objects. *Computer Graphics Forum* 30, 2 (2011), 287–296.

Kai Xu, Vladimir Kim, Qixing Huang, and Evangelos Kalogerakis. 2016. Data-Driven Shape Analysis and Processing. *Computer Graphics Forum* (2016), to appear.

Kai Xu, Honghua Li, Hao Zhang, Daniel Cohen-Or, Yueshan Xiong, and Zhiquan Cheng. 2010. Style-Content Separation by Anisotropic Part Scales. *ACM Trans. on Graphics* 29, 6 (2010), 184:1–184:9.

Kai Xu, Hao Zhang, Daniel Cohen-Or, and Baoquan Chen. 2012. Fit and Diverse: Set Evolution for Inspiring 3D Shape Galleries. *ACM Trans. on Graphics* 31, 4 (2012), 57:1–10.

Lihi Zelnik-Manor and Pietro Perona. 2004. Self-tuning spectral clustering. In *NIPS*. 1601–1608.

Hao Zhang, Alla Sheffer, Daniel Cohen-Or, Qingnan Zhou, Oliver van Kaick, and Andrea Tagliasacchi. 2008. Deformation-Driven Shape Correspondence. *Computer Graphics Forum (Proc. SGP)* 27, 5 (2008), 1431–1439.

Youyi Zheng, Daniel Cohen-Or, Melinos Averkiou, and Niloy J. Mitra. 2014. Recurring Part Arrangements in Shape Collections. *Computer Graphics Forum* 33, 2 (2014), 115–124.

Received January 2017; revised April 2017; final version April 2017; accepted April 2017