# MATCHING DISSIMILAR SHAPES

by

Oliver van Kaick

M.Sc., Federal University of Paraná, 2005

B.Sc., Federal University of Paraná, 2003

A Thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
in the
School of Computing Science
Faculty of Applied Sciences


SIMON FRASER UNIVERSITY
Fall 2011

## APPROVAL

| | |
|---|---|
| **Name:** | Oliver van Kaick |
| **Degree:** | Doctor of Philosophy |
| **Title of Thesis:** | Matching dissimilar shapes |
| **Examining Committee:** | Dr. Torsten Möller<br>Chair |
| | Dr. Hao (Richard) Zhang, Senior Supervisor |
| | Dr. Ghassan Hamarneh, Senior Supervisor |
| | Dr. Greg Mori, SFU Examiner |
| | Dr. Michael Kazhdan, External Examiner,<br>Department of Computer Science,<br>Johns Hopkins University |
| **Date Approved:** | 20 December 2011 |

# Declaration of
# Partial Copyright Licence

# Abstract

In this thesis, we address the challenge of computing correspondences between dissimilar shapes. This implies that, although the shapes represent the same class of object, there can be major differences in the geometry, topology, and part composition of the shapes as a whole. Additionally, the dissimilarity can also appear in the form of a shape that possesses additional parts that are not present in the other shapes. We propose three approaches for handling such shape dissimilarity. The first two approaches incorporate additional knowledge that goes beyond a direct geometric comparison of the shapes. In the first approach, of a supervised nature, the knowledge is provided by the user as a training set of manually segmented and labeled shapes. The training set is used in conjunction with shape descriptors to learn classifiers that distinguish different semantic classes of parts. The second approach, which is unsupervised, derives the knowledge automatically from a *set* of shapes. If all the shapes in the set roughly possess the same semantic part composition, we can derive their common structure by analyzing the shapes simultaneously, rather than individually. This is achieved by clustering shape segments in a descriptor space with a spectral method, which makes use of *third-party connections* between shape parts. We show that these approaches allow us to compute correspondences for shapes that differ significantly in their geometry and topology, such as man-made shapes. In the third approach, we compute *partial correspondences* between shapes that have additional parts in relation to each other. To address this challenge, we propose a new type of shape descriptor, called the *bilateral map*, whose region of interest is defined by *two* points. The region of interest adapts to the context of the two points and facilitates the selection of the scale and shape of this region, making this descriptor more effective for partial matching. We demonstrate the advantages of the bilateral map for computing partial and full correspondences between pairs of shapes.

**Keywords:** shape correspondence; co-segmentation; prior knowledge; shape descriptor

# Acknowledgments

I would like to thank all the people who helped me during this arduous journey which culminated in the successful defense of my PhD thesis. First of all, I would like to deeply thank my supervisors Dr. Hao (Richard) Zhang and Dr. Ghassan Hamarneh for their support, encouragement, criticism, and guidance throughout these years. Here, I would also like to extend my sincere thanks to Dr. Daniel Cohen-Or, who also supervised my work and provided various insightful suggestions on all the chapters of this thesis. It was a pleasure to work with this group of outstanding individuals who taught me many important points of being a good researcher and guided me to create work with which I am greatly satisfied. My gratitude also goes to my collaborators Oana Sidi, Yanir Kleiman, Dr. Lior Wolf, and Andrea Tagliasacchi, who were involved in the work presented in this thesis. I would also like to thank my examiners Michael Kazhdan and Greg Mori for reading my thesis and providing many thoughtful comments and suggestions. Moreover, I would also like to thank my colleagues at the GRaphics, Usability, and VIsualization (GrUVi) and Medical Image Analysis (MIAL) labs. I benefited from many fruitful discussions and reading groups with the lab members, while also enjoying their company during work and leisure times. A special thanks also to my friends who helped me to keep a different perspective on life. Finally, I am also grateful for all the support, help, and unconditional love that I received from my family, Baldur, Ilse, Joana, Christian and Patricia. Especially, Hisako, I thank you for your patience, support, love and understanding during all times.

# Contents

**Approval** ii

**Abstract** iii

**Acknowledgments** iv

**Contents** v

**List of Tables** vii

**List of Figures** viii

**Preface** x

**1 Introduction** 1

1.1 Overview . . . 3

1.1.1 Knowledge-driven correspondence methods . . . 3

1.1.2 Prior knowledge for part correspondence . . . 4

1.1.3 Unsupervised co-segmentation of a set of shapes . . . 5

1.1.4 Bilateral maps for partial matching . . . 7

1.2 Contributions . . . 10

1.3 Organization . . . 11

**2 Background and related work** 12

2.1 Problem definition . . . 12

2.2 Applications of correspondence . . . 15

2.3 Correspondence problems and related methods . . . 17

2.3.1 Similarity-based correspondence . . . 18
2.3.2 Rigid alignment . . . 20
2.3.3 Non-rigid alignment . . . 21
2.3.4 Time-varying registration . . . 22
2.4 Discussion . . . 23

**3 Prior knowledge for part correspondence 25**
3.1 Related work . . . 28
3.2 Overview . . . 29
3.3 Prior knowledge and probabilistic semantic labeling . . . 32
3.4 Part correspondence via joint labeling . . . 34
3.5 Experimental results . . . 37
3.6 Discussion . . . 43

**4 Unsupervised co-segmentation of a set of shapes 45**
4.1 Related work . . . 47
4.2 Overview . . . 52
4.3 Descriptors and per-object segmentation . . . 54
4.4 Descriptor-space spectral clustering . . . 55
4.5 Statistical model and co-segmentation . . . 57
4.6 Experimental Results . . . 59
4.7 Discussion . . . 66

**5 Bilateral maps for partial matching 69**
5.1 Related work . . . 72
5.2 Descriptor construction . . . 74
5.3 Experiments and results . . . 78
5.3.1 Shape correspondence . . . 78
5.3.2 Shape retrieval. . . . 86
5.4 Discussion . . . 87

**6 Conclusion and future directions 89**

**Bibliography 92**

# List of Tables

3.1 Statistical evaluation of the correspondence results . . . . . 41

4.1 Average co-analysis labeling accuracy . . . . . 62

5.1 Retrieval results on a dataset of complete and incomplete shapes . . . . . 86

# List of Figures


1.1 Correspondence between dissimilar man-made shapes . . . . . 2
1.2 Part correspondence results via our supervised approach . . . . . 4
1.3 Unsupervised co-segmentation . . . . . 5
1.4 Results of our unsupervised co-segmentation approach . . . . . 6
1.5 An example of partial matching . . . . . 7
1.6 Bilateral maps for partial matching . . . . . 8
1.7 Correspondence results obtained with our bilateral maps . . . . . 9

2.1 A meaningful correspondence between a sparse set of feature points on two shapes . . . . . 13
2.2 An example of a partial correspondence problem . . . . . 14
2.3 An example of a collection of man-made shapes (liquid containers) for which computing a correspondence is a challenging problem . . . . . 15
2.4 Different manners of solving the correspondence problem for two input shapes and their feature points: computing a correspondence without explicitly bringing the shapes into alignment, computing a global rigid transformation to align the two shapes, and computing local non-rigid transformations for the shape primitives to deform one shape into the other . . . . . 18
2.5 A progression of development in shape correspondence methods . . . . . 23

3.1 Meaningful correspondence between shape parts under significant geometric and topological discrepancies is made possible by incorporating prior knowledge 26
3.2 The set of man-made shapes used in our work . . . . . 27
3.3 Probabilistic semantic labeling . . . . . 30
3.4 Joint labeling . . . . . 34

3.5 Part correspondence results via joint labeling . . . . . 38
3.6 Joint labeling improves upon the use of prior knowledge alone . . . . . 39
3.7 Comparison to content-driven correspondence approach on geometrically dissimilar models . . . . . 42

4.1 Unsupervised co-segmentation of a highly varied set of container objects using our algorithm . . . . . 46
4.2 Challenge of unsupervised co-segmentation amid significant geometric variation and effectiveness of descriptor-space spectral clustering . . . . . 48
4.3 Co-segmentation in descriptor space vs. in spatial domain . . . . . 49
4.4 The power of third-party connections . . . . . 50
4.5 Overview of the steps in our co-analysis . . . . . 51
4.6 Result of the co-segmentation refinement applied on an initial co-segmentation 53
4.7 Results of our co-segmentation on a variety of shapes . . . . . 61
4.8 The power of the set demonstrated for all the classes . . . . . 63
4.9 Comparison to the approach in [Golovinskiy and Funkhouser 2009] . . . . . 65

5.1 Bilateral maps for partial matching . . . . . 70
5.2 Bilateral map construction . . . . . 71
5.3 Definition of the region of interest for the bilateral map . . . . . 75
5.4 Lesser sensitivity of the bilateral maps to topological changes . . . . . 76
5.5 Correspondence results obtained with our bilateral maps on a set of examples 79
5.6 Correspondence results obtained with the bilateral maps, compared to the results obtained with the single-point geodesic maps . . . . . 81
5.7 Examples of lesser sensitivity of the bilateral approach to topological changes 83
5.8 Quantitative evaluation of correspondences computed with the bilateral maps and geodesic maps at five different scales on a set of incomplete humans . . . 84

# Preface

The ideas presented in this thesis resulted from our observations in previous projects. I had chosen shape correspondence as the topic for my PhD research and worked on a project where we used ant colony optimization to compute correspondences between 2D contours [van Kaick et al. 2007]. We optimized what is now considered a *classical* objective function for correspondence: a function that seeks to maximize the similarity between shape descriptors and the distances between pairs of points (an isometry criterion).

However, it became clear from this project that to compute correspondences between dissimilar shapes, involving significant differences in geometry and topology, just comparing the geometry of the shapes is not sufficient. Thus, the idea arose of using some form of prior or additional knowledge to establish the correspondences. It was not obvious, at first, how to represent such knowledge, since a correspondence is a relation or mapping from one shape to another. However, we noticed that shape correspondence based on prior knowledge could be formulated as the problem of *recognizing* parts of the shapes.

With this realization, works such as *Inducing Semantic Segmentation from an Example* [Schnitman et al. 2006] served as an inspiration for developing our project on supervised part correspondence, where we use a training set of labeled shapes to establish correspondences between pairs of unknown shapes. Next, we verified that the need for a training set could be lifted by using the knowledge inherent to a set of shapes of the same class. The knowledge of the set is used to perform an unsupervised co-segmentation of the shapes.

Finally, it still remained a challenge to compute correspondences between pairs of dissimilar shapes without the aid of a set. In this case, we can assume that there is at least a partial match between the shapes. However, the difficulty then lies in the local shape descriptors, which are not well adapted for partial matching. Thus, the bilateral map was developed as a promising descriptor more suited for cases requiring partial correspondences.

This thesis includes material that was previously published in three separate papers and one additional paper that is currently under review. These papers correspond to the following chapters in the thesis:

- Chapter 2 contains a review of related work on shape correspondence. Parts of this chapter appeared in the paper "A Survey on Shape Correspondence", *Computer Graphics Forum*, 2011 [van Kaick et al. 2011a].
- Chapter 3 describes the supervised approach for part correspondence, which appeared in the paper "Prior Knowledge for Part Correspondence", *Computer Graphics Forum (Proc. Eurographics)*, 2011 [van Kaick et al. 2011b].
- Chapter 4 discusses the unsupervised approach for co-segmentation of a set of shapes, which appeared in "Unsupervised Co-Segmentation of a Set of Shapes via Descriptor-Space Spectral Clustering", *ACM Trans. on Graphics (Proc. SIGGRAPH Asia)*, 2011 [Sidi et al. 2011]. Note that this is a paper with equal contribution from the first two authors.
- Chapter 5 describes a new type of shape descriptor conceived for partial matching tasks; the *bilateral maps*. The content of this chapter also appears in a publication currently under review.

A few additional papers also resulted from this research on shape correspondence and related areas. However, these papers are not included in this work since their focus does not fall under the scope of the thesis:

- "Contour Correspondence via Ant Colony Optimization", *Proc. Pacific Graphics*, 2007 [van Kaick et al. 2007], where correspondences between 2D contours are computed by making use of the *ant colony* metaheuristic.
- "Learning Fourier Descriptors for Computer-Aided Diagnosis of the Supraspinatus", *Academic Radiology*, 2010 [van Kaick et al. 2010], where machine learning and local shape descriptors are used to classify muscles into different pathology classes.

# Chapter 1

# Introduction

Establishing a meaningful correspondence between two or more shapes is a fundamental task in several fields, including computer graphics, computer vision, and medical imaging. The problem consists in finding a meaningful mapping or relation between the elements of the shapes. Shape correspondence has been commonly studied in the context of mapping feature points detected on the shapes, however, it can also be posed as the problem of finding which *parts* of two shapes are related to each other. For example, in Figure 1.1 (a), the correspondence correctly maps segments representing the handles, bases and bodies of the vases across the pair. We call such a result a *part correspondence* or *coherent segmentation* (*co-segmentation*) of the shapes. Note that we can obtain a co-segmentation of a pair of shapes (Figure 1.1) or a set of several shapes (Figure 1.4).

Shape correspondence is a relevant problem since it lies at the core of crucial tasks such as surface registration [Gelfand et al. 2005; Jain et al. 2007; Aiger et al. 2008; Chang and Zwicker 2008] and reconstruction [Mitra et al. 2007; Wand et al. 2007; Sharf et al. 2008; Pekelny and Gotsman 2008]. It is also a requirement of several important applications, such as deformation [Sumner and Popović 2004], texture [Dinh et al. 2005], and style transfer [Xu et al. 2010], shape morphing [Alexa 2002], statistical shape modeling [Davies et al. 2008], and change detection [Mirzaalian et al. 2009]. Establishing a correspondence is also one option for solving the tasks of shape retrieval [Funkhouser and Shilane 2006; Tangelder and Veltkamp 2008] and recognition [Forsyth and Ponce 2003].

Traditional methods for computing a correspondence assume that the shapes are *similar*. We call these methods *content-driven*, since they compute a solution solely by comparing the geometry and structure of the shapes. One category of content-driven approaches obtain

(a) (b)

Figure 1.1: Correspondence between dissimilar man-made shapes. (a) A meaningful correspondence between shape parts is shown, where corresponding parts are implied by matching colors. Note the significant geometric and topological differences between the two shapes. The correspondence was obtained with our supervised knowledge-driven approach. (b) A feature-to-feature correspondence, where the matched feature points are shown with corresponding colors. The implied alignment is shown to the far right. This result was obtained with a state-of-the-art feature matching method (the deformation-driven method of Zhang et al. [2008]), which is not able to handle such a challenging case of dissimilar shapes.

a correspondence by *registering* the shapes: the two shapes are first aligned and then the correspondence is derived from the proximity of the aligned shapes. This can be done in a rigid [Rusinkiewicz and Levoy 2001; Gelfand et al. 2005; Aiger et al. 2008] or non-rigid [Elad and Kimmel 2003; Anguelov et al. 2004; Jain et al. 2007; Chang and Zwicker 2008] manner. On the other hand, the most common approach for more general shape correspondence is based on the *feature matching* approach, where feature points are detected on the shapes and matched according to their geometric similarity. This can be done by comparing local shape descriptors extracted at the feature points, by using an approximate isometry criterion, or a combination of both [Anguelov et al. 2004; Funkhouser and Shilane 2006; Bronstein et al. 2008b; Zhang et al. 2008; Huang et al. 2008; Tevs et al. 2009; Lipman and Funkhouser 2009; Kim et al. 2010]. Structural approaches proceed similarly by matching features on a *skeleton* of the shapes [Au et al. 2010; Shapira et al. 2008; Tang and Hamarneh 2008].

Content-driven shape correspondence is, of course, a reasonable and effective approach when the shapes possess enough similarity. However, a common requirement in recent applications is to compute correspondences between *dissimilar* shapes. That is, shapes that represent the same type of object or which possess common portions, but that are highly

dissimilar when compared as a whole. This can be the case when establishing correspondences between several objects of the same semantic category that are available in an on-line shape repository, or when we would like to simultaneously analyze a set of varied shapes. In such challenging cases involving shapes that are dissimilar, both in geometry and topology, content-driven methods are less effective, since traditional criteria commonly used in the solution search are less meaningful. For example, assumptions such as the similarity of shape descriptors or approximate isometry break down, since the large amount of variation in the shapes significantly reduces the usefulness of such criteria (Figure 1.1).

## 1.1 Overview

In this thesis, we propose methods for computing correspondences between *dissimilar* shapes, implying that these shapes can have pronounced differences in their geometry, topology and part composition. First, we propose two knowledge-driven methods that incorporate additional information about the shapes into the correspondence computation. These methods depart from traditional correspondence approaches and are able to match shapes with significant variability, which is typical of man-made objects. Next, we propose a new shape descriptor, the *bilateral map*, that is better suited for partial matching tasks and can be used with traditional feature matching approaches. The bilateral map is able to handle shapes that are dissimilar because they possess additional parts that should be ignored in the matching. We discuss these contributions in the following subsections.

### 1.1.1 Knowledge-driven correspondence methods

As motivated in Figure 1.1, traditional content-driven methods do not perform well in the task of matching dissimilar shapes. To solve such challenging correspondence cases, it becomes necessary to perform a *semantic analysis* of the shapes. That is, if we are able to detect and *recognize* the parts that compose the shapes, then we can also establish their correspondence. In this thesis, we propose to perform a semantic analysis of the shapes with the aid of additional *knowledge* about the shapes. We introduce two different methods to incorporate knowledge into the correspondence computation, obtaining one supervised and one unsupervised *knowledge-driven* approach.

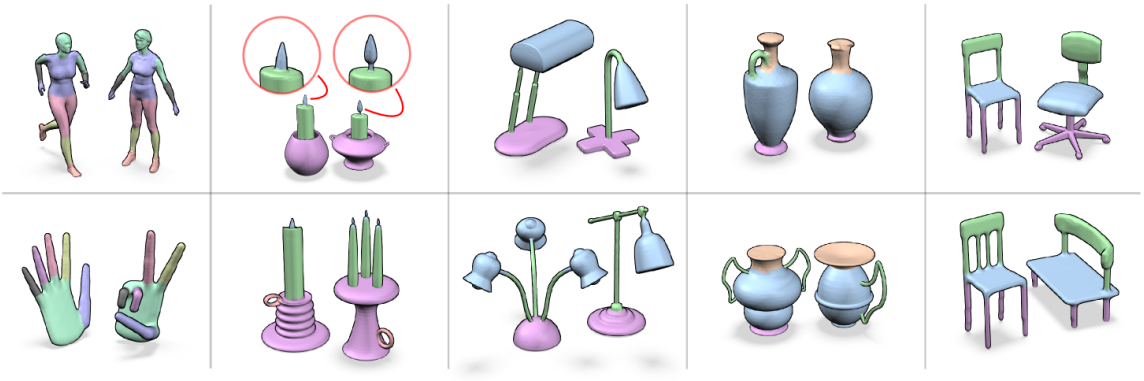


Figure 1.2: Part correspondence results via our supervised approach. Notice that our method succeeds even under significant geometric and topological dissimilarity between parts, while purely content-driven methods would fail.

### 1.1.2 Prior knowledge for part correspondence

In the supervised approach, prior knowledge about the class of shapes under consideration is given by a training dataset of manually segmented and labeled shapes. Thus, the knowledge is explicitly modeled by a human operator and provided in advance to the algorithm. Next, the correspondence between two unknown query shapes is obtained by comparing their parts with the parts of shapes in the dataset, and also by incorporating cues derived from the direct similarity of the shapes.

This comparison is done in practice by extracting shape descriptors from the training shapes and learning classifiers to distinguish the different classes of parts. Once we are given a pair of query shapes for which we want to compute a correspondence, we extract the same collection of shape descriptors and use the classifiers to perform a probabilistic semantic labeling of the shapes. We also add content information into the approach in the form of pairwise correspondences between faces with similar descriptors. Finally, the pairwise assignments are combined with the probabilistic labels and the connectivity of the meshes to perform a *joint labeling*, where we assign deterministic labels to the shapes according to a cost function. The part correspondence is then inferred from segments that were assigned the same label, as shown in Figure 1.1 (a).

Selected results of using this approach on a collection of organic and man-made shapes are shown in Figure 1.2. Notice the meaningful correspondences that are obtained for shapes

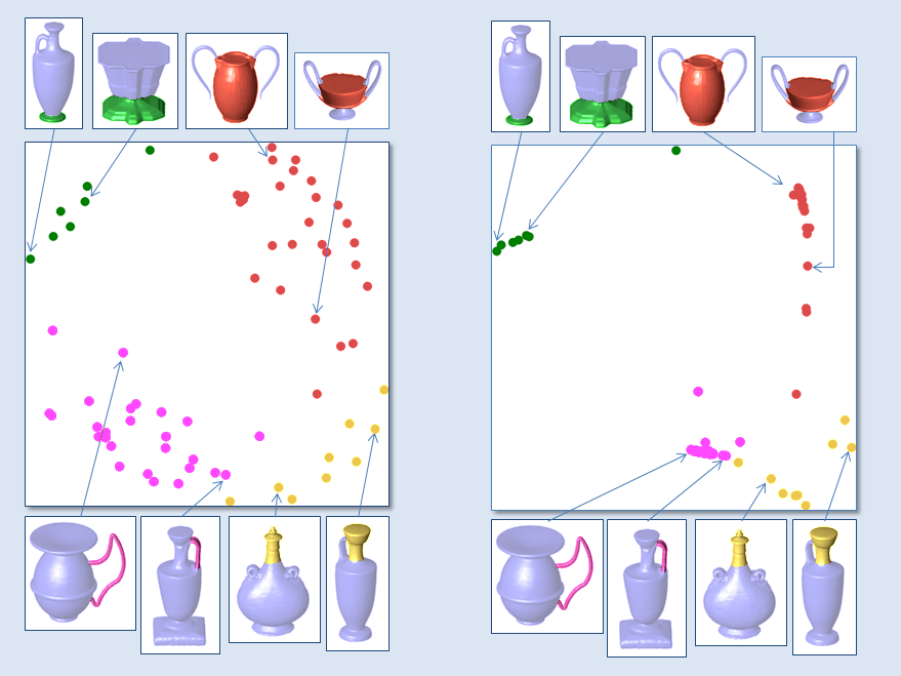


Figure 1.3: Unsupervised co-segmentation. In the original descriptor space (left), two segments in the same semantic class (two pink handles) can be far apart, while unrelated segments (a pink handle and a yellow neck) can be close. It is challenging to cluster this space without any knowledge of the semantic classes. However, the handles are drawn close in the diffusion map (right) through *third-party* connections. The third parties, which are all the segments lying in-between the two handles, establish several paths between the two segments, given by the high similarities between pairs of points. These multiple paths create a strong connection between the two handles.

with significant variations in their geometry (different shapes of light bulbs, vase bodies and necks, and candelabra parts) topology (chair legs, light bulb supports), and part composition (one vs. three candles, one vs. two vase handles). We discuss this method in Chapter 3.

### 1.1.3 Unsupervised co-segmentation of a set of shapes

In the unsupervised approach, the knowledge is not given explicitly by the user, but rather derived automatically from the shapes. If we are given a set of shapes that roughly belong to the same semantic class, it appears obvious that more knowledge can be inferred by analyzing the set as a whole, rather than analyzing each shape individually. Thus, this

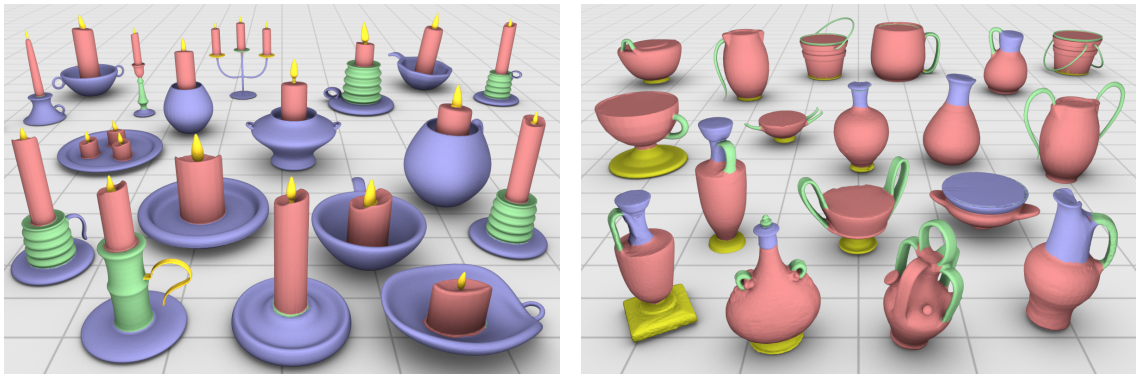


Figure 1.4: Results of our unsupervised co-segmentation approach. Corresponding segments in each class are shown with the same color. Notice how the segmentation and labeling is coherent for many of the parts in each set.

approach takes advantage exactly of this observation: We can extract knowledge about the shapes in an automatic manner by analyzing the shapes simultaneously and inferring their common structure. Next, the extracted knowledge can be used to obtain a co-segmentation of the shapes in the set.

More specifically, we start by obtaining an initial segmentation of the shapes in the set and computing a collection of shape descriptors for the resulting segments. Next, we perform clustering in the descriptor space to obtain the classes of potential shape parts.

However, performing the clustering directly in the descriptor space is problematic, since unrelated segments can be quite close, while two segments in the same class can be far away. This is illustrated in Figure 1.3. The left of the figure shows a 2D embedding of the descriptor collection for a set of example segments. As we can see, a simple clustering algorithm will not be able to find meaningful clusters in such a space. To resolve this problem, we first compute a diffusion map of the space of combined shape descriptors. Diffusion maps are a form of spectral embedding [Nadler et al. 2005] which unfold the non-linear and anisotropic structures that exist in the data. Then, a simple clustering algorithm can succeed by working in the embedded space (Figure 1.3, right). Unfolding the non-linear clusters is possible because of the *third-party connections* that exist in the data. That is, even if two shapes possess parts that are significantly dissimilar, we can still establish a link between them if there are other parts in the set (third parties) that create such a connection.

Finally, once the clusters of parts are obtained, we create a statistical model for each

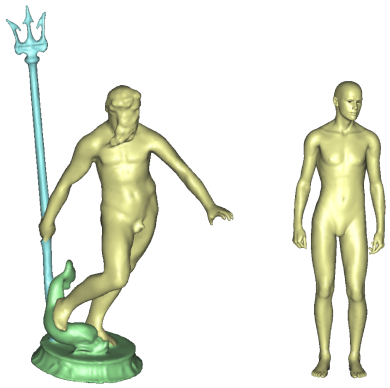


Figure 1.5: An example of partial matching. Our goal is to establish a correspondence between Neptune's statue (left) and the human (right), by matching the yellow parts and ignoring the extraneous parts shown in green and blue.

part class and use it to compute the final co-segmentation of the set. This is achieved by performing a graph cuts labeling of the shapes with the aid of the statistical models and traditional segmentation cues, such as the strength of concavities. The result is a segmentation of all the shapes that is consistent across the set (a co-segmentation), which also implies a correspondence between any pair of shapes.

Figure 1.4 shows results of our co-segmentation approach on two sets of man-made shapes. Notice that the important classes of parts are detected: candles, flames, and holders of the candelabra; and bases, bodies, handles, and necks of the vases. Notice also how the segmentations are consistent across each set. We describe this method in Chapter 4.

### 1.1.4 Bilateral maps for partial matching

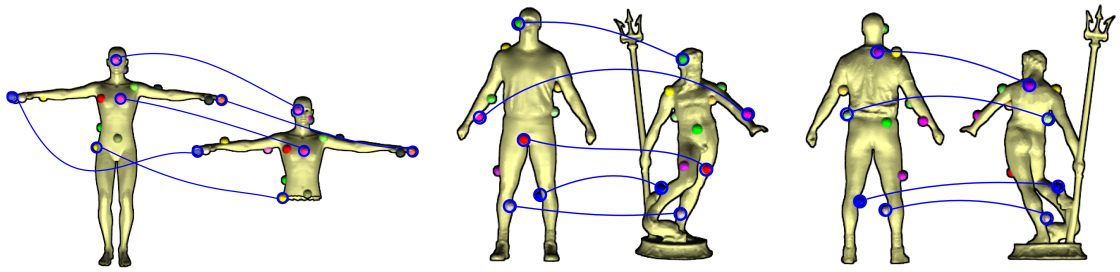
One specialization of the correspondence problem occurs when we require a correspondence only between the common portions of two shapes, as in the example shown in Figure 1.5. This can be a requirement when one of the shapes has additional or extraneous parts in relation to the other shape, and these parts should not be considered in the correspondence. For example, this is the case when dealing with incomplete shapes, shapes composed of a mixture of parts from multiple classes, or also when matching 3D scans that only partially overlap, such as different views of the same object.

(a) (b) (c)

Figure 1.6: Bilateral maps for partial matching. Descriptor bins are shown with alternating colors and the gray portions of the shapes are not part of the region of interest. Notice how the region captured between the head and right hand on both shapes is similar in (a) and (b), and does not include extraneous parts of the Neptune model, i.e., the spear and base. Compare to the traditional geodesic map for the right hand of the Neptune in (c), where extraneous parts are included even when using a reduced radius of coverage.

Partial matching is generally regarded as a harder problem than computing a full correspondence since it can be seen as composed of two sub-problems: first, we need to find the common parts of the shapes, and then we have to establish a correspondence between these parts [Gal and Cohen-Or 2006; Zhang et al. 2008]. As a consequence, this increases the size of the search space and also makes it difficult to define a proper objective function that describes our correspondence goal.

In this context, we can also use a feature matching approach to establish a correspondence, where detected feature points are characterized by local shape descriptors. A variety of such descriptors have been proposed in the literature, however, the most effective descriptors represent the *context* around a point, as in the popular shape context descriptor for images [Belongie et al. 2002]. This descriptor is commonly extended to 2D manifolds by laying out a grid on the surface around a point [Gatzke et al. 2005; Kalogerakis et al. 2010]. However, these descriptors are not ideal for partial matching, since they can be affected by changes in the topology of the shapes or by the existence of additional parts. Simply modifying the *scale* of the region of interest of the descriptors is not necessarily more effective, as shown in the example in Figure 1.6 (c).



(a) (b)

Figure 1.7: Correspondence results obtained with our bilateral maps. Corresponding points are shown with matching colors, and we connect some interesting matches with the blue curves. Notice that our descriptors enable a simple algorithm to find meaningful matches between shapes with missing parts (a), and shapes that include extraneous parts (b).

In this thesis, we propose a new type of local shape descriptor, the *bilateral map*, to handle partial matching cases. Instead of defining a region of interest around a single point and constraining it to a fixed radius, we compute a descriptor whose context is constrained by a pair of points, as shown in Figure 1.6 (a) and (b). More specifically, we compute the shortest path on the surface between the pair of points, and define a context region in the vicinity of the path. Our main observation is that it is more advantageous to define regions of interest anchored by two points instead of one point. This is demonstrated by several advantages that our bilateral map has over descriptors centered at a single point. First, the context region is *adaptive*. Since it is constrained to lie between the two reference points, it only includes portions of the shape that capture the structural relationship between these two points. Portions of the shape that are not relevant to the reference points and that may be potentially missing in other shapes are selectively ignored by the descriptor (contrast Figure 1.6 (b) to (c)). Secondly, given that the region of interest is adaptive, the selection of the scale of this region is greatly facilitated, since then the region extent parameter can be set to be proportional to the length of the path between the two points.

In Figure 1.7, we show two examples of correspondences computed with our bilateral maps and a simple matching algorithm. Notice how the descriptors are able to guide the algorithm to find meaningful matches for shapes with missing parts in (a) or differing topology in (b). The bilateral map is discussed in detail in Chapter 5.

## 1.2 Contributions

The contribution of this thesis is to introduce methods for matching *dissimilar shapes*. This problem has recently gained prominence in the fields of shape analysis and computer graphics, driven by applications that require the analysis of dissimilar shapes, such as assembly-based shape creation [Chaudhuri et al. 2011], which calls for a part correspondence across a set of dissimilar shapes, and deformation-transfer between shapes from different semantic categories [Baran et al. 2009] (e.g., transferring the motion from a hand to a human character). Matching dissimilar shapes has become a well-recognized challenging problem, and traditional methods are inadequate for this task, since they rely solely on the comparison of the local similarity of the shapes or on an isometry-preservation criterion. The impact of this thesis is to introduce three contributions to address this emerging problem. We believe that these contributions can also serve as the basis and inspiration for future work.

1. We propose one of the first methods to incorporate prior knowledge for matching dissimilar shapes (along with the method of Kalogerakis et al. [2010]).

   Knowledge-driven methods are advantageous when the shapes represent the same class of object but can differ significantly in their part composition, i.e., when the geometry of the parts and how they connect to each other can drastically change from one shape to the other. Moreover, our method is the first to address specifically the correspondence problem and also to incorporate direct content comparison. That is, we combine the prior knowledge with the classical geometric comparison between the shapes, so that we can benefit from this information if the shapes presented to the algorithm are in fact similar. The idea of incorporating prior knowledge for shape analysis is quite general, and we believe that other representations for adding such knowledge could be devised in the future.

2. We introduce the first method to analyze a set of dissimilar shapes in an unsupervised manner and compute their co-segmentation, where the shapes in the set can vary significantly in their geometry and part composition.

   By analyzing the set as a whole we are able to derive additional knowledge about the shapes and use this knowledge in the co-segmentation. Our method based on third-party connections in a descriptor space is more flexible than previous work based on geometric alignment [Golovinskiy and Funkhouser 2009; Xu et al. 2010], since

these methods assume that the shapes can be properly aligned and derive the part correspondence from such an alignment. This restricts the class of shapes that can be handled, and no learning on the part composition of the shapes is performed. Our method, on the other hand, is able to handle a richer variety of shapes. We believe that the concept of analyzing a set of shapes via third-party connections is general and could be used for other applications.

3. We deviate from the traditional definition of local shape descriptors, which are typically centered at a single point, and propose a new descriptor defined by two feature points; the *bilateral map*. We show that the bilateral map is a more promising alternative for partial matching tasks.

   We alleviate the problem of having to search for a common portion that can be matched between two shapes by constructing a descriptor that naturally captures partial regions of interest on the shapes. By confining the region of interest to the context between two feature points, we are able to capture partial regions without the need of part detection, while also simplifying the scale selection problem which is a difficulty when single-point descriptors are used. This descriptor is a first promising alternative for partial matching and could lead to the development of more refined descriptors that naturally capture partial regions of the shapes.

## 1.3 Organization

The reminder of this thesis is organized as follows. In Chapter 2, we discuss the problem of shape correspondence in more detail and review the techniques proposed in the literature to solve this problem. Chapter 3 and Chapter 4 respectively present our supervised and unsupervised methods for knowledge-driven correspondence. Chapter 5 discusses the bilateral maps for partial matching. Finally, Chapter 6 presents our conclusions and discusses directions for future work.

# Chapter 2

# Background and related work

In this chapter, we review the problem of shape correspondence and discuss the techniques that can be used to solve the problem in the classical setting. In subsequent chapters, we review works more closely related to each contribution proposed in this thesis.

## 2.1 Problem definition

Finding a meaningful correspondence between two or more shapes is a fundamental shape analysis task. The problem can be generally stated as: *given input shapes* $\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_N$*, find a meaningful relation (or mapping) between their elements*, e.g., see Figure 2.1. The elements can be primitives such as points, feature points, faces, skeletal features, or higher-level entities such as parts. Under different contexts, shape correspondence has also been referred to as registration, alignment, or simply matching. Shape correspondence is a key algorithmic component in tasks such as 3D scan alignment and space-time reconstruction, as well as an indispensable prerequisite in diverse applications including attribute transfer, shape interpolation and statistical modeling.

The correspondence problem has been traditionally studied in the image analysis communities, but in this work we focus on methods operating on geometric *shapes* represented by triangle meshes, as opposed to images or volumes. 3D shapes provide explicit geometry information, but generally lack a simple parameterization domain. Another distinction is that image analysis often benefits from rich local descriptors based on color and texture, while the descriptors computed for shapes are generally not as distinctive.

Several specializations of the correspondence problem can be considered. One may ask

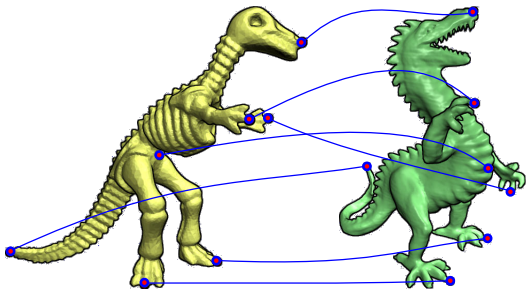


Figure 2.1: A meaningful correspondence (blue lines) between a sparse set of feature points on two shapes. Note the large amount of geometric variations between the shapes which make the computation of such a correspondence difficult.

whether it is meaningful to establish a correspondence in full, as in Figure 2.1, or only for part of the shapes, as in Figure 2.2. The additional need to find the common parts of the shapes not only increases the search space but also makes it difficult to define a proper objective function, making the partial correspondence problem harder [Gal et al. 2007; Zhang et al. 2008]. One may also consider the density of the correspondence computed. Sparse correspondence only seeks to identify and match a small set of landmark points. However, it is as difficult as its dense counterpart, since the challenging aspect of the correspondence search remains essentially the same — in both cases it is necessary to consider the global structures of the shapes and possibly the semantics of their parts to obtain a meaningful solution. Often, a dense correspondence is derived from a sparse one via some form of interpolation [Alexa 2002; Kraevoy and Sheffer 2004], although the computation of a dense correspondence can also present challenges in certain circumstances, e.g., in the case of partial matching or when there are topological differences between the shapes.

Defining what is a meaningful correspondence depends on the task at hand. The task can range from the simpler case of identifying portions of the shapes that are geometrically similar, to the more complex problem of relating elements that represent the same parts or serve the same function on the shapes. In the latter case, the matching parts may differ significantly in their geometry, structure within the context of the whole shape, or even topology. An example of such a semantic correspondence problem is shown in Figure 2.3. In general, computing such correspondences is difficult since it involves understanding the structure of the shapes at both the local and global levels, and possibly understanding the

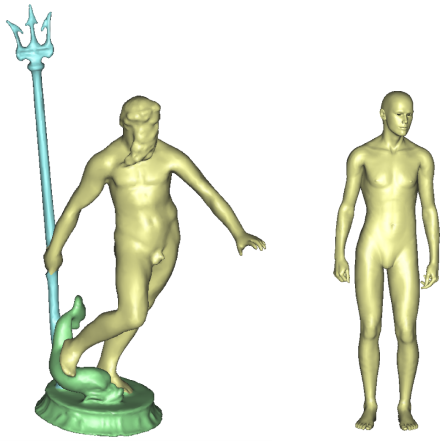


Figure 2.2: An example of a partial correspondence problem. The goal here is to establish a correspondence between Neptune's statue (left) and the human (right), by relating the parts in yellow and ignoring the extra parts shown in green and blue.

functionality of the shape parts.

Classical applications of shape correspondence, such as 3D scan alignment or shape morphing, call for traditional solutions which include rigid alignment and feature matching. Rigid alignment is typically solved with methods based on sampling and verifying candidate transformations [Irani and Raghavan 1996], or by applying the iterative closest point algorithm or one of its variations [Rusinkiewicz and Levoy 2001]. In the case of feature matching, shape descriptors are computed for representative points sampled from the shapes, and the correspondence is constructed by selecting the assignments that maximize the similarity between the descriptors. Additional constraints can also be incorporated, such as the preservation of distances between points (isometry assumption).

More recent works have attempted to deal with large variations in the shapes, to the point where rigid alignment is no longer suitable. On one hand, emerging techniques for surface deformation have been successfully applied to non-rigid registration of surfaces [Huang et al. 2008; Zhang et al. 2008]. On the other hand, progress in matching approximately isometric shapes has also been made in recent works [Bronstein et al. 2008a; Lipman and Funkhouser 2009]. Finally, there has been a recent trend to look beyond low-level geometric information and to incorporate high-level shape semantics into the shape analysis pipeline. Examples of such works include techniques for segmenting a mesh into parts [Shamir 2008], finding

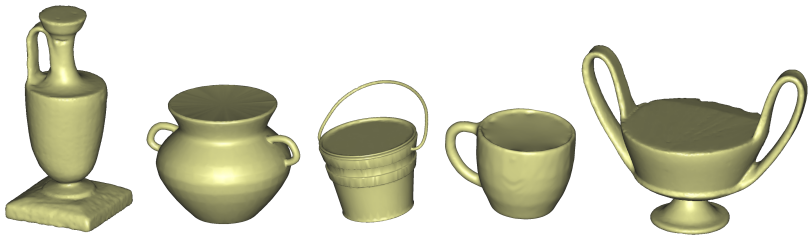


Figure 2.3: An example of a collection of man-made shapes (liquid containers) for which computing a correspondence is a challenging problem. Note how the shapes can be constituted by different types and numbers of parts (e.g., one or two handles, neck, base), how the parts of a same type can vary in their geometry (e.g., long vs. short handles), and how they can connect to each other in different manners.

analogies between these parts [Shalom et al. 2008], transferring information [Sumner and Popović 2004] or part styles [Xu et al. 2010] from one shape to another, extracting the high-level structure of the shapes for manipulation or deformation [Xu et al. 2009a; Gal et al. 2009], and using prior knowledge to learn how to label a shape [Kalogerakis et al. 2010] or establish a correspondence, as done in this thesis [van Kaick et al. 2011b; Sidi et al. 2011]. Computing a correspondence between shapes is one of the key problems that can benefit from semantically-driven techniques, since the goal is to understand the structure of the shapes in order to find a meaningful correspondence between their parts.

## 2.2 Applications of correspondence

In this section, we discuss what we view as the most important applications that make use of correspondence methods.

**Shape registration.** Given a number of scans in arbitrary initial positions, the goal of registration is to match regions that correspond across the scans, so that the scans can be aligned and the target object can be fully reconstructed. In certain situations, it can be assumed that the shapes do not change during the scanning (rigid registration) [Rusinkiewicz and Levoy 2001; Aiger et al. 2008]. However, it can also be the case that the acquisition process introduces non-linear distortions, or the shapes are free to deform during the acquisition, which creates the need for non-rigid registration [Anguelov et al. 2004; Jain et al.

2007; Brown and Rusinkiewicz 2007; Chang and Zwicker 2008; Huang et al. 2008].

**Time-varying surface reconstruction.** The goal in this task is to reconstruct a 3D shape that was scanned over time while moving and deforming. The challenge is to take the great amount of generated data and organize it into one single model that represents the deforming shape. Correspondence methods are central to time-varying surface reconstruction, since the point sets of different time frames have to be registered to yield the final model [Mitra et al. 2007; Wand et al. 2007; Sharf et al. 2008; Pekelny and Gotsman 2008; Li et al. 2009; Chang and Zwicker 2009; Tevs et al. 2009; Zheng et al. 2010].

**Shape interpolation.** In interpolation or morphing, one shape is gradually transformed into another. The transformation has to satisfy certain aesthetic requirements, so that the gradual change of the shape is visually pleasing [Alexa 2002]. One important property in this aspect is that the correspondence between the reference shape and the target shape should be meaningful, i.e., it should relate parts in the shapes that are semantically equivalent.

**Information transfer.** A task that is becoming common is to transfer information from a source 3D object to a target 3D object, especially to enable the reuse of attributes or motion information associated to the source shape. Examples include transferring a deformation or morph from one mesh to another [Sumner and Popović 2004], transferring textures while deforming a mesh [Dinh et al. 2005], and transferring the style of one group of shapes to another [Xu et al. 2010]. Such tasks clearly require a correspondence, since the motion, attribute or style of each element on the source shape has to be transferred to its corresponding element on the target shape.

**Symmetry detection.** The symmetries of a shape can act as an important cue when solving several tasks, such as registration, segmentation, compression, modeling and editing [Golovinskiy et al. 2007a]. Detecting the symmetries of a shape (a set of transformations that when applied to the shape do not modify its geometry) can be posed as the problem of finding a correspondence from the shape to itself. Therefore, it is natural that symmetry detection algorithms possess many similarities with correspondence methods, such as the use of transformation or point sampling for extrinsic [Mitra et al. 2006; Podolak et al. 2006] or intrinsic [Xu et al. 2009b; Kim et al. 2010] symmetry detection, or the use of spectral embeddings [Ovsjanikov et al. 2008; Lipman et al. 2010].

**Recognition and retrieval.** Understanding a scene described by a range image is one of the classic challenges in computer vision [Forsyth and Ponce 2003]. Shape correspondence is one of the approaches that can be used for this task. By establishing a correspondence between a query shape and the models in a dataset, the identity of the query shape is inferred from the best match to one of the models (according to a correspondence quality measure). A similar procedure can be utilized for retrieval [Funkhouser and Shilane 2006].

**Statistical shape modeling.** The analysis of anatomical structures such as organs or bones can be facilitated when a statistical shape model is available. These models are useful for extracting shapes from images or volumes, since they are able to describe the valid variations in the appearance and the size of a shape. Such models are typically constructed while computing a group correspondence for a collection of shapes that represent a common anatomical structure [Davies et al. 2008].

**Change detection.** Another application of correspondence is to track changes in a shape (e.g. displacements, growth) over time. In the medical field, an example application is to track the change in the number and density of moles on a patient's skin (for cancer prediction), which can be posed as a problem of point cloud correspondence [Mirzaalian et al. 2009]. In remote sensing, one example is to track the change over time in the layout of cities and their land usage [Leclerc et al. 2000].

## 2.3 Correspondence problems and related methods

Now, we give an overview of the different types of correspondence problems and describe representative methods that can be used to obtain a solution. We can derive a correspondence directly from the similarity of the elements, or we can first align the shapes and then derive a correspondence from the proximity of the aligned elements. Moreover, we can also iterate between the two procedures. These options directly affect which strategy should be selected to find the correspondence. It is worth noting that the alignment between the shapes is a side product of the computation which is useful and sometimes essential to the underlying application. The distinction between these cases is illustrated in Figure 2.4.

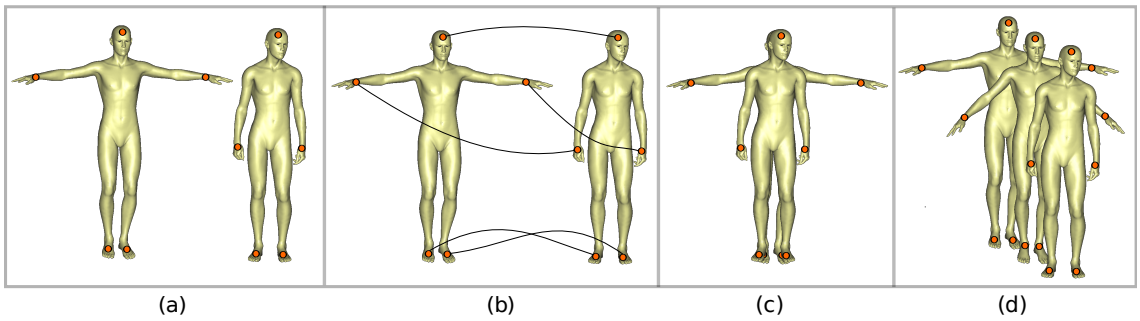



Figure 2.4: Different manners of solving the correspondence problem for the input shapes shown in (a) and their feature points (indicated by the dots): (b) computing a correspondence without explicitly bringing the shapes into alignment, (c) computing a global rigid transformation to align the two shapes, and (d) computing local non-rigid transformations for the shape primitives to deform one shape into the other.

### 2.3.1 Similarity-based correspondence

One of the fundamental ways of computing a correspondence is to estimate the similarity between shape elements or feature points collected from two different shapes and derive a correspondence from those estimates, which is sometimes called the *feature matching* approach. The elements are commonly characterized by shape descriptors. A correspondence is then obtained by selecting assignments between pairs of elements while optimizing an objective function composed of two terms. The first term seeks to maximize the similarity between the descriptors of corresponding elements, while the second term seeks to minimize the distortion that would be introduced in the shapes if they were deformed to align their corresponding elements. However, the second term is estimated without explicitly aligning the shapes. Ideally, satisfying these objectives should translate into a solution that is geometrically or semantically meaningful. Feature matching can be applied in any context where it is possible to compute descriptors for the elements. Example applications include registration of 3D scans [Castellani et al. 2008] and deforming surfaces [Anguelov et al. 2004], or skeleton matching [Biasotti et al. 2006].

A correspondence problem that takes into account only the similarity of shape descriptors can be solved effectively with the simplex algorithm by posing it as a linear assignment [Papadimitriou and Stieglitz 1982]. If the correspondence is constrained to a one-to-one mapping, the problem becomes that of finding an optimal matching in a weighted

bipartite graph, which can be solved more efficiently with the Hungarian algorithm in $O(n^3)$ time, where $n$ is the number of feature points in each shape [Papadimitriou and Stieglitz 1982]. In the general case, when a distortion term is also part of the objective, we have a Quadratic Assignment Problem (QAP) that is NP-hard. One group of methods in the literature computes quadratic assignments through integer optimization, which is relaxed to the continuous setting and solved with a continuous optimization technique. Examples include the softassign technique [Gold and Rangarajan 1995] (which iteratively normalizes rows and columns of an affinity matrix), concave programming [Maciel and Costeira 2003], approximations based on linear programming [Berg et al. 2005], spectral clustering [Leordeanu and Hebert 2005], or relaxation labeling [Zheng and Doermann 2006]. It can also be formulated in probabilistic terms and solved as a convex optimization problem [Zass and Shashua 2008].

Another group of methods solves the problem in the discrete setting without resorting to the continuous domain. One common solution approach in the discrete case is to solve the problem by computing an optimal labeling of a graph, e.g., the problem can be posed in terms of a Markov network where the set of labels corresponds to matching points on the target shape [Anguelov et al. 2004; Zheng et al. 2010]. Other methods make use of metaheuristics for combinatorial optimization, such as ant colony optimization [van Kaick et al. 2007]. One more option is to sample the space of correspondences in search of a solution, guided by an isometry criterion and importance sampling [Tevs et al. 2009].

Finally, one specific group of methods in discrete optimization finds a solution by making use of tree-based search techniques, such as branch-and-bound or priority search [Gelfand et al. 2005; Funkhouser and Shilane 2006; Zhang et al. 2008; Au et al. 2010; Xu et al. 2010]. During the tree expansion, each node represents a partial solution. A full solution is found by following the path from the root of the tree to one of its leaves. Although the specific strategy in which the tree is expanded differs from method to method, these techniques usually involve three important steps: expanding a node that represents a partial solution (branching), estimating how far the partial solution is from the optimum solution (bounding), and eliminating nodes that will not lead to the optimum solution (pruning).

Solutions are mainly represented as collections of assignments between pairs of feature points, and the expansion step involves adding a new pairwise assignment to a given solution. Bounding and pruning can be performed by verifying the quality of the registration given by the current solution, either by aligning the shapes [Gelfand et al. 2005] or by deforming one shape into the other [Zhang et al. 2008]. Other pruning methods include

testing the compatibility between pairwise assignments, such as quantifying the distortion introduced in the Euclidean [Gelfand et al. 2005; Funkhouser and Shilane 2006] or geodesic distances [Zhang et al. 2008; Au et al. 2010] between pairs of points, or testing the agreement in the spatial configuration of the shapes [Au et al. 2010]. Naturally, the descriptors computed for the feature points are also considered in the bounding and pruning steps.

### 2.3.2 Rigid alignment

Under certain assumptions, it is possible to pose the correspondence problem as a search for a geometric transformation that aligns the shapes. One example application is the rigid alignment of geometry scans used for shape acquisition. The goal here is to capture a real-world static 3D shape and obtain its digital representation. However, it may not be possible to capture the entire object in a single scanning pass due to self-occlusions and physical constraints of the scanner, so it is often necessary to acquire multiple scans and optimally align them to reconstruct the full object [Turk and Levoy 1994; Rusinkiewicz and Levoy 2001; Gelfand et al. 2005; Aiger et al. 2008]. The key characteristic of the rigid alignment problem is that the objects do not change from one scanning pass to another. Thus, it is assumed that each scan can be transformed with a *single rigid transformation* to align it perfectly with the other scans. Rigid transformations comprise translations and rotations, and one of their important properties is that they reside in a low-dimensional space.

Scan alignment is just one example of many applications that rely on the assumption of rigidity in the datasets. For two shapes given as 3D point sets $\mathcal{S}$ and $\mathcal{Z}$, the problem of rigid alignment can be posed as: find the rigid transformation that, when applied to $\mathcal{S}$, maximizes the number of points in $\mathcal{S}$ that align to points in $\mathcal{Z}$. This goal is usually dependent on a threshold $\epsilon$ that indicates when two points are close enough and can be considered as matching to each other [Irani and Raghavan 1996].

Following this formulation, alignment with rigid or even affine transformations is a problem with a clear objective function, and efficient algorithms also exist to optimally solve such instances, such as constrained transformation search [Huttenlocher 1991; Aiger et al. 2008] or geometric hashing [Wolfson and Rigoutsos 1997]. However, as the complexity of these algorithms is still at least quadratic in the number of input elements, heuristics can be used to speed up the solution search, e.g., by exploring candidate assignments suggested by the feature matching approach [Aiger et al. 2008], randomizing steps of the algorithms [Fischler and Bolles 1981; Irani and Raghavan 1996], or by making use of votes as in pose

clustering [Stockman 1987; Olson 1997]. Alternatively, a local-search algorithm such as the Iterative Closest Point (ICP) [Rusinkiewicz and Levoy 2001] can also be used.

### 2.3.3 Non-rigid alignment

Following our previous example of scan alignment, it might be necessary to lift the assumption that each scan can be perfectly aligned with a rigid transformation, e.g., when large amounts of noise are present in the scans. More significant examples of datasets that cannot be rigidly aligned include the correspondence of articulated shapes [Elad and Kimmel 2003; Anguelov et al. 2004; Jain et al. 2007; Chang and Zwicker 2008; Huang et al. 2008], where certain parts of the shapes can bend independently, the correspondence of anatomical shapes (e.g., organs) [Audette et al. 2000], which can deform in an elastic manner and introduce stretching to localized portions of the shape, and finally the correspondence of shapes with different geometries but that represent a class of objects with parts that are semantically related [Zhang et al. 2008]. In the latter case, we can see the problem as that of establishing a correspondence between shapes that can differ in both local stretching and bending.

In this setting, it becomes necessary to add more freedom to how the shapes can be brought into correspondence. This can be achieved by generalizing two aspects of the problem. First, non-rigid (possibly non-linear) transformations can be taken into consideration, e.g, thin-plate splines [Chui and Rangarajan 2003]. Secondly, these transformations can be applied separately to local portions of the shape. For example, the transformation applied to a shape can be represented as a set of per-vertex displacement vectors [Pauly et al. 2005] or per-vertex affine transformations [Sumner and Popović 2004]. Then, finding the best transformation amounts to computing the displacements that bring each vertex in correspondence with the target shape. The distinction from the rigid case is that the space of geometric transformations being considered is now inherently high-dimensional.

Due to the high-dimensional nature of the solution space, these problems are typically solved with a form of local or approximate search including a regularization term, e.g., gradient descent [Allen et al. 2003] or a combination of the non-linear transformations with the feature matching approach [Chui and Rangarajan 2003]. However, although heuristic solution methods are available, the quality of the results will typically depend on the complexity of the problem instance and the level of approximation introduced by the methods. As in the case of feature matching approaches, if there are more constraints available, they can be used to guide the algorithms more effectively to a correct solution.

Alternative methods for non-rigid correspondence include posing the problem as the piece-wise alignment of detected parts [Chang and Zwicker 2008], or using non-linear variants of the ICP algorithm, either by computing a warp function based on thin-plate splines [Chui and Rangarajan 2003; Brown and Rusinkiewicz 2007], or substituting the rigid transformation with a deformation based on rigid-body components [Huang et al. 2008]. Another possibility, especially for articulated shapes, is to first embed the shapes in a space where the configuration of the rigid parts is normalized, and then treat the problem simply as a case of rigid alignment in this embedding space. The embedding can be obtained with techniques such as Multi-Dimensional Scaling (MDS) [Elad and Kimmel 2003; Bronstein et al. 2006; Bronstein et al. 2008a], the GPS embedding [Rustamov 2007], or the spectral transform [Jain et al. 2007; Mateus et al. 2008; Sahillioğlu and Yemez 2010]. A comprehensive coverage of different forms of embeddings is given in [Zhang et al. 2010].

### 2.3.4 Time-varying registration

Due to recent technological advances, an application that is attracting attention is the reconstruction of 3D shapes acquired over time while moving and deforming. In this setting, a fixed number of scans is acquired per time step, and these scans have to be registered to allow the reconstruction of both the object and the motion sequence [Mitra et al. 2007; Wand et al. 2007; Sharf et al. 2008; Pekelny and Gotsman 2008; de Aguiar et al. 2008; Li et al. 2009; Gall et al. 2009; Chang and Zwicker 2009; Tevs et al. 2009; Zheng et al. 2010].

Although this may seem like another instance of the non-rigid alignment problem, there are certain particularities that make this problem unique. In the classic registration problem, it is assumed that all the scans can be registered to compose a single and coherent object. On the other hand, the time-varying setting introduces the additional difficulty that the shape might have deformed significantly from one frame to the other. Therefore, scans acquired later in time may only be registered to the earlier scans if the deformation is taken into account. Moreover, additional challenges are the large amount of missing data (due to occlusion) that can be present in each frame [Pekelny and Gotsman 2008], and datasets that were captured over sparse time frames [Chang and Zwicker 2009; Zheng et al. 2010]. However, the addition of temporal constraints can also help in reducing the size of the search space (e.g., kinematic constraints [Mitra et al. 2007]).

For obtaining correspondences in such a setting, variants of the ICP algorithm have been utilized [Wand et al. 2007; Pekelny and Gotsman 2008; Wand et al. 2009; Li et al. 2009].

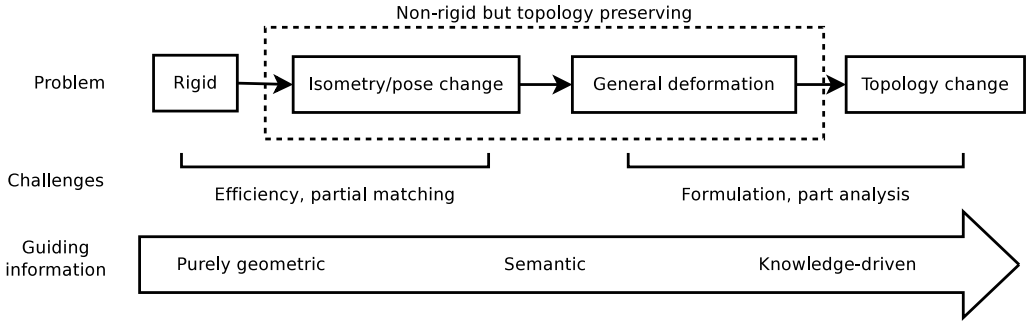



Figure 2.5: A progression of development in shape correspondence methods.

If a sufficient number of scans is acquired per time unit, it can be assumed that only small changes take place in the spatial configuration of the shapes (rigid-body components can be consistently tracked), which facilitates the computation. Other methods pose the problem as the reconstruction of a space-time surface [Mitra et al. 2007; Süßmuth et al. 2008; Sharf et al. 2008], or extract a skeleton that is coherent for all time frames [Zheng et al. 2010].

## 2.4 Discussion

In Figure 2.5, we show a progression in the development of solutions to shape correspondence. On one hand, significant progress has been made to compute rigid or pose-invariant alignments (through the correspondence of approximately isometric shapes). Since these problems have well-defined objectives, remaining work is mainly focusing on improving efficiency and accuracy, or on handling more difficult specializations of the problem, such as partial matching. In part, the success in these areas is also due to the fact that the correspondence can be obtained reliably from purely geometric information extracted from the shapes, and shape alignment can be described in terms of unambiguous transformations or objective functions.

On the other hand, finding a meaningful correspondence between shapes belonging to the same class but differing (sometimes significantly) in their geometry, structure, or even topology, remains a challenge. Traditional methods which rely on assumptions of rigidity, isometry, or sufficient geometric similarity between corresponding parts are simply inadequate. Man-made shapes such as the ones shown in Figure 2.3 are particularly challenging

to deal with, since these objects often differ not only by geometric deformations, but also by their part constitutions. Shape correspondence then departs from the low-level sphere of geometry analysis and becomes the higher-level problem of semantic reasoning, where we seek to recognize the parts of shapes and infer their semantics or functionality. A meaningful correspondence can then be established between the recognized parts. The utilization of prior knowledge is seen as a promising solution, where the main difficulty is how to model the knowledge and make use of it effectively.

Some recent works have taken the first steps towards knowledge-driven shape correspondence. Knowledge can be incorporated by utilizing a set of examples where a few landmark points have already been matched by an expert user [Ward and Hamarneh 2007], by using a set of examples already in full correspondence [Pitiot et al. 2007], or by relying on a set of pre-segmented and pre-labeled shapes, so that classifiers can be trained on these examples and applied to label the primitives of an unknown shape, as in the work of Kalogerakis et al. [2010] or the work in this thesis [van Kaick et al. 2011b] (Chapter 3). Another direction is to learn how the terms in the various objective functions should be weighted, depending on the restricted domain of the problem that we are considering [Caetano et al. 2009]. One more possibility is to consider group information when performing correspondence-related tasks, such as skeletonization [Ward and Hamarneh 2009] or consistent segmentation of a set of shapes [Golovinskiy and Funkhouser 2009]. The results of co-segmentation can be greatly improved by directly addressing the nonhomogeneous part scaling [Xu et al. 2010], or by performing an analysis of the part constitution of the shapes, as proposed in this thesis [Sidi et al. 2011] (Chapter 4).

# Chapter 3

# Prior knowledge for part correspondence

The majority of geometry processing methods in the graphics community have relied on low-level reasoning and operated on low-level features. Recently, there has been a research trend towards higher-level geometry processing, particularly the analysis of shapes at a more semantic level [Attene et al. 2009; Gal et al. 2009; Simari et al. 2009; Xu et al. 2009c; Kalogerakis et al. 2010; Mitra et al. 2010]. Shape correspondence is a fundamental problem that often requires a higher-level understanding of shapes. Applications such as attribute transfer [Sumner and Popović 2004; Baran et al. 2009], morphing [Alexa 2002], shape synthesis [Anguelov et al. 2005; Xu et al. 2010], and object recognition [Berg et al. 2005] are often meant to employ correspondences between shape parts which possess the same meaning or functionality rather than mere geometric similarity.

Classical approaches to shape correspondence are mainly *content-driven* [Funkhouser and Shilane 2006; Jain et al. 2007; Zhang et al. 2008; Lipman and Funkhouser 2009; Au et al. 2010; van Kaick et al. 2011a], focusing solely on the geometrical and structural similarities between the shapes to be matched. However, both criteria can be violated in challenging scenarios where there are large variations in the geometry or topology of the corresponding parts, as the examples in Figures 3.1 and 3.2 show. Shape correspondence under these circumstances is simply beyond pure geometry analysis and requires a semantic analysis of the shapes. Such an analysis necessitates the use of prior knowledge: to find a correspondence between parts that may be highly dissimilar geometrically, we need to invoke our memory of

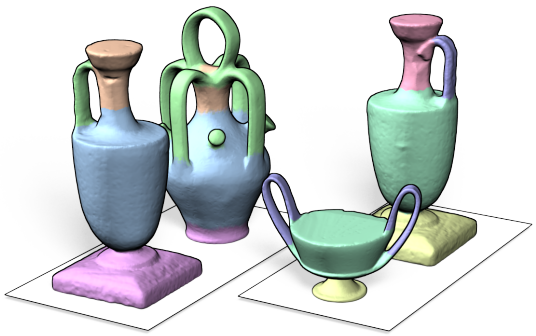


Figure 3.1: Meaningful correspondence between shape parts under significant geometric and topological discrepancies is made possible by incorporating prior knowledge. Notice the missing neck on one shape in the right pair, and the matching of one to multiple handles in both pairs. Corresponding parts on each pair are implied by matching colors.

similar parts that correspond to each other (a *recognition* process). This knowledge is then used to establish a correspondence between the unknown parts. Incorporating recognition into shape correspondence results in a *knowledge-driven* approach. The power of knowledge is exemplified when pure geometry analysis simply cannot succeed (Figure 3.1).

In this chapter, we introduce an approach to shape correspondence which incorporates prior knowledge. Aiming to mimic the human cognitive process where recognition is known to be primarily part-based [Marr 1982; Hoffman and Richards 1987], we compute part labels (a recognition process) in conjunction with correspondence. The result is thus a *part correspondence* in contrast to correspondences between low-level feature points, as done in most works to date (Chapter 2). The prior knowledge is imparted through a training set of pre-segmented models with semantic labels. The training set serves as a knowledge medium allowing to find the correspondence between geometrically dissimilar parts. With the prior knowledge, we learn a *probabilistic semantic label* for each face of a query shape (one of the shapes to be matched). The labeling is derived, with the aid of a classifier, from the similarity of the descriptors of the faces to those of a class in the training set.

Part correspondence may be established based solely on knowledge and individual labeling of the query shapes. However, this may lead to an unsatisfactory outcome when the knowledge set is incomplete or produces indeterminate recognition results. Such cases may have to rely, at least partially, on a direct comparison between the query shapes. We combine the use of knowledge with content-driven analysis, where the latter is based on

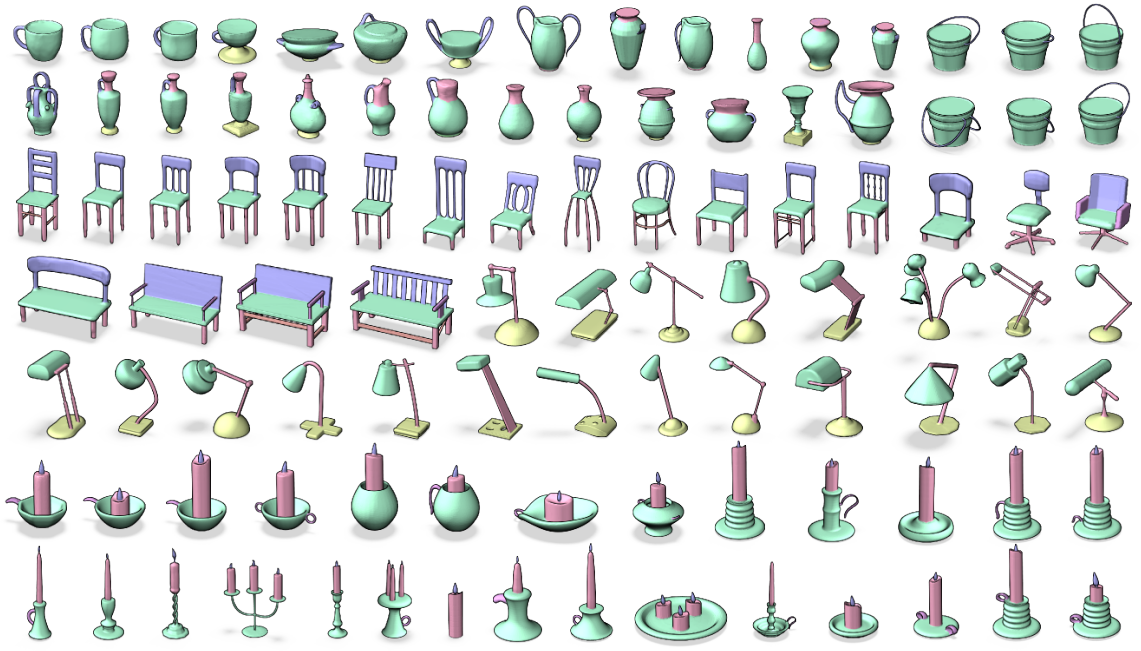


Figure 3.2: The set of man-made shapes used in our work. Note the large intra-class geometric and topological variability.

geometric similarity between the query shapes. The final correspondence is thus obtained through a *joint labeling* of the query shapes. The joint labeling makes use of the knowledge-driven probabilistic labels as well as feature pairings between query shapes to extract the actual parts that we seek to recognize and match across the shapes. The pairing of features incorporates the direct similarity between local regions of the query shapes.

Our contribution in this chapter is two-fold. First, we show that the incorporation of prior knowledge to shape correspondence is effective. In particular, it leads to significant improvement over classical approaches on query shapes exhibiting large intra-class part variability in geometry or topology (see Figures 3.1 and 3.2). Second, we show that the joint labeling approach is able to combine the knowledge-driven and content-driven analyses so that they complement each other in instances where using solely the knowledge or the local geometric similarity of the shapes is insufficient.

## 3.1 Related work

As discussed in detail in Chapters 1 and 2, content-driven approaches to shape correspondence compare a pair of shapes based on geometric similarities between matched features, an approximate isometry criterion, or a combination of both [Funkhouser and Shilane 2006; Bronstein et al. 2008a; Zhang et al. 2008; Lipman and Funkhouser 2009; van Kaick et al. 2011a]. The computational paradigm is an optimization or discrete search guided by these criteria. The recent methods of Zhang et al. [2008] and Au et al. [2010] both allow large shape variations, but only to a certain extent, as they are still confined by the premise of geometric similarity and do not model shape semantics.

Part correspondence brings relevance to the segmentation problem. Many approaches to meaningful shape segmentation [Shamir 2008] follow the minima rule [Hoffman and Richards 1987], where segment boundaries are defined near concave regions. Other methods identify shape parts based on their geometric characteristics such as convexity and compactness [Kraevoy et al. 2007]. Clustering using an intrinsic surface metric or curvature is also common [Katz and Tal 2003; Liu and Zhang 2007]. Structural approaches mainly focus on skeleton topology [Au et al. 2010; Shapira et al. 2008]. While satisfactory individual segmentation results can be obtained, these approaches are not designed to find a consistent segmentation between the shapes, i.e., a segmentation that partitions the models into similar parts that correspond across the shapes.

Golovinskiy and Funkhouser address the consistent segmentation problem [2009] where the connection between matching parts is initially built by a global rigid alignment using iterative closest point. No shape semantics are incorporated and their approach is not designed to handle large intra-class geometric variations such as stretching. Recent work of Xu et al. [Xu et al. 2010] handles non-homogeneous part stretching by grouping the shapes based on their style and then performing part correspondence. Perhaps the first work on explicitly incorporating prior knowledge into shape segmentation is that of Simari et al. [2009], where a multi-objective optimization is performed to segment and label a given shape. However, the user is required to provide semantic knowledge specific to the shape or shape parts and formulate such knowledge to fit the optimization framework; no training set or learning is used.

The ability to tag, annotate, or label a shape lies at the heart of semantic shape analysis. Manual annotation allows the user to either semantically label parts or to apply an

ontology to the structure of the shapes [Attene et al. 2009]. Another group of approaches is based on correspondence analysis guided by shape geometry. As an application of their skeleton-driven correspondence algorithm, Au et al. [2010] assign semantic tags to the skeleton features of a subject shape by fully matching it to each labeled shape in a training set and then applying a simple majority voting to determine the tags. Shapira et al. [2009] tag parts in a similar manner, but rely on a part-in-whole contextual signature to retrieve the most relevant parts.

Recently, Kalogerakis et al. [2010] present a method to learn the semantic labels of a shape based on training data. This work is significant as it is the first generic learning-based method for semantic shape segmentation. Our work shares similarities with this method, such as the use of a training set of pre-segmented and pre-labeled shapes and classifiers to recognize the shape parts. However, we apply prior knowledge to solve the correspondence problem and augment knowledge-driven analysis with content-driven analysis via the joint labeling approach.

Recognition by correspondence is a classical paradigm in computer vision [Basri and Jacobs 1997], as are semantic segmentation, labeling, and classification in images based on learning [Levin and Weiss 2006; Schnitman et al. 2006]. Such approaches are rare in shape analysis however. Images are typically feature-rich, with color and texture cues as well as foreground and background contrasts, which exemplify the usefulness of local feature patches. However, for typical 3D models of the kind we consider (Figure 3.2), the distinctiveness of local surface features is significantly reduced. Also, one may be confined to a limited training set which still contains diverse intra-class shape variations. While image-based methods can typically benefit from the availability of large data collections to form adequate training sets or the knowledge base, the same cannot be said about 3D model collections. These are some of the challenges we wish to address in our work.

## 3.2 Overview

In this section, we present an overview of our approach to incorporate prior knowledge for part correspondence; see Figures 3.3 and 3.4 for an illustration. Full details of the algorithmic components are covered in subsequent sections.

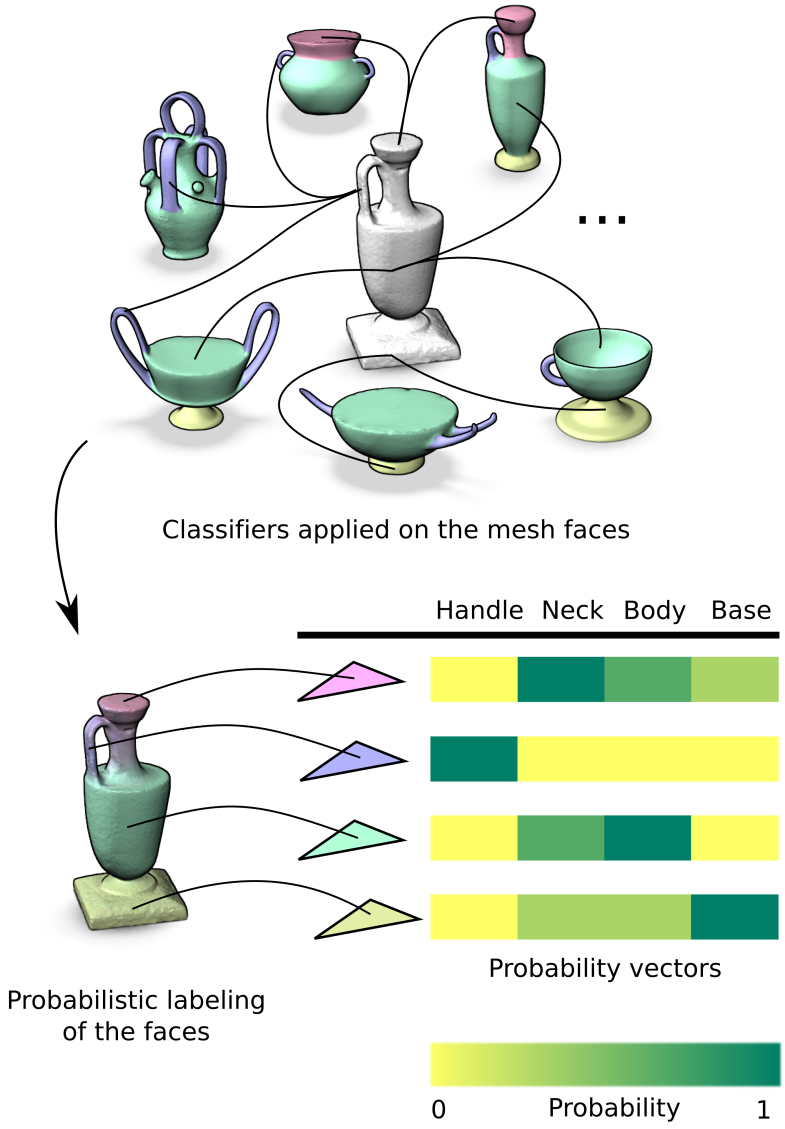



Figure 3.3: Probabilistic semantic labeling. Given a query shape (the gray vase), we first label each of its faces with per-label classifiers. The result is a label probability vector per face (shown as a color-blend on the surface of the bottom shape). The classifiers are learned from the training set (shown around the query with their ground-truth labeling).

**Problem setting.** The central problem that we are addressing here is the computation of a meaningful correspondence between a source shape $\mathcal{S}_1$ and a target shape $\mathcal{S}_2$ (the query shapes). In our case, we require a correspondence that is defined at the part level, i.e., it maps groups of triangular faces on $\mathcal{S}_1$ to groups of faces of equivalent parts on $\mathcal{S}_2$, as opposed to a mapping of feature points from one shape to the other.

**Content-driven analysis.** A straightforward way to establish the correspondence between $\mathcal{S}_1$ and $\mathcal{S}_2$ is by computing a set of shape descriptors and matching the faces based on these descriptors, which we refer to as the feature matching approach. In this context, one extracts for each face on each shape a set of descriptors that capture information about its geometric properties and context, resulting in a vector of scalar values associated to each face. Ideally, by measuring the distance between two such vectors with an appropriate metric we obtain an indication of the similarity of their corresponding faces (or the surrounding regions of the faces). Now, a correspondence between $\mathcal{S}_1$ and $\mathcal{S}_2$ can be computed by any algorithm that computes a correspondence based on the vector distances (examples for such an algorithm are bipartite matching or quadratic optimization, as discussed in Chapter 2).

**Incorporation of semantic knowledge.** One of our goals is to show that a more meaningful correspondence is obtained when semantic information is added to the solution. This is especially effective when the shapes vary greatly in geometry and part constitution. In the knowledge-driven component of our approach, semantic information is derived from a set of training shapes and characterized by classifiers for part labels. The shapes in the training set are pre-segmented and tagged with semantic labels. The labeling is defined at the face level. Next, we compute shape descriptors for the faces of each training shape and, based on the descriptors, we train a classifier $\mathcal{K}_l$ for each label $l$, using the faces labeled $l$ as positive examples of the label, and the remaining faces as negative examples. The learning phase results in one model per part type, each capturing the discriminative properties of the faces that belong to the label.

**Probabilistic semantic labeling.** Now, given a face $f_i$ on a query shape, we compute the same set of descriptors and estimate $p(l|f_i)$, the probability of $f_i$ having the label $l$. $p(l|f_i)$ is obtained by applying the classifier $\mathcal{K}_l$ to the descriptors of face $f_i$. By applying all the classifiers, the face receives a label probability vector composed of the probabilities for all possible labels $l$ (Figure 3.3). In Section 3.3, we elaborate on the probabilistic labeling.

**Joint labeling and part correspondence.** Each face belonging to each of the query shapes has now an associated probabilistic label. It would be possible to use the labels in place of our original descriptors and obtain the correspondence via a feature matching algorithm. However, we are interested in combining knowledge and geometric shape content, and we also require a correspondence that is defined at the part level. Therefore, we find the solution with a method that extracts segments from the shapes while simultaneously considering their correspondence. This step of the algorithm is achieved with our *joint labeling* (illustrated in Figure 3.4).

The joint labeling takes into simultaneous consideration the probabilistic semantic labels of each face (based on knowledge/training), the mesh connectivity (captured by intra-mesh arcs between pairs of faces on the same shape), and connections between the faces of the two query shapes (feature pairing by inter-mesh arcs). The output of the joint labeling is a set of parts for each shape and their correspondence. The inter-mesh arcs come from candidate assignments extracted from the similarity of shape descriptors. Details on the joint labeling are described in Section 3.4.

The joint labeling has two advantages over the labeling based only on knowledge. First, there are practical limitations for the knowledge representation, such as the size and variability of the training set and accuracy of the shape descriptors. In this case, the inter-mesh edges complement the knowledge with extra information on the direct similarity between portions of the shapes. Moreover, even if the training set were large enough, covering an infinitude of variations, there would still be cases where a purely knowledge-driven approach could fail, such as when identical parts appear on different locations of a shape, or when there exist ambiguities in the functional role of the parts. The addition of content analysis contributes to the disambiguation of such cases.

## 3.3 Prior knowledge and probabilistic semantic labeling

**Training set.** The first step towards the development of our correspondence approach is to create a knowledge base from a dataset of training shapes. This dataset contains manually segmented shapes and the semantic labeling of each part (e.g., labels such as "leg" and "seat" for chairs). Thus, each mesh face is associated to a semantic label. An example dataset, used in part of our experiments, is shown in Figure 3.2.

**Shape descriptors.** Next, we compute a collection of descriptors for each face of all the training shapes. These descriptors should capture different properties of the faces, such as the local geometry of the shape around the face and its context in relation to the whole shape. The purpose of using a collection of descriptors is that their union should be rich enough to distinguish the faces of different classes. We extract descriptors based on principal component and curvature analysis in the neighborhood of a face at multiple scales, the shape diameter function [Shapira et al. 2008], average of geodesic distances starting from each face, and the binning of face areas into geodesic maps. Some of these descriptors are similar to those that appear in the learning approach of Kalogerakis et al. [2010].

**Classifier training.** Finally, we group the faces from the training set according to their labels. Suppose that for each label $l$, we are given $n_l$ training faces coming from different shapes and grouped into the set $F_l$. We compute the descriptors for each face in this set and denote the full set of descriptor vectors as $D_l$. Then, for each label $l$, we train a classifier $\mathcal{K}_l$ with the descriptors $D_l$ as positive examples of the label, and with the descriptors $D_{\bar{l}} = \cup_{D_{l'}}$, with $l' \neq l$, as negative examples. Notice that instead of training a set of per-label classifiers, a single multi-class classifier could also be used to take advantage of shared decision rules.

To train a classifier, we use the "gentleboost" algorithm, which has several advantages in relation to other choices, such as performing automatic feature selection, being a time-efficient training algorithm, and attaching confidence values to each classification decision [Kalogerakis et al. 2010]. By adjusting the importance weight given to each training sample, it is also possible to account for unbalanced datasets. Details on this algorithm can be found in [Friedman et al. 2000]. The unnormalized confidence values returned by each classifier can be transformed into probabilities with the softmax activation function (a generalization of the logistic function to multiple variables).

**Probabilistic semantic labeling.** Now, given an unknown face $f_i$ on a query shape, we compute its associated set of descriptors $D_i$ and use the classifier $\mathcal{K}_l$ to estimate $p(l|D_i)$, the probability that face $f_i$ should belong to class $l$ based on its descriptors $D_i$. The probabilistic labeling is performed for all faces on the two query shapes $\mathcal{S}_1$ and $\mathcal{S}_2$. One of our key contributions here is to show that computing a correspondence based on the semantic labeling provides superior results when compared to those computed directly from the similarity of the descriptors.

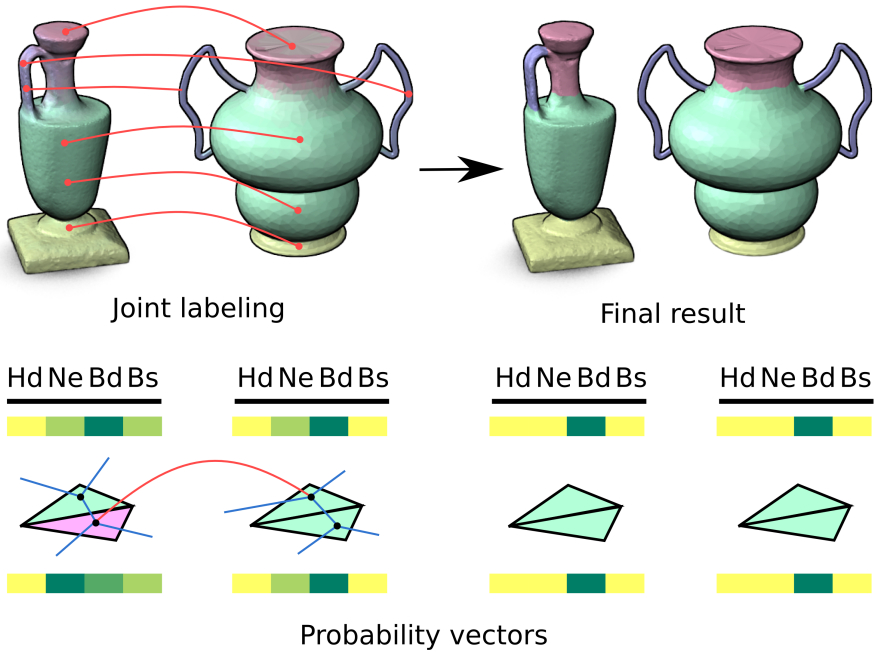



Figure 3.4: Joint labeling. Given a pair of shapes (to the left), we obtain their optimized labeling by making use of their probabilistic labels (Figure 3.3), intra-mesh arcs (blue lines) coming from the shapes' connectivity, and feature pairings given by the inter-mesh arcs (red lines) added between faces with similar descriptors. The result (to the right) is a segmentation of the shapes and correspondence at the segment level (indicated by matching colors). The probability vectors and the final deterministic labels are shown for two pairs of faces (before and after the joint labeling). The classes are: Handle (Hd), Neck (Ne), Body (Bd), and Base (Bs).

## 3.4 Part correspondence via joint labeling

After obtaining the probabilistic labeling of the faces on the query shapes, we utilize this information in a joint labeling scheme (i.e. labeling both query shapes simultaneously) to obtain the final result. By posing the correspondence problem as that of label optimization, we are able to simultaneously incorporate both the semantic information and local similarity of the two query shapes into the computation, while also extracting as a result semantic parts from the query shapes and their correspondence. This process is inspired by methods for consistent segmentation and labeling [Golovinskiy and Funkhouser 2009; Xu et al. 2010; Ng et al. 2010], although we deviate from their scheme and do not make use of rigid alignment.

**Joint labeling.** The label optimization problem is defined as the assignment of labels to the nodes of a graph such that a given energy is minimized [Ng et al. 2010; Shapira et al. 2009; Kalogerakis et al. 2010]. The set of assigned labels is the same one used in the definition of the semantic information. Given a query shape $\mathcal{S}_1$, we define a graph $G_{\mathcal{S}_1} = \{V_{\mathcal{S}_1}, E_{\mathcal{S}_1}\}$, where the nodes $V_{\mathcal{S}_1}$ are the faces of the mesh and the arcs in $E_{\mathcal{S}_1}$ connect two faces if they are adjacent on the mesh. A similar graph can be defined for a shape $\mathcal{S}_2$. We perform the joint labeling on a graph $G = \{V, E\}$, where $V = V_{\mathcal{S}_1} \cup V_{\mathcal{S}_2}$ and the connectivity of the graph is given by $E = E_{\text{intra}} \cup E_{\text{inter}}$, where $E_{\text{intra}}$ and $E_{\text{inter}}$ are two types of arcs: intra-mesh and inter-mesh arcs. The intra-mesh arcs are simply given by $E_{\text{intra}} = E_{\mathcal{S}_1} \cup E_{\mathcal{S}_2}$. The inter-mesh arcs $E_{\text{inter}}$ connect faces in $\mathcal{S}_1$ to faces in $\mathcal{S}_2$.

**Feature pairing.** We select a set of pairwise assignments from $\mathcal{S}_1$ to $\mathcal{S}_2$, based on the similarity of shape descriptors, to constitute the inter-mesh arcs. Here, an assignment is simply a correspondence from a face in $\mathcal{S}_1$ to a face in $\mathcal{S}_2$. However, to increase the quality of the assignments, we also incorporate a learning procedure into this step. We learn which shape descriptors are most discriminative in distinguishing correct from incorrect assignments. The learning is performed on assignments derived from the training set. Since a large collection of descriptors is available, the learning also has the advantage of weighting the influence of each descriptor in the similarity computation.

First, we derive a set of training assignments $\mathcal{A}$ from the dataset of training shapes. For each type of shape descriptor, we compute the similarity between all possible pairs of faces in a sample of training shape pairs. Suppose that two training shapes are denoted as $\mathcal{T}_1$ and $\mathcal{T}_2$. For each face in $\mathcal{T}_1$, we select the $k$ first pairwise assignments with the highest similarity for each descriptor and add them to the set $\mathcal{A}$. The similarity is given by the inverse of the Euclidean distance between the descriptors of the two faces, while $k$ is set to 10 throughout our experiments. We choose a small $k$ so that we can capture the relation between the highest descriptor similarities and the correctness of the assignments. Each assignment in $\mathcal{A}$ is either labeled as *true*, if it maps two faces with the same label, or *false*, if the labels are different. Next, we associate to each assignment in $\mathcal{A}$ a vector that stores the similarity of each descriptor. We use such vectors to train a classifier using the "gentleboost" algorithm.

The outcome of the training is a classifier that allows us to label a candidate assignment with *true* or *false*, based on the vector of similarities obtained by considering multiple shape descriptors, while also assigning a confidence to this decision. Finally, given a pair of query

shapes, we select candidate assignments with the same procedure used for the training data. We apply the classifier and select the top 20% assignments with the most confidence of having the label *true*, to obtain a small yet reliable set of inter-mesh arcs.

**Labeling energy.** The energy to be minimized by the labeling is composed of two types of terms: unary and binary. The unary term (or data term) takes into consideration how likely it is that a given node has a specific label. This is encoded by a labeling cost assigned to each face. The binary terms consider the connectivity between faces (intra- and inter-mesh) and quantify how likely it is that two neighboring nodes have a specific pair of labels, according to a pairwise cost. We define the energy of the labeling $\mathbf{l}$ as

$$\begin{aligned}\mathcal{E}(\mathbf{l}) \;=\; & \sum_{u \in V} \mathcal{E}_{\text{data}}(u, l_u) \;+ \sum_{uv \in E_{\text{intra}}} \mathcal{E}_{\text{intra}}(u, v, l_u, l_v) \;+ \\ & \sum_{uv \in E_{\text{inter}}} \mathcal{E}_{\text{inter}}(u, v, l_u, l_v),\end{aligned} \tag{3.1}$$

where $l_u$ and $l_v$ are the labels of nodes $u$ and $v$, respectively, and $\mathcal{E}_{\text{data}}$ and $\mathcal{E}_{\text{intra}}$, $\mathcal{E}_{\text{inter}}$ are the unary and binary terms.

The unary (data) term is given by

$$\mathcal{E}_{\text{data}}(u, l_u) = -a_u \, \log p(l_u | D_u), \tag{3.2}$$

where $p(l_u|D_u)$ is the probability of node $u$ having label $l_u$, based on the classifiers applied to the face descriptors $D_u$, and $a_u$ is the area of the face corresponding to node $u$. The weight $a_u$ ensures that the cost is given in terms of labeling the total shape area.

The intra-mesh binary term is defined as

$$\mathcal{E}_{\text{intra}}(u, v, l_u, l_v) = L(l_u, l_v)[\omega_\theta \, \theta_{uv} \;+\; \omega_\ell \, \ell_{uv}], \tag{3.3}$$

where we take into account the compatibility $L(l_u, l_v)$ between two labels, as well as the edge length $\ell_{uv}$ and dihedral angle $\theta_{uv}$ between faces $u$ and $v$, similarly as done in [Shapira et al. 2009] and [Kalogerakis et al. 2010]. The label compatibility term $L$ is derived from the training data in the form of statistics that quantify how likely it is that two labels appear neighboring each other. $L(l, l) = 0$ if two faces share the same label $l$. The parameters $\omega_\theta$ and $\omega_\ell$ regulate how much the angle and edge length contribute to the total energy.

Finally, the inter-mesh term is given by

$$\mathcal{E}_{\text{inter}}(u, v, l_u, l_v) = L(l_u, l_v)[\omega_\eta \, \eta_{uv}], \tag{3.4}$$

where $\eta_{uv}$ is the confidence that the assignment between faces $u$ and $v$ is correct, and $\omega_\eta$ regulates the influence of the inter-mesh term to the total energy. Thus, the higher the confidence value attached to the assignment, the more the cost is increased if the assigned labels are different.

**Graph-cut optimization.** We use multi-label graph-cuts to assign labels to the nodes in an optimal manner. More specifically, the $\alpha$-$\beta$ swap algorithm is utilized [Boykov et al. 2001], since the pairwise costs $L$ do not define a metric. By minimizing the given energy, we obtain the most likely label for each face while also avoiding the creation of small disconnected segments. The optimal parameters $\omega_\theta$, $\omega_\ell$, and $\omega_\eta$ are obtained by performing a grid search on training data separated for this purpose. This procedure is explained in Section 3.5.

## 3.5 Experimental results

In this section, we present a set of experiments aimed at evaluating our approach for shape correspondence. We also contrast our results to other state-of-the-art methods.

**Datasets.** We utilize two datasets in our experiments. All the shapes are pre-segmented and labeled, implying that the ground-truth label for each mesh face is known. We designed the first dataset, shown in Figure 3.2, composed of four classes of man-made shapes with large geometric and topological variability. Notice that the presence of some of the semantic parts is optional, and certain parts can appear more than once on each shape. The second dataset consists of a selected subset of classes from the mesh segmentation benchmark [Chen et al. 2009]. The selected classes are listed in Table 3.1 and examples appear in Figure 3.5 (a)-(e). We utilize the segmentations and labelings created for these shapes by Kalogerakis et al. [2010]. We selected classes that possess shapes in different poses (e.g., Human or Hand) and shapes with considerable structural and geometric variability (e.g., FourLeg), as opposed to models with a predictable structure. Also note that the segmentations for the first dataset were created by a single user with a pre-defined goal, while the second dataset is given by the average segmentation for each class. With this setting, we demonstrate that our method is robust against variations in the dataset design process.

**Correspondence results.** Figure 3.5 shows a set of visual results. To generate these correspondences, we selected for each class a random subset of 60% of the shapes as the

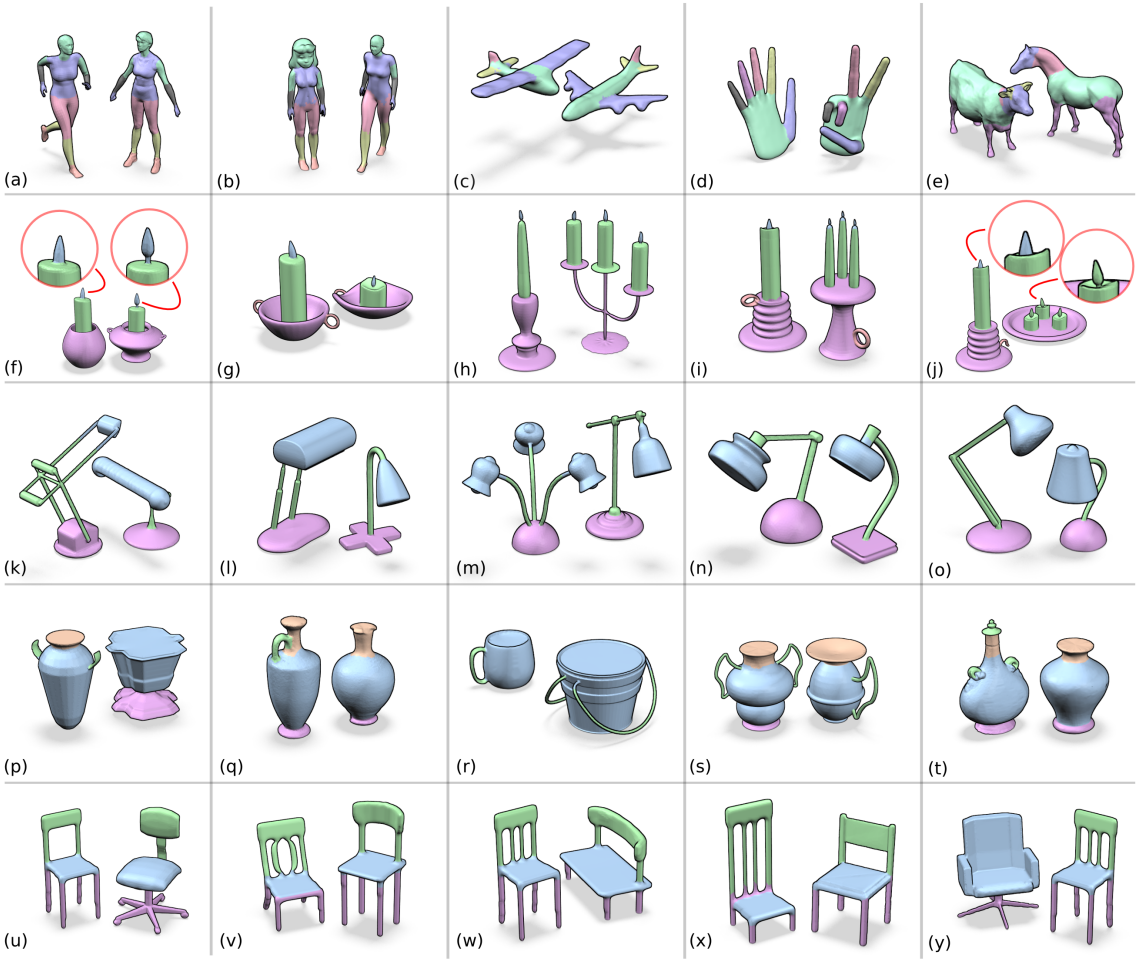



Figure 3.5: Part correspondence results via joint labeling. Aside from the failure cases in (j), (t), and (y), our method succeeds even under significant geometric and topological variations between the matched parts, while purely content-driven methods would fail (see Figure 3.7 for a comparison). Corresponding segments between a pair of shapes are shown with the same color.

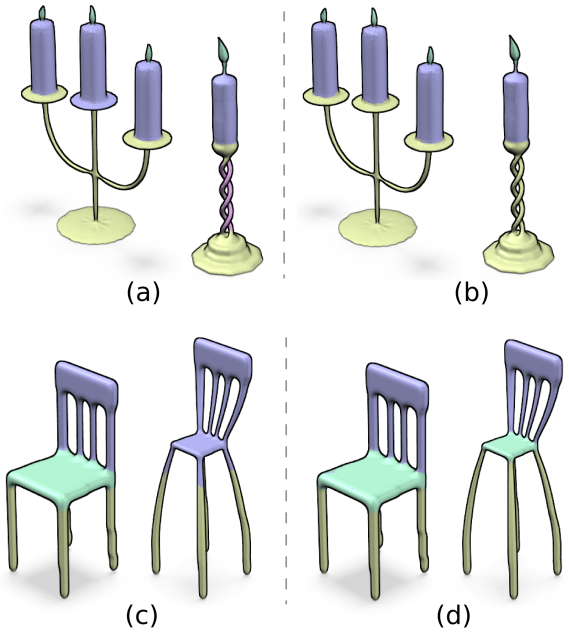



Figure 3.6: Joint labeling improves upon the use of prior knowledge alone. (a) and (c): correspondences obtained using probabilistic labels only, without content analysis. (b) and (d): results from joint labeling. The latter is effective in implying the correct correspondence for mislabeled parts when there is sufficient part-to-part similarity.

training set, and delegated the remaining shapes to be test cases. A random subset of 3/4 of the training shapes was used to learn the classifiers, while the remaining 1/4 of shapes was used to select the best parameters for the joint labeling. We perform a two-level grid search on the three parameters $\omega_\theta$, $\omega_\ell$, and $\omega_\eta$ and select those that result in the best labeling of the selected 1/4 of the training data, according to the ground truth. Next, we compute the joint labeling for all the pairs of test shapes. Some of these pairs of test shapes are shown in Figure 3.5.

Firstly, we see in Figure 3.5 (a)-(e) that our method is able to establish meaningful correspondences for queries that are also handled by content-driven methods. In addition to that, and more importantly, we also observe that the prior knowledge is effective in matching shapes whose parts differ by significant geometric changes. Examples include matching different types of candelabra in (f)-(j), different lights and bases in (k)-(o), variations in the recipients and bases in (p)-(t), and seat rests with different railings in (u)-(x). Our method is also able to handle topological variations, as the different lamp supports in (k)-(m), and

variations in chair legs as in (u) and (y). Finally, our method is also successful in matching shapes with different numbers of parts, such as the multiple wax candles in (h)-(j), and the different numbers of handles in (r) and (s). Notice that the presence of additional parts, such as the handles in (p) and (q), do not affect the accuracy of the results. Figure 3.5 also shows cases where our method did not provide the most accurate correspondence, mainly due to the lack of sufficient prior knowledge and limitations in the shape descriptors. Examples are the missed candle flame in (j), the extra handle added to the top of the vase in (t), and the rest and arms of the chair in (y) that were not properly separated from the seat.

In cases of insufficient knowledge, an accurate correspondence can still be obtained via joint labeling when there is enough similarity between parts of the shapes. Such examples can be seen in Figure 3.6. In (a), the correspondence is computed based purely on the knowledge, i.e., only the unary and intra-mesh terms enter the labeling. Part of the support of the candelabrum on the right is mistakenly labeled as a handle and, therefore, does not have a corresponding part on the shape on the left. However, when the content of the shapes is also taken into consideration (i.e., the inter-mesh term is added), we obtain the more accurate correspondence in (b), since the descriptors of the thin structures on both shapes are similar. We observe an analogous situation in (c), where the seat of the chair is mistakenly labeled as a backrest, due to the geometric distortion present in the model. In (d), the joint labeling is also able to obtain the correct part correspondence, guided by the similarity of the seats. We point out that both knowledge and content are essential for finding the correct correspondence in these cases.

In Table 3.1, we show a statistical evaluation of the correspondence results we obtained. For each shape class, we performed 5 experiments in the same manner as described before (60% training shapes and 40% test shapes). We average the accuracy over all the pairs of test shapes in each experiment. The accuracy for a single pair of shapes $\mathcal{S}_1$ and $\mathcal{S}_2$ is calculated with a function that measures the quality of the correspondence between the two shapes. It is given by

$$\text{Accuracy}(\mathbf{l}, \mathbf{t}) = \frac{\sum_{i \in \mathcal{S}_1} \sum_{j \in \mathcal{S}_2} a_i \, a_j \, |\delta(l_i = l_j) + \delta(t_i = t_j) - 1|}{\sum_{i \in \mathcal{S}_1} \sum_{j \in \mathcal{S}_2} a_i \, a_j}, \tag{3.5}$$

where $a_i$ is the area of face $i$, $\mathbf{l}$ is the labeling of the shapes returned by our method, $\mathbf{t}$ is the ground-truth labeling, and $\delta(x = y)$ is 1 only if $x = y$. This quantity measures how many of the faces that have the same label in the ground-truth are also found in corresponding segments, independently of which labels were assigned to the faces. It also captures the

Table 3.1: Statistical evaluation of the correspondence results.

| Class | Corr. |
|---|---|
| Candelabra | 88% |
| Chairs | 87% |
| Lamps | 97% |
| Vases | 86% |

| Class | Corr. |
|---|---|
| Airplanes | 92% |
| Ants | 96% |
| Birds | 86% |
| Fish | 92% |

| Class | Corr. |
|---|---|
| Four-legged | 82% |
| Hands | 80% |
| Humans | 90% |
| Octopuses | 96% |

notion that two faces that do not possess the same label in the ground-truth should not be assigned to corresponding segments. The weighting is chosen so that the correspondence accuracy is given in terms of the total shape area.

We can see that the average accuracy for the first dataset (classes of man-made shapes with significant variability) is 90%, while for the second dataset (organic shapes), it is 89%. We attribute the failure cases to a few factors. Firstly, shapes that have parts with no counterparts in the prior knowledge can occur in the test set. An example is the chair shown before in Figure 3.5 (y), contrasted to the other shapes in the dataset (Figure 3.2). Secondly, the descriptors may not be sufficient to distinguish certain parts of the shapes, such as the different fingers in the hands. Finally, a small fraction of the errors is also due to imperfections in the labeling, such as dislocated borders which happen when the inter-mesh and intra-mesh terms compete in the optimization. Such cases can be adjusted, for example, with post-processing that displaces the borders to concave regions [Shamir 2008].

When comparing the statistical results of the approach using only prior knowledge (unary and intra-mesh terms) to that of the joint labeling (incorporating the inter-mesh term), we find that both approaches are comparable, with a deviation of $\pm 5\%$ in the accuracies. The joint labeling does not lead to a significant increase in the accuracies since the inter-mesh term is primarily designed to handle cases such as those in Figure 3.6, where the content aids in improving the correspondence for certain portions of the shapes that do not appear in the knowledge. Such cases are not very frequent in the datasets, but could be prevalent in certain situations (e.g., a collection of shapes modeled by reusing existing parts).

**Comparison to classical approach.** We compare our approach with two state-of-the-art content-driven methods that are considered to be comparatively competitive at handling general deformation between models. In Figure 3.7, we show a comparison with the deformation-driven approach of Zhang et al. [Zhang et al. 2008], which finds a matching

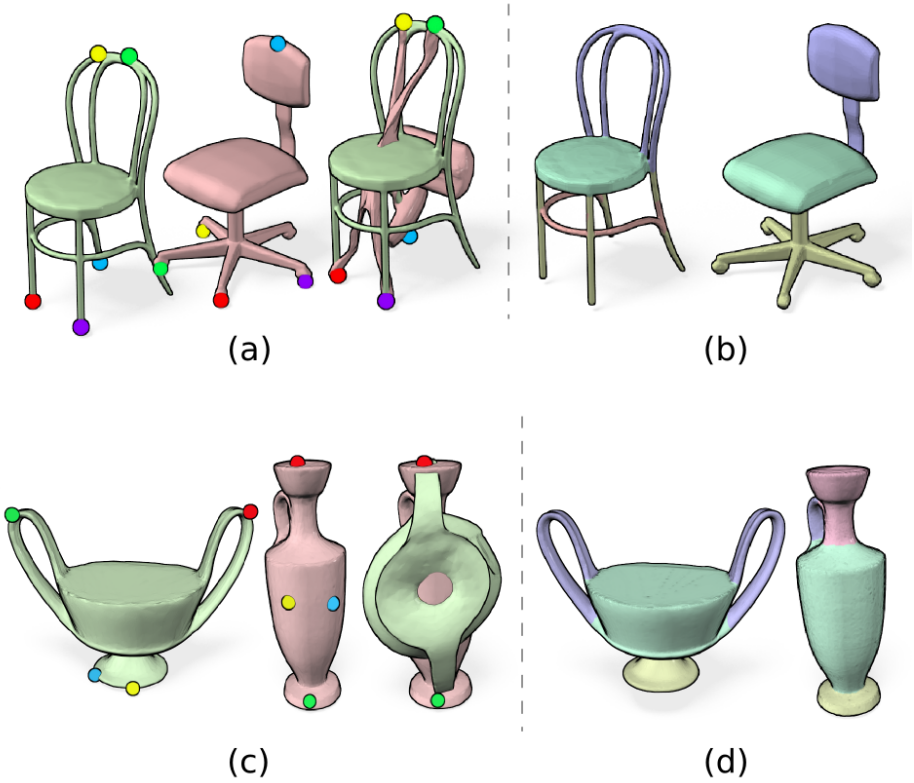



Figure 3.7: Comparison to content-driven correspondence approach on geometrically dissimilar models. (a) and (c): results from the deformation-driven method of Zhang et al. 2008, including the induced deformation and the matching feature points. (b) and (d): results from our joint labeling method.

between sparse sets of feature points. As we can see, when the corresponding parts differ sufficiently in their scales or geometric properties, this method fails as it is still formulated in terms of matching the geometry of the shapes. Notice that these two shapes are incorrectly matched not only because the method tries to find the best non-rigid alignment between the two shapes; as can be seen, it is also difficult to extract coherent feature points on these two shapes, and the shape descriptors of corresponding points are dissimilar due to the different part composition of the shapes (especially if the descriptors capture the context or neighborhoods of the points). Our method on the other hand finds a correspondence between two shapes by using knowledge as the medium; it succeeds since the resulting correspondences have sufficient support from the training set. We also compared our method to that of Shapira et al. [2009]. However, when presented with the query examples shown

in Figure 3.7, the partitioning of the shapes obtained with this method differed significantly from one shape to the other, preventing the method from establishing a meaningful correspondence. Other recent works on content-driven shape correspondence, e.g., [Lipman and Funkhouser 2009; Au et al. 2010], are not expected to succeed on these examples either as they are also based on geometric similarities between matched features, an approximate isometry criterion, or a combination of both.

**Timing.** The most expensive procedure in the knowledge-driven framework is classifier learning. For a training set ranging from 20 to 30 models with average size of 30K triangles, this step can take 10 hours in an AMD Opteron 1GHz processor. Then, applying the classifiers on two queries and performing the joint labeling runs in the order of minutes.

## 3.6 Discussion

The key idea presented in this chapter is that challenging cases of 3D shape correspondence can be solved effectively by incorporating prior knowledge. At the same time, considering the direct geometric similarity between the query shapes is still advantageous, particularly when the knowledge base is incomplete or leads to indeterminate recognition results. Thus, we introduced an effective approach via joint labeling, which combines knowledge-driven probabilistic semantic labeling with content-driven analysis. The content analysis is incorporated in the form of assignments between similar local regions of the shapes. We demonstrated significant improvement on shape correspondence results over classical approaches, particularly when the query shapes exhibit large geometric or topological variations.

**Limitations.** While the idea of joint labeling is quite general, the content analysis component of our approach is still fairly primitive in terms of the feature similarity employed. Also, there still remain failure cases as shown in Figure 3.5. We believe that this can be attributed to our current reliance on low-level shape descriptors in the content-driven analysis as well as in the recognition step, where a query shape is compared to shapes in the training set. More advanced geometric analysis tools incorporating criteria such as shape symmetry [Golovinskiy et al. 2007b] or style-content separation [Xu et al. 2010] may lead to improvements.

**Future directions.** It would be interesting to explore the possibility of avoiding the need for pre-classifying the training shapes. Either we could integrate the recognition of the training parts with the recognition of the shapes, or we could avoid it altogether by developing generic labels and classifiers that capture the general notion of handle and base, rather than a cup handle or a lamp base, for example. In hindsight, what ultimately makes the correspondence approach effective under challenging circumstances, particularly for many classes of man-made shapes, is the ability to learn the *functionality* of the parts. Recognizing functionality is clearly a difficult problem. It calls for intermediate-level descriptors that are able to capture properties such as flatness, concavity, and symmetry that are required in order to achieve a certain functionality. Learning functionalities of parts for shape analysis is certainly an interesting direction for future work.

# Chapter 4

# Unsupervised co-segmentation of a set of shapes

As motivated in Chapters 2 and 3, high-level analysis of 3D shapes has received increased attention in recent works. There are methods that infer high-level knowledge about a given shape from its geometry [Fu et al. 2008; Mitra et al. 2010]. There are also works which utilize semantic knowledge to segment a given shape [Simari et al. 2009; Kalogerakis et al. 2010] or establish a correspondence between a pair of shapes [van Kaick et al. 2011b] (described in Chapter 3). The problem of analyzing a *set* of shapes as a whole has received less attention [Golovinskiy and Funkhouser 2009; Xu et al. 2010]. The interesting question about *co-analysis* of a set of shapes is whether more knowledge can be inferred from the set rather than from an individual or pairs of shapes alone. For example, can we better segment the shapes given as a set rather than as individuals? While it seems obvious that a set of shapes contains more knowledge than each individual, it remains a challenge, particularly in the *unsupervised* setting, to extract appropriate knowledge inherent to the set to facilitate fundamental analysis tasks such as segmentation and correspondence.

In this chapter, we investigate the problem of unsupervised co-segmentation of a set of shapes, where our goal is to reveal the semantic shape parts and establish their correspondence across the set. In our setting, the input shapes belong to a common family, that is, loosely speaking, they share the same functionality and general form. However, their corresponding parts are not necessarily similar; see Figure 4.1. The co-segmentation is unsupervised in that there is no training set which provides any knowledge to assist the

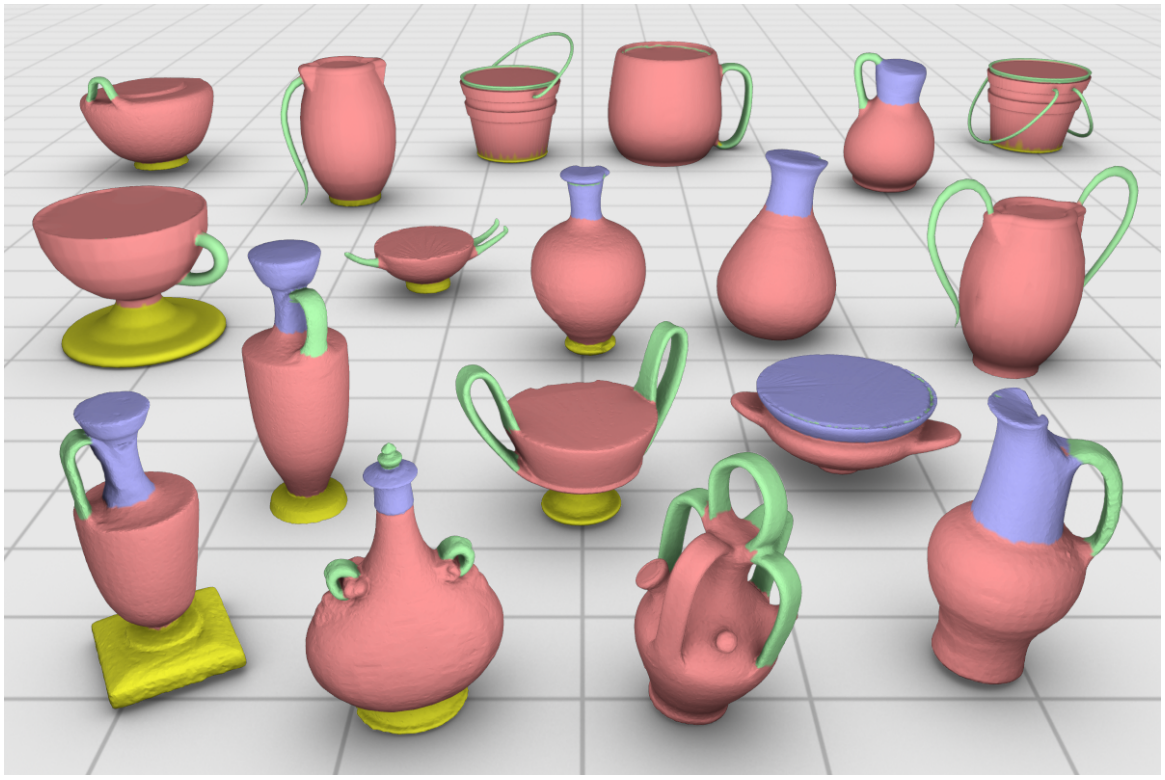


Figure 4.1: Unsupervised co-segmentation of a highly varied set of container objects using our algorithm. Corresponding parts differ in their shape, pose, position, and cardinality.

analysis, as opposed to works such as [Kalogerakis et al. 2010] and the work described in Chapter 3, where the ability to properly match geometrically dissimilar parts is critically supported by the existence of relevant prior knowledge in the training set. It should also be noted that the analyses performed in the above works do not represent a co-analysis of a *set*. The work in this chapter can be seen to complement these knowledge-driven approaches with knowledge extracted from a target set.

The setting and objective of our analysis share similarities with the recent works of Golovinskiy and Funkhouser [2009] and Xu et al. [2010], which both compute an unsupervised co-segmentation of a set of shapes. The method in [Golovinskiy and Funkhouser 2009] pre-aligns the set of shapes and then combines criteria for intra-shape segmentation and inter-shape proximity to cluster mesh faces across the set. Xu et al. [2010] apply a similar co-segmentation scheme with the focus being to remove non-homogeneous part scales from

the analysis equation. In our setting, the set consists of shapes with a variety of non-rigid, geometric, and even topological differences; see Figure 4.1. The dissimilarity between corresponding parts is so pronounced that simply spatially aligning the shapes is not effective. Towards this end, our approach allows correspondence to be inferred indirectly through *third parties* in the set and the analysis is performed in a *descriptor space*; see Figure 4.2.

Specifically, we treat the unsupervised co-segmentation of a set of shapes as a clustering problem. The clustering is performed in a space of shape descriptors rather than on the spatial coordinates of the shapes themselves. This allows the handling of corresponding parts which may differ in pose, location, and even cardinality. Obviously, such variations would challenge any technique based on spatial alignment or direct clustering of shape geometry, e.g., [Golovinskiy and Funkhouser 2009], as shown in Figure 4.3. In addition, the descriptor clustering approach allows us to make use of a key enabling feature of the input set, namely, third-party connections. Even if two shapes possess parts that are significantly dissimilar, we can still establish a link between them if there are other parts in the set (third parties) that create such a connection, resulting in a successful co-segmentation. Figure 4.4 provides a concrete example. In contrast, spatial-domain alignment and clustering alone is unable to fully utilize the existence of third-party connections.

When performing the analysis in a descriptor space, the clusters that characterize the different shape parts do not necessarily take on an isotropic form; they may be elongated so that two corresponding parts which should belong to the same cluster are relatively far apart, e.g., see Figure 4.2. To this end, we perform spectral clustering with the aid of diffusion maps, which take the non-linear and anisotropic structures in the data and unfold them into a new space, so that the similarities between data points are translated into geometric proximity. A simple clustering algorithm can then succeed in the embedded space by simply considering Euclidean distances.

We show that our unsupervised approach is able to co-segment families of shapes with significant geometric variations, achieving results that are competitive to supervised approaches [Kalogerakis et al. 2010; van Kaick et al. 2011b].

## 4.1 Related work

Shape segmentation [Shamir 2008] and correspondence [van Kaick et al. 2011a] are two of the most fundamental problems in high-level shape analysis. To segment a shape into

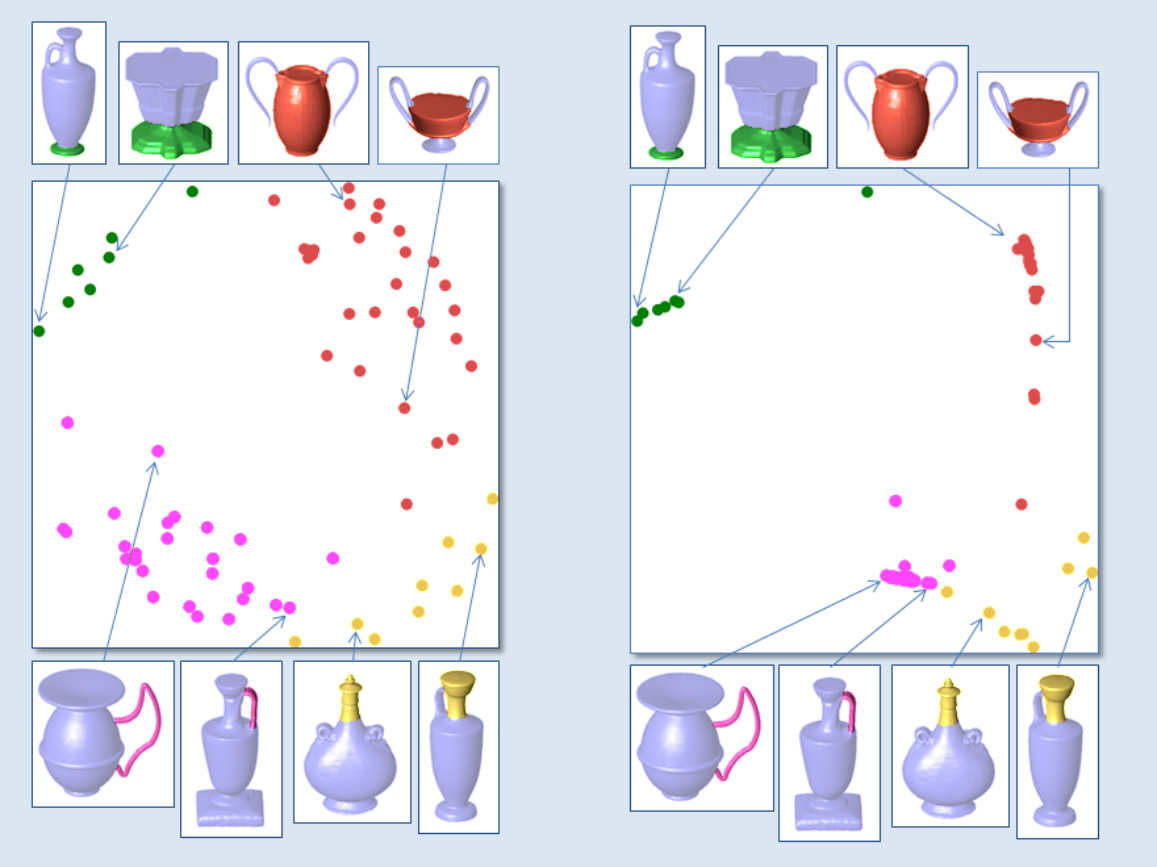
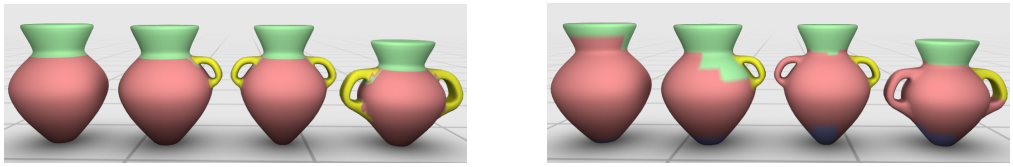


Figure 4.2: Challenge of unsupervised co-segmentation amid significant geometric variation (left) and effectiveness of descriptor-space spectral clustering (right). In the original descriptor space (left), two segments in the same semantic class (two pink handles) can be far apart, while unrelated segments (a pink handle and a yellow neck) can be closer. It is challenging to resolve this without any knowledge of the semantic classes. However, the handles are drawn close in the diffusion map (right) through *third-party* connections. The third parties, which are all the segments lying in-between the two handles, establish several paths between the two segments, given by the high similarities between pairs of points. These multiple paths create a strong connection between the two handles. Note that the two plots are 2D embeddings of the descriptor space, obtained with multidimensional scaling.



(a) (b)

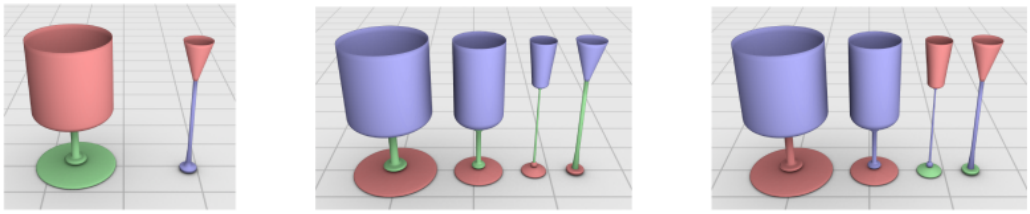
Figure 4.3: Co-segmentation in descriptor space (left) vs. in spatial domain [Golovinskiy and Funkhouser 2009] (right). The use of descriptor-space clustering enables our method to handle variations in part placement and cardinality (see the handles). Co-segmentation results via spatial alignment are less meaningful.

meaningful parts, and to compare or find a correspondence between these parts, we need to understand the high-level structure of the shapes. Much effort has been devoted to solving these problems, however, the endeavor has been mainly focused on analyzing only one shape or a pair of shapes at a time.

The question that naturally follows is whether we can benefit from simultaneously analyzing a set of shapes from the same family, since intuitively more information is then available. In the image domain, positive results for co-segmentation have been demonstrated, e.g., by using a generative model to match the appearance histogram of two images and enforce spatial coherency [Rother et al. 2006], or by making use of discriminative clustering, which seeks maximal separation of the classes [Joulin et al. 2010].

In the case of shapes, a first indication of the advantage of set analysis appears in the works of Kalogerakis et al. [2010] and van Kaick et al. [2011b] (described in Chapter 3), where the knowledge is represented in the form of discriminative part models which are learned from a set of manually segmented and labeled shapes. The models can then be used to segment and label an unknown shape from the same class [Kalogerakis et al. 2010], or to establish a part correspondence between a pair of shapes [van Kaick et al. 2011b]. Although the knowledge is defined in terms of a set of shapes, it is still manually created and not automatically inferred from the set.

The works of Golovinskiy et al. [2009] and Xu et al. [2010] take a concrete step towards co-analysis of shapes and propose methods that infer knowledge from the set alone. In [Golovinskiy and Funkhouser 2009], the co-segmentation is posed as a graph clustering problem. In this graph, each node corresponds to a face in one of the meshes, and the edges



(a) (b) (c)

Figure 4.4: The power of third-party connections. When presented with a set containing only two rather dissimilar shapes (a), our co-segmentation scheme returns a less-than-satisfactory result. However, when the set is augmented with the two "in-between" shapes (b), they serve as third-parties which help establish a link between the dissimilar parts, providing a meaningful co-segmentation of the set. In contrast, a method based on spatial alignment and clustering [Golovinskiy and Funkhouser 2009] does not benefit from this property and leads to a less meaningful co-segmentation (c).

come from: (1) the individual connectivity of the shapes, and (2) from a set of correspondence edges connecting faces that are geometrically close. The correspondence edges are added after the shapes were aligned to each other. The clustering of this graph naturally provides a per-shape segmentation that is coherent across the group; the method is however limited to shapes that can be spatially aligned; see Figures 4.3 and 4.4.

To overcome limitations of global alignment and scaling [Kazhdan et al. 2004; Golovinskiy and Funkhouser 2009], Xu et al. [2010] classify the input shapes into different *styles* according to the scales of the shape parts. Co-segmentation based on graph clustering is then applied only to the shapes within each style cluster, and is modified to take into consideration the parts and derive the correspondence edges from part similarity. This modification allows the co-analysis to succeed for a larger variety of shapes, especially those whose parts differ by non-homogeneous part scaling.

We are interested in co-segmenting shapes with more variability than those in [Golovinskiy and Funkhouser 2009; Xu et al. 2010], such that corresponding parts can be rather dissimilar geometrically as well as topologically. Therefore, we deviate from the scheme of spatial alignment or clustering the shape parts in the spatial domain where the shapes reside. Instead, we automatically derive statistical models that describe the different parts of the shapes in a space of shape descriptors and utilize spectral clustering to account for

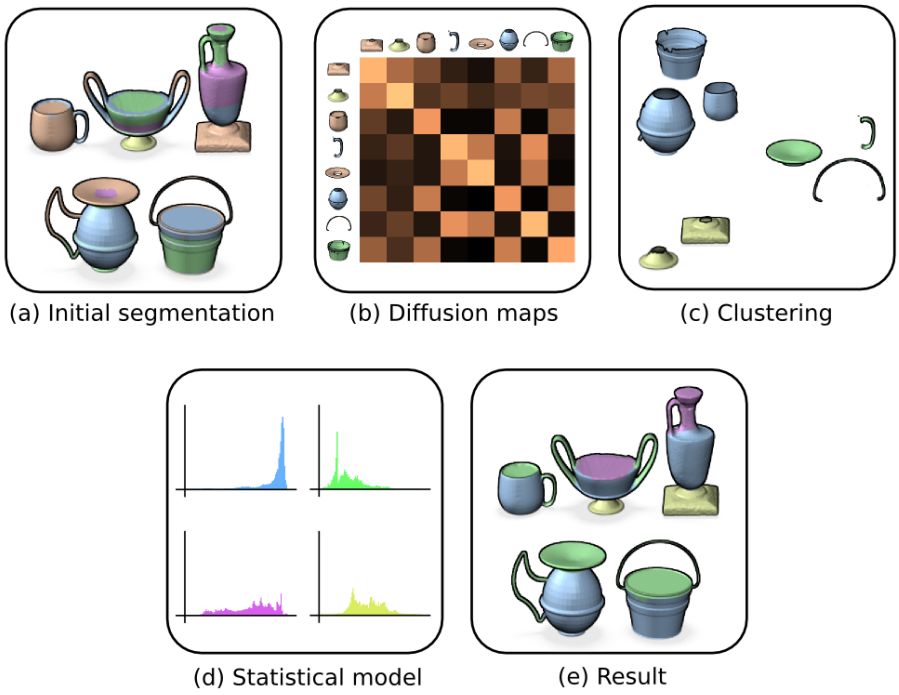



Figure 4.5: Overview of the steps in our co-analysis. (a) An individual segmentation is computed for each shape. (b) The segments from all the shapes are embedded into a common space by using diffusion maps based on a similarity matrix, where darker colors in the matrix entries indicate higher similarities. (c) The segments are clustered in the embedded space. (d) A statistical model is created to describe each cluster. (e) The statistical model is used to label the shapes and obtain the final co-segmentation of the set.

clusters of arbitrary shapes.

In an independent work, Huang et al. [2011] describe an unsupervised algorithm that jointly segments shapes in a heterogeneous shape library, obtaining results comparable to supervised approaches on a benchmark test. However, unlike our approach, this technique does not guarantee that the segmentations of all shapes within a shape class are consistent, that is, parts that share the same semantic meaning do not necessarily carry the same label.

There has been a large body of works in geometry processing using spectral methods [Zhang et al. 2010]. Spectral clustering via diffusion maps is also not new. The mesh segmentation work of de Goes et al. [2008] specifically applies the diffusion distance, as we do in our work. All these solutions exploit special properties of the spectral embedding.

However, the computed embeddings have always been a transformation from the spatial coordinates of individual input shapes. Our work computes spectral embeddings of shape descriptors. The descriptors of all the shapes in the set take part in the embedding, enabling us to perform a co-analysis.

## 4.2 Overview

Our co-analysis method takes as input a set of meshes from a given family and computes their co-segmentation and labeling. The co-analysis also provides a correspondence among the segments of any pair or group of shapes in the set, since the segments corresponding to the same part class will possess a common label. The label does not necessarily carry a semantic meaning, but serves more as a part index. For each family of shapes, the user also provides the *maximum* number of labels $L$ that should be recovered from the set. This number loosely corresponds to the number of different kinds of semantic parts that constitute the shapes, e.g., $L = 4$, for a set of vases that can have a base, body, handle, and neck. Note that some types of parts can repeat or be omitted on the shapes, e.g., there can exist vases with multiple handles and vases without a base.

To carry out the co-analysis, we start with a per-object segmentation of each shape in the input set. Then we extract shape descriptors for the initial sets of segments. Next, based on the descriptors, we cluster the segments using diffusion maps. Finally, we build a statistical model for each cluster, which is used to obtain the final co-segmentation and labeling of the shapes in the set. We describe these steps as follows (see Figure 4.5).

**Per-object segmentation.** The first step in the co-analysis is to obtain an individual segmentation for each shape in the input set. We achieve this by grouping the mesh faces with mean-shift clustering based on shape descriptors defined for the faces, although any reasonable alternative can be used here. More details are given in Section 4.3. The outcome of this procedure is a set of candidate segments per shape. The purpose of the per-object segmentation is to facilitate the co-analysis, since we are interested in analyzing shape parts, rather than lower-level primitives. However, notice that the final co-segmentation given by our algorithm is a refined labeling performed at the face level (Section 4.5), which allows us to correct imperfections that appear in the per-object segmentation.

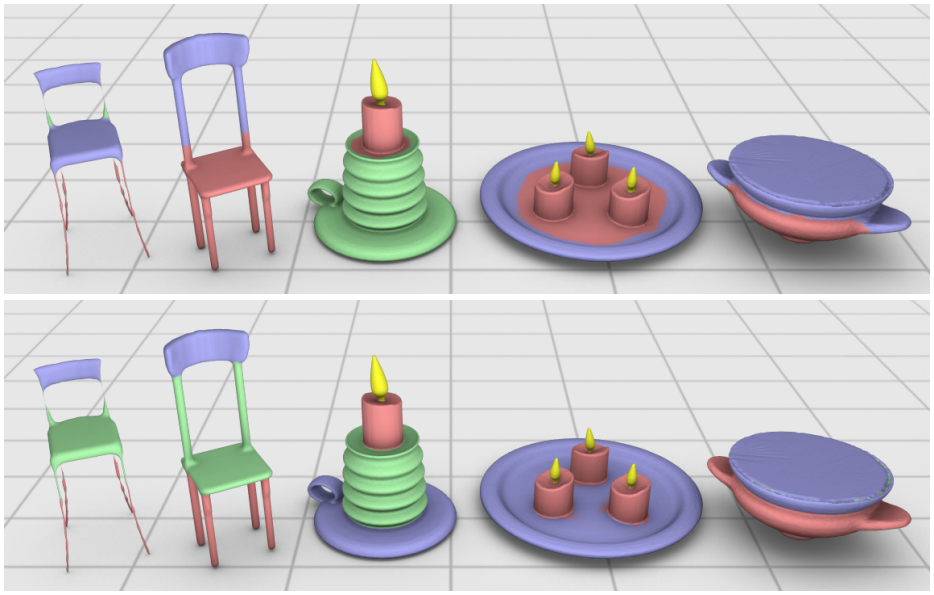


Figure 4.6: Result of the co-segmentation refinement (bottom) applied on an initial co-segmentation (top). Notice how boundaries are displaced to better locations (candles) and some mislabeled segments receive the correct label after the refinement (chairs). The co-segmentation is performed independently for each class.

**Descriptor-space spectral clustering.** All the candidate segments produced by the per-object segmentation are embedded into a common space via diffusion maps. The embedding translates the similarity between segments into spatial proximity, so that a clustering method based on Euclidean distance will be able to find meaningful clusters in this space; see Figure 4.2 for an example. To define the embedding, an affinity matrix is constructed from the similarities between pairs of segments. The similarities are given by the distance between descriptors defined at the segment level. The embedding is then given by the first few scaled eigenvectors of the affinity matrix. Note that the embedding is constructed from the similarities among *all* the segments in the set. This implies that, even if two corresponding parts are far apart in the original descriptor space, they may end up close-by in the embedding if other parts in the set (third parties) imply their correspondence by transitivity. We obtain an initial co-segmentation by clustering the segments in the embedded space. All the segments in a single cluster potentially represent a certain class of parts, e.g., base, body, handle, or neck of a vase. This step is carried out with a hierarchical

clustering scheme. We elaborate on descriptor-space clustering in Section 4.4.

**Statistical model and refined co-segmentation.** A statistical model is then built to describe each cluster of parts and used to label the shapes. The advantage of this final step of the method is that it allows us to correct errors that appear in the initial segmentation, since now we perform a more detailed labeling of the shapes (Figure 4.6). More specifically, the statistical model is defined in terms of Gaussian models learned from the shape descriptors. We derive the probability of labeling each face with a given class by how well the descriptors of the face fit the part model. Next, we obtain the best segmentation for each shape by applying graph cuts labeling based on these probabilities. The final result is a refined co-segmentation and coherent labeling of all the shapes in the set. More details on this step of the algorithm are given in Section 4.5.

## 4.3 Descriptors and per-object segmentation

**Shape descriptor extraction.** We normalize the shapes for overall scale, and compute a set of shape descriptors that are used in the different steps of the co-analysis. The algorithm works with faces and segments (a group of connected faces). Thus, we compute descriptors both at the face-level and at the segment-level. First, we obtain an upright orientation of each shape [Fu et al. 2008] and define a subset of the face-level descriptors based on the reoriented shapes. More specifically, we record for each face, the geodesic distance from the base of the shape to the face, and the angle between the normal of the face and the upright orientation vector. Finally, we also make use of the shape diameter function [Shapira et al. 2009], which gives an estimate of the thickness of the shape at the face.

For the segment-level descriptors, we compute histograms that capture the distribution of each face-level descriptor for all the faces in the segment. For a segment $s_i$ and the face-level descriptor $d$, we denote the histogram as $h_i^d$. We also include two descriptors only defined at the segment level. $a_i$ is the segment area normalized by the total shape area. $g_i$ is a vector of three components that describes the overall geometry of the segment. We have $g_i = [\gamma_l \, \gamma_p \, \gamma_s]$, where

$$\gamma_l = \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2 + \lambda_3}, \quad \gamma_p = \frac{2(\lambda_2 - \lambda_3)}{\lambda_1 + \lambda_2 + \lambda_3}, \quad \text{and} \tag{4.1}$$

$$\gamma_s = \frac{3\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3}, \quad \text{with} \quad \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0. \tag{4.2}$$

These terms give an indication of how linear (cigar-shaped), planar and spherical the shape of the segment is. $\lambda_1$, $\lambda_2$, and $\lambda_3$ are the three eigenvalues obtained when applying principal component analysis to all the vertices that are part of the segment.

**Per-object segmentation.** Using the shape descriptors, we compute a per-object segmentation for each shape in the set. Although any reasonable segmentation algorithm can be used for this step, we opted to use the *mean-shift algorithm* [Comaniciu and Meer 2002], where we cluster the mesh faces into larger segments. Mean-shift operates by finding the modes (local maxima of density or cluster centers) of points in feature space. The advantage of using mean-shift lies in its non-parametric nature, i.e., the number of clusters does not have to be known in advance. Instead, the algorithm requires an estimation of the clustering *bandwidth* parameter, which is the radius of support around the neighborhood of a point used to compute the point's mean. However, the bandwidth can be typically estimated from the data, e.g., as done in shape analysis [Shamir et al. 2006]. Following a similar procedure, we manually fix the bandwidth in our experiments to a percentage of the range of each descriptor. The actual distance measure used for the clustering is derived from the descriptors, but defined in terms of diffusion distances (Section 4.4). Finally, disconnected clusters are broken into separate segments according to their connectivity.

The output of this step is a set of candidate segments for each object. Notice that the result can constitute an over-segmentation of the shapes, i.e., a semantic part might be composed of more than one segment. However, our goal here is to obtain candidate segments as suggested by the shape descriptors. These segments will be utilized in the co-analysis of the set, which will then provide a refined segmentation and coherent labeling of the shapes.

## 4.4 Descriptor-space spectral clustering

We recall that one of our goals in the co-analysis is to extract information on what types of parts compose the shapes in the set. With this objective in mind, we take the segments computed individually for each shape (as described in the previous section) and cluster them into groups of similar segments. Next, we derive a statistical model for each cluster to represent each type of part that appears in the set. More details are given in Section 4.5.

However, it is important to notice that the objects that we consider can have a large amount of variability. Thus, a simple clustering algorithm will not group the segments into

the proper classes, since the segments are non-uniformly distributed in descriptor space. The shape of the clusters can be highly anisotropic or even non-linear. To overcome this difficulty, we first embed the segments into a new space with the aid of diffusion maps, where the Euclidean distance between two segments will better reflect their similarity. Then, clustering in the embedding will provide a more accurate grouping of segments. The usefulness of applying the diffusion maps for co-segmentation is illustrated in Figure 4.2.

**Embedding computation.** We start with the set of segments obtained from all the shapes, $S = \{s_1, \ldots, s_n\}$. The dissimilarity between two segments $s_i$ and $s_j$ is given by

$$\mathcal{D}(s_i, s_j) = \sqrt{\sum_{d=1}^{n_d} \mathrm{EMD}^2(h_i^d, h_j^d) + |a_i - a_j|^2 + \|g_i - g_j\|_2^2}, \tag{4.3}$$

where $h_i^d$, $a_i$, and $g_i$ are the segment-level descriptors discussed in Section 4.3 and $n_d = 3$. EMD is the earth-mover's distance, a common measure of the dissimilarity between two probability distributions, since the $h_i^d$ are histograms.

Next, we construct an affinity matrix $W$, with

$$W_{i,j} = \exp(-\mathcal{D}(s_i, s_j)/2\sigma^2). \tag{4.4}$$

Note that, in our method, we obtain the pairwise affinities by applying a Gaussian kernel to the segment dissimilarities, but other choices for the kernel are also possible [Nadler et al. 2005]. By defining a diagonal matrix $D_{i,i} = \sum_j W_{i,j}$, we obtain the normalized $M = D^{-1}W$. $M$ can be seen as a stochastic matrix, with $M_{i,j}$ being the probability of a transition from segment $s_i$ to segment $s_j$ in one time step. The transition probability can be interpreted as the strength of the connection between the two segments.

Finally, we compute the eigendecomposition of the matrix $M$, obtaining eigenvalues $\lambda_0 = 1 > \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{n-1} \geq 0$ and eigenvectors $\psi_0, \ldots, \psi_{n-1}$ [Nadler et al. 2005; Coifman and Lafon 2006]. The diffusion map at time $t$ is then given by

$$\Psi_t(s) = (\lambda_1^t \psi_1(s), \ldots, \lambda_{n-1}^t \psi_{n-1}(s)), \tag{4.5}$$

where $\Psi_t(s)$ defines the coordinates of segment $s$ in the embedding or map. The eigenvector $\psi_0$ is constant and is thus discarded (according to the Perron-Frobenius theorem, every stochastic matrix has such a vector if all row sums are equal to one).

**Interpretation.** The main result regarding diffusion maps is that the Euclidean distance between two points $x$ and $y$ on the map is equal to the *diffusion distance* between the two points [Nadler et al. 2005]. The diffusion distance is given by

$$\mathscr{D}_t^2(x,y) = \sum_z (p(t,z|x) - p(t,z|y))^2 w(z), \tag{4.6}$$

where $p(t,z|x)$ is the probability of transition from $x$ to $z$ in $t$ time steps, and $w(z)$ weights the local density at $z$, giving more weight to low density points. The larger the number of short paths that exist between $x$ and $y$, the more the distance will decrease, giving an indication of how strongly the two points are connected. The time parameter $t$ can be varied to analyze the structure of the points at different scales [Coifman and Lafon 2006]. An alternative interpretation is to see the map as the state of a dynamic system after $t$ steps of a diffusion process have taken place.

**Implementation.** We compute the diffusion map using only the first three eigenvectors, since the diffusion distance can be well approximated in this manner [Nadler et al. 2005]. Moreover, we select $t = 3$ for computing the embedding of all the segments and $t = 5$ for the initial per-object segmentation (as described in Section 4.3).

**Clustering.** After obtaining the diffusion maps, we cluster the segments in the embedded space with an agglomerative hierarchical algorithm. We start with each segment as an initial cluster and, during the incremental construction of the hierarchy, we merge the current pair of clusters with minimal distance. The distance between two clusters is given by the Euclidean distance between their centroids. The final number of clusters is provided by the user and corresponds approximately to the number of semantic parts that constitute the shapes. The result of the clustering is a grouping of the segments into the potential classes of parts that exist in the set.

## 4.5 Statistical model and co-segmentation

**Statistical model.** We now derive a statistical model for each class of parts. The models are constructed from the clusters of segments, based on the shape descriptors.

For each cluster $c_i$, we collect the descriptor values for all the faces of all the segments in the cluster. Based on the observed values, we estimate a multi-dimensional Gaussian to

model the class,

$$p(f|c_i) = p(D_f, \mu_i, \Sigma_i) = C\, e^{-\frac{1}{2}(D_f-\mu_i)^T \Sigma_i^{-1}(D_f-\mu_i)}, \tag{4.7}$$

where $\mu_i$ and $\Sigma_i$ are the parameters that model the $i$-th class, $D_f$ are the descriptors of face $f$, and $C$ is the Gaussian's normalization constant. The parameter dimensions are $3 \times 1$ and $3 \times 3$, respectively, as we use three face-level descriptors. The parameters are estimated with a standard expectation-maximization approach. We observed that this scheme provides a simple, yet effective, class model, e.g., as opposed to more complex models such as mixtures of Gaussians.

Finally, the probability that an unknown face $f$ belongs to class $c_i$ is given by Bayes' Theorem,

$$p(c_i|f) \propto p(f|c_i)p(c_i), \tag{4.8}$$

where the prior $p(c_i)$ is taken as the sum of the area of the segments that are part of cluster $c_i$, normalized by the total area of all segments in the set.

**Refined co-segmentation.** The statistical model for each class is used to perform the final co-segmentation. We pose the co-segmentation as a labeling optimization which is solved individually for each shape. The group information enters the optimization through the data term of the labeling energy.

Given a mesh, we define the graph $G = \{V, E\}$, where the nodes $V$ are given by the faces of the mesh and an arc $\{u, v\} \in E$ if the faces $u$ and $v$ are neighbors on the mesh. The optimization is then posed as finding the labeling $\mathbf{l}$ that minimizes the energy

$$\mathcal{E}(\mathbf{l}) = \sum_{u \in V} \mathcal{E}_{\text{data}}(u, l_u) \; + \sum_{uv \in E} \mathcal{E}_{\text{smooth}}(u, v, l_u, l_v) \; , \tag{4.9}$$

where $l_u$ and $l_v$ are the labels assigned to nodes $u$ and $v$, respectively, and $\mathcal{E}_{\text{data}}$ and $\mathcal{E}_{\text{smooth}}$ are the data and smoothness energy terms.

The data term is given by

$$\mathcal{E}_{\text{data}}(u, l_u) = -\omega_{\text{data}} \log(p(c_{l_u}|u)), \tag{4.10}$$

where $p(c_{l_u}|u)$ is the probability that node $u$ is part of cluster $c_{l_u}$, given by the statistical model of cluster $c_{l_u}$, and $\omega_{\text{data}}$ is a constant that regulates the influence of the data term in the total energy. The cost of assigning a specific label to the node increases according to how unlikely it is that the face belongs to the corresponding class.

Similarly to [Shapira et al. 2009], the smoothness term is defined as

$$\mathcal{E}_{\text{smooth}}(u, v, l_u, l_v) = \begin{cases} 0, & \text{if } l_u = l_v \\ -\log(\theta_{uv}/\pi)\,\ell_{uv}, & \text{otherwise,} \end{cases} \tag{4.11}$$

where $\ell_{uv}$ is the length of the edge between the faces corresponding to $u$ and $v$, and $\theta_{uv}$ is the dihedral angle between the two faces.

To obtain the labeling that minimizes the energy $\mathcal{E}$, we use graph cuts optimization [Boykov et al. 2001]. More specifically, the multi-label $\alpha$-expansion algorithm is utilized, since the smoothness term defines a metric in the space of labels.

The result of the labeling optimization is a co-segmentation of the set, since the labeling energy is based on the statistical models obtained from the co-analysis. Thus, in addition to an individual segmentation for each shape, we also obtain a correspondence among the segments, so that two segments on two different shapes correspond to each other if they possess the same label. Notice also that this step allows us to correct errors that appear in the initial co-segmentation, such as displaced boundaries and mislabeled segments. An example of the refinement is shown in Figure 4.6.

## 4.6 Experimental Results

In this section, we evaluate our unsupervised co-segmentation method and present qualitative and quantitative results. We also compare our method with the state-of-the-art in co-segmentation and with a supervised approach. Since our goal in this work is to present a fully unsupervised approach, all the results were obtained with a fixed set of parameters.

**Datasets and methodology.** We use seven classes of shapes in our experiments: candelabra, chairs, four-legged animals, goblets, guitars, lamps, and vases. The sets of man-made shapes are composed of objects that possess significant variability, i.e., a common type of part can appear with different topologies and geometries across the set, and it can be absent or appear multiple times on a shape. We modeled two new classes of man-made shapes for this work, while the remaining classes appeared in [van Kaick et al. 2011b] (Chapter 3). We manually segmented and labeled each shape, according to a specific labeling scheme for each class. This provides a ground-truth label for each face. We also selected one set of organic shapes (four-legged animals) from the Princeton Segmentation Benchmark [Chen et al. 2009], and use the ground-truth labeling created by Kalogerakis et al. [2010]. Note

that the ground truth is only used for a statistical evaluation and is not utilized by our algorithm. The co-segmentation is performed separately for each class.

**Co-segmentation and labeling.** Visual results of our co-segmentation are shown in Figure 4.7. Notice how, despite the great variability in the shape parts, our co-analysis is able to extract the common parts in the set and yield a coherent labeling. We point out illustrative examples. For candelabra, the method is able to identify the flames and wax candles of different sizes across the set, and separate them from the bases and holders, which appear in rich varieties. Also, the multiplicity of the flames and candles does not pose a problem to the method. Moreover, the lamps and goblets are successfully segmented and labeled into their three constituent parts, even though the geometry or thickness of the corresponding parts varies. Notice also the different topologies of the lamp supports that are detected. Chairs are also co-segmented into their main constituent parts, and we obtain a correspondence between legs of different topologies, including star-shaped legs and legs with railings. In the vases, handles of different sizes and shapes are identified, as well as bodies with very different geometries, including spherical, cylindrical, or flat bodies. We also see how the various types of guitar bodies are separated from the necks. Finally, we observe that the co-analysis is also successful when applied to a set of organic shapes with significant variability. Heads with short or long necks are correctly labeled, as well as the various animal bodies and legs.

We also notice a few shortcomings in the results. The small handles that appear on the candelabra are not properly separated from the holders, and similarly the small guitar headstocks are not separated from the fretboards. The cylindrical sections in the chair backrests are assigned to the same clusters as the seats. And, the animal tails are fused with the segments that represent the bodies, while some of the ears and horns are fused with the heads and necks. These problems appear due to imperfections in the clustering, which assigns an incorrect label to these parts. Another shortcoming includes the chair made up of thin wires on the front row, which is mislabeled. We attribute the problem to the lack of information available to the co-segmentation: this shape has unique parts that do not have similar counterparts in the set. This is an intrinsic limitation of our approach.

To assess the quality of the results in a quantitative manner, we show a statistical evaluation in Table 4.1. The first column indicates the number of shapes in the class, the second column corresponds to the labeling accuracy of the initial co-segmentation, and the

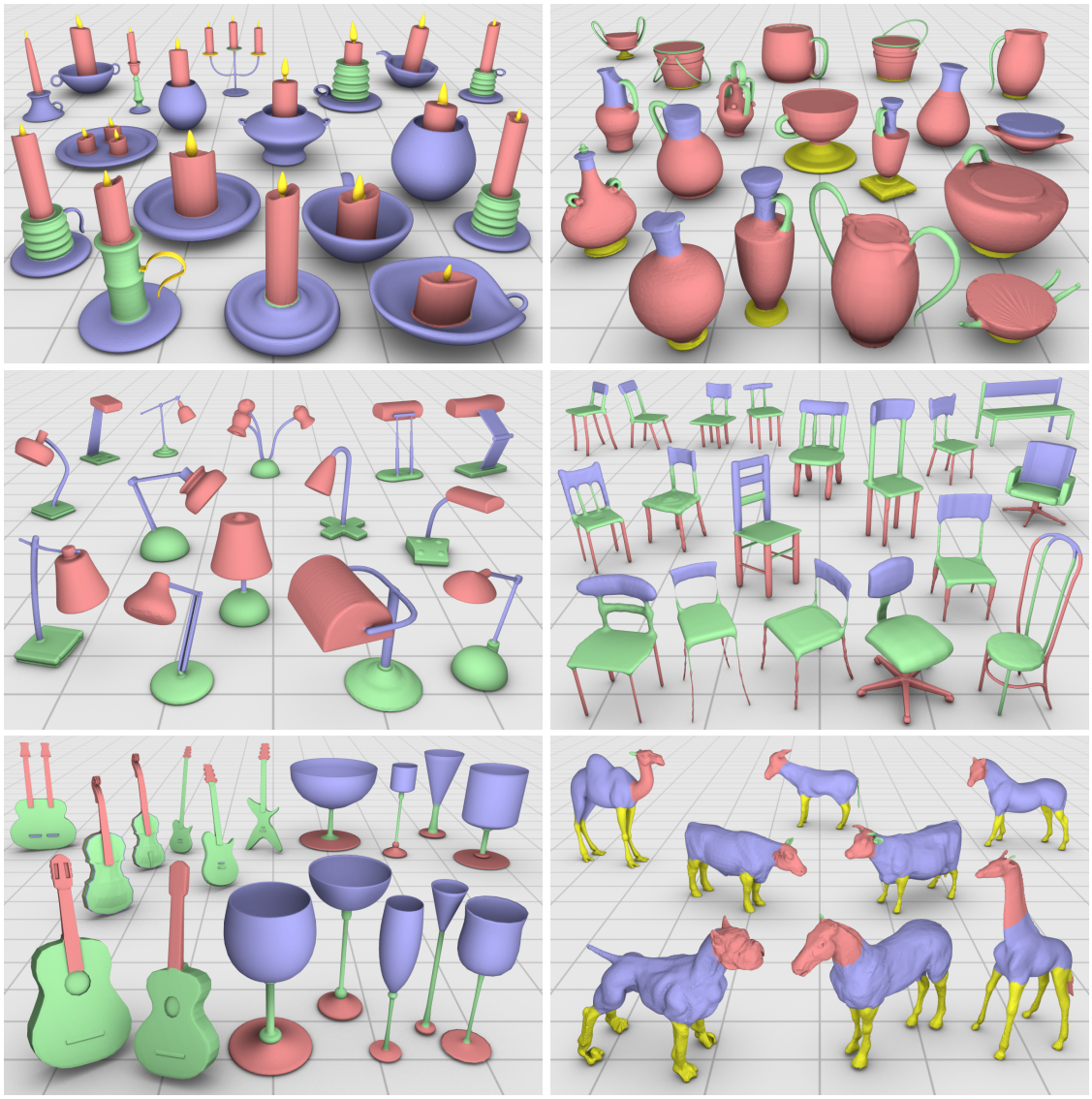


Figure 4.7: Results of our co-segmentation on a variety of shapes. Corresponding segments in each class are shown with the same color. Notice how the segmentation and labeling is coherent for many of the parts in each set. The results for all the sets were obtained with the same parameters and shape descriptors. Labeling accuracy statistics for each class are shown in Table 4.1.

Table 4.1: Average co-analysis labeling accuracy.

| Class | Num. shapes | Initial lab. | Refined lab. |
|---|---|---|---|
| Candelabra | 28 | 73.0 | 84.4 |
| Chairs | 20 | 78.6 | 84.8 |
| Four-legged | 20 | 75.9 | 77.3 |
| Goblets | 12 | 98.0 | 98.2 |
| Guitars | 44 | 86.4 | 87.2 |
| Lamps | 20 | 93.8 | 94.3 |
| Vases | 28 | 84.5 | 87.4 |

third column shows the labeling accuracy after the refinement. Each entry is the average accuracy for all the shapes in the class. The accuracy for a single shape is given by

$$\text{Accuracy}(\mathbf{l}, \mathbf{t}) = \frac{\sum_i a_i \, \delta(l_i = t_i)}{\sum_i a_i}, \tag{4.12}$$

where $a_i$ is the area of face $i$, $\mathbf{l}$ is the labeling returned by the co-segmentation, $\mathbf{t}$ is the ground-truth labeling, and $\delta(x = y)$ is 1 only if $x = y$. This measure captures the amount of area of the shape that is labeled correctly by the co-segmentation [Kalogerakis et al. 2010]. Since our co-segmentation does not return labels associated to specific semantic classes, before computing the accuracy we find the best one-to-one matching between our labels and the ground-truth labels. The matching is used coherently for the whole set.

The average labeling accuracy for all the classes is about 84% for the initial co-segmentation, and 88% for the refined labeling. We notice that, with the exception of one class, the accuracies are at least 84% or higher. We attribute the 10% accuracy gap between certain classes, e.g., candelabra vs. lamps, to the greater part variability that appears on the candelabra, chairs, and vases. Notice also a 4% improvement from the initial to the refined co-segmentation. This difference arises as a result of the boundary refinement, where parts that were oversegmented or wrongly labeled in the initial co-segmentation are properly labeled and have their segmentation refined with the statistical models (Figure 4.6).

**Effect of the set.** In Figure 4.8, we demonstrate the power of the set by evaluating how the co-segmentation accuracy improves when the set is enriched. For each class, we start with a pair of shapes and incrementally increase the set size by adding one shape at a time

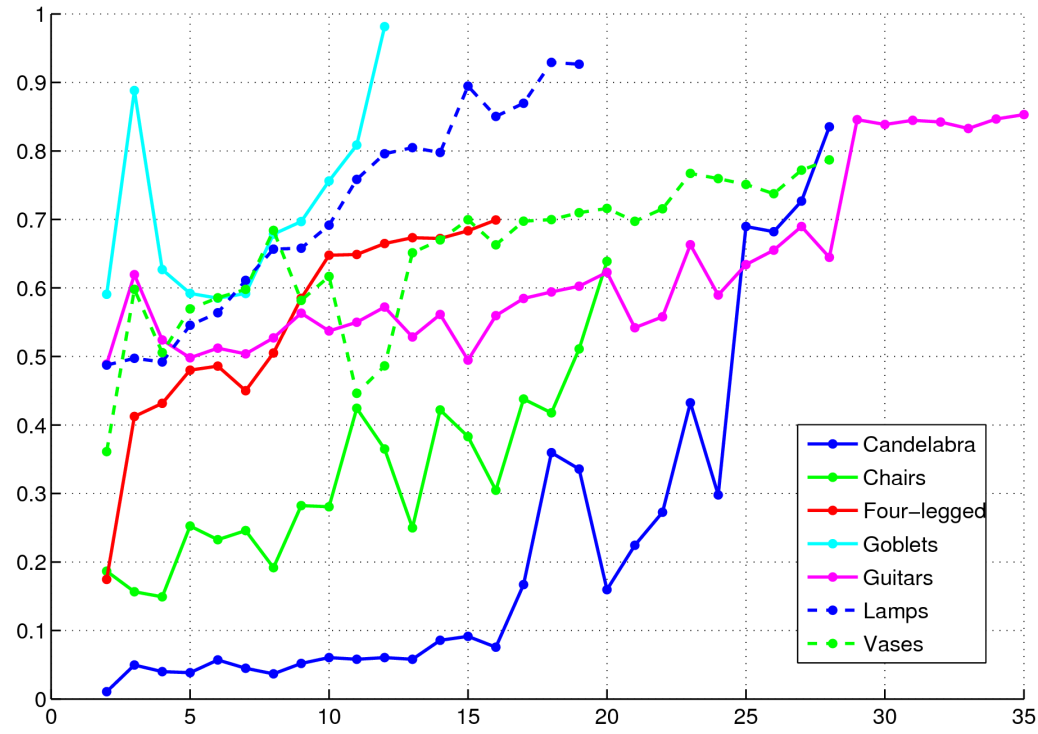



Figure 4.8: The power of the set demonstrated for all the classes. As the number of shapes in the set increases ($x$-axis), the accuracy of the co-segmentation ($y$-axis) shows a clear trend of improvement.

to the set. To isolate the effect of the graph cuts refinement, we perform only the initial labeling for each subset and report the labeling accuracy according to (4.12). The $x$-axis denotes the subset size, while the $y$-axis is the accuracy for processing the subset.

For all the classes, the general trend of the curves clearly demonstrates that the co-segmentation accuracy is improved as more shapes are added to the set. The non-monotonicity of the accuracy plots occurs when a difficult or unique exemplar is added to the set. The accuracy is then recovered or improved once additional shapes providing more information are added to the set. Note that the curves can change according to the specific order in which the shapes are added to the set. We particularly chose orderings that better reveal the monotonic behavior of the algorithm. Random orderings lead to more points that break the monotonicity, however, the overall upward trend of the curves remains the same.

**Comparison to the state-of-the-art.** Figure 4.9 shows the results obtained by applying the method of Golovinskiy and Funkhouser [2009] on the same sets of shapes[1]. Notice that this method obtains satisfactory results for sets that can be properly aligned with similarity transformations, e.g., chairs. However, when the shapes possess significant variability in topology (vases and candelabra) or pose (four-legged animals), the segmentations either miss important parts of the shapes (the animal heads), or are not meaningful at all (as in the vases and candelabra). This is due to the fact that this method derives the relation between the parts of different shapes from their proximity after alignment. It is not possible, even after a perfect alignment, to derive a correspondence between the parts of some of the vases just from proximity. Our method, on the other hand, is more robust in this regard, since it makes use of shape descriptors to handle more shape variability, and derives the inter-shape relations from third parties in the set.

**Comparison to a supervised approach.** We also compare our unsupervised co-segmentation to the supervised approach of Kalogerakis et al. [2010]. The authors of this approach kindly provided the results of applying their method on four sets that we use in this work. In each experiment, 70% of the shapes in a set were randomly selected as training data, while the remaining 30% were used as test shapes on which the labeling accuracy was evaluated. These experiments were repeated 5 times for each class, and the results were averaged. The accuracy for a single shape was also computed with the measure in (4.12). The accuracies obtained are: 80.9%, 91.6%, 97.3%, and 96.0%, for candelabra, chairs, lamps, and vases, respectively. Although a direct comparison of the accuracies of both approaches does not represent a meticulous evaluation, due to the splitting of the dataset into train and test sets with the supervised approach, we believe that in this manner we are nevertheless capturing the accuracy of the supervised approach in an average case.

By comparing these numbers with Table 4.1, we see a difference of at most 10% between the two approaches. We conclude that such close results demonstrate the high potential of the unsupervised co-segmentation. Although the results of the supervised approach are more accurate on average, we recall that it requires the preparation of a reasonably-sized training set, since the shapes have to be manually segmented and labeled. Additionally, a separate training set is necessary for each different class of shapes, and the approach only extracts a labeling that follows the pattern of the training examples. If a different training

[1]The authors kindly provided their implementation to us.

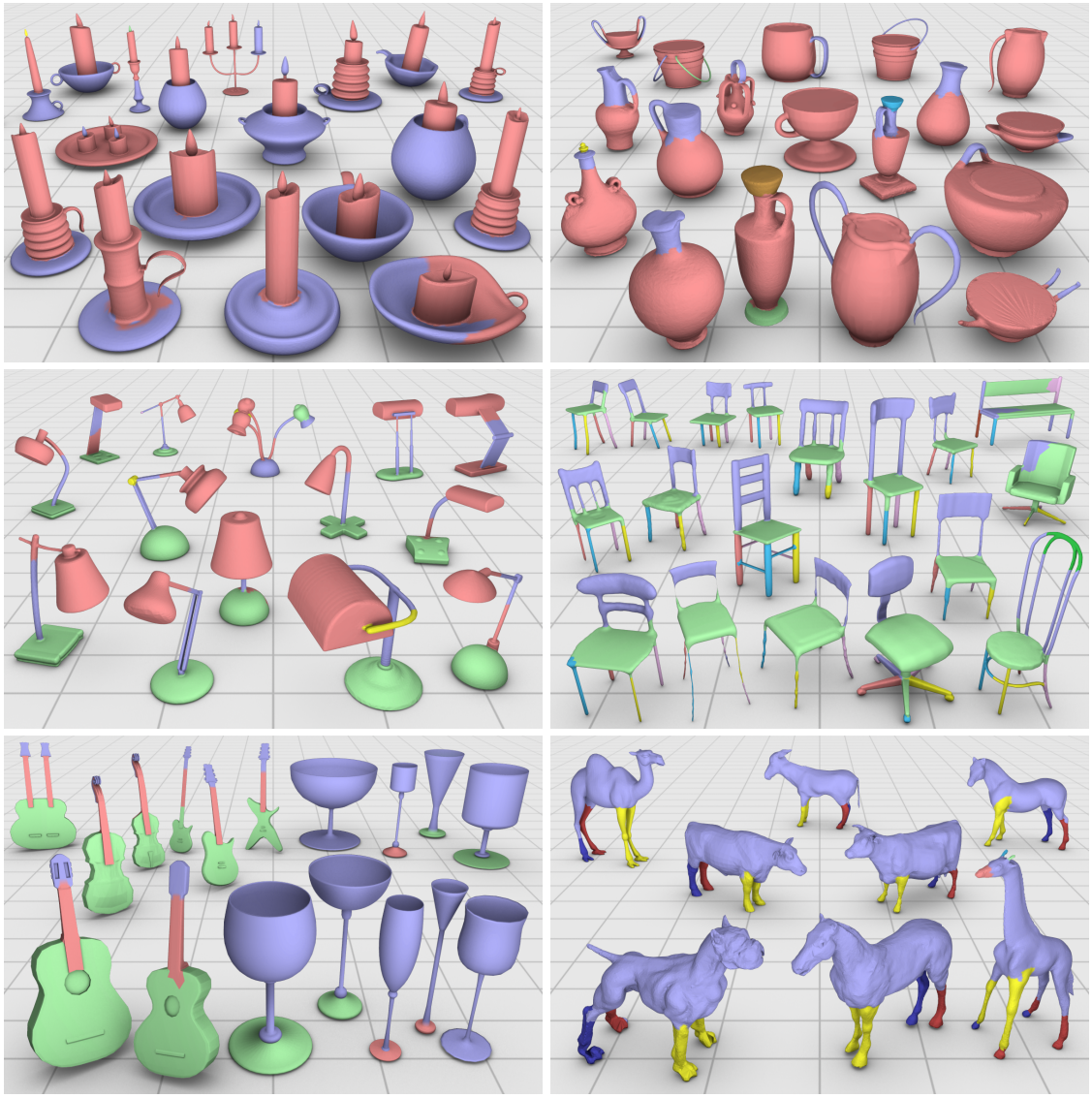


Figure 4.9: Comparison to the approach in [Golovinskiy and Funkhouser 2009]. In contrast to our co-analysis results in Figure 4.7, we observe that this method provides good results for sets of shapes that can be spatially aligned (chairs), but less meaningful co-segmentations for shapes with great part variability (candelabra).

set is given, the results can change considerably. The unsupervised approach, on the other hand, derives its knowledge automatically from the set. Finally, we conjecture that the accuracies for the unsupervised approach can possibly be increased with the enhancement of certain components of the method, such as the shape descriptors and statistical models.

**Performance.** Our implementation is fairly efficient, executing in 10 minutes for a set of 30 shapes in an AMD Opteron 2.4GHz with 8GB of memory. In contrast, supervised approaches such as those in [Kalogerakis et al. 2010] and [van Kaick et al. 2011b] (Chapter 3) can take on the order of hours for similarly-sized training sets and hardware configuration. The faster performance of the unsupervised co-segmentation is mainly due to two aspects: 1. The supervised approaches involve classifier training with complex algorithms such as *boosting*. 2. We perform most of the analysis at the segment-level, rather than face-level, a natural approach to greatly reduce the complexity of the method.

## 4.7 Discussion

We presented a method for co-segmentation of a set of shapes via descriptor-space spectral clustering. The question inherent to co-analysis is whether we can extract more information by analyzing a set simultaneously, instead of analyzing only individual shapes or pairs of shapes. Our investigation leads to an affirmative answer as we have shown that the set makes it possible for semantically related parts that differ in geometry and even topology to be linked via third-party connections. By performing the analysis in a descriptor space and exploiting the power of spectral clustering, we are able to properly co-segment sets of shapes exhibiting significant variability.

Ultimately, we would like to subject our co-segmentation approach to large-scale tests. Existing benchmarks for shape segmentation do not quite serve the purpose however. The Princeton Segmentation Benchmark (PSB) [Chen et al. 2009] consists of multiple segmentations per shape that were created individually via automatic methods or manual efforts without taking the sets into account, i.e., they are not necessarily consistent across each shape class. While Kalogerakis et al. [2010] provide a ground-truth labeling of the PSB data which is consistent across the sets, it is not obvious that this is the proper ground-truth to compare to. The ground-truth is prepared to guide the supervised algorithm to what the user desires to extract from the shapes. As Kalogerakis et al. clearly showed in

their paper, if an alternative ground-truth is prepared, the algorithm extracts a different segmentation. Hence a comparison to such a user-designed ground-truth is problematic. The ideal ground-truth would reflect the part composition of the set as given by the geometry of the shapes and the variability of the parts. The preparation of rigorous large-scale tests for co-segmentation of sets requires significant effort and we leave that for future work.

Nevertheless, we evaluated our approach on a moderately-sized dataset with seven classes of shapes, and compared the co-segmentation to one possible ground-truth prepared by a human. We showed that the result is close to the user-designed segmentations, implying that it does possess semantic meaning.

**Limitations.** The main limitation of our approach stems from the obvious fact that the quality of the results is entirely dictated by the input set. It may be possible that there is no link between two semantically related segments since the third parties that may establish such a link are missing, e.g., the wired chair and peculiar vases in Figure 4.7 are unique entities; there are no third parties that relate the parts of these shapes across the set. Also, while related parts can be properly linked via the set, imperfect clustering results may cause unrelated segments to be assigned to the same label, e.g., the small candelabra handles and animal parts in Figure 4.7.

Moreover, the success of our approach is inherently tied to the quality or usefulness of the shape descriptors. Firstly, one limitation of our current descriptors is that their computation requires the input models to be manifold meshes. Secondly, and more importantly, the dependence on shape descriptors leads to the difficult question of what information should be extracted to provide sufficient knowledge about the shapes. Nevertheless, our approach will naturally benefit from incorporating more sophisticated descriptors, e.g., of a structural nature [Biasotti et al. 2008; Shapira et al. 2009]. Another point for practical improvement is the incorporation of more advanced statistical models to represent the clusters of parts.

**Future directions.** Our current results do not yet surpass those of supervised approaches that are supported by well-built training sets [Kalogerakis et al. 2010; van Kaick et al. 2011b]. However, our set-driven co-segmentation has the potential of outperforming a supervised approach when the knowledge of the latter is insufficient. Perhaps more effective would be a semi-supervised approach, where the user needs to provide only a reduced amount of knowledge and the algorithm can maximally exploit the knowledge in the input set, or a setting where the user only corrects the parts erroneously labeled so that the system actively

learns and adapts the co-segmentation of the set.

The question of how much we can learn from a set is still open for further study, and could lead to novel approaches that take advantage of group information in innovative ways. Another direction for further research is to develop domain-specific shape descriptors, so that the co-analysis can be specialized to specific classes, e.g., humanoid characters, creatures, tools, or vehicles.

# Chapter 5

# Bilateral maps for partial matching

Feature analysis forms the basis of many shape analysis techniques, where feature points are detected on the shapes and characterized by local shape descriptors. The descriptors are *local* since they are used to represent the individual feature points, not to compare two shapes as a whole. Important tasks such as segmentation [Shamir 2004], shape retrieval [Tangelder and Veltkamp 2008], correspondence [van Kaick et al. 2011a], and symmetry detection [Mitra et al. 2006; Xu et al. 2009b], can all be solved with the aid of local shape descriptors. A variety of such descriptors have been proposed in the literature, where a point can be represented either by a *scalar* property (e.g., curvature [Manay et al. 2006] or local volume [Shapira et al. 2009]), or more effectively by representing the *context* around the point, as in the popular shape context descriptor [Belongie et al. 2002]. The latter is commonly extended to 2D manifolds by laying out a concentric grid on the surface around a point and then aggregating the geometric properties of points or faces that fall within each grid cell or bin, e.g., curvature [Gatzke et al. 2005] or area [Kalogerakis et al. 2010] can be summed. We call this descriptor a *geodesic map*. The main characteristic of all of these local shape descriptors is that they capture a region of interest centered at a *single* feature point.

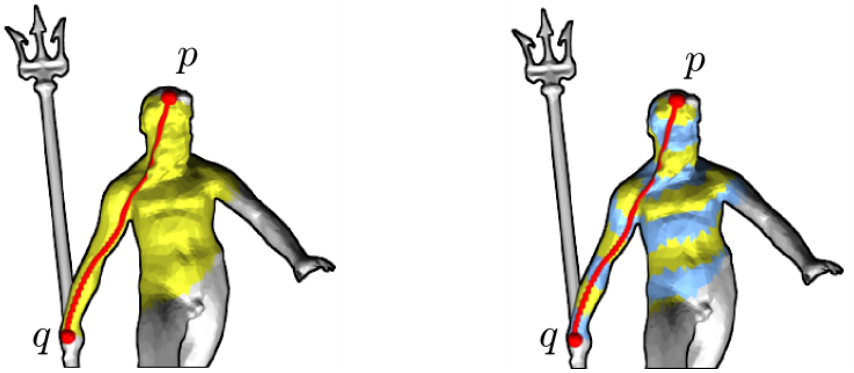
When dealing with incomplete shapes, shapes composed of a mixture of parts from multiple classes, or shapes that possess significant topological variability, it becomes necessary to use descriptors that enable *partial matching*. As shown with the example in Figure 5.1, if we wish to match a human model to the Neptune, the extraneous parts of the Neptune (spear and base) should not be included in the context. Otherwise, the effectiveness of the descriptors is reduced, affecting the quality of retrieval and correspondence results. One

(a) (b) (c)

Figure 5.1: Bilateral maps for partial matching. Descriptor bins are shown with alternating colors and the gray portions of the shapes are not part of the region of interest. Notice how the region captured between the head and right hand on both shapes is similar in (a) and (b), and does not include extraneous parts of the Neptune model, i.e., the spear and base. Compare to the traditional geodesic map for the right hand of the Neptune in (c), where extraneous parts are included even when using a reduced radius of coverage.

solution is to first segment the shape into meaningful parts and then eliminate the extraneous parts from the descriptor's context. However, obtaining a meaningful segmentation of a shape is a difficult problem [Shamir 2004], as well as determining what regions are extraneous. Thus, we would rather make the descriptors independent of such a requirement. A more straightforward solution for this problem is to assign a *scale* or *radius* parameter to the descriptor, to reduce its region of coverage (context), as shown in Figure 5.1 (c). However, automatically selecting the proper scale is not a trivial problem, and it is clear that this solution still has deficiencies as, no matter what radius is selected, the context of certain feature points will always include undesired portions of the models (as happens with Neptune's right hand).

In this chapter, we propose a new type of local shape descriptor that we call the *bilateral map* (Figure 5.1), which is designed to circumvent these problems. Instead of defining a region of interest around a single point and constraining it to a fixed radius, we compute a descriptor whose context is constrained by a pair of points. More specifically, we compute the shortest path on the surface between the pair of points, and define a context region in the vicinity of the path (Figure 5.2 (a)). Next, we define equally spaced bins in this region,




(a) (b)

Figure 5.2: Bilateral map construction. (a) Starting from the shortest geodesic path between the two reference points (in red), we define the region of interest around the path (in yellow). (b) The region is then divided into equally spaced bins (shown in alternating colors) along the shortest path.

with bin boundaries along the path (Figure 5.2 (b)). Finally, we aggregate a geometric property of all the faces or points that fall into each bin, as in the geodesic maps. For example, the area of the faces or curvature of the points.

Our main observation is that it is more advantageous to define regions of interest anchored by two points instead of one point. This is demonstrated by several advantages that our bilateral map has over descriptors centered at a single point. First, the context region is *adaptive*. That is, since it is constrained to lie between the two reference points, it only includes portions of the shape that capture the structural relationship between these two points. Portions of the shape that are not relevant to the reference points and that may be potentially missing in other shapes are ignored by the descriptor (contrast Figure 5.1 (b) to (c)). Secondly, given that the region of interest is adaptive, the selection of the scale of this region is facilitated, since then the region extent parameter can be set to be proportional to the length of the path between the two points. Finally, the bilateral approach is also less sensitive to moderate topological changes, in comparison to other descriptors that also depend on geodesics. Notice also that the descriptors are constructed without the need for part detection or a meaningful segmentation of the models.

Despite these advantages, one may expect an increase of cost in using the bilateral maps,

since given a set of feature points, we now need to examine a quadratic number of descriptors instead of a linear number of descriptors as for the single-point approach. Nevertheless, it turns out that the most successful correspondence approaches already require an analysis of the compatibility between *pairs* of matches, such as recent search-based methods [Gelfand et al. 2005; Zhang et al. 2008; Au et al. 2010], sampling-based algorithms [Tevs et al. 2011], or spectral matching methods [Leordeanu and Hebert 2005]. Thus, the additional cost of the bilateral maps is mainly their computation, since the quadratic complexity of the correspondence algorithm remains the same. Moreover, our descriptors naturally fit into such approaches, since they can be added directly to the estimate of pairwise compatibility, along with other constraints such as isometry preservation (Chapter 2).

We demonstrate the effectiveness of the new descriptors with several experiments on shape correspondence and retrieval, and evaluate the quality of the results in a qualitative and quantitative manner. We show that, by making use of the bilateral maps in a simple framework, we obtain better results than when using a similar descriptor centered at a single point. Moreover, we also obtain improved results for partial and complete retrieval when comparing to a state-of-the-art approach; the persistent heat signature of Dey et al. [Dey et al. 2010]. Although the design of the bilateral maps was motivated by partial matching problems, they can also be used for full matching, since in this case their performance is similar to that of single-point descriptors.

## 5.1 Related work

**Shape matching.** A prominent problem in shape analysis is the development of means to compare the geometry of shapes, be them the full models or parts. This problem is at the heart of shape retrieval [Iyer et al. 2005; Tangelder and Veltkamp 2008] and correspondence [van Kaick et al. 2011a] (Chapter 2). When matching shapes with significant variability and missing data, it is important that the above applications are able to perform *partial matching*. Partial matching is difficult since, before computing the similarity of the shapes, we first need to find the common portions of the shapes. This requires the careful design of descriptors that are less sensitive to variations in the part composition of the models and also a mechanism to search for a partial match. The latter can be achieved, for example, by detecting a sharp increase in the objective function when outlier regions are excluded from the match [Gelfand et al. 2005; Zhang et al. 2008], or by making use of voting

methods, which select only the most plausible correspondences [Lipman and Funkhouser 2009; Au et al. 2010].

There are also approaches that compute a partial correspondence without relying on shape descriptors, such as the methods of Bronstein and Bronstein for rigid [Bronstein and Bronstein 2008b] and non-rigid [Bronstein and Bronstein 2008a] matching. Their approach optimizes a type of Mumford-Shah functional, commonly used for image segmentation. This results in a numerical optimization solved with a quasi-Newton minimization algorithm.

**Local shape descriptors.** A great variety of local shape descriptors have been proposed in the literature, since the most common approach to address shape correspondence and retrieval is to utilize descriptors as a more suitable representation for shape comparison. Simple descriptors capture scalar properties of points, such as the integral invariants [Manay et al. 2006], or the shape diameter function [Shapira et al. 2009].

Descriptors that go beyond scalar properties can be obtained by explicitly representing a region of interest or *context* around the points. The shape context descriptor of Belongie et al. [2002] is based on laying out a grid on this region and then counting the number of feature points that fall into the bins implied by the grid. A direct extension of this descriptor to 3D has been proposed [Körtgen et al. 2003], as well as a version that defines the grid relative to the surface orientation; the spin images [Johnson and Hebert 1999]. Kazhdan et al. [2003] propose to encode such contextual descriptors in a rotationally-invariant manner by using spherical harmonics. A natural extension of shape context to manifolds is to lay out the grid on the surface of the models, and then aggregate the curvature [Gatzke et al. 2005] or area [Kalogerakis et al. 2010] of the surface portion that falls into the bins [Heider et al. 2011]. More elaborate representations propose to encode multiple local properties of the region of interest with a statistical model [Castellani et al. 2008].

However, these descriptors provide little flexibility to capture partial regions of the shapes, since what can be regulated is mainly the size of the context region. Thus, multi-scale descriptors have been proposed that are more suitable in this type of scenario, such as the multi-scale features of Li and Guskov [2005], or integral invariants captured at different scales [Manay et al. 2006]. However, the context region captured by these descriptors is still isotropic and non-adaptive (the shape of the region is fixed, independently of where it lies on the shape).

Gal and Cohen-Or propose a descriptor designed specifically for partial matching, where

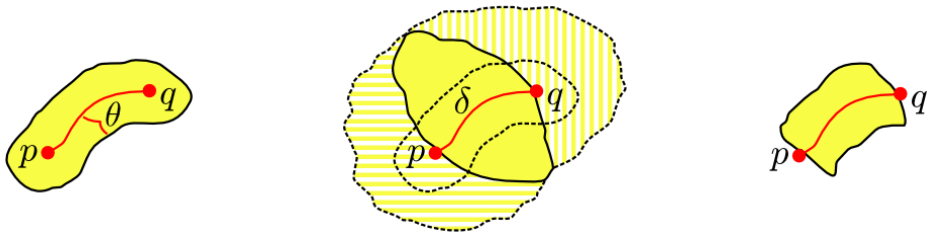
salient regions are extracted from the models and stored as a full geometry that is later matched with geometric hashing [Gal and Cohen-Or 2006]. Another possibility for partial matching is to include part information to limit the context of the descriptors, e.g., by making use of the isophotic metric [Pottmann et al. 2004] or a part-aware metric [Liu et al. 2009]. Finally, a representation that has achieved considerable success in shape matching is the heat kernel signature [Dey et al. 2010; Ovsjanikov et al. 2010], which can tolerate significant variability and missing data. However, this signature also suffers from the scale selection problem, as it is based on isotropic heat diffusion starting from one point, and hence can be seen as a multi-scale approach.

In contrast to these works, the bilateral maps go beyond the multi-scale or fixed-size context representations, but do not require part detection, since the partial regions of the shapes are defined by pairs of points.

## 5.2 Descriptor construction

In this section, we discuss in detail the construction of the bilateral maps, exemplified in Figure 5.2. First, given a pair of reference points $(p, q)$, we define a region of interest around the shortest geodesic path between the two points (Figure 5.2 (a)). The way in which we define this region is explained below. Next, we compute the shortest geodesic distance from each face within the region of interest to the point $p$, defining a scalar distance field. By dividing the range of possible scalar values into $b$ equal intervals, we are able to divide the region into $b$ segments or bins. The boundaries of the bins cross perpendicularly to the path (Figure 5.2 (b)). Finally, for each bin, we aggregate a local property of all the faces or vertices that fall into the bin and use the result as the scalar value that represents the bin. In this work, we sum the area of all the faces in the bin, although other properties, such as curvature, can be utilized. The resulting descriptor is a $b$-dimensional vector. To normalize the descriptor, we divide each vector entry by the sum of all entries.

Notice that our construction provides a directional descriptor. That is, the descriptor for $(p, q)$ is different from the one computed for $(q, p)$. This is in fact a desired and important property, since an undirected construction could result in the incorrect switching of corresponding points.

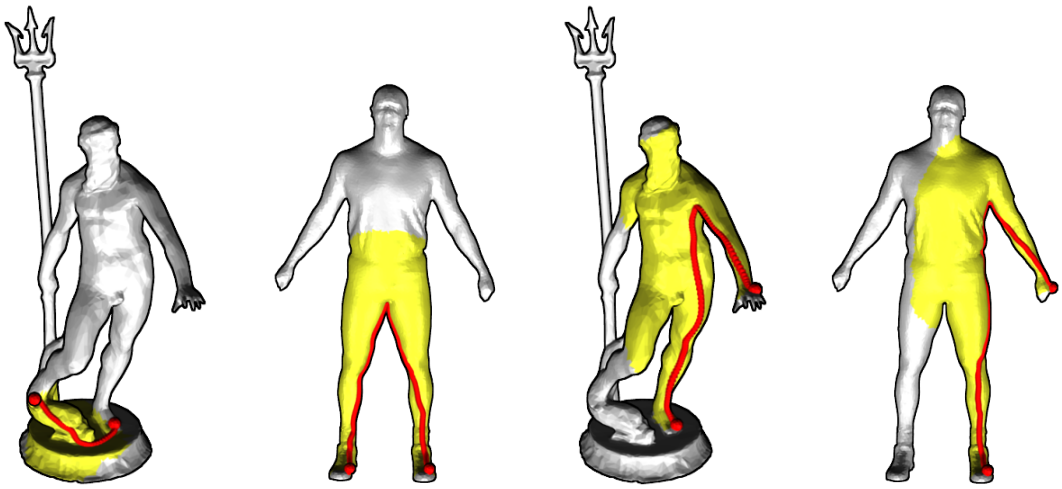



(a) (b) (c)

Figure 5.3: Definition of the region of interest for the bilateral map. (a) An initial region (in yellow) is defined as an offset from the shortest path (in red). (b) A filtering region (in yellow at the center) is defined by intersecting the expanding fronts from the two feature points. (c) The final region of interest is given by the intersection of (a) and (b).

**Region of interest.** Our goal is to capture a region around the geodesic path, and to avoid including extraneous portions of the models, especially those that lie beyond the reference points. Thus, we proceed as shown in Figure 5.3. First, we create a scalar field given by the smallest distance from each face to the path. All the faces with the distance below a threshold $\theta$ define our *initial* region (Figure 5.3 (a)). Before thresholding, we divide all the distances by the shortest geodesic distance between $p$ and $q$, so that the threshold becomes relative to the path length. The parameter $\theta$ can be seen as the relative *width* of the initial region. We use a fixed $\theta$ throughout our experiments and show that this constrained scheme is robust. Next, we define a *filtering* region to exclude portions of the shape that lie beyond the reference points. Given that $\delta$ is the shortest geodesic distance between $p$ and $q$, we compute the intersection of all the faces that are at most a distance of $\delta$ away from $p$ and all the faces that are at most $\delta$ away from $q$. In this manner, we obtain a region that excludes the faces "behind" $p$ and $q$ (Figure 5.3 (b)) – assuming that there are no small topological tunnels or handles. Finally, we intersect the initial and filtering regions to define our region of interest (Figure 5.3 (c)).

**Adaptiveness and scale selection.** Our bilateral maps have several advantages over approaches centered at a single point. Firstly, instead of selecting a free parameter to define the scale of the region of interest, which could assume any arbitrary value, we only have to select the relative *width* of the bilateral maps. Since the threshold that we require



(a) (b) (c) (d)

Figure 5.4: Lesser sensitivity of the bilateral maps to topological changes. When a topological shortcut exists in (a), the descriptors capturing the region between the two feet are significantly different in (a) and (b). However, notice that many of the remaining descriptors are still intact, e.g., between the foot and left hand in (c) and (d), between the foot and points on the head, etc.

for the construction is relative to the distance between the two points, such a constrained parameter can be easily set to work well for a variety of shapes, as we show in our experiments in Section 5.3. Secondly, the geometry of the region of interest is *adaptive* and captures mainly the parts of the shape that lie between the reference points (Figure 5.1 (a) and (b)). In contrast to other descriptors, the region does not simply expand isotropically towards all directions; it is *anisotropically* constrained by the geodesic path.

**Lesser sensitivity to topological changes.** Since our construction is based on the geodesic path between two reference points, the bilateral maps are affected when the path is drastically changed. This will happen when a topological shortcut is created on the shape (Figure 5.4). Note that this is an inherent limitation of all descriptors relying on geodesic distances. Nevertheless, we observe that most of the bilateral maps will still remain intact when such a shortcut is created, since it affects predominantly the descriptor corresponding to the two reference points that are short-circuited with the change (Figure 5.4). In the bilateral approach, each feature point is associated to multiple descriptors. Thus, if one

of these descriptors is affected, there still remain several other descriptors to guide the algorithm. This property makes the full set of bilateral maps of similar shapes less sensitive to moderate topological changes. Note that this is not the case for single-point descriptors based on a geodesic neighborhood, since each feature point has a single associated descriptor which is then modified and unable to provide a proper match for the point.

**Suitability to existing correspondence methods.** Another advantage of our bilateral maps is that they can be used with existing correspondence approaches without the need of a major modification to these methods. Recent correspondence methods work by minimizing an objective similar to

$$\begin{aligned} \pi^* = \quad & \underset{\pi}{\operatorname{argmin}} \sum_{p \in \mathcal{S}_1} \operatorname{dissim}(D_p, D_{\pi(p)}) + \\ & \alpha \sum_{p,q \in \mathcal{S}_1} \left| \operatorname{dist}_{\mathcal{S}_1}(p,q) - \operatorname{dist}_{\mathcal{S}_2}(\pi(p), \pi(q)) \right|, \end{aligned} \tag{5.1}$$

where $\pi^*$ is the optimal correspondence that we seek, $p, q$ are points on shape $\mathcal{S}_1$, $\pi(p), \pi(q)$ are their corresponding points on shape $\mathcal{S}_2$, dissim is the dissimilarity between the descriptors $D_p$ and $D_{\pi(p)}$, and $\operatorname{dist}_{\mathcal{S}_i}$ is the geodesic distance on shape $\mathcal{S}_i$. This objective function captures the notion that two pairs of points should be matched if the distance between the points is similar on each shape, and the descriptors of the matching points are also similar [Leordeanu and Hebert 2005; Zhang et al. 2008]. $\alpha$ regulates the balance between the geodesic inconsistency and the descriptor dissimilarity in the objective function.

Thus, to use our descriptors, we can simply replace the first term of the objective by

$$\sum_{p,q \in \mathcal{S}_1} \operatorname{dissim}(D_{(p,q)}, D_{(\pi(p),\pi(q))}) \tag{5.2}$$

where $D_{(p,q)}$ are now the bilateral maps defined over pairs of points. Notice that, if we assume that the method uses the same set of points to test for distance preservation, then the method already had to consider $O(n^2)$ pairs with the objective in (5.1), where $n$ is the number of feature points on $\mathcal{S}_1$. Thus, the general complexity of optimizing (5.1) is not increased with this modification; only the time required to compute the descriptors is increased. However, also note that, when computing geodesic distances for $n$ feature points, we have a time complexity of $O(n(|E| + |V|log|V|))$, where $|V|$ and $|E|$ are respectively the total number of vertices and edges in the mesh. Thus, when using descriptors based on geodesic neighborhoods, this computation might dominate over the matching complexity.

Similarly, for other applications such as shape retrieval, we will also have an increase in the time required to compute the descriptors. However, as we show in Section 5.3, typically a simple scheme is used to match the signatures of two shapes. Thus, in practice, we may be able to find a balance between the number of features extracted from the shapes and the quadratic number of descriptors generated.

## 5.3 Experiments and results

In this section, we demonstrate the effectiveness of the bilateral maps in the contexts of shape correspondence and retrieval. For all the experiments, the same descriptor parameters are used: number of bins $b = 20$, and descriptor width $\theta = 0.35$.

### 5.3.1 Shape correspondence

**Correspondence algorithm.** A variety of methods that make use of local shape descriptors for correspondence have been proposed in the literature (Chapter 2). We chose to evaluate the bilateral maps within a simple framework, so that the effectiveness of the descriptors can be easily isolated from the correspondence algorithm and compared in a more direct manner to a compatible single-point counterpart.

Given two meshes $\mathcal{M}_1$ and $\mathcal{M}_2$, we sample 50 feature points uniformly across each surface and compute the bilateral maps for all the pairs of points. Next, we apply a two-step scheme to compute a correspondence between the shapes. The first step is a simple voting procedure. For each bilateral map on $\mathcal{M}_1$, we find the most similar map on $\mathcal{M}_2$, and place a vote on the two pairwise matches implied by the two descriptors. We utilize the $\ell 1$-norm to measure the dissimilarity between two descriptors. At the end of the voting, we select for each feature point in $\mathcal{M}_1$, the matching point in $\mathcal{M}_2$ with most votes.

Due to the descriptiveness of the bilateral maps, this simple procedure already provides a good set of candidate correspondences. However, there is no enforcement of global consistency, as the descriptors for symmetric regions of the shapes can be similar. Thus, in the second step, we filter the pairwise matches returned by the voting according to a consistency

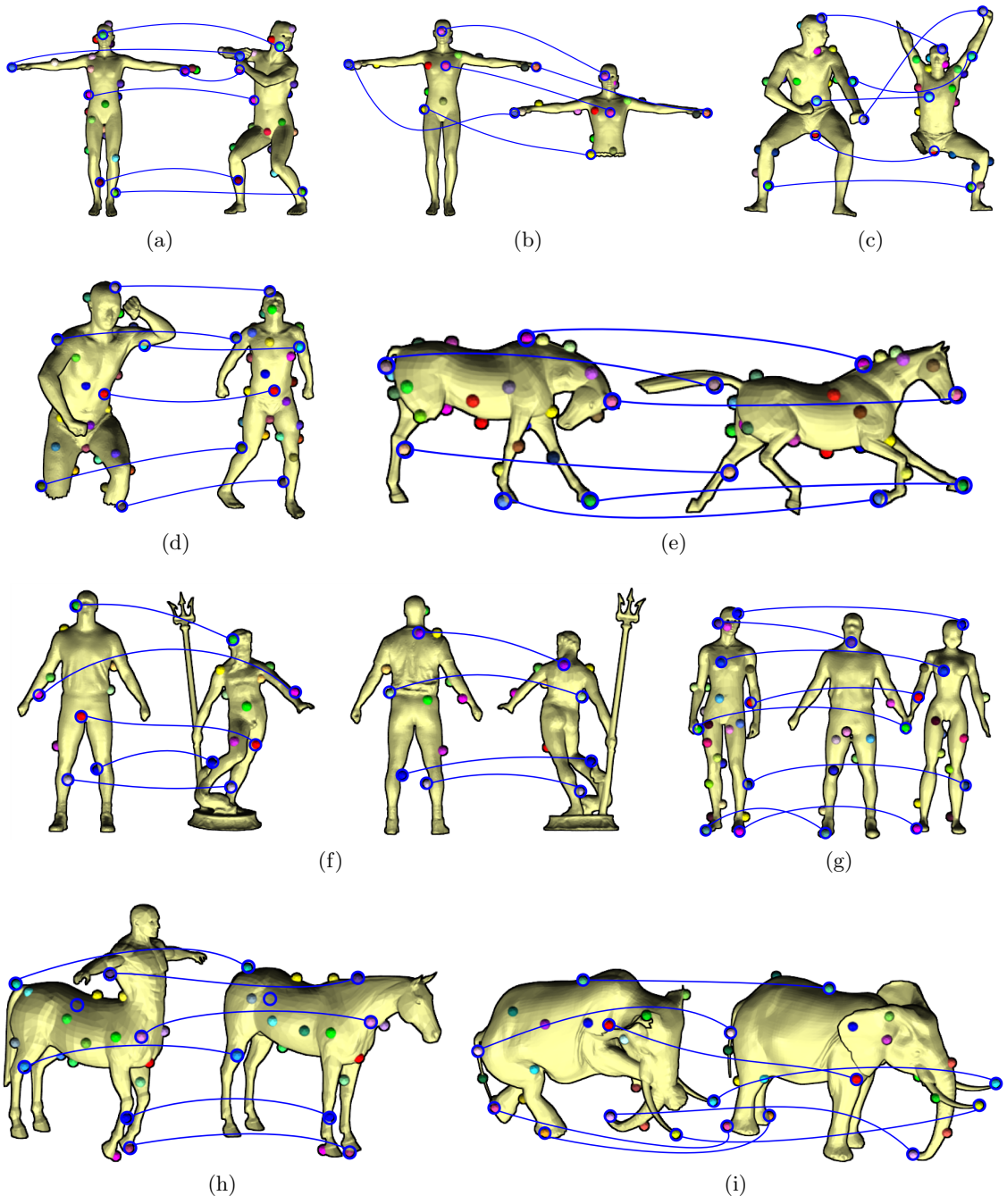



Figure 5.5: Correspondence results obtained with our bilateral maps on a set of example pairs. Corresponding points are shown with matching colors, and we connect some interesting matches with the blue curves. Notice that our descriptors enable a simple algorithm to find meaningful matches between complete shapes in different poses (a), shapes with missing parts (b)-(e), and shapes that include extraneous parts or possess different topology (f)-(i). The same region offset $\theta$ was used in all examples.

criterion similar to that in equations 5.1 and 5.2,

$$\begin{aligned}\text{consistency}(p, \pi(p)) = &\sum_{q \in \mathcal{M}_1} \text{dissim}(D_{(p,q)}, D_{(\pi(p),\pi(q))}) \\ &+ \alpha \sum_{q \in \mathcal{M}_1} \left| \text{dist}_{\mathcal{M}_1}(p, q) - \text{dist}_{\mathcal{M}_2}(\pi(p), \pi(q)) \right| .\end{aligned} \tag{5.3}$$

We set $\alpha = 0.7$ in the experiments, after normalizing both terms to the range $[0, 1]$, and select only the top 70% matches that pass the filtering criterion. We repeat this filtering procedure twice, since the consistency criterion for certain points changes after a first filtering pass.

**Results.** In Figure 5.5, we show visual examples of correspondences computed with our bilateral maps and the algorithm described above. Notice how the descriptors are able to guide the algorithm to find meaningful matches for shapes with missing parts or even differing topology. In (a), the method is applied to a pair of complete shapes, to show that the bilateral maps are also suitable for computing full correspondences between shapes in different poses, as they are intrinsic descriptors. The examples in (b)-(i) present cases requiring partial matching. We see that in a simple case where a shape is matched to a version of itself cut in half, shown in (b), the descriptors are able to provide the correct matches between all the selected feature points. In (c)-(e), we see that the bilateral maps are effective for matching incomplete shapes, with missing arms or legs.

Finally, in (f)-(i), we show the effectiveness of the bilateral maps for matching hybrid shapes or shapes with topological differences. In (f), we see a correspondence computed for our motivating example of the human vs. Neptune, where the front and back of the shapes are shown. The correspondence correctly ignores the feature points on the extraneous regions, although the right arm of the human is matched to Neptune's left arm. In (g), we see a human matched to a single manifold composed of two humans holding hands. We see that all the correspondences are meaningful up to symmetry switching, although our simple algorithm does not enforce that the human is matched to the parts of a single human in the composed shape. In (h), we see another meaningful correspondence computed for two shapes with extraneous regions (human body and horse's head). In the example in (i), the left elephant has a different topology from the one on the right (its head, trunk, and tusks are connected to a front leg), but correct matches are still obtained between the trunks, tusks, and the bodies of the elephants.

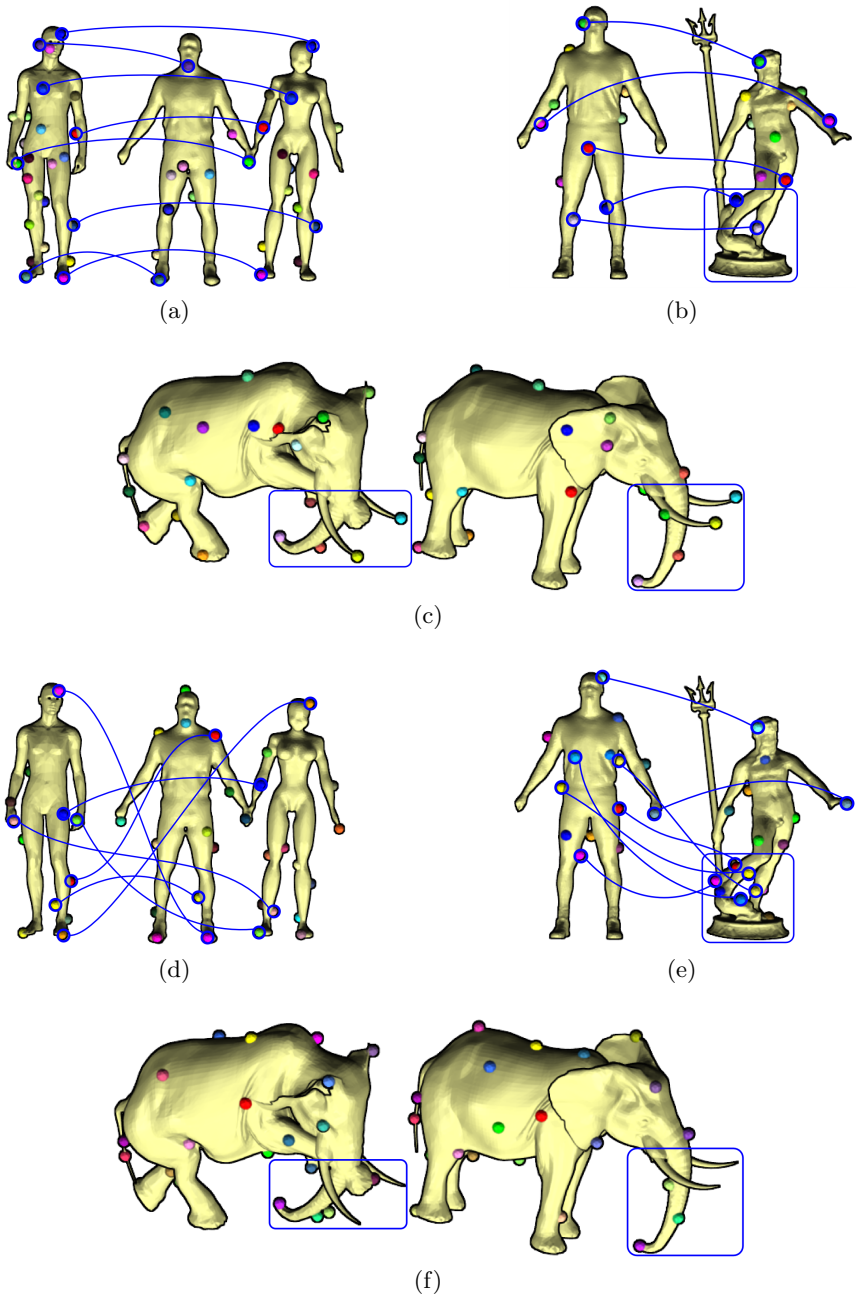



Figure 5.6: Correspondence results obtained with the bilateral maps in (a)-(c), compared to the results obtained with the single-point geodesic maps in (d)-(f). The single-point descriptors require the selection of the best scale for computing the correspondence, and the results are not as meaningful as those obtained with the bilateral maps, especially in the highlighted regions.

**Comparison to single-point descriptors.** Figure 5.6 presents a comparison of the bilateral maps to a traditional single-point geodesic map. There certainly exist more advanced local shape descriptors in the literature, for example, the persistent heat signature of Dey et al. [2010]. However, we chose to compare to the geodesic maps since they use the same type of intrinsic binning as the bilateral maps, and they are also based on collecting the area of the faces that fall within each bin. In this manner, we are able to evaluate more directly the effect of using pairs of points instead of single points. (But note that we compare to the heat signature in the next section in terms of shape retrieval).

For the single-point descriptors, we compute the geodesic maps at five different scales. The extent of the descriptors is set to 20%, 40%, 60%, 80%, and 100% of the longest geodesic path on each shape. Each descriptor is composed of 20 concentric bins. Next, we compute correspondences with each scale and select the best result for each example in Figure 5.6 (d)-(f). To establish the correspondence, we select for each feature point, the matching point with the greatest similarity. Then, we also filter the correspondence with the criterion in Equation 5.3, using the average dissimilarity of the two single-point descriptors that are part of each pair.

By contrasting the results of the bilateral maps in Figure 5.6 (a)-(c) to the results of the single-point descriptors in (d)-(f), we see that the bilateral maps are more effective in performing partial matching, without the need to change the scale parameter. When selecting the proper scale, the single-point descriptors are able to match important features such as the head and arms of the humans, and the body of the elephants. However, we see that incorrect matches are established between some of the extraneous parts of the shapes; for example, between the base of the Neptune and various parts of the human, or the front leg of the first elephant and the body of the second elephant. The existence of extraneous regions and topological changes have a stronger impact on the single-point descriptors than on the bilateral maps. The single-point approach completely fails in (d), since the geodesic maps of the two shapes are significantly different.

**Topological changes.** One advantage of the bilateral maps is that, if two shapes possess moderate topological differences, we can still obtain a meaningful correspondence between the two. As discussed in Section 5.2, although some of the descriptors are modified when one of the shapes suffers topological changes, enough of the multiple bilateral maps associated to each feature still remain intact if only a few topological shortcuts are created.

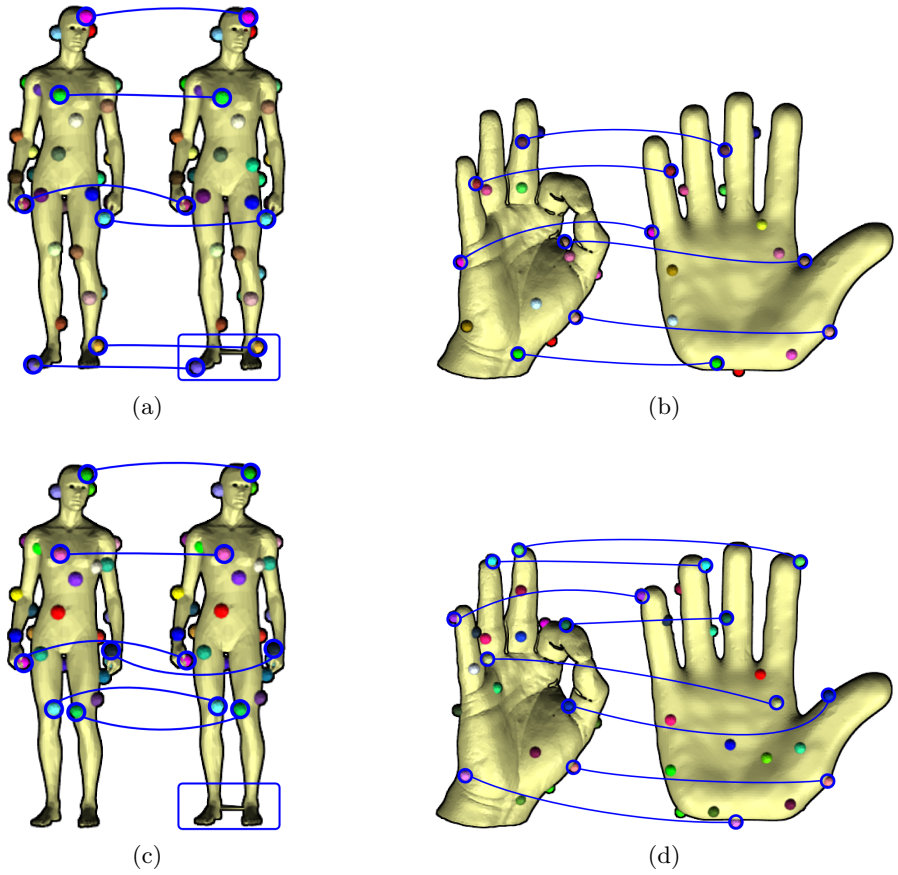



Figure 5.7: Examples of lesser sensitivity of the bilateral approach to topological changes. Note that more meaningful correspondences are obtained with the bilateral map in (a)-(b), compared to the geodesic map in (c)-(d), on shapes with moderate topological differences.

We already illustrated the lesser sensitivity of the bilateral maps to topological differences with the example of matching a human to Neptune in Figure 5.5 (f) and the elephants in Figure 5.5 (i). We show two additional examples in Figure 5.7, contrasting them to the single-point geodesic map. In (a), we see that creating a topological shortcut between the two legs of a human has no noticeable effect in the correspondence results when this human is matched to itself with the bilateral maps. The geodesic maps also provide a meaningful correspondence, although there are no matches at the end of the legs, since the descriptors at any scale are different in this region. In (b), the bilateral maps are able to match two hands with different topologies (the index finger and thumb are connected in the first hand),

although the specific parts where the differences occur are not matched, since many of the descriptors are dissimilar there. The geodesic maps also provide many good correspondences. However, since we have to select a smaller scale for this descriptor to be able to ignore the topological differences, this also creates some confusion in the correspondences, such as the index finger being matched to the middle finger.

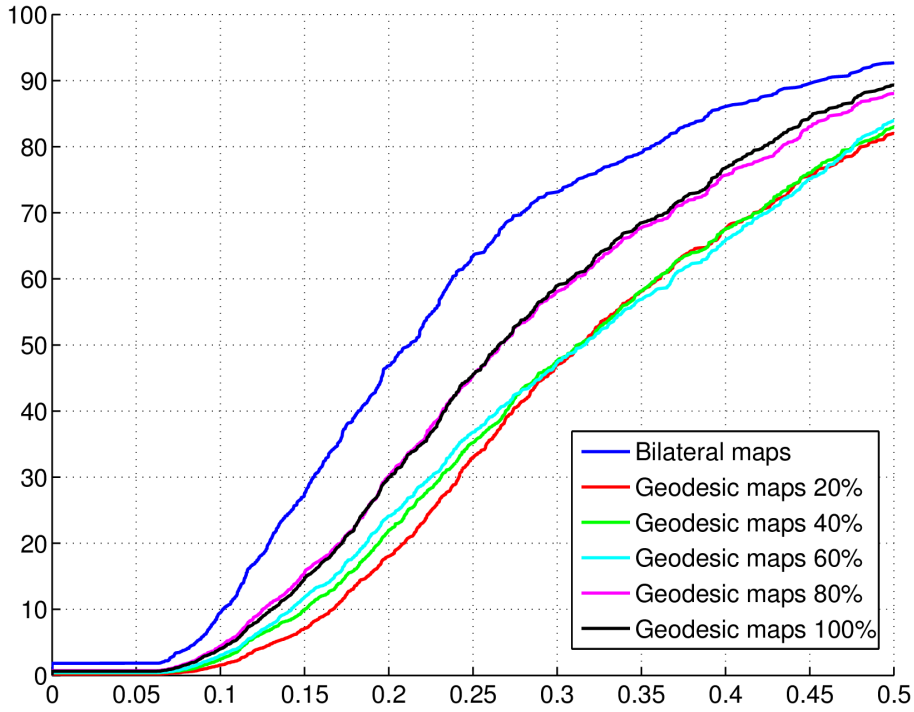



Figure 5.8: Quantitative evaluation of correspondences computed with the bilateral maps and geodesic maps at five different scales on a set of incomplete humans. The $x$-axis denotes the correspondence error given in terms of geodesic distances, while the $y$-axis denotes the percentage of correspondences with such an error. Note the overall better performance of the bilateral maps.

**Quantitative evaluation.** We also compare the bilateral maps to the single-point geodesic maps in a quantitative manner. We selected 13 incomplete humans from the dataset for partial shape retrieval of Dey et al. [2010]. These humans have missing parts or missing data in the form of multiple holes on the meshes. Next, we computed correspondences between each human and 8 other randomly selected humans. We used the same

algorithms described above for the bilateral maps and the geodesic maps and the same sets of feature points for each shape.

To evaluate a resulting correspondence, we first created a ground-truth for each shape. We selected 36 consistent landmark points on each shape, as in the benchmark of Kim et al. [Kim et al. 2011]. In the case of features appearing in missing parts, we placed the ground-truth at the location nearest to the landmark. For example, if a human is missing its left leg from the knee downwards, we placed the landmark corresponding to the left toe at the location of the left knee. Since the descriptors are computed on automatically-extracted features, we measure the distance of these feature points to the landmarks to measure the accuracy of a correspondence. This is done by representing each feature point by a vector of geodesic distances to the landmarks. By finding the best match for such vectors between two shapes, we establish the ground-truth for the feature points.

Once we are given a correspondence $\pi$ computed with one of the descriptors, we compute its error in relation to the ground-truth correspondence $\pi_{\text{gt}}$ by adding up the geodesic distance from each selected point to the ground-truth point,

$$\text{Error}(\pi, \pi_{\text{gt}}) = \sum_{p \in \mathcal{M}_1} d_{\mathcal{M}_2}(\pi(p), \pi_{\text{gt}}(p)). \tag{5.4}$$

The geodesic distances $d_{\mathcal{M}_2}$ are normalized by $\sqrt{\text{Area}(\mathcal{M}_2)}$ [Kim et al. 2011].

Figure 5.8 shows the results of this experiment. Each curve denotes the percentage of correspondences that have an error below a given geodesic distance. It can be clearly seen that the bilateral maps have more correspondences with lower errors than the geodesic maps computed at any of the 5 given scales. 70% of the correspondences given by the bilateral maps have an error of 0.27 or less, while the error is 0.37 for the same percentage of the best scale geodesic maps. Note also that, for this dataset, the geodesic maps with larger scales (80% and 100% of the maximum geodesic distance on each mesh) gave the best results for single-point descriptors, since these larger scales better capture the human structure.

**Limitations.** The bilateral maps also possess some limitations, which can be observed in Figure 5.5. As common with intrinsic descriptors [Kim et al. 2011], when the shapes have strong intrinsic symmetries, e.g., left- and right-halves of a human, the bilateral maps and geodesic consistency are unable to tell the symmetries apart. Moreover, as seen in Figure 5.5 (f) and (i), matches are not established for one arm of the Neptune, or the front legs of the

Table 5.1: Retrieval results on a dataset of complete and incomplete shapes. Each table entry shows top-3 and top-5 hit rates for our bilateral maps, PHS, and EVD.

| **Queries** | **Ours** | **PHS** | **EVD** |
|---|---|---|---|
| 32 incompl. | 91% / 94% | 88% / 91% | 62% / 62% |
| 18 compl. | 100% / 100% | 78% / 83% | 100% / 100% |
| 50 total | 94% / 96% | 84% / 88% | 76% / 76% |

elephants. This happens since there are several topological shortcuts in these regions, and many of the descriptors become distinct.

### 5.3.2 Shape retrieval.

**Dataset and methodology.** We show results of using our bilateral maps for shape retrieval and also compare to the method of Dey et al. [2010]. Their method utilizes the persistent heat kernel signature (PHS) to detect interest points on the shapes and also to define shape signatures for retrieval. Their work is the first to evaluate results specifically on incomplete shapes and also to build a dataset for this purpose. We perform the comparison by evaluating the bilateral maps on Dey et al's dataset. The dataset is composed of 300 shapes organized into 21 classes (humans, horses, chairs, etc.). Its query set is composed of 18 complete shapes and 32 incomplete shapes, while the target set contains 197 complete and 101 incomplete shapes.

To perform the retrieval, we also utilize a procedure similar to that used by Dey et al. First, we compute the bilateral maps for 20 interest points uniformly sampled on each shape. We take the resulting 380 descriptors to create the signature of each shape. Next, during the retrieval phase, the matching score between two shapes $\mathcal{M}_1$ and $\mathcal{M}_2$ is given by the expression [Dey et al. 2010]

$$\sum_{f_1 \in F_1} \min_{f_2 \in F_2} \|f_1 - f_2\|_1 + \sum_{f_2 \in F_2} \min_{f_1 \in F_1} \|f_1 - f_2\|_1, \tag{5.5}$$

where $F_1$ and $F_2$ are the sets of signatures for $\mathcal{M}_1$ and $\mathcal{M}_2$, respectively, $f_1 \in F_1$ and $f_2 \in F_2$ denote single descriptor vectors in the signature set, and $\| \dots \|_1$ is the $\ell 1$-norm. This score is simple and efficient to compute for large datasets.

For each query, the target shapes are ordered according to the score in (5.5) and the results are evaluated in terms of the top-3 and top-5 hit rates. Basically, given a query shape, we have a top-$k$ hit if a shape from the same class is retrieved within the top $k$ matches. For a class with $N$ shapes, the top-$k$ hit rate is the percentage of top-$k$ hits with respect to $N$.

**Results.** Table 5.1 shows the retrieval hit rates for our descriptors, compared to the method of Dey et al. [2010] (PHS) and the Eigenvalue descriptor method (EVD) of Jain and Zhang [2007]. We see that, although PHS and EVD have different performances on complete and incomplete shapes, our descriptors provide better results than these two methods on both types of shapes. Thus, although the bilateral maps were designed for partial matching, their performance is not dropped when they are used to match complete shapes.

## 5.4 Discussion

The key idea in this chapter is to show that it is worthwhile to consider two points instead of one point in defining regions of interest for local shape descriptors. To demonstrate this point, we proposed the *bilateral map*, which anisotropically adapts its shape to the region comprised between two points. We showed that the bilateral map offers a promising alternative to the classic descriptor definition for tasks such as shape correspondence and retrieval. The bilateral map was designed to deal better with the partial matching problem, but we showed that this does not hinder its performance on the full matching problem. Also, if we assume that the same feature points are used for testing the preservation of distances, then the increase in complexity when going from one point to a pair of points is often absorbed by the complexity of the correspondence algorithm that utilizes the descriptors.

We have shown results when using the bilateral maps within a simple correspondence framework, to more easily evaluate the effectiveness of the descriptors. However, these descriptors can potentially help to improve the results of more elaborate methods based on the feature analysis approach. For example, they can be used within recent search-based algorithms [Gelfand et al. 2005; Zhang et al. 2008; Au et al. 2010], sampling-based algorithms [Tevs et al. 2011], or even to guide transformation-based [Lipman and Funkhouser 2009; Kim et al. 2011] or registration-based methods [Bronstein and Bronstein 2008a] more rapidly towards a good solution. Moreover, we evaluated the bilateral maps by collecting the

area of the faces that fall within each bin. However, the descriptor can be potentially used with other geometric properties, such as curvature [Gatzke et al. 2005], the shape diameter function [Shapira et al. 2009], or even with more sophisticated scalar fields, such as the heat diffusion field [Dey et al. 2010].

**Future directions.** It would be interesting to investigate higher-order generalizations of the proposed descriptors. For example, by going one step further and considering three points on a surface, we can define a region that can also be potentially used for matching. A two-dimensional grid can then be defined for binning this region. This generalization lifts the need to specify a threshold $\theta$ for the width of the descriptors, as in the case of the bilateral maps, although the complexity of the matching algorithm then becomes cubic on the number of feature points.

In this work, the comparisons between single-point descriptors and the bilateral maps have been performed by using the same sets of feature points. This might give an advantage to the bilateral maps, since more descriptors or signatures are generated. It would be informative to perform a comparison where the same number of descriptors is used in both paradigms, although a meaningful criterion has to be devised to select which feature points should be used in each paradigm.

The last point leads to another interesting topic for future investigation, *feature sampling*, which is a difficult problem when the goal is to capture features that appear consistently across two different shapes. We evaluated the proposed bilateral maps with feature points uniformly sampled across the surfaces. However, an interesting question is whether there is a feature sampling approach that is more suitable for the new descriptors. Previous works proposed to use sampling schemes based on extremity selection [Zhang et al. 2008], saliency [Castellani et al. 2008] or information criteria such as entropy [Tevs et al. 2011], but the particular choice of method is highly correlated with the category of shapes being considered. Perhaps the best way of selecting feature points on surfaces is to use an approach similar to that of Tevs et al. [2011], where the feature sampling is part of the correspondence computation and, potentially, could even be made part of the descriptor construction.

# Chapter 6

# Conclusion and future directions

In this thesis, we addressed the challenge of establishing a correspondence between dissimilar shapes. That is, although the shapes represent the same class of object or possess similar subportions, the shapes as a whole can have significant differences in their geometry, topology and part composition. Our contribution is to propose solutions to this problem on two different fronts.

First, we proposed two algorithms, one supervised and one unsupervised, that incorporate additional knowledge about the shapes. The knowledge is used to establish a part correspondence or co-segmentation of the shapes. In the supervised approach, the additional knowledge is obtained from a training set of segmented and labeled shapes. In the unsupervised method, the knowledge is extracted from the co-analysis of a set of shapes, which reveals the common semantic types of parts that exist in the set. We showed that such approaches allow us to obtain meaningful co-segmentation results for dissimilar shapes.

Secondly, we proposed a new type of shape descriptor, the *bilateral map*, to compute partial correspondences between shapes. The assumption is that the shapes possess a similar portion, although each shape can have additional parts that do not exist in the other shape. The key feature of the bilateral map is that its region of interest is defined by two points. In that manner, portions of the shapes that are not important for the context of these two points are selectively ignored in the matching. We showed that this descriptor is a promising alternative to classical descriptors for establishing partial matches.

**Future directions.** There are several directions in which our contributions could be improved or extended to address new types of problems.

Firstly, both knowledge-driven approaches can benefit from incorporating more advanced shape descriptors. For example, descriptors that capture more structural information, such as symmetry [Golovinskiy et al. 2007b], part [Liu et al. 2009] or skeletal [Au et al. 2008] information, as well as descriptors that would allow these methods to be applied to non-manifold meshes. Another possibility is to add descriptors that capture partial shape information, such as our bilateral map. The methods could also be adapted to particular fields with the use of domain-specific descriptors, e.g., when we know the expected part constitution of the shapes. Ultimately, the goal in a semantic analysis of shapes would be to obtain descriptors that capture the *functionality* of the shape parts.

While designing shape descriptors is one of the important tasks in shape analysis, if we are given multiple descriptors, another challenge is to select the set of descriptors that best capture the properties of a part class. In this context, the knowledge-driven approaches can also benefit from incorporating more advanced *feature selection* methods to aid in this task.

Secondly, each of the knowledge-driven approaches has its specific applicability. For example, the supervised approach can be steered to segment the shapes in the exact manner that the user specifies through the training set. The unsupervised approach does not require the construction of such a training set, but automatically derives the part composition that is implied by the set of shapes, although the result does not necessarily conform to what the user desires. Thus, it seems advantageous to combine these two methods into a semi-supervised or active learning framework [Settles 2010], where the user needs to provide less input (perhaps interactively), and the method derives all the remaining knowledge from the set, e.g., as applied to the segmentation of individual 3D shapes [Top et al. 2011].

A related challenge is to extend the unsupervised co-segmentation to a multi-class scenario, where the input consists of more than one class of shapes. In this case, the co-analysis would have to classify the shapes into their semantic categories and compute the co-segmentation of each class. Ideally, these two steps would be performed simultaneously, so that the shape classification and part analysis could benefit from each other.

Finally, we proposed a promising descriptor for partial matching with the bilateral map. However, there are still aspects of the problem that can be addressed more directly. First, the bilateral map could be made more effective by defining its region of interest based on other types of distance fields, e.g., incorporating part-specific information [Liu et al. 2009;

Dey et al. 2010]. Secondly, we are still using an isometry criterion to filter correspondences in our experiments. It would be interesting to propose an improved isometry criterion that is more specific to partial matching, where the distances between feature points are normalized relative to the partial match. Finally, the sampling of feature points could be integrated with the descriptor construction and computation of the partial correspondence.

# Bibliography


Aiger, D., Mitra, N. J., and Cohen-Or, D. 2008. 4-points congruent sets for robust surface registration. *ACM Trans. on Graphics (Proc. SIGGRAPH) 27*, 3, 1–10.

Alexa, M. 2002. Recent advances in mesh morphing. *Computer Graphics Forum 21*, 2, 173–198.

Allen, B., Curless, B., and Popović, Z. 2003. The space of human body shapes: reconstruction and parameterization from range scans. *ACM Trans. on Graphics (Proc. SIGGRAPH) 22*, 3, 587–594.

Anguelov, D., Srinivasan, P., Pang, H.-C., Koller, D., Thrun, S., and Davis, J. 2004. The correlated correspondence algorithm for unsupervised registration of nonrigid surfaces. In *NIPS*.

Anguelov, D., Srinivasan, P., Koller, D., Thrun, S., Rodgers, J., and Davis, J. 2005. SCAPE: shape completion and animation of people. *ACM Trans. on Graphics (Proc. SIGGRAPH) 24*, 3, 408–416.

Attene, M., Robbiano, F., Spagnuolo, M., and Falcidieno, B. 2009. Characterization of 3D shape parts for semantic annotation. *Computer-Aided Design 41*, 10, 756–763.

Au, O. K.-C., Tai, C.-L., Chu, H.-K., Cohen-Or, D., and Lee, T.-Y. 2008. Skeleton extraction by mesh contraction. *ACM Trans. on Graphics (Proc. SIGGRAPH) 27*, 3, 44:1–44:10.

Au, O. K.-C., Cohen-Or, D., Tai, C.-L., Fu, H., and Zheng, Y. 2010. Electors voting for fast automatic shape correspondence. *Computer Graphics Forum (Proc. EUROGRAPHICS) 29*, 2, 645–654.

Audette, M. A., Ferrie, F. P., and Peters, T. M. 2000. An algorithmic overview of surface registration techniques for medical imaging. *Medical Image Analysis 4*, 3, 201–217.

Baran, I., Vlasic, D., Grinspun, E., and Popović, J. 2009. Semantic deformation transfer. *ACM Trans. on Graphics (Proc. SIGGRAPH) 28*, 3, 1–6.

Basri, R., and Jacobs, D. W. 1997. Recognition using region correspondences. *Int. J. Comput. Vision 25*, 2, 145–166.

Belongie, S., Malik, J., and Puzicha, J. 2002. Shape matching and object recognition using shape context. *IEEE PAMI 24*, 4, 509–522.

Berg, A. C., Berg, T. L., and Malik, J. 2005. Shape matching and object recognition using low distortion correspondence. In *Proc. IEEE Conf. on CVPR*, vol. 1, 26–33.

Biasotti, S., Marini, S., Spagnuolo, M., and Falcidieno, B. 2006. Sub-part correspondence by structural descriptors of 3D shapes. *Computer-Aided Design 38*, 9, 1002–1019.

Biasotti, S., Giorgi, D., Spagnuolo, M., and Falcidieno, B. 2008. Reeb graphs for shape analysis and applications. *Theoretical Computer Science 392*, 1–3, 5–22.

Boykov, Y., Veksler, O., and Zabih, R. 2001. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell. 23*, 11, 1222–1239.

Bronstein, A. M., and Bronstein, M. M. 2008. Not only size matters: regularized partial matching of nonrigid shapes. In *Proc. Workshop on Nonrigid Shape Analysis and Deformable Image Alignment (NORDIA)*, 1–6.

Bronstein, A. M., and Bronstein, M. M. 2008. Regularized partial matching of rigid shapes. In *Proc. Euro. Conf. on Comp. Vis. (ECCV)*, 143–154.

Bronstein, A. M., Bronstein, M. M., and Kimmel, R. 2006. Efficient computation of isometry-invariant distances between surfaces. *SIAM J. Scientific Computing 28*, 5, 1812–1836.

Bronstein, A. M., Bronstein, M. M., Bruckstein, A. M., and Kimmel, R. 2008. Partial similarity of objects, or how to compare a centaur to a horse. *Int. J. Comput. Vision 84*, 2, 163–183.

Bronstein, A. M., Bronstein, M. M., and Kimmel, R. 2008. *Numerical Geometry of Non-Rigid Shapes*. Springer.

Brown, B. J., and Rusinkiewicz, S. 2007. Global non-rigid alignment of 3-D scans. *ACM Trans. on Graphics (Proc. SIGGRAPH) 26*, 3.

Caetano, T. S., McAuley, J. J., Cheng, L., Le, Q. V., and Smola, A. J. 2009. Learning graph matching. *IEEE PAMI 31*, 6, 1048–1058.

Castellani, U., Cristani, M., Fantoni, S., and Murino, V. 2008. Sparse points matching by combining 3D mesh saliency with statistical descriptors. *Computer Graphics Forum (Proc. EUROGRAPHICS) 27*, 2, 643–652.

Chang, W., and Zwicker, M. 2008. Automatic registration for articulated shapes. *Computer Graphics Forum (Proc. SGP) 27*, 5, 1459–1468.

Chang, W., and Zwicker, M. 2009. Range scan registration using reduced deformable models. *Computer Graphics Forum (Proc. EUROGRAPHICS) 28*, 2.

Chaudhuri, S., Kalogerakis, E., Guibas, L., and Koltun, V. 2011. Probabilistic reasoning for assembly-based 3D modeling. *ACM Trans. on Graphics (Proc. SIGGRAPH) 30*, 4.

Chen, X., Golovinskiy, A., , and Funkhouser, T. 2009. A benchmark for 3D mesh segmentation. *ACM Trans. on Graphics (Proc. SIGGRAPH) 28*, 3, 1–12.

Chui, H., and Rangarajan, A. 2003. A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding 89*, 114–141.

Coifman, R. R., and Lafon, S. 2006. Diffusion maps. *Applied and Computational Harmonic Analysis 21*, 1, 5–30.

Comaniciu, D., and Meer, P. 2002. Mean shift: a robust approach towards feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell. 24*, 603–619.

Davies, R., Twining, C., and Taylor, C. 2008. *Statistical Models of Shape: Optimisation and Evaluation*. Springer. ISBN: 978-1-84800-137-4.

de Aguiar, E., Stoll, C., Theobalt, C., Ahmed, N., Seidel, H.-P., and Thrun, S. 2008. Performance capture from sparse multi-view video. *ACM Trans. on Graphics (Proc. SIGGRAPH) 27*, 3.

de Goes, F., Goldenstein, S., and Velho, L. 2008. A hierarchical segmentation of articulated bodies. *Computer Graphics Forum (Proc. SGP) 27*, 5, 1349–1356.

Dey, T. K., Li, K., Luo, C., Ranjan, P., Safa, I., and Wang, Y. 2010. Persistent heat signature for pose-oblivious matching of incomplete models. *Computer Graphics Forum (Proc. SGP) 29*, 5, 1545–1554.

Dinh, H. Q., Yezzi, A., and Turk, G. 2005. Texture transfer during shape transformation. *ACM Trans. on Graphics 24*, 2, 289–310.

Elad, A., and Kimmel, R. 2003. On bending invariant signatures for surfaces. *IEEE PAMI 25*, 10, 1285–1295.

Fischler, M. A., and Bolles, R. C. 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM 24*, 6, 381–395.

Forsyth, D. A., and Ponce, J. 2003. *Computer Vision: A Modern Approach*. Prentice Hall.

Friedman, J., Hastie, T., and Tibshirani, R. 2000. Additive logistic regression: a statistical view of boosting. *Annals of statistics 28*, 2, 337–374.

Fu, H., Cohen-Or, D., Dror, G., and Sheffer, A. 2008. Upright orientation of man-made objects. *ACM Trans. on Graphics (Proc. SIGGRAPH) 27*, 3, 1–8.

Funkhouser, T., and Shilane, P. 2006. Partial matching of 3D shapes with priority-driven search. In *Proc. Symp. on Geom. Processing (SGP)*, 131–142.

Gal, R., and Cohen-Or, D. 2006. Salient geometric features for partial shape matching and similarity. *ACM Trans. on Graphics 25*, 1, 130–150.

Gal, R., Shamir, A., Hassner, T., Pauly, M., and Cohen-Or, D. 2007. Surface reconstruction using local shape priors. In *Proc. Symp. on Geom. Processing (SGP)*.

Gal, R., Sorkine, O., Mitra, N. J., and Cohen-Or, D. 2009. iWIRES: an analyze-and-edit approach to shape manipulation. *ACM Trans. on Graphics (Proc. SIGGRAPH) 28*, 3, 1–10.

Gall, J., Stoll, C., de Aguiar, E., Theobalt, C., Rosenhahn, B., and Seidel, H.-P. 2009. Motion capture using joint skeleton tracking and surface estimation. In *Proc. IEEE Conf. on CVPR*.

Gatzke, T., Grimm, C., Garland, M., and Zelinka, S. 2005. Curvature maps for local shape comparison. In *Proc. IEEE Conf. on Shape Modeling and Applications*, 244–253.

Gelfand, N., Mitra, N. J., Guibas, L. J., and Pottmann, H. 2005. Robust global registration. In *Proc. Symp. on Geom. Processing (SGP)*, 197–206.

Gold, S., and Rangarajan, A. 1995. Softmax to softassign: neural network algorithms for combinatorial optimization. *J. Artif. Neural Netw. 2*, 4, 381–399.

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

Golovinskiy, A., Podolak, J., and Funkhouser, T. 2007. Symmetry-aware mesh processing. *Princeton University TR-782-07*.

Golovinskiy, A., Podolak, J., and Funkhouser, T. 2007. Symmetry-aware mesh processing. Tech. Rep. TR-782-07, Princeton University.

Heider, P., Pierre-Pierre, A., Li, R., and Grimm, C. 2011. Local shape descriptors, a survey and evaluation. In *Proc. Eurographics Workshop on 3D Object Retrieval*, 49–56.

Hoffman, D. D., and Richards, W. A. 1987. Parts of recognition. *Readings in computer vision: issues, problems, principles, and paradigms*, 227–242.

Huang, Q.-X., Adams, B., Wicke, M., and Guibas, L. J. 2008. Non-rigid registration under isometric deformations. *Computer Graphics Forum (Proc. SGP) 27*, 5, 1449–1457.

Huang, Q., Koltun, V., and Guibas, L. 2011. Joint shape segmentation with linear programming. *ACM Trans. on Graphics (Proc. SIGGRAPH Asia) 30*, 6, 1–11.

Huttenlocher, D. P. 1991. Fast affine point matching: an output-sensitive method. In *Proc. IEEE Conf. on CVPR*, 263–268.

Irani, S., and Raghavan, P. 1996. Combinatorial and experimental results for randomized point matching algorithms. In *Proc. Symp. on Computational geometry*, 68–77.

Iyer, N., Jayanti, S., Lou, K., Kalyanaraman, Y., and Ramani, K. 2005. Three-dimensional shape searching: state-of-the-art review and future trends. *Computer-Aided Design 37*, 5, 509–530.

Jain, V., and Zhang, H. 2007. A spectral approach to shape-based retrieval of articulated 3D models. *Computer-Aided Design 39*, 5, 398–407.

Jain, V., Zhang, H., and van Kaick, O. 2007. Non-rigid spectral correspondence of triangle meshes. *International Journal on Shape Modeling 13*, 1, 101–124.

Johnson, A., and Hebert, M. 1999. Using spin-images for efficient multiple model recognition in cluttered 3D scenes. *IEEE PAMI 29*, 5, 433–449.

Joulin, A., Bach, F., and J.Ponce. 2010. Discriminative clustering for image co-segmentation. In *Proc. IEEE Conf. on CVPR*, 1943–1950.

Kalogerakis, E., Hertzmann, A., and Singh, K. 2010. Learning 3D mesh segmentation and labeling. *ACM Trans. on Graphics (Proc. SIGGRAPH) 29*, 3, 1–11.

Katz, S., and Tal, A. 2003. Hierarchical mesh decomposition using fuzzy clustering and cuts. *ACM Trans. on Graphics (Proc. SIGGRAPH) 22*, 3, 954–961.

Kazhdan, M., Funkhouser, T., and Rusinkiewicz, S. 2003. Rotation invariant spherical harmonic representation of 3D shape descriptors. In *Proc. Symp. on Geom. Processing (SGP)*, 156–164.

Kazhdan, M., Funkhouser, T., and Rusinkiewicz, S. 2004. Shape matching and anisotropy. *ACM Trans. on Graphics 23*, 3, 623–629.

Kim, V. G., Lipman, Y., Chen, X., and Funkhouser, T. 2010. Möbius transformations for global intrinsic symmetry analysis. *Computer Graphics Forum (Proc. SGP) 29*, 5.

Kim, V., Lipman, Y., and Funkhouser, T. 2011. Blended intrinsic maps. *ACM Trans. on Graphics (Proc. SIGGRAPH) 30*, 4, 1–12.

Körtgen, M., Park, G.-J., Novotni, M., and Klein, R. 2003. 3D shape matching with 3D shape contexts. In *Proc. 7th Central European Seminar on Computer Graphics.*

KRAEVOY, V., AND SHEFFER, A. 2004. Cross-parameterization and compatible remeshing of 3D models. *ACM Trans. on Graphics (Proc. SIGGRAPH) 23*, 3, 861–869.

KRAEVOY, V., JULIUS, D., AND SHEFFER, A. 2007. Model composition from interchangeable components. In *Proc. Pacific Graphics*, 129–138.

LECLERC, Y. G., LUONG, Q.-T., FUA, P. V., AND MIYAJIMA, K. 2000. Detecting changes in 3-D shape using self-consistency. In *Proc. IEEE Conf. on CVPR*, vol. 1, 395–402.

LEORDEANU, M., AND HEBERT, M. 2005. A spectral technique for correspondence problems using pairwise constraints. In *Proc. Int. Conf. on Comp. Vis. (ICCV)*, vol. 2, 1482–1489.

LEVIN, A., AND WEISS, Y. 2006. Learning to combine bottom-up and top-down segmentation. *Int. J. Comput. Vision (Proc. ECCV) 81*, 1, 105–118.

LI, X., AND GUSKOV, I. 2005. Multi-scale features for approximate alignment of point-based surfaces. In *Proc. Symp. on Geom. Processing (SGP)*.

LI, H., ADAMS, B., GUIBAS, L. J., AND PAULY, M. 2009. Robust single-view geometry and motion reconstruction. *ACM Trans. on Graphics (Proc. SIGGRAPH Asia)*.

LIPMAN, Y., AND FUNKHOUSER, T. 2009. Möbius voting for surface correspondence. *ACM Trans. on Graphics (Proc. SIGGRAPH) 28*, 3, 1–12.

LIPMAN, Y., CHEN, X., DAUBECHIES, I., AND FUNKHOUSER, T. 2010. Symmetry factored embedding and distance. *ACM Trans. on Graphics (Proc. SIGGRAPH) 29*, 4.

LIU, R., AND ZHANG, H. 2007. Mesh segmentation via spectral embedding and contour analysis. *Computer Graphics Forum (Proc. EUROGRAPHICS) 26*, 3, 385–394.

LIU, R., ZHANG, H., SHAMIR, A., AND COHEN-OR, D. 2009. A part-aware surface metric for shape analysis. *Computer Graphics Forum (Proc. EUROGRAPHICS) 28*, 2, 397–406.

MACIEL, J., AND COSTEIRA, J. P. 2003. A global solution to sparse correspondence problems. *IEEE PAMI 25*, 2, 187–199.

MANAY, S., CREMERS, D., HONG, B.-W., YEZZI, A. J., AND SOATTO, S. 2006. Integral invariants for shape matching. *IEEE PAMI 28*, 10, 1602–1618.

Marr, D. 1982. *Vision - A Computational Investigation into the Human Representation and Processing of Visual Information.* W. H. Freeman.

Mateus, D., Horaud, R., Knossow, D., Cuzzolin, F., and Boyer, E. 2008. Articulated shape matching using Laplacian eigenfunctions and unsupervised point registration. In *Proc. IEEE Conf. on CVPR*, 1–8.

Mirzaalian, H., Hamarneh, G., and Lee, T. 2009. Graph-based approach to skin mole matching incorporating template-normalized coordinates. In *Proc. IEEE Conf. on CVPR*, 2152–2159.

Mitra, N. J., Guibas, L., and Pauly, M. 2006. Partial and approximate symmetry detection for 3D geometry. *ACM Trans. on Graphics (Proc. SIGGRAPH) 25*, 3, 560–568.

Mitra, N. J., Flory, S., Ovsjanikov, M., Gelfand, N., Guibas, L., and Pottmann, H. 2007. Dynamic geometry registration. In *Proc. Symp. on Geom. Processing (SGP)*, 173–182.

Mitra, N. J., Yang, Y.-L., Yan, D.-M., Li, W., and Agrawala, M. 2010. Illustrating how mechanical assemblies work. *ACM Trans. on Graphics (Proc. SIGGRAPH) 29*, 4, 1–12.

Nadler, B., Lafon, S., Coifman, R. R., and Kevrekidis, I. G. 2005. Diffusion maps, spectral clustering and eigenfunctions of Fokker-Planck operators. In *NIPS*, 1–8.

Ng, B., Abugharbieh, R., and Hamarneh, G. 2010. Group MRF for fMRI activation detection. In *Proc. IEEE Conf. on CVPR*, 2887–2894.

Olson, C. F. 1997. Efficient pose clustering using a randomized algorithm. *Int. J. Comput. Vision 23*, 2, 131–147.

Ovsjanikov, M., Sun, J., and Guibas, L. 2008. Global intrinsic symmetries of shapes. *Computer Graphics Forum (Proc. SGP) 27*, 5, 1341–1348.

Ovsjanikov, M., Mérigot, Q., Mémoli, F., and Guibas, L. 2010. One point isometric matching with the heat kernel. *Computer Graphics Forum (Proc. SGP) 29*, 5.

Papadimitriou, C., and Stieglitz, K. 1982. *Combinatorial Optimization: Algorithms and Complexity.* Pretice Hall.

Pauly, M., Mitra, N. J., Giesen, J., Gross, M., and Guibas, L. J. 2005. Example-based 3D scan completion. In *Proc. Symp. on Geom. Processing (SGP)*.

Pekelny, Y., and Gotsman, C. 2008. Articulated object reconstruction and markerless motion capture from depth video. *Computer Graphics Forum (Proc. EUROGRAPHICS) 27*, 2, 399–408.

Pitiot, A., Delingette, H., and Thompson, P. M. 2007. Learning shape correspondence for n-d curves. *International Journal of Computer Vision 71*, 1, 71–88.

Podolak, J., Shilane, P., Golovinskiy, A., Rusinkiewicz, S., and Funkhouser, T. 2006. A planar-reflective symmetry transform for 3D shapes. *ACM Trans. on Graphics (Proc. SIGGRAPH) 25*, 3, 549–559.

Pottmann, H., Steiner, T., Hofer, M., Haider, C., and Hanbury, A. 2004. The isophotic metric and its application to feature sensitive morphology on surfaces. In *Proc. Euro. Conf. on Comp. Vis. (ECCV)*, 560–572.

Rother, C., Kolmogorov, V., Minka, T., and Blake, A. 2006. Cosegmentation of image pairs by histogram matching – incorporating a global constraint into MRFs. In *Proc. IEEE Conf. on CVPR*, 993–1000.

Rusinkiewicz, S., and Levoy, M. 2001. Efficient variants of the ICP algorithm. In *Proc. 3rd Int. Conf. on 3D Digital Imaging and Modeling*, 145–152.

Rustamov, R. 2007. Laplace-Beltrami eigenfunctions for deformation invariant shape representation. In *Proc. Symp. on Geom. Processing (SGP)*, 225–233.

Sahillioğlu, Y., and Yemez, Y. 2010. 3D shape correspondence by isometry-driven greedy optimization. In *Proc. IEEE Conf. on CVPR*.

Schnitman, Y., Caspi, Y., Cohen-Or, D., and Lischinski, D. 2006. Inducing semantic segmentation from an example. In *Proc. of ACCV*, vol. 2, 373–384.

Settles, B. 2010. Active learning literature survey. Tech. Rep. 1648, University of Wisconsin-Madison.

Shalom, S., Shapira, L., Shamir, A., and Cohen-Or, D. 2008. Part analogies in sets of objects. In *Proc. of Eurographics Symposium on 3D Object Retrieval*, 33–40.

Shamir, A., Shapira, L., and Cohen-Or, D. 2006. Mesh analysis using geodesic mean-shift. *The Visual Computer 22*, 99–108.

Shamir, A. 2004. A formulation of boundary mesh segmentation. In *Proc. Int. Symp. on 3D Data Processing, Visualization and Transmission (3DPVT)*, 82–89.

Shamir, A. 2008. A survey on mesh segmentation techniques. *Computer Graphics Forum 27*, 6, 1539–1556.

Shapira, L., Shamir, A., and Cohen-Or, D. 2008. Consistent mesh partitioning and skeletonization using the shape diameter function. *The Visual Computer 24*, 4, 249–259.

Shapira, L., Shalom, S., Shamir, A., Cohen-Or, D., and Zhang, H. 2009. Contextual part analogies in 3D objects. *Int. J. Comput. Vision 89*, 2–3, 309–326.

Sharf, A., Alcantara, D. A., Lewiner, T., Greif, C., Sheffer, A., Amenta, N., and Cohen-Or, D. 2008. Space-time surface reconstruction using incompressible flow. *ACM Trans. on Graphics (Proc. SIGGRAPH Asia) 27*, 5, 1–10.

Sidi, O., van Kaick, O., Kleiman, Y., Zhang, H., and Cohen-Or, D. 2011. Unsupervised co-segmentation of a set of shapes via descriptor-space spectral clustering. *ACM Trans. on Graphics (Proc. SIGGRAPH Asia) 30*, 6, to appear.

Simari, P., Nowrouzezahrai, D., Kalogerakis, E., and Singh, K. 2009. Multi-objective shape segmentation and labeling. *Computer Graphics Forum (Proc. SGP) 28*, 5, 1415–1425.

Stockman, G. 1987. Object recognition and localization via pose clustering. *Comput. Vision Graph. Image Process. 40*, 3, 361–387.

Sumner, R. W., and Popović, J. 2004. Deformation transfer for triangle meshes. *ACM Trans. on Graphics (Proc. SIGGRAPH) 23*, 3, 399–405.

Süssmuth, J., Winter, M., and Greiner, G. 2008. Reconstructing animated meshes from time-varying point clouds. *Computer Graphics Forum (Proc. SGP) 27*, 5, 1469–1476.

Tang, L., and Hamarneh, G. 2008. SMRFI: Shape matching via registration of vector-valued feature images. In *Proc. Computer Vision and Pattern Recognition (CVPR)*, 1–8.

Tangelder, J. W. H., and Veltkamp, R. C. 2008. A survey of content based 3D shape retrieval methods. *Multimedia Tools and Applications 39*, 3, 441–471.

Tevs, A., Bokeloh, M., Wand, M., Schilling, A., and Seidel, H.-P. 2009. Isometric registration of ambiguous and partial data. In *Proc. IEEE Conf. on CVPR*, 1185–1192.

Tevs, A., Berner, A., Wand, M., Ihrke, I., and Seidel, H.-P. 2011. Intrinsic shape matching by planned landmark sampling. *Computer Graphics Forum (Proc. EUROGRAPHICS) 30*, 543–552.

Top, A., Hamarneh, G., and Abugharbieh, R. 2011. Active learning for interactive 3D image segmentation. *Lecture Notes in Computer Science (Proc. MICCAI)*, 603–610.

Turk, G., and Levoy, M. 1994. Zippered polygon meshes from range images. In *Proc. SIGGRAPH*, 311–318.

van Kaick, O., Hamarneh, G., Zhang, H., and Wighton, P. 2007. Contour correspondence via ant colony optimization. In *Proc. Pacific Graphics*, 271–280.

van Kaick, O., Ward, A., Hamarneh, G., Schweitzer, M., and Zhang, H. 2010. Learning fourier descriptors for computer-aided diagnosis of the supraspinatus. *Academic Radiology 17*, 8, 1040–1049.

van Kaick, O., Zhang, H., Hamarneh, G., and Cohen-Or, D. 2011. A survey on shape correspondence. *Computer Graphics Forum 30*, 6, 1681–1707.

van Kaick, O., Tagliasacchi, A., Sidi, O., Zhang, H., Cohen-Or, D., Wolf, L., and Hamarneh, G. 2011. Prior knowledge for part correspondence. *Computer Graphics Forum (Proc. EUROGRAPHICS) 30*, 2, 553–562.

Wand, M., Jenke, P., Huang, Q.-X., Bokeloh, M., Guibas, L., and Schilling, A. 2007. Reconstruction of deforming geometry from time-varying point clouds. In *Proc. Symp. on Geom. Processing (SGP)*, 49–58.

Wand, M., Adams, B., Ovsjanikov, M., Berner, A., Bokeloh, M., Jenke, P., Guibas, L., Seidel, H.-P., and Schilling, A. 2009. Efficient reconstruction of non-rigid shape and motion from real-time 3D scanner data. *ACM Trans. on Graphics 28*, 2, 1–15.

Ward, A. D., and Hamarneh, G. 2007. Statistical shape modeling using MDL incorporating shape, appearance, and expert knowledge. *Lecture Notes in Computer Science (Proc. MICCAI) 4791*, 278–285.

Ward, A. D., and Hamarneh, G. 2009. The groupwise medial axis transform for fuzzy skeletonization and pruning. *IEEE PAMI Accepted for future publication*.

Wolfson, H. J., and Rigoutsos, I. 1997. Geometric hashing: an overview. *IEEE Computational Science & Engineering 4*, 4, 10–21.

Xu, K., Zhang, H., Tagliasacchi, A., Liu, L., Li, G., Meng, M., and Xiong, Y. 2009. Partial intrinsic reflectional symmetry of 3d shapes. *ACM Trans. on Graphics (Proc. SIGGRAPH Asia) 28*, 5, 1–10.

Xu, K., Zhang, H., Tagliasacchi, A., Liu, L., Li, G., Meng, M., and Xiong, Y. 2009. Partial intrinsic reflectional symmetry of 3D shapes. *ACM Trans. on Graphics (Proc. SIGGRAPH Asia) 28*, 5, 1–10.

Xu, W., Wang, J., Yin, K., Zhou, K., van de Panne, M., Chen, F., and Guo, B. 2009. Joint-aware manipulation of deformable models. *ACM Trans. on Graphics (Proc. SIGGRAPH) 28*, 3, 1–9.

Xu, K., Li, H., Zhang, H., Cohen-Or, D., Xiong, Y., and Cheng, Z. 2010. Style-content separation by anisotropic part scales. *ACM Trans. on Graphics (Proc. SIGGRAPH Asia) 29*, 5, 1–9.

Zass, R., and Shashua, A. 2008. Probabilistic graph and hypergraph matching. In *Proc. IEEE Conf. on CVPR*.

Zhang, H., Sheffer, A., Cohen-Or, D., Zhou, Q., van Kaick, O., and Tagliasacchi, A. 2008. Deformation-driven shape correspondence. *Computer Graphics Forum (Proc. SGP) 27*, 5, 1431–1439.

Zhang, H., van Kaick, O., and Dyer, R. 2010. Spectral mesh processing. *Computer Graphics Forum 29*, 6, 1865–1894.

Zheng, Y., and Doermann, D. 2006. Robust point matching for nonrigid shapes by preserving local neighborhood structures. *IEEE PAMI 28*, 4, 643–649.

Zheng, Q., Sharf, A., Tagliasacchi, A., Chen, B., Zhang, H., Sheffer, A., and Cohen-Or, D. 2010. Consensus skeleton for non-rigid space-time registration. *Computer Graphics Forum (Proc. EUROGRAPHICS) 29*, 2, 635–644.