

# Tensor Maps for Synchronizing Heterogeneous Shape Collections

QIXING HUANG, The University of Texas at Austin, USA  
ZHENXIAO LIANG, The University of Texas at Austin, USA  
HAOYUN WANG, Tsinghua University, USA  
SIMIAO ZUO, The University of Texas at Austin, USA  
CHANDRAJIT BAJAJ, The University of Texas at Austin, USA

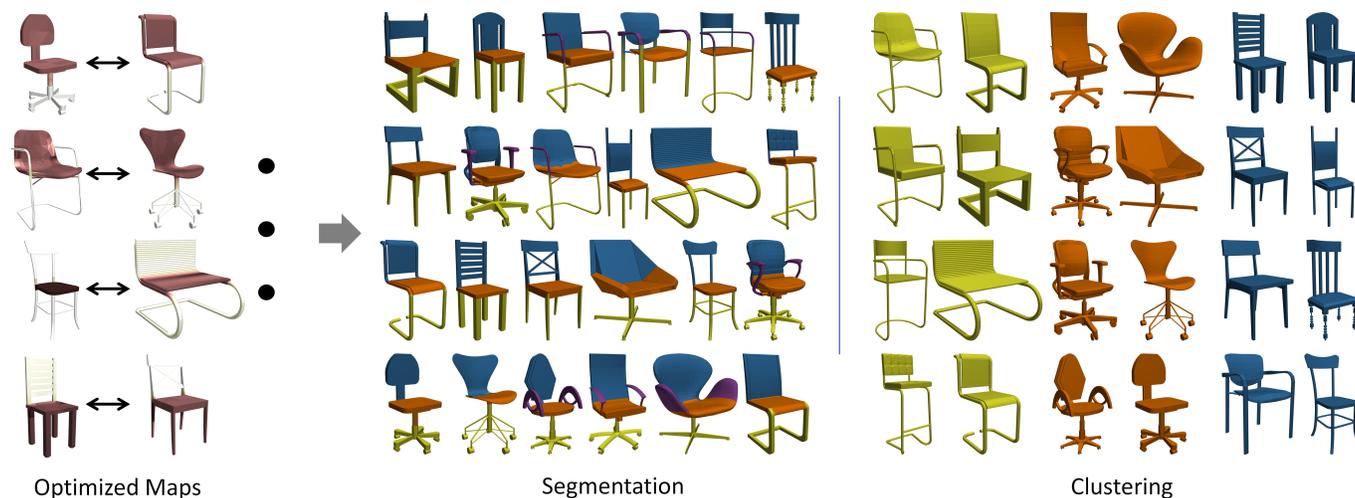


Fig. 1. A prototypical example of a heterogeneous shape model collection with highly consistent correspondence maps generated from our tensor approach and optimized across the entire collection. These maps simultaneously encodes structural and functional similarities and variances. The corresponding regions of each object in the shape collections is shown with matching color. The induced shape maps additionally enable applications in shape segmentation and sub-region weighted shape co-clustering.

Establishing high-quality correspondence maps between geometric shapes has been shown to be the fundamental problem in managing geometric shape collections. Prior work has focused on computing efficient maps between pairs of shapes, and has shown a quantifiable benefit of joint map synchronization, where a collection of shapes are used to improve (denoise) the pairwise maps for consistency and correctness. However, these existing map synchronization techniques place very strong assumptions on the input shapes collection such as all the input shapes fall into the same category and/or the majority of the input pairwise maps are correct. In this paper, we present a multiple map synchronization approach that takes a heterogeneous shape collection as input and simultaneously outputs consistent

dense pairwise shape maps. We achieve our goal by using a novel tensor-based representation for map synchronization, which is efficient and robust than all prior matrix-based representations. We demonstrate the usefulness of this approach across a wide range of geometric shape datasets and the applications in shape clustering and shape co-segmentation.

CCS Concepts: • **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability;

Additional Key Words and Phrases: machine learning, numerical analysis, optimization, object scanning/acquisition, shape analysis and shape matching and retrieval

## ACM Reference Format:

Qixing Huang, Zhenxiao Liang, Haoyun Wang, Simiao Zuo, and Chandrajit Bajaj. 2019. Tensor Maps for Synchronizing Heterogeneous Shape Collections. *ACM Trans. Graph.* 38, 4, Article 106 (July 2019), 18 pages. <https://doi.org/10.1145/3306346.3322944>

## 1 INTRODUCTION

Digital shape collections are a rich resource of information for diverse data driven applications. Developing effective tools to analyze and organize them is a central research problem in geometry processing and machine learning. Prior related papers have focused on computing shape correspondence maps across all pairs of shapes in

Authors' addresses: Qixing Huang, The University of Texas at Austin, 2317 Speedway, Austin, TX, 78712, USA; Zhenxiao Liang, The University of Texas at Austin, 2317 Speedway, Austin, TX, 78712, USA; Haoyun Wang, Tsinghua University, 2317 Speedway, Austin, TX, 78712, USA; Simiao Zuo, The University of Texas at Austin, 2317 Speedway, Austin, TX, 78712, USA; Chandrajit Bajaj, The University of Texas at Austin, 2317 Speedway, Austin, TX, 78712, USA.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2019 Association for Computing Machinery.  
0730-0301/2019/7-ART106 \$15.00  
<https://doi.org/10.1145/3306346.3322944>

the collection. Such shape maps facilitate the propagation and aggregation of shape information, enabling diverse data driven processing for texture and animation transfer [Kraevoy and Sheffer 2004; Schreiner et al. 2004; Sumner and Popović 2004], browsing [Gao et al. 2015; Huang et al. 2014; Kim et al. 2012; Xu et al. 2013], co-analysis [Fish et al. 2016; Hu et al. 2012; Huang et al. 2011; Sidi et al. 2011; van Kaick et al. 2013], abstraction [Yumer and Kara 2012], and modeling/synthesis [Funkhouser et al. 2004; Kraevoy et al. 2007]. Yet existing techniques have primarily focused on homogeneous shape collections, e.g., a collection of human body models [Giorgi et al. 2007] or a collection of chair models of similar style (c.f. [Kim et al. 2013, 2012]).

In this paper, we present a framework for analyzing heterogeneous shape collections, e.g., those downloaded from different internet model repositories. In contrast to homogeneous shape collections, heterogeneous shape collections usually exhibit significant geometric variabilities, e.g., different part configurations. We demonstrate that establishing high-quality maps across heterogeneous shape collections provides unique opportunities for analyzing and hierarchically organizing such shape collections. Consider shape segmentation, a fundamental task in shape analysis, we show that one can identify parts of a shape through shape differences induced from shape maps. For example, chair back is a part because it is added to a stool. As another example, we can recognize chair legs as a part through the difference between a swivel basis and a four-leg basis (See Figure 1).

Besides variability-driven part discovery, we also show how to use shape maps to derive meaningful cluster structures of a heterogeneous shape collection. Specifically, when computing maps between pairs of shapes in isolation (we call them initial maps in this paper) using an off-the-shelf shape matching method, intra-cluster maps tend to be more accurate than inter-cluster maps. In other words, if we can accurately recover the underlying ground-truth maps, then the differences between the initial maps and the recovered maps provide meaningful affinity scores for shape clustering. Compared to standard shape clustering approaches that are based on shape descriptors and/or shape similarity scores, this approach is particularly suitable for fine-grained classification, e.g., different poses of the same person. Another unique feature of this approach is that clustering can be performed region-wise, e.g., map differences within backs/seats/legs of chairs, providing great flexibility in exploring shape collections.

To make such applications possible, it is vital to establish high-quality maps across heterogeneous shape collections. This is a quite challenging task because we need both to link corresponding regions that undergo significant geometric changes (e.g., chair backs under different fine-grained classes) and to identify regions that should not be matched (e.g., from swivel basis to rocking basis). Despite significant advances in shape matching (c.f. [van Kaick et al. 2011]) and particularly more recent works on matching a shape collection jointly to improve the maps between pairs of shapes (or map synchronization) [Chen et al. 2014; Cosmo et al. 2017; Huang et al. 2014; Huang and Guibas 2013; Huang et al. 2012, 2019; Kim et al. 2012; Nguyen et al. 2011; Wang et al. 2013; Zhang et al. 2019], the outputs of state-of-the-art approaches remain insufficient for high-quality analysis of heterogeneous shape collections.

In this paper, we introduce a novel method for map synchronization that nicely addresses the performance barrier among existing map synchronization approaches. Our approach is motivated from a recent trend that utilizes high-order tensor decompositions to solve challenging problems in machine learning and relevant fields [Anandkumar et al. 2014; Cichocki et al. 2015; Kolda and Bader 2009; Lahat et al. 2015; Lu et al. 2016; Sidiropoulos et al. 2017]. In particular, we show how to formulate map synchronization as solving a low canonical polyadic (CP) rank (c.f. [Ashraphijuo and Wang 2017]) tensor recovery problem. This formulation is further enhanced by a pre-filtering operation that enforces the cycle-consistency constraint along 3-cycles [Huang and Guibas 2013; Nguyen et al. 2011]. Combing them together, our approach can recover accurate maps among a heterogeneous shape collection from highly noisy and incomplete maps computed between pairs of shapes in isolation. We provide an analysis to justify the effectiveness of our approach against matrix-based map synchronization techniques. The resulting maps enable the applications described above, i.e., analyzing shape variability for shape segmentation and utilizing map residuals for shape clustering.

Compared with recent works on shape segmentation that focused on supervised learning [Guo et al. 2015; Kalogerakis et al. 2017, 2010; Yi et al. 2017a,b], our data-driven shape segmentation approach is completely unsupervised. We argue that such unsupervised data-driven approaches are useful in many settings not only because labeling shape parts are costly and error-prone, but also how to define meaningful parts is quite subjective (c.f. [Yi et al. 2016]).

We have evaluated our map synchronization approach on two benchmark datasets SHREC07 [Giorgi et al. 2007] and ShapeNetCore [Chang et al. 2015]. Experimental results show that our approach significantly outperforms state-of-the-art matrix based map synchronization techniques. In particular, on heterogeneous data sets such as ShapeNetCore, the performance gain of our approach is salient against existing approaches. Moreover, for the task of shape segmentation on ShapeNetCore, our approach achieves state-of-the-art performance, particular on categories that show large geometric and topological variabilities. For the task of clustering heterogeneous shape collections, our approach outperforms alternative approaches that are based on shape descriptors and popular shape distance metrics.

In summary, we present the following contributions in this paper:

- We introduce a novel tensor map synchronization approach for establishing high-quality correspondence maps across a heterogeneous shape collection;
- We introduce a novel data-driven shape segmentation approach that utilizes maps to explore shape variabilities for identifying meaningful shape parts and decompositions;
- We introduce a novel shape clustering approach that leverages map residuals to cluster heterogeneous shape collections;

## 2 RELATED WORKS

The problems considered in this paper are relevant to four sub-research areas, namely, map synchronization, tensor decomposition, data-driven shape segmentation, and shape clustering. Due to space constraints, we focus on the most relevant prior works.

*Map Synchronization.* Formally speaking, map synchronization concerns the task of optimizing maps among a shape collection jointly to improve the maps computed between each of the pairs of shapes. Existing works on map synchronization fall into two categories: combinatorial optimization based techniques and matrix optimization based techniques. Combinatorial optimization based techniques follow the general methodology of applying cycle-consistency criterion (namely, composition of maps along cycles of length three (3-cycles) shall equal to identity maps) to improve maps computed between each pair of shapes [Huang et al. 2006; Huber 2002; Nguyen et al. 2011; Zach et al. 2010]. A limitation of these approaches is that one has to sample many inconsistent cycles to enforce the cycle-consistency constraint. In contrast, matrix optimization based approaches utilize the equivalence between the cycle-consistency constraint and the positive semidefinite or low-rank structure of the matrix that stores pair-wise maps in blocks (c.f. [Huang and Guibas 2013]). This leads to simple and effective formulations of map synchronization as low-rank matrix recovery. From the optimization point of view, people have introduced convex optimization techniques [Chen et al. 2014; Huang and Guibas 2013; Leonardos et al. 2017; Wang and Singer 2013] and non-convex optimization techniques such as alternating minimization [Zhou et al. 2015], reweighted least squares [Chatterjee and Govindu 2013; Huang et al. 2017], reweighted factorization [Arrigoni et al. 2018], and spectral techniques [Pachauri et al. 2013; Shen et al. 2016]. However, existing matrix-based map synchronization techniques are rather insufficient for establishing high-quality maps across a heterogeneous shape collection. Our approach combines and then extends both types of approaches. Specifically, we enforce the consistency along 3-cycles to pre-filter incorrect maps. The results are then fed into a low-CP-rank tensor recovery framework that is generalized from low-rank matrix recovery techniques.

*Tensor Decomposition.* Tensor decomposition [Kolda and Bader 2009; Sidiropoulos et al. 2017] emerged as a powerful tool in revealing intrinsic and coherent structure of high dimensional data. It has been successfully applied to many data science fields, including signal processing [De Lathauwer et al. 2007; Muti and Bourennane 2005; Sidiropoulos et al. 2000], computer vision [Shashua and Hazan 2005; Vasilescu and Terzopoulos 2002; Vlasic et al. 2005], and data mining [Anandkumar et al. 2014; Sun et al. 2006, 2005]. Tensors provide a natural way to capture coherencies across all dimensions. Such dependencies are often lost in flattened matrix formulations, especially when some form of vectorization is deployed. Our tensor approach is similarly motivated, and we show how to design tensor representations and algorithms to recover low-rank tensors for map synchronization. Our approach is mostly relevant to [Sharan and Valiant 2017], which proposes an alternating optimization scheme for tensor decomposition that preserves orthogonality among factors. In this paper, we extend the approach to low-CP-rank tensor recovery with only partial observations.

*Data-driven Shape Segmentation.* Jointly segmenting a collection of similar shapes has been studied extensively in the literature [Hu et al. 2012; Huang et al. 2011, 2014; Sidi et al. 2011; van Kaick et al. 2013; Wang et al. 2012; Yumer and Kara 2012]. Most of these techniques enforce consistency of segmentations across similar shapes

to boost the performance of segmenting individual shapes. However, these techniques are mostly suitable for shape collections that show small geometrical and topological variabilities (e.g. organic shapes), and are not designed for large-scale man-made shape collections (e.g. ShapeNetCore) that show large structural variabilities. In contrast, our approach is based on analyzing geometrical and topological variabilities of 3D shapes in order to derive meaningful shape parts (e.g. the main difference between chairs and stools indicate that chair back is a meaningful part). [Huang et al. 2014] presents a first attempt in this direction, we introduce a novel approach based on analyzing null spaces of functional maps [Ovsjanikov et al. 2012].

*Shape Clustering.* Existing works on shape clustering fall into two major categories, namely descriptor based and graph-clustering based. Descriptor based approaches compute a descriptor for each shape so that similar shapes have similar descriptors. Shape clustering is then formulated as clustering data points in the descriptor space via single-linkage clustering [Gan et al. 2007], k-means [Ohbuchi et al. 2008], and sub-space clustering [Hu et al. 2012]. These techniques work well for classifying shapes into general categories but they perform poorly for fine-grained classification tasks where the shape differences are subtle. Graph-clustering builds a similarity graph that links similar shapes. Our approach falls into this category. However, we use the difference between each input map and corresponding recovered map to establish the similarity graph. Our approach does not require an external shape similarity metric and the underlying clusters are derived solely from input maps. Recently, [Bajaj et al. 2018] introduced an approach that performs simultaneous clustering and mapping using a matrix-based map representation. Our approach shares the similar idea of using maps to extract shape clusters. However, our approach explicitly utilizes the fact that intra-cluster maps are more accurate than inter-cluster maps, leading to additional performance gains.

### 3 PROBLEM STATEMENT AND APPROACH OVERVIEW

In this section, we present an overview of our tensor map synchronization approach (Section 3.1) as well as its applications in shape segmentation and shape clustering (Section 3.2).

#### 3.1 Overview of Tensor Map Synchronization

We first formally describe our tensor map synchronization problem. We then present an overview of our approach.

*Problem Statement.* We study map synchronization under partial similarity. Specifically, consider a collection of shapes  $\mathcal{S} = \{S_1, \dots, S_n\}$  that are partially similar to each other. Following the convention in [Huang et al. 2014], we assume that there exists an universal shape  $\bar{S}$ , and each shape  $S_i$  is a partial observation of  $\bar{S}$ . Without losing generality, we assume the input shapes and the universal shape are given by sets. In the following, we denote the sizes of  $S_i$  and  $\bar{S}$  as  $|S_i| = \bar{m}_i$  and  $|\bar{S}| = \bar{m}$ , respectively (note that  $m$  is reserved for the dimension of functional spaces, which will be used later). With  $Q_i^*$  we denote the underlying embedding map from  $S_i$  to  $\bar{S}$ . Note that these  $Q_i^*$  are unknowns and to be determined. The input to map synchronization is given by maps computed along an observation graph  $\mathcal{G} = (\mathcal{S}, \mathcal{E})$ , where every edge  $(i, j) \in \mathcal{E}$  is

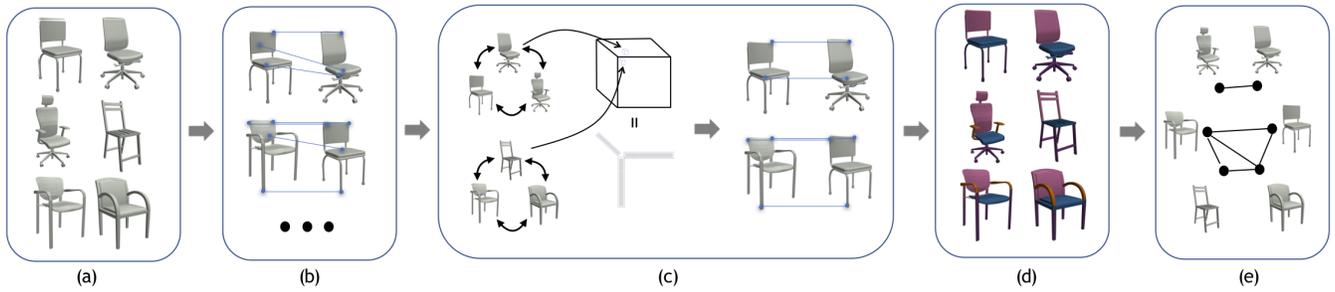


Fig. 2. **Approach overview.** (a) Input heterogeneous shape collection. (b) Initial pairwise (possibly noisy) shape correspondence maps computed using any off-the-shelf shape matching algorithm. Note that not all pairwise correspondences are required, furthermore these initial maps may be inaccurate and possibly link sub-regions that should not be matched (e.g. a tubular back frame support of one chair is linked to the tubular leg of another). (c) Our tensor map synchronization approach combines a pre-filtering operation by enforcing consistency along 3-cycles and a low-CP-rank tensor recovery procedure to generate all consistent correspondences. (d) Our resulting optimized and consistent maps can be directly used to perform accurate shape segmentation. (e) We use the difference between initial maps and recovered maps to perform weighted and consistent shape clustering (i.e. by exploiting that intra-cluster initial maps possess greater accuracy than inter-cluster initial maps, suitable weights are assigned).

associated with a partial map  $P_{ij}^{in}$  that links a subset of elements of  $S_i$  and a subset of elements of  $S_j$ . Each  $P_{ij}^{in}$  is pre-computed using an off-the-shelf algorithm and may contain incorrect correspondences. Following the convention, we represent the partial map  $P_{ij}^{in} \in \{0, 1\}^{\bar{m}_j \times \bar{m}_i}$  as a binary matrix of dimension  $\bar{m}_j \times \bar{m}_i$ <sup>1</sup>. Likewise, we represent the embedding map  $Q_i^* \in \{0, 1\}^{\bar{m} \times \bar{m}_i}$  as a matrix of dimension  $\bar{m} \times \bar{m}_i$ . Our goal is to recover the universal shape  $\bar{S}$  and the embedding maps  $Q_i^*$ ,  $1 \leq i \leq n$  from  $P_{ij}^{in}, \forall (i, j) \in \mathcal{E}$ . The final output consists of recovered maps  $P_{ij} = Q_j^T Q_i$  between all pairs of shapes that are induced from the embedding maps.

As we will discuss shortly, we employ the functional map representation [Ovsjanikov et al. 2012] for efficient coding and inference of shape maps across the input shape collection. Let  $\mathcal{F}(S_i)$  and  $\mathcal{F}(\bar{S})$  be the functional spaces associated with  $S_i$  and  $\bar{S}$ , respectively. In this setting, our goal is to compute  $Y_i : \mathcal{F}(S_i) \rightarrow \mathcal{F}(\bar{S}), 1 \leq i \leq n$ , each of which is a functional representation of the point-based embedding map  $Q_i, 1 \leq i \leq n$ . Note that similar to the point-based setting described above, we seek to recover the latent functional space  $\mathcal{F}(\bar{S})$  and the functional embedding maps  $Y_i$  together. In contrast,  $\mathcal{F}(S_i)$  are pre-computed.

*Approach Overview.* As illustrated in Figure 2, the central idea of our approach is to consider triplets of shapes. Specifically, we enforce the consistency of maps along each cycle of length 3. The resulting correspondences are regarded as partial and noisy observations of an underlying 3D tensor that admits a low-CP-rank decomposition. We show how to recover the underlying low-CP-rank tensor through continuous optimization under the L2-norm, which is followed by a rounding procedure. In particular, we leverage alternating minimization to recover the underlying low-CP-rank tensor effectively. Compared with matrix-based map synchronization techniques [Bajaj et al. 2018; Chen et al. 2014; Huang and Guibas 2013; Shen et al. 2016], our tensor-based map synchronization approach exhibits two

<sup>1</sup>This notation is consistent with the fact that if  $\mathbf{e}_p \in \mathbb{0}, \mathbf{1}^{\bar{m}_i}$  is the indicator vector of  $p$ -th element on  $S_i$ , then  $P_{ij}^{in} \mathbf{e}_p$  gives the indicator vector of its corresponding element on  $S_j$

advantages. First, the 3-cycle consistency filtering step can greatly reduce the noise among the input maps. Second, each slice of the low-CP-rank tensor corresponds to a data matrix for matrix-based map synchronization techniques. Our tensor-based synchronization has the advantage that it enforces consistency among these matrix-based map synchronizations. The technical details are explained in Section 4.1.

In Section 4.2, we show how to adapt our approach under the functional map representation [Huang et al. 2014; Ovsjanikov et al. 2012; Wang et al. 2013], which effectively addresses issues of generating consistent samples and scaling to large-scale datasets.

To further enhance the scalability of our approach to large-scale datasets, we introduce a transductive approach that first performs tensor map synchronization on a coreset of shapes sampled from the input shape collection, and then establishes maps between the remaining shapes and this coreset. This approach enables us to compute high-quality maps among many thousands of shapes. Technical details are explained in Section 4.3.

## 3.2 Overview of Applications

The output of our map synchronization algorithm enables many downstream applications, and we focus on two of them in this paper, namely, shape segmentation and shape clustering. In this section, we present an overview of these two applications. Section 7 and Section 6 elaborate some technical details.

*Application in Shape Segmentation.* We present a novel approach for shape segmentation that utilizes shape maps to explore shape differences. The key idea is that if there is structural variability between two shapes, then their difference naturally corresponds to a shape part or a collection of shape parts (e.g. the difference between a chair and a stool indicates that the chair back is a part). Under functional map representation, such shape differences can be easily encoded as null spaces of linear maps. We present a simple approach that aggregates null spaces between a source shape to other shapes into a part decomposition of the source shape.

*Application in Shape Clustering.* Our shape clustering approach is based on the observation that when input shape collection falls into multiple clusters, intra-cluster maps tend to be more accurate than inter-cluster maps. In other words, if we define the affinity score between two shapes based on the difference between each input map and the recovered map (i.e. a small difference between maps leads to a large score), then when recovered maps are accurate, such affinity scores provide informative cues for identifying underlying clusters, i.e. intra-cluster pairs usually have large affinity scores and inter-cluster pairs usually have small affinity scores. Specifically, we show that spectral clustering on affinity scores recovers the underlying clusters.

## 4 TENSOR MAP SYNCHRONIZATION

In this section, we introduce our tensor-based map synchronization approach in details. We begin with describing our approach under the point-based setting in Section 4.1. We then show how to modify the approach under the functional representation for scalable map synchronization in Section 4.2. Finally, we show how to modify our approach to perform map synchronization on very large shape collections in Section 4.3.

### 4.1 Tensor Point-Based Map Synchronization

The key idea of our approach is to formulate map synchronization as recovering a low-CP-rank tensor, and the measurements for recovery come from analyzing the consistency of maps along 3-cycles.

*3-Cycle Consistency Filtering.* We evaluate the consistency of initial maps along triplets of shapes. Specifically, for each triplet of shapes  $(S_i, S_j, S_k)$ , where  $(i, j), (j, k), (i, k) \in \mathcal{E}$ , we introduce a binary diagonal matrix

$$C_{ijk}^{in} := \text{diag}(c_{ijk}) \in \{0, 1\}^{\bar{m}_i \times \bar{m}_i}$$

where  $c_{ijk} \in \{0, 1\}^{\bar{m}_i}$  is a vector derived from  $p_{ij}^{in}$ ,  $p_{jk}^{in}$  and  $p_{ki}^{in}$ . Specifically, for the partial self-map  $p_{ki}^{in} p_{jk}^{in} p_{ij}^{in}$  on  $S_i$  that is induced from the 3-cycle  $i \rightarrow k \rightarrow j \rightarrow i$ , we set  $c_{ijk}^{in}(p) = 1$  if and only if  $p$  is mapped to  $p$  itself in  $p_{ki}^{in} p_{jk}^{in} p_{ij}^{in}$ . Otherwise, we set  $c_{ijk}^{in}(p) = 0$ . The resulting vector  $c_{ijk}^{in}$  essentially indicates which elements of  $S_i$  possess consistent correspondences along this 3-cycle.

Note that such consistency check can effectively remove incorrect correspondences, particularly on shape collections that exhibit large geometric variabilities. As shown in Figure 3, the fraction of incorrect correspondences in the filtered maps  $P_{ij}^{in} C_{ijk}^{in}$  is usually significantly reduced from that in  $P_{ij}^{in}$ , while the majority of the input correspondences are preserved in the filtered maps.

Moreover, compared with the methodology of computing fully consistent maps along 3-cycles [Nguyen et al. 2011], our approach enforces 3-cycle consistency at the correspondence level. This shows a clear advantage because our approach is still effective even when every input map contains some incorrect correspondences. On the other hand, we will later show how to aggregate such partially consistent correspondences along 3-cycles (some of which are still

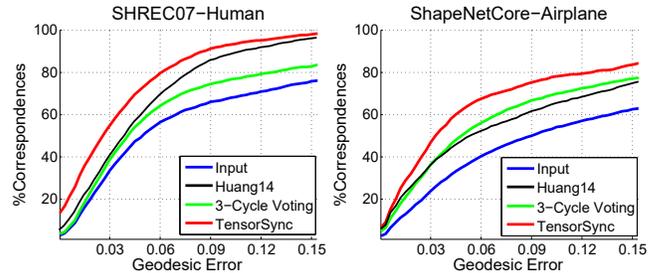


Fig. 3. Cumulative distribution functions of the initial maps  $\{P_{ij}^{in}\}$  and the filtered correspondences  $\{R_{ijk}^{in}\}$ . As empirical comparisons, we also show results of our tensor map synchronization approach and matrix-based technique described in [Huang et al. 2014]. Note that the low-CP-rank tensor recovery formulation further enhances map quality. (Left) SHREC07-Human [Giorgi et al. 2007] 85.2% input correspondences are preserved in the filtered maps. (Right) ShapeNetCore-Chair [Chang et al. 2015] 82.3% input correspondences are preserved in the filtered maps.

incorrect as shown in Figure 3) into consistent full maps across the entire shape collection.

*Low-CP-rank Tensor for Map Synchronization.* We recover the underlying maps  $P_{ij}$ ,  $1 \leq i, j \leq n$  by solving a low-CP-rank tensor recovery problem, where  $P_{ij}^{in} C_{ijk}^{in}$ ,  $1 \leq i, j, k \leq n$  are treated as noisy and incomplete measurements of this tensor. Consider a  $n \times n \times n$  block tensor  $R \in \mathbb{R}^{N \times N \times N}$ , where  $N = \sum_i \bar{m}_i$ . The  $(j, i, k)$ -th block  $R_{jik} \in \mathbb{R}^{\bar{m}_j \times \bar{m}_i \times 1}$  of  $R$ , which we treat it as a matrix of dimension  $\bar{m}_j \times \bar{m}_i$ , is given by

$$R_{jik} := \begin{cases} P_{ij}^{in} C_{ijk}^{in} & (i, j), (j, k), (i, k) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

To motivate our low-CP-rank recovery approach, let us gain some insights about the structure of  $R$  in the case where we have pairwise maps between all pairs of shapes, and where all the input pairwise maps  $P_{ij}^{in}$  are correct. Let matrix  $Q \in \mathbb{R}^{\bar{m} \times n}$  collect all the underlying maps  $Q_i$ ,  $1 \leq i \leq n$ , i.e.  $Q = (Q_1, \dots, Q_n) = (\mathbf{q}_1, \dots, \mathbf{q}_n)^T$ , where  $\mathbf{q}_i \in \{0, 1\}^{\bar{m}}$ ,  $1 \leq i \leq n$  indicates which shapes contain the  $i$ -th element of the universal shape as well as the corresponding indices. As each column of  $Q$  only contains one non-zero element, it is easy to see that  $\mathbf{q}_i$  are orthogonal with each other. Let  $\bar{Q} = (Q_1 \mathbf{1}, \dots, Q_n \mathbf{1}) \in \mathbb{R}^{\bar{m} \times n}$ . We again write down the rows of  $\bar{Q} = (\bar{\mathbf{q}}_1, \dots, \bar{\mathbf{q}}_m)^T$ . The following proposition characterizes the structure of  $R$  when the measurements are complete and correct:

**PROPOSITION 4.1.** *Suppose we have input maps between all pairs of shapes and these input maps are correct. Then the tensor  $R$  introduced in (1) admits the following CP decomposition [Kolda and Bader 2009]:*

$$R = \sum_{l=1}^{\bar{m}} \mathbf{q}_l \otimes \mathbf{q}_l \otimes \bar{\mathbf{q}}_l, \quad (2)$$

where  $\otimes$  denotes the tensor operator.

**Proof:** The proof is straight-forward because an element  $R_{jik}(p', p) = 1$  if and only if (1) the  $p$ -th element of  $S_i$  and the  $p'$ -th element of  $S_j$

correspond to the same element of the universal shape, and (2)  $S_k$  also contains this element.  $\square$

*Map Synchronization via Alternating Minimization.* Generally speaking, when a tensor  $R$  admits a CP decomposition of the form  $R = \sum_{l=1}^{\bar{m}} \mathbf{a}_l \otimes \mathbf{b}_l \otimes \mathbf{c}_l$  for some underlying vectors  $\mathbf{a}_l, \mathbf{b}_l, \mathbf{c}_l, 1 \leq l \leq \bar{m}$ , a common approach for recovering  $\mathbf{a}_l, \mathbf{b}_l$  and  $\mathbf{c}_l$  from  $R$  is to solve a L2-minimization problem [Sharan and Valiant 2017; Zhang et al. 2018], i.e.,

$$\min_{\{\mathbf{a}_l, \mathbf{b}_l, \mathbf{c}_l\}} \|R - \sum_{l=1}^{\bar{m}} \mathbf{a}_l \otimes \mathbf{b}_l \otimes \mathbf{c}_l\|_{\mathcal{F}}^2 \quad (3)$$

where  $\|\cdot\|_{\mathcal{F}}$  is the tensor Frobenius norm [Golub and Van Loan 1996]. Our setting differs from this standard setting in three ways. First, we have additional constraints among  $\mathbf{a}_l, \mathbf{b}_l, \mathbf{c}_l, 1 \leq l \leq \bar{m}$ , e.g.  $\mathbf{a}_l = \mathbf{b}_l$ , and  $\mathbf{a}_l$  are orthogonal with each other. Second, the elements in  $R$  are noisy. Third, we only have partial measurements of  $R$ .

To adapt (3) in our setting, we consider the following modifications. First, since the number of non-zero elements of  $R$  is generally significantly larger than the number of variables in  $\mathbf{a}_l, \mathbf{b}_l, \mathbf{c}_l$ , we simply relax the dependencies among  $\mathbf{a}_l, \mathbf{b}_l, \mathbf{c}_l$ . We will enforce these constraints after recovering these latent factors. Such a relaxation aligns with the lifting methodology in compressed sensing and low-rank matrix recovery [Candes et al. 2011, 2012]. Moreover, to address noisy measurements, we propose to still use the L2-norm. In contrast to the popular choice of using the L1-norm for recovery [Candes et al. 2011], we found that using the L2-norm works quite well in our setting. This can be understood from the facts that (1) noisy pairwise maps, no matter correctness, have bounded norms, i.e. the elements of  $R$  are either 1 or 0; 2) the optimal solution to L2-norm minimization is highly relevant to leading eigenvector computation, which has proven to be effective for matrix-based map synchronization techniques (c.f. [Pachauri et al. 2013; Shen et al. 2016]); and 3) one can still perform rounding for recovery. Finally, instead of using tensor Frobenius norm directly, we enforce the partial observations specified by  $C_{ijk}^{in}, 1 \leq i, j, k \leq n$ , which encode results of the 3-cycle consistency filtering step. Combining these modifications together, we arrive at the following optimization problem for tensor-based map synchronization:

$$\begin{aligned} & \underset{A, B, C}{\text{minimize}} && \sum_{1 \leq i, j, k \leq n} \|\mathcal{B}_{jik}(R - \sum_{l=1}^{\bar{m}} \mathbf{a}_l \otimes \mathbf{b}_l \otimes \mathbf{c}_l) \cdot C_{ijk}^{in}\|_{\mathcal{F}}^2 && (4) \\ & \text{subject to} && A = (\mathbf{a}_1, \dots, \mathbf{a}_{\bar{m}}), \\ & && B = (\mathbf{b}_1, \dots, \mathbf{b}_{\bar{m}}), \\ & && C = (\mathbf{c}_1, \dots, \mathbf{c}_{\bar{m}}). && (5) \end{aligned}$$

where for every tensor  $R' \in \mathbb{R}^{N \times N \times N}$  that admits the same block structure as  $R$ ,  $\mathcal{B}_{jik}(R') \in \mathbb{R}^{\bar{m}_j \times \bar{m}_i}$  extracts the  $(j, i, k)$ -th block of  $R'$  as a matrix of dimension  $\bar{m}_j \times \bar{m}_i$ . Note that the value of  $\bar{m}$  is automatically inferred from the input and will be discussed shortly. Motivated from the success of using alternating minimization for low-CP-rank tensor factorization [Kolda and Bader 2009; Sharan and Valiant 2017], we solve (5) via alternating minimization.

The initial values of  $A$  and  $B$  are given by spectral decomposition of a  $n \times n$  block data matrix  $P \in \mathbb{R}^{N \times N}$  that encodes  $P_{ij}^{in}, (i, j) \in \mathcal{E}$  in blocks:

$$P_{ji} := \begin{cases} P_{ij}^{in} & (i, j) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

Let  $(\lambda_l(P), \mathbf{u}_l(P)), 1 \leq l \leq N$  be the eigenvalues and corresponding eigenvectors of  $P$  sorted in increasing order. Following [Chen et al. 2014; Keshavan et al. 2010], we set  $\bar{m}$  so that it has the largest gap in  $\lambda_{\bar{m}}(P) - \lambda_{\bar{m}+1}(P)$ . Accordingly, we let  $A = B = (\mathbf{u}_1(P), \dots, \mathbf{u}_{\bar{m}}(P)) \cdot \text{diag}(\lambda_1^{\frac{1}{2}}(P), \dots, \lambda_{\bar{m}}^{\frac{1}{2}}(P))$ , or in other words,  $AB^T$  gives the low-rank approximation of  $P$  with respect to the L2-norm. In addition, we set  $\mathbf{c}_l = \mathbf{1}, 1 \leq l \leq \bar{m}$ , which are optimal when the input shapes are fully similar with each other.

Given the initial values of  $A, B$  and  $C$ , we alternate between optimizing one of them to minimize (4) while fixing the remaining two. In this case, (4) becomes quadratic in the active variables whose optimal values can be obtained by solving a linear system. We apply the same procedure for  $B$  and  $C$  as well. The details are left to Appendix A.

Since we have relaxed the constraints among  $A$  and  $B$  when performing low-CP-rank tensor recovery (See (2)), we have to enforce these constraints when computing  $Q$ . After obtaining the optimal solutions of  $A$  and  $B$ , we compute  $Q$  by optimizing

$$\min_Q \|AB^T - Q^T Q\|_{\mathcal{F}}^2 \quad (7)$$

Same as spectral map synchronization [Pachauri et al. 2013; Shen et al. 2016], the optimal  $Q$  is given by the top  $\bar{m}$  eigenvectors of  $\frac{AB^T + BA^T}{2}$ . We then apply the procedure described in [Chen et al. 2014] to obtain an approximate binary solution to (7).

The computational cost of this alternating minimization procedure is  $O(N^3 \bar{m} n_{al})$ , where  $n_{al}$  is the total number of alternating minimizations (We set  $n_{al} = 100$  in this paper). Please refer to Appendix A for a detailed analysis.

*Comparison to Matrix-Based Synchronization Techniques.* The central idea of matrix-based synchronization approaches is to recover a low-rank matrix from noisy measurements of its elements that are given by the input maps, e.g. matrix  $P$  in (6) (c.f. [Bajaj et al. 2018; Chen et al. 2014; Huang and Guibas 2013; Shen et al. 2016]). The advantage of our approach is two-fold. First, our 3-cycle consistency filtering scheme can effectively improve the signal-to-noise ratio among the remaining correspondences. Suppose we divide the input maps into the cluster of correct maps and the cluster of incorrect maps. For simplicity, we assume maps across different clusters are inconsistent. Let  $p_{cor}$  and  $p_{incor}$  be the average numbers of input maps that are associated with each shape in the correct cluster and the incorrect cluster, respectively. Then before 3-cycle consistency filtering, the signal-to-noise ratio is  $\frac{p_{cor}}{p_{incor}}$ . After 3-cycle consistency filtering, the signal-to-noise ratio becomes at least  $\frac{p_{cor}^2}{p_{incor}^2}$  (the actual ratio depends on how consistent the maps in the incorrect cluster are). In addition, although 3-cycle consistency filtering reduces the total number of input correspondences, the number of remaining correspondences is still significantly higher than what is required for recovery (which is linear in the total number of sample points).

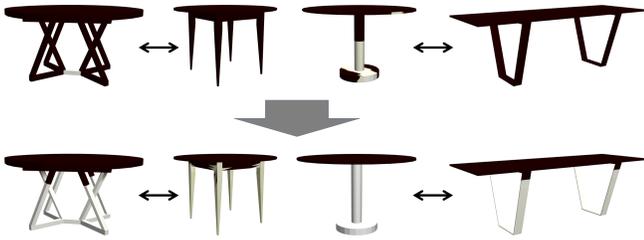


Fig. 4. Our tensor map approach recovers high-quality consistent maps across a heterogeneous shape collection with large structural variability. (Top-row) Noisy initial maps estimated between pairs of shapes in isolation. We show two pairs from the Table category in ShapeNetCore [Chang et al. 2015]. Corresponding sub-regions are colored same. The multi-color structural sub-regions show mismatch correspondence (i.e. leg and base of the tables). (Bottom) Jointly optimized shape maps between the same pairs, showing more accurate correspondence maps (i.e. table tops are matched and even leg supports are matched consistently). Note that our joint tensor map computation utilizes all the shapes in the collection to achieve this greater accuracy.

Hence the improved signal-to-noise ratio provides a foundation for low-rank based techniques to recover the underlying ground-truth maps (c.f. [Chen et al. 2014; Huang and Guibas 2013; Huang et al. 2017; Shen et al. 2016]).

We can understand the additional advantage of our tensor formulation by rewriting the objective function in (5) as

$$f(A, B, C) = \sum_{k=1}^n g_k(A, B, \mathbf{c}_k^{\text{row}}),$$

$$g_k(A, B, \mathbf{c}_k^{\text{row}}) := \sum_{i=1}^n \sum_{j=1}^n \|\mathcal{B}_{jik}(R) - \sum_{l=1}^{\bar{m}} c_{kl} \mathbf{a}_l \mathbf{b}_l^T\|_{\mathcal{F}}^2$$

where  $\mathbf{c}_k^{\text{row}} = (c_{k1}, \dots, c_{k\bar{m}})$  denotes the  $k$ -th row of  $C$ . Intuitively, each  $g_k(A, B, \mathbf{c}_k^{\text{row}})$  seeks to perform matrix-based map synchronization from each slice of  $R$ . The advantage of (5) comes from the fact that instead of performing matrix-based map synchronizations independently, it enforces the consistency among all slices (i.e.  $A$  and  $B$  are shared among different slices) and is more resilient to noisy measurements in  $R$ . In other words, while matrix-based recovery techniques can leverage the uncorrelated random noise within each slice of  $R$  (c.f. [Chen et al. 2014; Huang and Guibas 2013; Huang et al. 2017; Shen et al. 2016]), our approach utilizes the uncorrelated random noise across the entire  $R$  for robust recovery.

## 4.2 Tensor Functional Map Synchronization

A major limitation of the approach described in the preceding section is that typically one can only place dozens of sample points per shape, leading to a scalability issue on large-scale datasets. To address this issue, we propose to use a functional map representation [Ovsjanikov et al. 2012], which has proven to be quite effective for map synchronization [Huang et al. 2014; Wang et al. 2013]. In the following paragraph, we give a brief introduction to the functional map representation. Please refer to [Huang et al. 2014; Ovsjanikov et al. 2012; Wang et al. 2013] and the references therein for more details.

---

### ALGORITHM 1: High level algorithm flow for tensor-based map synchronization.

---

- input:** Pre-computed functional space  $\mathcal{F}(S_i)$  for each shape  $S_i$ . Initial maps  $P_{ij}^{\text{in}}$  associated with an observation graph  $\mathcal{G} = (\mathcal{S}, \mathcal{E})$ .
- output:** A latent function space  $\mathcal{F}(\bar{\mathcal{S}})$  and embedding maps  $Y_i : \mathcal{F}(S_i) \rightarrow \mathcal{F}(\bar{\mathcal{S}})$ ,  $1 \leq i \leq n$
- 1: Convert  $P_{ij}^{\text{in}}$  into its corresponding functional map  $X_{ij}^{\text{in}}$ .
  - 2: Perform 3-cycle consistency filtering to obtain a mask  $M_{ijk}^{\text{in}}$  for each 3-cycle.
  - 3: Perform low-CP-rank tensor recovery from  $X_{ij}^{\text{in}} M_{ijk}^{\text{in}}$  to obtain latent variables  $A, B$  and  $C$  via (9).
  - 4: Recover  $Y_i$  from latent variables  $A$  and  $B$  via (10).
- 

Functional maps provide effective low-dimensional encodings of maps between pairs of shapes. This is done by associating each shape  $S_i$  with a low-dimensional linear functional space  $\mathcal{F}(S_i)$  of dimension  $m$  ( $m = 30$  in this paper). The basis of  $\mathcal{F}(S_i)$  is usually given by the leading  $k$  eigenvectors of the Laplacian matrices on  $S_i$  [Ovsjanikov et al. 2012; Wang et al. 2013].  $\mathcal{F}(S_i)$  provides a platform to effectively approximate indicators of primal elements such as points, feature points and segments (c.f. [Ovsjanikov et al. 2012; Wang et al. 2013]). This representation allows us to represent the relation between shape  $S_i$  and shape  $S_j$  by a linear map  $X_{ij} \in \mathbb{R}^{m \times m} : \mathcal{F}(S_i) \rightarrow \mathcal{F}(S_j)$  between these two functional spaces. We can convert the initial map  $P_{ij}^{\text{in}}$  into a functional map  $X_{ij}^{\text{in}}$  by solving a linear system (c.f. [Ovsjanikov et al. 2012]). In the other direction, we can convert a linear map into a point-based map through nearest neighbor search in the embedding space [Ovsjanikov et al. 2012]. Another key advantage of the functional map representation is to address inconsistent sampling, which is an issue under the point-based setting described in the preceding section.

A similarity between point-based maps and functional maps is that both of them can be represented as matrices. In fact, point-based maps can be considered special functional maps, under delta functional basis (c.f. [Ovsjanikov et al. 2012]). It turns out we can extend our tensor-based map synchronization approach under the point-based setting for functional maps by following three simple modifications (see Algorithm 1 for the overall flow). First, we modify the 3-cycle consistency filtering step to accommodate approximate cycle-consistency. Specifically, for each correspondence  $(p, q)$  induced from partial self-map  $P_{ki}^{\text{in}} P_{jk}^{\text{in}} P_{ij}^{\text{in}}$ , we set  $\mathbf{c}_{ijk}^{\text{in}}(p) = 1$  if and only if

$$d_{S_i}(p, q) \leq \epsilon \cdot \text{diam}(S_i),$$

where  $d_{S_i}(\cdot, \cdot)$  is the distance metric on  $S_i$  (i.e. geodesic distance for organic shapes and Euclidean distance for man-made shapes).  $\text{diam}(S_i)$  is the diameter of  $S_i$  with respect to  $d_{S_i}$ . We use  $\epsilon = 0.05$  for all of our experiments. As there are many vertices on a given 3D model and the initial maps are usually inaccurate, allowing approximate cycle-consistency brings more signals for map computation. On the other hand, the continuity of basis functions can indeed fuse approximately consistent correspondences.

In the second modification, we convert each initial map  $P_{ij}^{in}$  and the mask  $C_{ijk}^{in}$  into their functional forms as

$$X_{ij}^{in} = F_j^T P_{ij}^{in} F_i, \quad M_{ijk}^{in} = F_i^T C_{ijk}^{in} F_i,$$

where  $F_i \in \mathbb{R}^{\bar{m}_i \times m}$  is the matrix that stores the functional basis of  $\mathcal{F}(S_i)$  in its columns. Consider a  $n \times n \times n$  block tensor  $Z \in \mathbb{R}^{(nm) \times (nm) \times n}$  whose  $(j, i, k)$ -th block, which we treat it as a matrix of dimension  $m \times m$ , is given by

$$Z_{jik} := \begin{cases} X_{ij}^{in} M_{ijk}^{in} & (i, j), (j, k), (i, k) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

We apply the same procedure to recover the functional representations  $Y_i \in \mathbb{R}^{m \times m}$ ,  $1 \leq i \leq n$  of the latent maps  $Q_i$ ,  $1 \leq i \leq n$ . We first solve

$$\min_{A, B, C} \sum_{1 \leq i, j, k \leq n} \|\mathcal{B}_{jik}(Z - \sum_{l=1}^{\bar{m}} a_l \otimes b_l \otimes c_l) \cdot M_{ijk}^{in}\|_{\mathcal{F}}^2 \quad (9)$$

We then solve the following optimization problem to obtain the latent functional maps  $Y = (Y_1, \dots, Y_n)$ :

$$\min_Y \|AB^T - Y^T Y\|_{\mathcal{F}}^2 \quad (10)$$

In the same spirit as the projection operator under the point-based setting, we convert the induced functional map  $X_{ij}^r := Y_j^T Y_i$  from  $S_i$  to  $S_j$  into a point-based map  $P_{ij}^*$  following [Rodolà et al. 2017]. This operation can be considered as a counterpart of the projection operation in the point-based setting (c.f. [Chen et al. 2014]). Finally, we convert  $P_{ij}^*$  into the final functional map  $X_{ij}^*$  using [Rodolà et al. 2017]. As shown in Figure 4, our approach can identify accurate correspondences under large structural variability. In addition, our approach can also remove extraneous correspondences in the initial maps.

### 4.3 Transductive Map Synchronization

In this section, we present an efficient map synchronization approach for very large-shape collections. Instead of applying our tensor-based map synchronization approach on the entire input shape collection  $\mathcal{S}$ , we decompose  $\mathcal{S} = \mathcal{S}_c \cup \mathcal{S}_r$  into a coreset  $\mathcal{S}_c$  and a remaining set  $\mathcal{S}_r$ . In our experiments, we compute  $\mathcal{S}_c$  by applying farthest-point-sampling ([Eldar et al. 1997]) on  $\mathcal{S}$  with respect to a shape descriptor (i.e., GPS[Rustamov 2007] for organic shapes and D2[Osada et al. 2002] for man-made shapes). We then apply our tensor map synchronization approach on  $\mathcal{S}_c$ , obtaining a latent shape space  $\bar{\mathcal{S}}$  and for each shape  $S_i \in \mathcal{S}_c$  an embedding map  $Y_i : \mathcal{F}(S_i) \rightarrow \mathcal{F}(\bar{\mathcal{S}})$ . We then fix  $Y_i, i \in \mathcal{S}_c$  and compute for each shape  $S_i \in \mathcal{S}_r$  its embedding map  $Y_i : \mathcal{F}(S_i) \rightarrow \mathcal{F}(\bar{\mathcal{S}})$ .

Specifically, for each shape  $S_i \in \mathcal{S}_r$ , let  $\mathcal{N}_i \subset \mathcal{S}_c$  be a random set of core shapes to which we will link  $S_i$ . When  $\mathcal{S}_c$  is small, we simply treat  $\mathcal{N}_i = \mathcal{S}_c$ . When  $\mathcal{S}_c$  is large, we compute  $\mathcal{N}_i$  as a random subset of  $\mathcal{S}_c$ . Let  $X_{ij}^{in}$  be the initial functional map from  $S_i$  to  $S_j$ , where  $S_j \in \mathcal{N}_i$ . We solve the following optimization problem to obtain the embedding map  $Y_i, \forall S_i \in \mathcal{S}_r$ :

$$\min_{Y_i} \sum_{S_j \in \mathcal{N}_i} \|Y_j^T Y_i - X_{ij}^{in}\|_{\mathcal{F}}^2 \quad (11)$$

The optimal solution to (11) is given by a closed-form expression

$$Y_i = \left( \sum_{S_j \in \mathcal{N}_i} Y_j Y_j^T \right)^\dagger \left( \sum_{S_j \in \mathcal{N}_i} Y_j X_{ij}^{in} \right).$$

We apply the same procedure as Section 4.2 to round the induced functional maps  $X_{ij}^* = Y_j^T Y_i$  into point-based maps and then apply (11) to solve the embedding map again, which gives the final embedding map  $Y_i$  for shape  $S_i$ .

## 5 EXPERIMENTAL EVALUATION

In this section, we present an experimental evaluation of our tensor map synchronization approach.

### 5.1 Experimental Setup

*Dataset.* We consider two benchmark datasets that are widely used for evaluating shape maps. The first benchmark dataset is SHREC07 [Giorgi et al. 2007]. SHREC07 consists of 20 categories, where each category has 20 shapes. Same as [Kim et al. 2011], we pick 11 categories (Human, Glasses, Airplane, Ant, Teddy, Hand, Plier, Fish, Bird, Armadillo and Fourleg) that are suitable for inter-shape matching. For each category, we compute the initial maps between all pairs of shapes using blended intrinsic map (or BIM), which is a state-of-the-art method for inter-shape matching. These maps are converted into functional maps using the technique described in [Ovsjanikov et al. 2012].

The second benchmark dataset is ShapeNetCore [Chang et al. 2015], which contains 50K man-made shapes in 55 categories. We pick 9 popular categories (Aeroplane, Bicycle, Boat, Bus, Car, Chair, Motorbike, Sofa, Train). These categories exhibit larger variabilities in geometry and topology than categories in SHREC07. For categories with more than 400 shapes, we perform farthest point sampling using the D2 descriptor [Osada et al. 2002] to sample 400 shapes (which form the coreset) for experimental evaluation. We first use our tensor map synchronization approach to compute consistent maps within each coreset. Since it is still too costly to compute consistent maps within the coresets, we randomly connect each shape with 64 other shapes for map computation. Note that we do not connect each shape with adjacent shapes (e.g. with similar shape descriptors), as we would like to establish maps between diverse shapes to explore shape variability. We will utilize such shape variability to derive shape parts. Given the consistent maps computed within each coreset, we then use the approach described in 4.3 to establish consistent maps for the remaining shapes. Regarding the initial maps, we employ the FFD alignment procedure described in [Huang et al. 2014]. We then convert these correspondences into functional maps using [Ovsjanikov et al. 2012].

*Baseline Approaches.* For experimental evaluation, we compare our approach against four state-of-the-art map synchronization approaches. The first baseline approach is cycle-voting [Nguyen et al. 2011]. Our approach differs from cycle-voting in terms of the tensor formulation for map recovery and the fact that we leverage partially consistent maps. The second baseline is FuncSync [Huang et al. 2014], which is a state-of-the-art map synchronization approach that leverages low-rank matrix recovery under the functional map representation. The third baseline is [Cosmo et al. 2017], which is

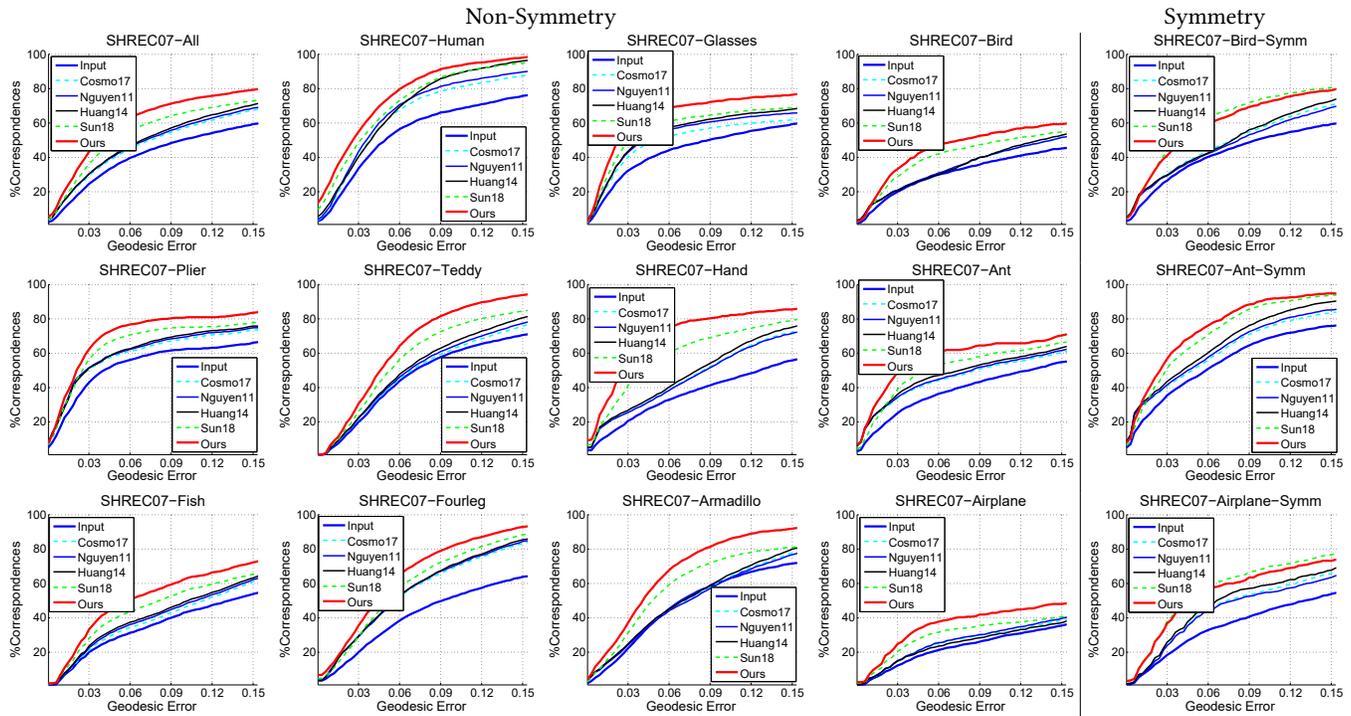


Fig. 5. **SHREC07-Evaluation.** Quantitative evaluation of our tensor map approach and prior baseline approaches on 11 categories of the SHREC07 dataset [Giorgi et al. 2007]. We compare against four baseline approaches: Cosmo17 [Cosmo et al. 2017], Nguyen11 [Nguyen et al. 2011], Huang14 [Huang et al. 2014], and Sun18 [Sun et al. 2018]. We show both (Left) results without factoring out the underlying symmetry and (Right) results after factoring out the underlying symmetry. The correspondence accuracy charts are best visualized in color and digitally zoomed in.

a state-of-the-art approach under the point-based representation. We use 128 sample points across all experiments. The last baseline is [Sun et al. 2018], which leverages the Kronecker product operator to synchronize maps among symmetric objects. As a by-product, it also promotes consistent correspondence pairs across the input shapes, which ultimately boost the quality of synchronized shape maps.

*Evaluation Protocol.* For both SHREC07 and ShapeNetCore, we evaluate the geodesic distance between each predicted corresponding point and its corresponding annotated feature point. This geodesic distance is normalized by the diameter of the target shape (in geodesic distance as well). The same as [Cosmo et al. 2017; Huang et al. 2014; Kim et al. 2011], we report the percentage of correspondences whose errors fall within a varying threshold, i.e. the cumulative probability.

## 5.2 Analysis of Results

*Map Synchronization Quality.* As shown in Figure 5 and Figure 6, our approach can significantly improve the quality of maps on SHREC07 and ShapeNetCore. The improvements are roughly consistent when varying the threshold for defining the cumulative probability. So to simplify the discussion, we evaluate the cumulative probability by setting  $\epsilon_{\min} = 0.1$ . Based on the level of improvements, we can roughly classify the categories in SHREC07 and ShapeNetCore

into three groups. The first group collects categories where our approach shows the salient improvements and the recovered maps are mostly correct (the cumulative probability of recovered maps is close to 100% for the selected range of errors). These categories include Human, Armadillo, Teddy and Fourleg from SHREC07, and Car, Chair, Bus and Sofa from ShapeNetCore. The cumulative probability of recovered maps/corresponding absolute improvement on these categories are 90%/23% (Human), 83%/20% (Armadillo), 82%/20% (Teddy), 82%/31% (Fourleg), 88%/28% (Car), 85%/23% (Chair), 83%/20% (Bus), and 80%/20% (Sofa), respectively. The second group collects categories where our approach yields salient improvements while there are some incorrect maps in the recovered set (the cumulative probability of recovered maps is less than 100%). These categories include Glasses, Ant, and Fish from SHREC07, and Aeroplane, Train, Bicycle, and Motorbike from ShapeNetCore. The cumulative probability of recovered maps/corresponding absolute improvement on these categories are 70%/20% (Glasses), 65%/18% (Ant), 62%/21% (Fish), 75%/22% (Aeroplane), 70%/19% (Train), 65%/25% (Bicycle), 62%/20% (Motorbike), respectively. The third group collects categories where our approach only exhibits modest improvements. These categories are Bird and Airplane from SHREC07, and Boat from ShapeNetCore. The cumulative probability of recovered maps/corresponding absolute improvement on these categories are 53%/15% (Bird), 42%/14% (Airplane), 55%/17% (Boat), respectively. We can observe two patterns from these categories. First, the cumulative probabilities of

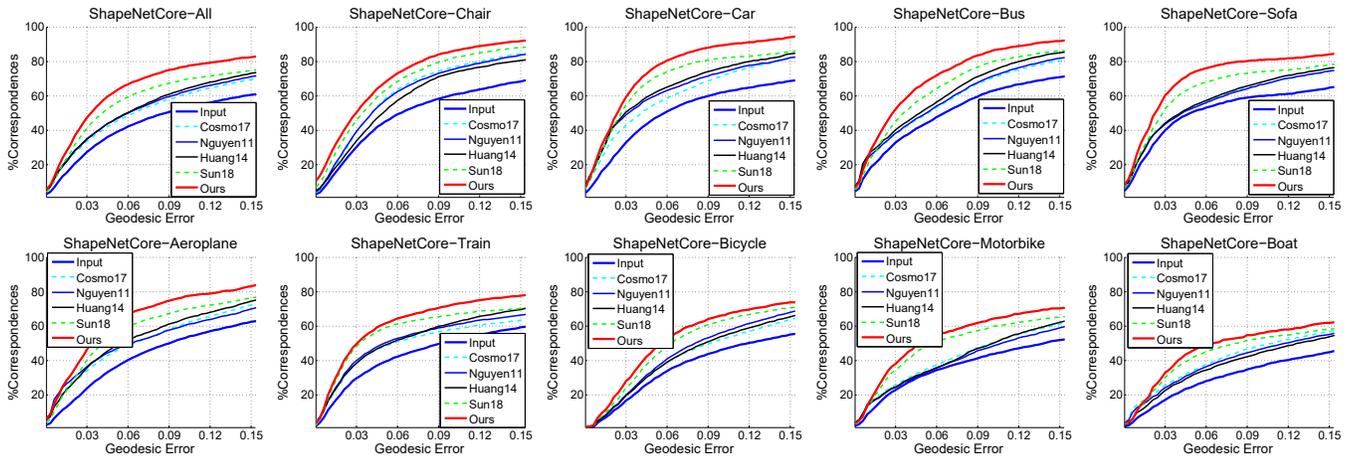


Fig. 6. **ShapeNetCore-Evaluation.** Quantitative evaluation of the proposed approach and baseline approaches on 9 categories of the ShapeNetCore dataset [Chang et al. 2015]. We compare our approach against four baseline approaches: Cosmo17 [Cosmo et al. 2017], Nguyen11 [Nguyen et al. 2011], Huang14 [Huang et al. 2014], and Sun18 [Sun et al. 2018].

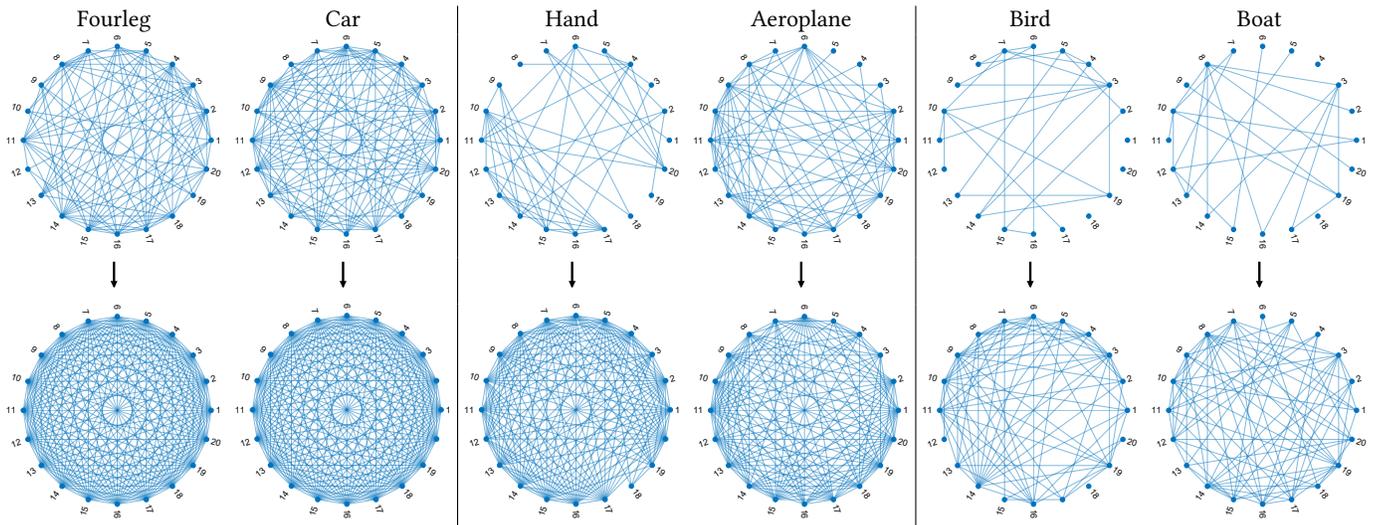


Fig. 7. **Map graph visualization.** We show the sub-graph of correct maps among the input maps and the recovered maps. We say a map is correct if the mean error among annotated features is below 0.1. For categories from ShapeNetCore (e.g. Car, Aeroplane and Boat), we draw a vertex sub-graph among 20 randomly picked shapes to make the visualization uncluttered. (Top) Sub-graph of correct input maps. (Bottom) Sub-graph of correctly recovered maps.

recovered maps are correlated with the quality of initial maps. This is expected, as our approach requires that a good portion of the input maps are correct in order to recover the underlying ground-truth maps. Second, the improvements are also related to whether the shapes in a particular category is symmetric or not. On those categories (e.g. Bird, Ant, Airplane), the absolute improvements become salient after factoring out the underlying symmetries (c.f. [Kim et al. 2011]) (See Figure 5(Right) and Figure 6(Right)). In particular, on Airplane and Bird, the absolute improvements become 24%(Airplane) and 20%(Bird) after factoring out the underlying symmetry.

To further understand the performance of our approach across these groups, we plot map quality measured using ground-truth

annotations for both the input maps and the recovered maps. To simplify the analysis, we say a map is good if the average geodesic error over the annotated feature points is below 0.1. In addition, we pick one exemplar category per dataset from each group (see Figure 7) for additional analysis. For Fourleg in the first group, its sub-graph of correct maps already forms a strongly connected graph, which provides a strong foundation for suppressing the noise in the input maps. As a result, the graph of correct recovered maps becomes a clique. The correct initial maps of Car also forms a strongly connected graph. Likewise, the graph of correct recovered maps is a clique as well.

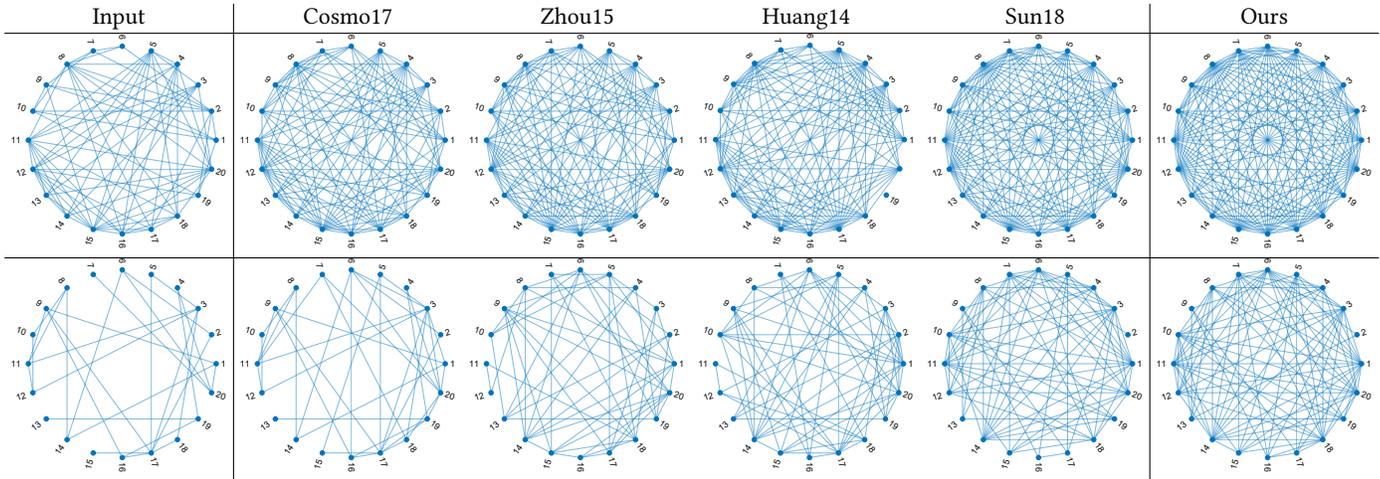


Fig. 8. **Map graph visualization.** We show the sub-graph (top) of only consistent (correct) maps amongst all the pairwise input maps and also the recovered correct maps (below). We say a map is correct if the mean error among annotated features is below 0.1. For categories from ShapeNetCore, we draw a vertex sub-graph among 20 randomly picked shapes to make the visualization un-cluttered. Note that each row shows the results of using different methods on the same category. (Top) Sub-graphs of correct maps when applying different approaches on ShapeNetCore-Motorbike. (Bottom) Sub-graph of correct maps when applying different approaches on SHREC07-Ant.

For Hand in the second group, its sub-graph of correct maps becomes less strongly connected. In fact, we can view this sub-graph as a collection of two densely connected components while the edges between these two connected components are significantly sparser. In this regime, our approach can still recover most of the underlying maps. This again shows the power of the tensor formulation, which is able to separate the signal from the noise despite the fact that a significant portion of the maps between these components are incorrect. The sub-graph of Aeroplane shows a similar behavior, and our approach can still recover most of the underlying maps.

Finally, for Bird in the third category, the graph of correct input maps consists of mostly disconnected components. In this case, many recovered maps between these components are still incorrect. However, our approach can still nicely recover the underlying ground-truth maps within each component and some maps between the components. Note that we utilize a hard threshold to determine a map is correct or not. It follows that some correct recovered maps are attributed to the fact that we synchronize a few maps that are close to be correct.

**Baseline Comparison.** Our approach outperforms baseline approaches across all categories. On categories in the first group, our approach yields slightly better results than that of [Sun et al. 2018] and [Huang et al. 2014]. This is due to the fact that the fraction of correct initial maps is significant, and matrix based map synchronization techniques are already delivering good results. Still, our approach produces better results, particularly in the regime when the cut-off threshold is small. We can also observe the advantage of our approach from the sub-graph of correct initial maps and the sub-graph of correctly recovered maps on Motorbike (see Figure 8(Top)).

On categories in the second group, our approach outperforms baseline approaches significantly. The relative performance gains

on Glasses, Ant, Fish, Aeroplane, Train, Bicycle, Motorbike are 9%, 5%, 7%, 10%, 5%, 2%, and 3%, respectively. This shows the clear advantage of the tensor formulation for separating signals from noise in the regime of relatively low signal-to-noise ratio. We can also observe the advantage of our approach in Figure 8(Bottom). Our approach can nicely recover most of the underlying ground-truth maps between shapes in different clusters. In contrast, all baseline approaches fail to recover many such inter-component maps. On the other hand, all approaches can nicely recover the underlying ground-truth maps within each component.

On categories in the third group, our approach still outperforms baseline approaches by a considerable margin. The relative performance gains on Bird, Airplane, and Boat are 7%, 8%, and 2%, respectively.

**Runtime Analysis.** Our approach is fairly efficient. The computational cost on ShapeNetCore with 400 shapes and 30 basis functions is 3 hours 40 minutes on a machine with 8-core 3.4GHZ CPU and 128G main memory.

### 5.3 Ablation Study

We proceed to compare our approach with two alternative approaches. In the first alternative approach, we employ [Huang et al. 2014] but replace the input map of each block with  $X_{ij}^c = \sum_k X_{ijk}$ .

Note that this is a strong baseline as  $X_{ij}^c$  is derived from enforcing the cycle-consistency constraint among triplets of objects. In the second alternative approach, we use a sequential approach to perform the low-rank tensor decomposition. Specifically, we first reshape  $Z$  into a matrix of dimension  $\text{vec}(Z) \in n \times (nm)^2$  (i.e. along the dimension of  $c$ ) and perform SVD to derive the rank-1 approximation  $c \cdot f^T$  of  $Z$  (i.e. with respect to the L2-norm), where  $c \in \mathbb{R}^n$  and  $f \in \mathbb{R}^{n^2 m^2}$ . We then reshape  $f$  into a matrix  $\text{mat}(f) \in \mathbb{R}^{nm \times nm}$ .

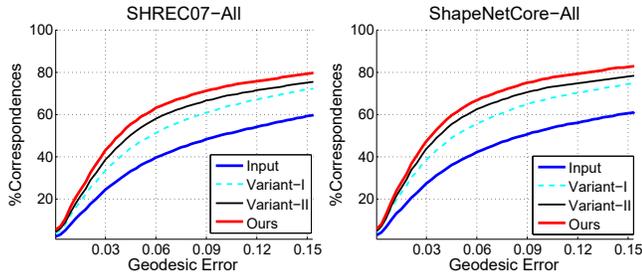


Fig. 9. **Ablation study.** Comparison between our approach and two alternative approaches on benchmark datasets. Our approach delivers the best performance. (Baseline-I) 3-cycle pre-filtering + matrix-based map recovery. (Baseline-II) 3-cycle pre-filtering + sequential low-CP-rank tensor recovery.

and perform SVD to derive the decomposition  $A \cdot B^T$  for  $\mathbf{f}$ . We apply the same approach in Section 4.2 to obtain the latent maps from  $A$  and  $B$ .

As shown in Figure 9, our approach outperforms two alternative approaches. The improvements are consistent across all three groups of categories. This shows the advantage of using the full CP decomposition for map synchronization. Moreover, our approach yields the largest performance gains on the second group of categories. This is not surprising due to low signal-to-noise ratios for categories within this group. To maximize the quality of the recovered maps, one has to utilize the full tensor formulation. Furthermore, both our approach and two alternative approaches are superior to baseline approaches. In other words, it is beneficial to combine 3-cycle consistency prefiltering and low-rank matrix/tensor recovery techniques.

## 6 APPLICATION IN SHAPE SEGMENTATION

In this section, we describe how to apply our map synchronization approach for the application of joint shape segmentation. We begin with introducing our approach in Section 6.1. We then evaluate our approach in Section 6.2.

### 6.1 Approach

A popular criterion for jointly segmenting a collection of shapes is to enforce the consistency of segmentations across the entire shape collection [Huang et al. 2011; Sidi et al. 2011; Wang et al. 2012; Yumer and Kara 2012]. However, this approach is not ideal for heterogeneous shape collections that exhibit large geometric and topological variabilities. We propose to use a different criterion that is inspired from the variability of shapes, namely, we call a region of a shape  $S_1$  a part if either it is replaced on another shape  $S_2$  (four-leg basis to swivel basis) or it is removed on  $S_2$  (chair to stool).

Under the functional map setting, we can capture such variability by analyzing the null space of a functional map  $X_{12} \in \mathbb{R}^{m \times m} : \mathcal{F}(S_1) \rightarrow \mathcal{F}(S_2)$ . In practice, even for the recovered maps of our approach, their null spaces are not perfect, meaning we have to determine their null spaces numerically. In this paper, we compute these null spaces using the following procedure. First, we compute

the singular value decomposition of  $X_{12}$  as

$$X_{12} = U_{12} \Sigma_{12} V_{12}^T, \quad \Sigma_{12} = \text{diag}(\sigma_{12}^{(1)}, \dots, \sigma_{12}^{(m)}).$$

We then determine the dimension of the null space of  $X_{12}$  by detecting the largest gap  $l^* = \max_{1 \leq l \leq m} \sigma_{12}^{(l)} - \sigma_{12}^{(l+1)}$ , where  $\sigma_{12}^{(m+1)} = 0$ .

The dimension of the null space is then given by  $m - l^*$ . With  $\hat{V}_{12}$  we denote the resulting null space. We say a null space non-trivial if its dimension is bigger than 0.

Intuitively, each non-trivial null space shall correspond to indicator functions of a missing part. Formally speaking, we can define the distance between a point on a shape to this null space as

$$d(p, \hat{V}_{12}) = \min_{\mathbf{x}_1} \|\mathbf{f}_p - \hat{V}_{12} \mathbf{x}_1\| / \|\mathbf{f}_p\|,$$

where  $\mathbf{f}_p$  is the projection of the delta function of  $p$  on this functional space. It is clear that when  $p$  is within the underlying missing part, then  $d(p, \hat{V}_{12})$  is small. Otherwise  $d(p, \hat{V}_{12})$  is large. We cut shape  $S_1$  into two parts using the median of  $d(p, \hat{V}_{12})$ .

Each non-trivial null space generates a cut of the input shape. We aggregate these segmentation cues by adapting the approach of RandomizedCut [Golovinskiy and Funkhouser 2008]. Figure 10 illustrates segmenting a chair model based on its differences to other chair models.

### 6.2 Experimental Evaluation

We have evaluated our approach on five categories of the shape segmentation dataset described in [Yi et al. 2016]. For experimental evaluation, we employ the Rand index score [Rand 1971]. In addition, we consider three baseline approaches for unsupervised segmentation, namely, co-segmentation [Sidi et al. 2011], sub-space clustering [Hu et al. 2012], and consistent latent space (or consistentLB) [Wang et al. 2013].

As shown in Table 1, Our tensor map approach achieves state-of-the-art results on Chair, Airplane, Table and Guitar shape collections. A characteristic of these categories is that their geometric variations are big, which provide sufficient signals for identifying the underlying segments. However, for the Car collection, where the shape

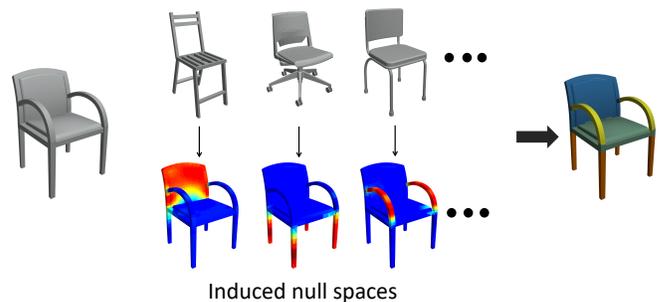


Fig. 10. We compute the segmentation of each shape based on its difference to other shapes. These differences are captured by the null spaces of functional maps. Each induced null space is visualized by the the distance from the indicator of each point to this null space. Red means small and blue means large.

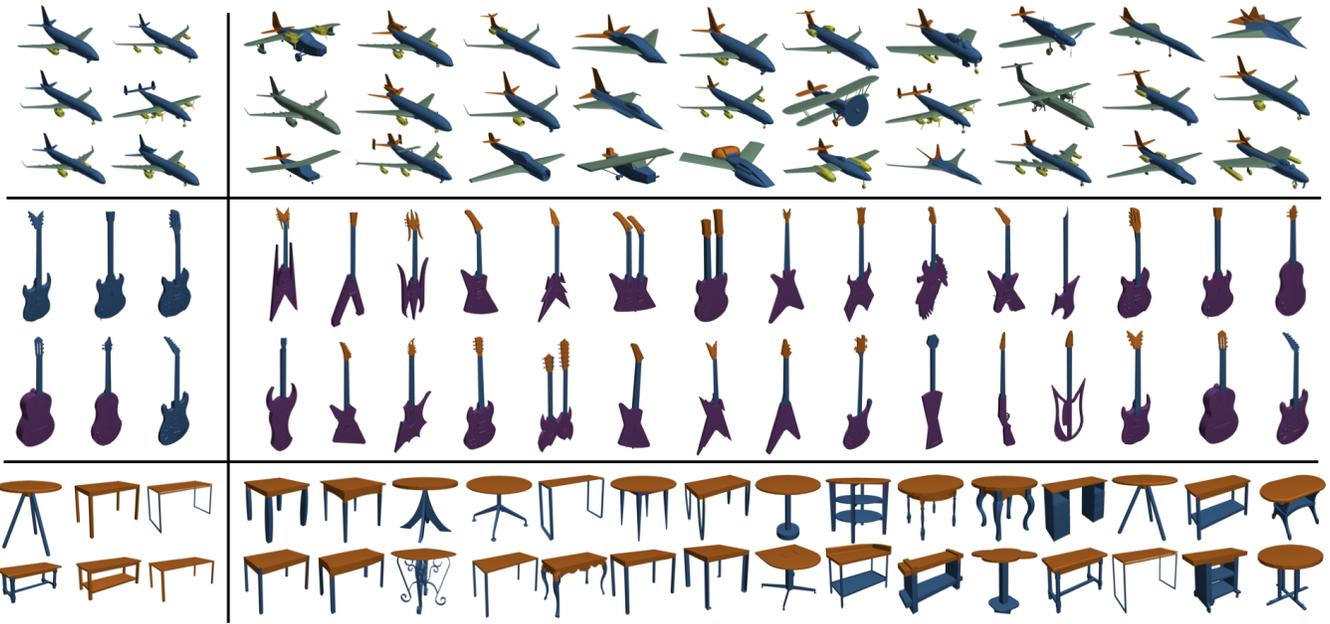


Fig. 11. **Joint shape segmentation.** We show joint shape segmentation results when increasing the size of the input shape collection. (Left) On a relatively small shape collection, our approach leads to inconsistent results. (Right) on a large shape collection, our approach leads to consistent segmentation results. Note that our approach does not utilize geometric segmentation cues on each shape. The segmentations are induced from shape differences derived from consistent shape maps.

Table 1. **Shape segmentation result.** Rand index score [Rand 1971] of three baseline approaches and our approach on five rigid classes [Yi et al. 2016].

	Chair	Table	Airplane	Guitar	Car
Co-segmentation	19.12%	17.34%	16.85%	14.37%	16.27%
Sub-space	16.59%	13.21%	10.17%	12.09%	<b>9.38%</b>
ConsistentLB	13.21%	14.15%	12.33%	11.20%	10.53%
Ours	<b>11.03%</b>	<b>9.52%</b>	<b>9.13%</b>	<b>7.96%</b>	19.20%

variation is small, our approach is not as competitive as the baseline approach of [Hu et al. 2012]. These experiments demonstrate the effectiveness of exploring shape variability as a compliment of existing approaches for joint shape segmentation.

We have also tested the performance of our approach while increasing the size of the input shape collection. As shown in Figure 11, the performance of our approach improves as the size of input shape collection increases. For small shape collections, our approach may yield under-segmentations and/or inconsistent segmentations. We can understand this from the fact that a small shape collection may not provide sufficient variability cues for segmentation, and such cues tend to be unstable. In our experiments, we found that the segmentation results become steady with more than 30 shapes on the categories we have tested.

## 7 APPLICATION IN SHAPE CLUSTERING

In this section, we describe how to apply our map synchronization approach for the application of shape clustering. We begin

with introducing our approach in Section 7.1. We then evaluate our approach in Section 7.2.

### 7.1 Approach

Our goal is to divide a heterogeneous shape collection into clusters of similar shapes (e.g., those that belong to the same categories). Existing approaches for shape clustering (or object clustering in general) fall into two categories (c.f. [Berkhin 2006; Xu and Wunsch 2005]). The first category of approaches leverages suitable object descriptors and computes object clusters so that objects in the same clusters have similar object descriptors (c.f. [Berkhin 2006; Xu and Wunsch 2005]). The second category of approaches constructs a similarity graph by connecting similar objects with respect to some affinity scores (c.f. [Berkhin 2006; Xu and Wunsch 2005]). These methods then perform graph clustering to obtain the shape clusters. The key to the success of the second category of approaches is to determine a meaningful affinity score. Our approach falls into the second category. However, our innovation is to define the affinity scores based on the input maps associated with pairs of objects. As we will demonstrate immediately, such affinity scores are more powerful and flexible than state-of-the-art approaches introduced in the literature.

Specifically, our affinity score is based on the intuition that when considering object maps among objects that fall into multiple clusters, intra-cluster maps are likely to be more accurate than inter-cluster maps. In fact, for optimization-based pairwise shape matching techniques (e.g., preservation of geodesic distances [Kim et al.

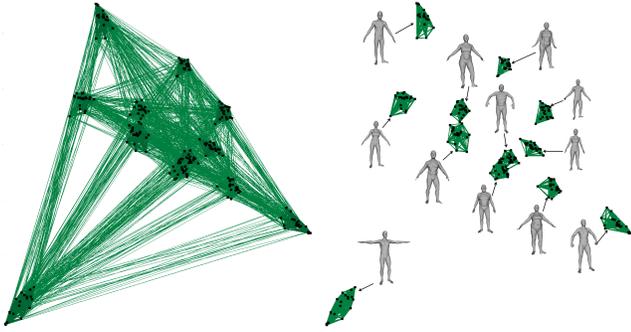


Fig. 12. **Human Shape Clustering.** (Left) Visualization of the affinity scores among the input shapes. We draw an edge between two shapes if their score is above 0.1. The induced edge lengths are given by the first two eigenvectors of the Graph Laplacian of the affinity matrix. We can see that shapes are densely connected within each cluster than across different clusters. (Right) The resulting clusters. We visualize each cluster by removing inter-cluster edges. For each cluster, we show the rest pose of the person that is most represented in that cluster. Please refer to Table 2 for a quantitative evaluation.

2011]), the optimal map is usually close to the ground-truth when the two input shapes are close. In contrast, when the two input shapes are less similar, the corresponding optimization problem becomes hard to solve (e.g., leading to local minimums) and the optimal map tends to drift away from the underlying ground-truth. This motivates us to define the affinity score for each object pair based on the difference between the input map and the recovered map. As our map synchronization approach can accurately recover the underlying ground-truth, this affinity score turns out to provide powerful signals for separating objects from different clusters.

More precisely, we first compute for each edge  $(i, j) \in \mathcal{E}$  a residual error  $r_{ij}$  defined as

$$r_{ij} := \left( \|X_{ij} - X_{ij}^{\text{in}}\|_{\mathcal{F}}^2 + \|X_{ji} - X_{ji}^{\text{in}}\|_{\mathcal{F}}^2 \right)^{\frac{1}{2}}. \quad (12)$$

We then define the affinity score  $w_{ij}$  as

$$w_{ij} := \exp\left(-\frac{r_{ij}^2}{2\sigma^2}\right), \quad \sigma := \text{median}_{j \in \mathcal{N}} \min_{i \in \mathcal{N}} r_{ij}. \quad (13)$$

Finally, we use the spectral graph clustering approach described in [Belkin and Niyogi 2001], which determines the number of underlying clusters by detecting the spectral gap and performs k-means clustering on the leading eigenvectors to recover the underlying clusters. Since this is a standard graph-based clustering approach, we refer to [Belkin and Niyogi 2001] for the technical details.

## 7.2 Experimental Results

We have applied our shape clustering approach to two challenging datasets that exhibit interesting cluster structures.

**Human Dataset.** The first dataset is a human dataset that consists of 12 different persons with 20 poses per person. The first 10 persons are taken from the testing set of the FAUST dataset [Bogo et al. 2014]. The 11th category is given by the Male category of the TOSCA dataset [Bronstein et al. 2008]. The last category is given by

Table 2. **Rand index score.** We compare our approach with three baseline approaches with respect to the Rand-index score [Rand 1971] (in percentages, and the lower the better): Shape-Des [Wu et al. 2015], which is based on comparing geodesic shape distributions; Image-Des [Su et al. 2015], which is based on comparing concatenated multi-view shape descriptors; SMAC [Bajaj et al. 2018], which is based on spectral clustering in map-based embedding spaces.

	Human	Chair	Chair-Back	Chair-Leg
Shape-Des	24.13%	20.17%	25.12%	16.43%
Image-Des	9.78%	8.78%	7.51%	6.93%
Dis Distance	12.13%	14.23%	11.89%	13.21%
SMAC	8.42%	6.23%	7.86%	8.72%
Ours	<b>2.79%</b>	<b>3.19%</b>	<b>4.39%</b>	<b>3.82%</b>

randomly sampling 20 shapes from the SCAPE dataset [Angelou et al. 2005]. We use blended intrinsic maps to compute initial maps between all pairs of shapes. We then apply our approaches to recover consistent shape maps and derive the underlying clusters. Figure 12 shows the affinity scores derived from pairwise shape maps. We can see that the affinity scores nicely indicate the underlying structures. In addition, our approach accurately recovers different poses of the individual persons except that we mis-clustered four shapes.

Table 2 compares our approach with four baseline approaches. The first baseline approach performs k-means clustering with respect to the GPS shape descriptor [Rustamov 2007], where we use [Pelleg and Moore 2000] to determine the number of clusters. The second baseline performs k-means clustering with respect to the multi-view image descriptor [Su et al. 2015] via calibrated front views. Again, we use [Pelleg and Moore 2000] to determine the number of clusters. The third baseline replaces the map residual error by the conformal Wasserstein neighborhood dissimilarity distance (CWN) [Boyer et al. 2011]. Finally, the fourth baseline is based on [Bajaj et al. 2018], which performs simultaneous mapping and clustering in the eigen-space spanned by the leading eigen-vectors of a data matrix. It turns out our approach significantly outperforms descriptor based shape clustering approaches. We can understand this from the perspective that the differences between shapes are subtle, and pure geodesic and image-based descriptors cannot differentiate them. CWN performs slightly better than descriptor based approaches, yet it still possesses a gap to our approach. The major reason is that many inter-shape maps are inaccurate, and thus they do not fully reflect the actual shape variations. Finally, our approach also outperforms [Bajaj et al. 2018], thanks to the power of our tensor formulation for accurate recovery of the underlying ground-truth maps.

**Chair Dataset.** The second dataset is the chair dataset, which collects 5 fine-grained chair classes (swivel, cantilever, sofa, windsor, and fourleg) from the chair category of ShapeNetCore. To keep the dataset balanced, we uniformly sample 20 shapes from a category if the size of that category is bigger than 20. In total, we obtained 100 shapes. We perform three clustering tasks, where the first task is based on clustering entire objects, and where the second task is based on clustering with respect to individual parts (i.e., chair back and chair legs). For the task of clustering entire objects, we follow a similar procedure for the Human dataset to compute consistent

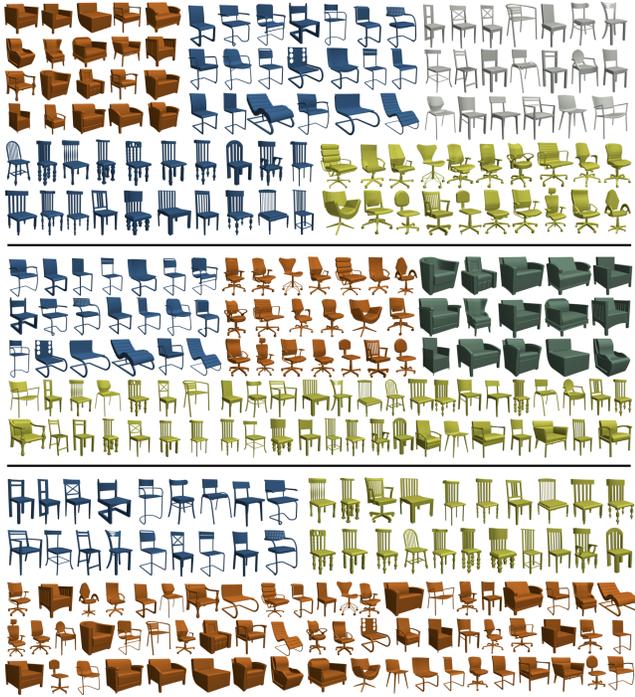


Fig. 13. **Chair Clustering.** The top block shows the clustering results with respect to the entire chair shapes. The middle block shows the clustering results with respect to chair legs. The bottom block shows the clustering results with respect to chair backs. Shapes in the same cluster are visualized with the same color. Note that our approach identifies all underlying cluster structures in each case.

shapes maps and use map residuals to obtain the underlying clusters. The only difference is that we replace the input maps by FFD alignments [Huang et al. 2014].

We again compare our approach to a similar set of four baseline approaches as the Human dataset. For the first baseline, we use the pre-trained PointNet++ [Qi et al. 2017] model to compute a shape descriptor for each shape. We apply the same procedure to determine the number of clusters and perform k-means to determine the resulting clusters. The second baseline remains the same. The third baseline is given by the conformal Wasserstein distance (or CWD) with respect to the Euclidean distance. The last baseline is again given on [Bajaj et al. 2018]. As shown in Figure 13 and Table 2, our approach leads to better results than baseline approaches. Specifically, our approach outperforms descriptor-based approaches considerably. Note that both descriptor-based approaches utilize cutting-edge deep learning techniques. This shows the advantage of using map residuals for classifying fine-grained classes. The third baseline leads to the lowest performance, due to relatively low quality initial maps on such a heterogeneous dataset. Likewise, our approach leads to better results than [Bajaj et al. 2018], as we can recover the underlying maps more accurately.

We now extend the approach to perform part-based clustering. Specifically, we mark a 3D region specified by a 3D bounding box.

when evaluating the map residual, we only consider the map residuals within the region of each shape that is inside each 3D bounding box. This is done by modifying the map residual as

$$r_{ij} := \left( \|(X_{ij} - X_{ij}^{in}) \cdot P_i\|_{\mathcal{F}}^2 + \|(X_{ji} - X_{ji}^{in}) \cdot P_j\|_{\mathcal{F}}^2 \right)^{\frac{1}{2}}, \quad (14)$$

where  $P_i \in \mathbb{R}^{m \times m}$  and  $P_j := X_{ij} P_i \in \mathbb{R}^{m \times m}$  are the projection matrices in the functional spaces associated with  $S_i$  and  $S_j$ , respectively.

As illustrated in Figure 13, by varying the specified 3D bounding boxes, we can cluster the input shapes with respect to different regions such as chair back and chair legs. For example, with respect to chair backs our approach identifies three dominant modes such as solid, open and Windsor backs. Regarding chair legs, our approach clusters with distinctive styles such as fourlegs, swivel legs, cantilever legs, and sofa legs. In addition, our approach also outperforms baseline approaches (see Table 2).

## 8 DISCUSSION, LIMITATIONS AND FUTURE WORK

In this paper, we have introduced a new approach for map synchronization. The key to our approach is a tensor-based representation for explicitly encoding the consistency of maps among a collection of relevant objects. This leads to a simple formulation of map synchronization as a low-rank tensor recovery. We have shown significant advantages of this approach against existing matrix-based approaches in terms of empirical performance on large-scale benchmark datasets. The effectiveness of this approach is demonstrated on applications of clustering and co-segmentation of heterogeneous shape collections. In each of the cases the structural matches are vastly improved through our improved correspondence maps.

Our approach possesses a few limitations. First, our approach is not particularly designed for symmetric objects. In particular, for rotational symmetric objects, our approach only yields modest results. To address this issue, it would be interesting to combine the lifting approach described in [Sun et al. 2018] and the tensor formulation described in this paper. Another limitation of our approach is that the input graph  $\mathcal{E}$  is favored to be a random graph or a geometric graph. Otherwise, there may not be a sufficient number of 3-cycles from  $\mathcal{E}$  to perform effective tensor map synchronization.

There are also many opportunities for future research. First of all, it is interesting to explore other low-rank tensor recovery techniques based on tensor decomposition and with additional low-rank regularization. Moreover, it would also be interesting to explore other representations for map synchronization, e.g., a mixture of matrix-based and tensor-based representations. Finally, one could explore other tasks related to heterogeneous shape collections such as hierarchical shape decomposition and shape synthesis using multi-dimensional tensors.

## ACKNOWLEDGMENTS

The research of CB was supported in part by NIH-R01GM117594. Qixing Huang would like to acknowledge support from NSF DMS-1700234, NSF CIP-1729486, NSF IIS-1618648, a gift from Snap Research and a GPU donation from Nvidia Inc.

## REFERENCES

- Animashree Anandkumar, Rong Ge, Daniel Hsu, Sham M. Kakade, and Matus Telgarsky. 2014. Tensor Decompositions for Learning Latent Variable Models. *J. Mach. Learn. Res.* 15, 1 (Jan. 2014), 2773–2832. <http://dl.acm.org/citation.cfm?id=2627435.2697055>
- Dragomir Anguelov, Praveen Srinivasan, Daphne Koller, Sebastian Thrun, Jim Rodgers, and James Davis. 2005. SCAPE: Shape Completion and Animation of People. *ACM Trans. Graph.* 24, 3 (July 2005), 408–416. <https://doi.org/10.1145/1073204.1073207>
- Federica Arrigoni, Beatrice Rossi, Pasqualina Fragneto, and Andrea Fusiello. 2018. Robust synchronization in SO(3) and SE(3) via low-rank and sparse matrix decomposition. *Computer Vision and Image Understanding* 174 (2018), 95 – 113. <https://doi.org/10.1016/j.cviu.2018.08.001>
- Morteza Ashraphijuo and Xiaodong Wang. 2017. Fundamental conditions for low-CP-rank tensor completion. *The Journal of Machine Learning Research* 18, 1 (2017), 2116–2145.
- Chandrajit Bajaj, Tingran Gao, Zihang He, Qixing Huang, and Zhenxiao Liang. 2018. SMAC: Simultaneous Mapping and Clustering Using Spectral Decompositions. In *Proceedings of the 35th International Conference on Machine Learning (Proceedings of Machine Learning Research)*, Jennifer Dy and Andreas Krause (Eds.), Vol. 80. PMLR, Stockholm, Sweden, 324–333. <http://proceedings.mlr.press/v80/bajaj18a.html>
- Mikhail Belkin and Partha Niyogi. 2001. Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering. In *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic (NIPS'01)*. MIT Press, Cambridge, MA, USA, 585–591. <http://dl.acm.org/citation.cfm?id=2980539.2980616>
- P. Berkhin. 2006. A Survey of Clustering Data Mining Techniques. In *Grouping Multidimensional Data: Recent Advances in Clustering*, Jacob Kogan, Charles Nicholas, and Marc Teboulle (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 25–71. [https://doi.org/10.1007/3-540-28349-8\\_2](https://doi.org/10.1007/3-540-28349-8_2)
- James C. Bezdek and Richard J. Hathaway. 2003. Convergence of alternating optimization. *Neural, Parallel Sci. Comput.* 11 (December 2003), 351–368. Issue 4.
- Federica Bogo, Javier Romero, Matthew Loper, and Michael J. Black. 2014. FAUST: Dataset and Evaluation for 3D Mesh Registration. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, Columbus, OH, USA, 3794–3801. <https://doi.org/10.1109/CVPR.2014.491>
- Doug M. Boyer, Yaron Lipman, Elizabeth St. Clair, Jesus Puente, Biren A. Patel, Thomas Funkhouser, Jukka Jernvall, and Ingrid Daubechies. 2011. Algorithms to automatically quantify the geometric similarity of anatomical surfaces. *Proceedings of the National Academy of Sciences* 108, 45 (2011), 18221–18226. <https://doi.org/10.1073/pnas.1112822108> arXiv:<https://www.pnas.org/content/108/45/18221.full.pdf>
- Alexander Bronstein, Michael Bronstein, and Ron Kimmel. 2008. *Numerical Geometry of Non-Rigid Shapes* (1 ed.). Springer Publishing Company, Incorporated, New York City.
- E. J. Candes, X. Li, Y. Ma, and J. Wright. 2011. Robust principal component analysis? *Journal of the ACM (JACM)* 58, 3 (2011), 11.
- E. J. Candes, T. Strohmer, and V. Voroninski. 2012. PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics* 66 (august 2012), 1241–1274. Issue 8.
- Angel X. Chang, Thomas A. Funkhouser, Leonidas J. Guibas, Pat Hanrahan, Qi-Xing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiang Xiao, Li Yi, and Fisher Yu. 2015. ShapeNet: An Information-Rich 3D Model Repository. *CoRR* abs/1512.03012 (2015).
- Avishek Chatterjee and Venu Madhav Govindu. 2013. Efficient and Robust Large-Scale Rotation Averaging. In *ICCV*. IEEE Computer Society, Sydney, Australia, 521–528.
- Yuxin Chen, Leonidas J. Guibas, and Qi-Xing Huang. 2014. Near-Optimal Joint Object Matching via Convex Relaxation. In *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014*. JMLR, Inc., Beijing, China, 100–108.
- Andrzej Cichocki, Danilo Mandic, Lieven De Lathauwer, Guoxu Zhou, Qibin Zhao, Cesar Caiifa, and Huy Anh Phan. 2015. Tensor decompositions for signal processing applications: From two-way to multiway component analysis. *IEEE Signal Processing Magazine* 32, 2 (2015), 145–163.
- Luca Cosmo, Emanuele Rodolà, Andrea Albarelli, Facundo Mémoli, and Daniel Cremers. 2017. Consistent Partial Matching of Shape Collections via Sparse Modeling. *Comput. Graph. Forum* 36, 1 (2017), 209–221. <https://doi.org/10.1111/cgf.12796>
- Lieven De Lathauwer, Josphine Castaing, and Jean-François Cardoso. 2007. Fourth-order cumulant-based blind identification of underdetermined mixtures. *IEEE Transactions on Signal Processing* 55, 6 (2007), 2965–2973.
- Yuval Eldar, Michael Lindenbaum, Moshe Porat, and Yehoshua Y. Zeevi. 1997. The farthest point strategy for progressive image sampling. *IEEE Trans. Image Processing* 6, 9 (1997), 1305–1315.
- Noa Fish, Oliver van Kaick, Amit Bermanto, and Daniel Cohen-Or. 2016. Structure-oriented Networks of Shape Collections. *ACM Trans. Graph.* 35, 6, Article 171 (Nov. 2016), 14 pages. <https://doi.org/10.1145/2980179.2982409>
- Thomas Funkhouser, Michael Kazhdan, Philip Shilane, Patrick Min, William Kiefer, Ayellet Tal, Szymon Rusinkiewicz, and David Dobkin. 2004. Modeling by Example. *ACM Trans. Graph.* 23, 3 (Aug. 2004), 652–663. <https://doi.org/10.1145/1015706>
- 1015775
- Guojun Gan, Chaoqun Ma, and Jianhong Wu. 2007. *Data Clustering: Theory, Algorithms, and Applications (ASA-SIAM Series on Statistics and Applied Probability)*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.
- Lin Gao, Yan-Pei Cao, Yu-Kun Lai, Hao-Zhi Huang, Leif Kobbelt, and Shi-Min Hu. 2015. Active Exploration of Large 3D Model Repositories. *IEEE Trans. Vis. Comput. Graph.* 21, 12 (2015), 1390–1402.
- Daniela Giorgi, Silvia Biasotti, and Laura Paraboschi. 2007. Shape retrieval contest 2007: Watertight models track. *SHREC competition 8*, 7 (2007).
- Aleksey Golovinskiy and Thomas Funkhouser. 2008. Randomized Cuts for 3D Mesh Analysis. *ACM Trans. Graph.* 27, 5, Article 145 (Dec. 2008), 12 pages. <https://doi.org/10.1145/1409060.1409098>
- G.H. Golub and C.F. Van Loan. 1996. *Matrix computations*. Johns Hopkins University Press, Baltimore, MD.
- Kan Guo, Dongqing Zou, and Xiaowu Chen. 2015. 3D Mesh Labeling via Deep Convolutional Neural Networks. *ACM Trans. Graph.* 35, 1, Article 3 (Dec. 2015), 12 pages. <https://doi.org/10.1145/2835487>
- Ruizhen Hu, Lubin Fan, and Ligang Liu. 2012. Co-Segmentation of 3D Shapes via Subspace Clustering. *Comput. Graph. Forum* 31, 5 (Aug. 2012), 1703–1713. <https://doi.org/10.1111/j.1467-8659.2012.03175.x>
- Qixing Huang, Vladlen Koltun, and Leonidas Guibas. 2011. Joint Shape Segmentation with Linear Programming. *ACM Trans. Graph.* 30, 6, Article 125 (Dec. 2011), 12 pages. <https://doi.org/10.1145/2070781.2024159>
- Qixing Huang, Fan Wang, and Leonidas Guibas. 2014. Functional Map Networks for Analyzing and Exploring Large Shape Collections. *ACM Transactions on Graphics* 33, 4, Article 36 (July 2014), 11 pages. <https://doi.org/10.1145/2601097.2601111>
- Qi-Xing Huang, Simon Flöry, Natasha Gelfand, Michael Hofer, and Helmut Pottmann. 2006. Reassembling Fractured Objects by Geometric Matching. *ACM Trans. Graph.* 25, 3 (July 2006), 569–578. <https://doi.org/10.1145/1141911.1141925>
- Qi-Xing Huang and Leonidas Guibas. 2013. Consistent Shape Maps via Semidefinite Programming. In *Proceedings of the Eleventh Eurographics/ACMSIGGRAPH Symposium on Geometry Processing (SGP '13)*. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 177–186. <https://doi.org/10.1111/cgf.12184>
- Qi-Xing Huang, Guo-Xin Zhang, Lin Gao, Shi-Min Hu, Adrian Butscher, and Leonidas Guibas. 2012. An Optimization Approach for Extracting and Encoding Consistent Maps in a Shape Collection. *ACM Trans. Graph.* 31, 6, Article 167 (Nov. 2012), 11 pages. <https://doi.org/10.1145/2366145.2366186>
- Xiangru Huang, Zhenxiao Liang, Chandrajit Bajaj, and Qixing Huang. 2017. Translation Synchronization via Truncated Least Squares. In *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.). Curran Associates, Inc., Long Beach, CA, USA, 1459–1468. <http://papers.nips.cc/paper/6744-translation-synchronization-via-truncated-least-squares.pdf>
- Xiangru Huang, Zhenxiao Liang, Xiaowei Zhou, Yao Xie, Leonidas J. Guibas, and Qixing Huang. 2019. Learning Transformation Synchronization. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Piscataway, NJ, USA.
- Daniel Huber. 2002. *Automatic Three-dimensional Modeling from Reality*. Ph.D. Dissertation. Robotics Institute, Carnegie Mellon University, Pittsburgh, PA.
- Evangelos Kalogerakis, Melinos Averkiou, Subhransu Maji, and Siddhartha Chaudhuri. 2017. 3D Shape Segmentation with Projective Convolutional Networks. In *CVPR*. IEEE Computer Society, 6630–6639.
- Evangelos Kalogerakis, Aaron Hertzmann, and Karan Singh. 2010. Learning 3D Mesh Segmentation and Labeling. In *ACM SIGGRAPH 2010 Papers (SIGGRAPH '10)*. ACM, New York, NY, USA, Article 102, 12 pages. <https://doi.org/10.1145/1833349.1778839>
- Raghunandan H. Keshavan, Andrea Montanari, and Sewoong Oh. 2010. Matrix Completion from a Few Entries. *IEEE Trans. Inf. Theor.* 56, 6 (June 2010), 2980–2998. <https://doi.org/10.1109/TIT.2010.2046205>
- Vladimir G. Kim, Wilmot Li, Niloy J. Mitra, Siddhartha Chaudhuri, Stephen DiVerdi, and Thomas Funkhouser. 2013. Learning Part-based Templates from Large Collections of 3D Shapes. *ACM Trans. Graph.* 32, 4, Article 70 (July 2013), 12 pages.
- Vladimir G. Kim, Wilmot Li, Niloy J. Mitra, Stephen DiVerdi, and Thomas Funkhouser. 2012. Exploring Collections of 3D Models Using Fuzzy Correspondences. *ACM Trans. Graph.* 31, 4, Article 54 (July 2012), 11 pages. <https://doi.org/10.1145/2185520.2185550>
- Vladimir G. Kim, Yaron Lipman, and Thomas Funkhouser. 2011. Blended Intrinsic Maps. *ACM Trans. Graph.* 30, 4 (July 2011), 79:1–79:12.
- Tamara G. Kolda and Brett W. Bader. 2009. Tensor Decompositions and Applications. *SIAM Rev.* 51, 3 (2009), 455–500. <https://doi.org/10.1137/07070111X>
- Vladislav Kraevoy and Alla Sheffer. 2004. Cross-parameterization and Compatible Remeshing of 3D Models. *ACM Trans. Graph.* 23, 3 (Aug. 2004), 861–869. <https://doi.org/10.1145/1015706.1015811>
- Vladislav Kraevoy, Dan Julius, and Alla Sheffer. 2007. Model Composition from Interchangeable Components. In *Proceedings of the 15th Pacific Conference on Computer Graphics and Applications*. IEEE Computer Society, Washington, DC, USA, 129–138. <https://doi.org/10.1109/PG.2007.43>

- Dana Lahat, Tülay Adalı, and Christian Jutten. 2015. Multimodal Data Fusion: An Overview of Methods, Challenges and Prospects. *Proceedings of the IEEE* 103, 9 (Aug. 2015), 1449–1477. <https://doi.org/10.1109/JPROC.2015.2460697>
- Spyridon Leonardos, XiaoWei Zhou, and Kostas Daniilidis. 2017. Distributed consistent data association via permutation synchronization. In *ICRA*. IEEE, Singapore, 2645–2652.
- Canyi Lu, Jiashi Feng, Yudong Chen, Wei Liu, Zhouchen Lin, and Shuicheng Yan. 2016. Tensor Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Tensors via Convex Optimization. In *CVPR*. IEEE Computer Society, Las Vegas, NV, USA, 5249–5257.
- Damien Muti and Salah Bourennane. 2005. Multidimensional filtering based on a tensor approach. *Signal Processing* 85, 12 (2005), 2338–2353.
- Andy Nguyen, Mirela Ben-Chen, Katarzyna Welnicka, Yinyu Ye, and Leonidas Guibas. 2011. An Optimization Approach to Improving Collections of Shape Maps. *Computer Graphics Forum* 30 (2011), 1481–1491. Issue 5. <https://doi.org/10.1111/j.1467-8659.2011.02022.x>
- Ryutarou Ohbuchi, Kunio Osada, Takahiko Furuya, and Tomohisa Banno. 2008. Salient local visual features for shape-based 3D model retrieval. In *Shape Modeling and Applications, 2008. SMI 2008. IEEE International Conference on*. IEEE Computer Society, Berkeley, CA, USA, 93–102. <https://doi.org/10.1109/SMI.2008.4547955>
- Robert Osada, Thomas Funkhouser, Bernard Chazelle, and David Dobkin. 2002. Shape Distributions. *ACM Trans. Graph.* 21, 4 (Oct. 2002), 807–832. <https://doi.org/10.1145/571647.571648>
- Maks Ovsjanikov, Mirela Ben-Chen, Justin Solomon, Adrian Butscher, and Leonidas J. Guibas. 2012. Functional maps: a flexible representation of maps between shapes. *ACM Trans. Graph.* 31, 4 (2012), 30:1–30:11.
- Deepthi Pachauri, Risi Kondor, and Vikas Singh. 2013. Solving the multi-way matching problem by permutation synchronization. In *Advances in Neural Information Processing Systems 26*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (Eds.). Curran Associates, Inc., Lake Tahoe, Nevada, US, 1860–1868. <http://papers.nips.cc/paper/4987-solving-the-multi-way-matching-problem-by-permutation-synchronization.pdf>
- Dau Pelleg and Andrew Moore. 2000. X-means: Extending K-means with Efficient Estimation of the Number of Clusters. In *Proceedings of the 17th International Conf. on Machine Learning*. Morgan Kaufmann, San Francisco, CA, USA, 727–734.
- Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J. Guibas. 2017. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. *CoRR* abs/1706.02413 (2017).
- W.M. Rand. 1971. Objective criteria for the evaluation of clustering methods. *J. Amer. Statist. Assoc.* 66, 336 (1971), 846–850.
- E. Rodolà, L. Cosmo, M. M. Bronstein, A. Torsello, and D. Cremers. 2017. Partial Functional Correspondence. *Comput. Graph. Forum* 36, 1 (Jan. 2017), 222–236. <https://doi.org/10.1111/cgf.12797>
- Raif M. Rustamov. 2007. Laplace-Beltrami Eigenfunctions for Deformation Invariant Shape Representation. In *Proceedings of the Fifth Eurographics Symposium on Geometry Processing (SGP '07)*. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 225–233. <http://dl.acm.org/citation.cfm?id=1281991.1282022>
- John Schreiner, Arul Asirvatham, Emil Praun, and Hugues Hoppe. 2004. Inter-surface Mapping. *ACM Trans. Graph.* 23, 3 (Aug. 2004), 870–877. <https://doi.org/10.1145/1015706.1015812>
- Vatsal Sharan and Gregory Valiant. 2017. Orthogonalized ALS: A Theoretically Principled Tensor Decomposition Algorithm for Practical Use. In *Proceedings of the 34th International Conference on Machine Learning (Proceedings of Machine Learning Research)*, Doina Precup and Yee Whye Teh (Eds.), Vol. 70. PMLR, International Convention Centre, Sydney, Australia, 3095–3104.
- Amnon Shashua and Tamir Hazan. 2005. Non-negative Tensor Factorization with Applications to Statistics and Computer Vision. In *Proceedings of the 22nd International Conference on Machine Learning (ICML '05)*. ACM, New York, NY, USA, 792–799. <https://doi.org/10.1145/1102351.1102451>
- Yanyao Shen, Qixing Huang, Nati Srebro, and Sujay Sanghavi. 2016. Normalized Spectral Map Synchronization. In *Advances in Neural Information Processing Systems 29*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett (Eds.). Curran Associates, Inc., Barcelona, Spain, 4925–4933. <http://papers.nips.cc/paper/6128-normalized-spectral-map-synchronization.pdf>
- Oana Sidi, Oliver van Kaick, Yanir Kleiman, Hao Zhang, and Daniel Cohen-Or. 2011. Unsupervised Co-segmentation of a Set of Shapes via Descriptor-space Spectral Clustering. *ACM Trans. Graph.* 30, 6, Article 126 (Dec. 2011), 10 pages. <https://doi.org/10.1145/2070781.2024160>
- Nicholas D Sidiropoulos, Rasmus Bro, and Georgios B Giannakis. 2000. Parallel factor analysis in sensor array processing. *IEEE transactions on Signal Processing* 48, 8 (2000), 2377–2388.
- Nicholas D. Sidiropoulos, Lieven De Lathauwer, Xiao Fu, Kejun Huang, Evangelos E. Papalexakis, and Christos Faloutsos. 2017. Tensor Decomposition for Signal Processing and Machine Learning. *Trans. Sig. Proc.* 65, 13 (July 2017), 3551–3582. <https://doi.org/10.1109/TSP.2017.2690524>
- Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik Learned-Miller. 2015. Multi-view Convolutional Neural Networks for 3D Shape Recognition. In *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV) (ICCV '15)*. IEEE Computer Society, Washington, DC, USA, 945–953. <https://doi.org/10.1109/ICCV.2015.114>
- Robert W. Sumner and Jovan Popović. 2004. Deformation transfer for triangle meshes. *ACM Trans. Graph.* 23, 3 (Aug. 2004), 399–405. <https://doi.org/10.1145/1015706.1015736>
- Jimeng Sun, Dacheng Tao, and Christos Faloutsos. 2006. Beyond Streams and Graphs: Dynamic Tensor Analysis. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '06)*. ACM, New York, NY, USA, 374–383. <https://doi.org/10.1145/1150402.1150445>
- Jian-Tao Sun, Hua-Jun Zeng, Huan Liu, Yuchang Lu, and Zheng Chen. 2005. CubeSVD: A Novel Approach to Personalized Web Search. In *Proceedings of the 14th International Conference on World Wide Web (WWW '05)*. ACM, New York, NY, USA, 382–390. <https://doi.org/10.1145/1060745.1060803>
- Yifan Sun, Zhenxiao Liang, Xiangru Huang, and Qixing Huang. 2018. Joint Map and Symmetry Synchronization. In *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part V*. Springer, Munich, Germany, 257–275. [https://doi.org/10.1007/978-3-030-01228-1\\_16](https://doi.org/10.1007/978-3-030-01228-1_16)
- Oliver van Kaick, Kai Xu, Hao Zhang, YanZhen Wang, Shuyang Sun, Ariel Shamir, and Daniel Cohen-Or. 2013. Co-hierarchical Analysis of Shape Structures. *ACM Trans. Graph.* 32, 4, Article 69 (July 2013), 10 pages. <https://doi.org/10.1145/2461912.2461924>
- Oliver van Kaick, Hao Zhang, Ghassan Hamarneh, and Daniel Cohen-Or. 2011. A Survey on Shape Correspondence. *Comput. Graph. Forum* 30, 6 (2011), 1681–1707.
- M. A. O. Vasilescu and Demetri Terzopoulos. 2002. Multilinear Analysis of Image Ensembles: TensorFaces. In *Proceedings of the 7th European Conference on Computer Vision-Part I (ECCV '02)*. Springer-Verlag, London, UK, UK, 447–460. <http://dl.acm.org/citation.cfm?id=645315.649173>
- Daniel Vlastic, Matthew Brand, Hanspeter Pfister, and Jovan Popović. 2005. Face transfer with multilinear models. *ACM transactions on graphics (TOG)* 24, 3 (2005), 426–433.
- Fan Wang, Qixing Huang, and Leonidas J. Guibas. 2013. Image Co-segmentation via Consistent Functional Maps. In *Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV '13)*. IEEE Computer Society, Washington, DC, USA, 849–856. <https://doi.org/10.1109/ICCV.2013.110>
- Lanhui Wang and Amit Singer. 2013. Exact and Stable Recovery of Rotations for Robust Synchronization. *Information and Inference: A Journal of the IMA* 2 (2013), 145–193. Issue 2.
- Yunhai Wang, Shmulik Asafi, Oliver van Kaick, Hao Zhang, Daniel Cohen-Or, and Baoquan Chen. 2012. Active Co-analysis of a Set of Shapes. *ACM Trans. Graph.* 31, 6, Article 165 (Nov. 2012), 10 pages. <https://doi.org/10.1145/2366145.2366184>
- Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 2015. 3D ShapeNets: A deep representation for volumetric shapes. In *CVPR*. IEEE Computer Society, 1912–1920.
- Kun Xu, Kang Chen, Hongbo Fu, Wei-Lun Sun, and Shi-Min Hu. 2013. Sketch2Scene: Sketch-based Co-retrieval and Co-placement of 3D Models. *ACM Trans. Graph.* 32, 4, Article 123 (July 2013), 15 pages. <https://doi.org/10.1145/2461912.2461968>
- Rui Xu and D. Wunsch, II. 2005. Survey of Clustering Algorithms. *Trans. Neur. Netw.* 16, 3 (May 2005), 645–678. <https://doi.org/10.1109/TNN.2005.845141>
- Li Yi, Leonidas Guibas, Aaron Hertzmann, Vladimir G. Kim, Hao Su, and Ersin Yumer. 2017a. Learning Hierarchical Shape Segmentation and Labeling from Online Repositories. *ACM Trans. Graph.* 36, 4, Article 70 (July 2017), 12 pages. <https://doi.org/10.1145/3072959.3073652>
- Li Yi, Vladimir G. Kim, Duygu Ceylan, I-Chao Shen, Mengyan Yan, Hao Su, Cewu Lu, Qixing Huang, Alla Sheffer, and Leonidas Guibas. 2016. A Scalable Active Framework for Region Annotation in 3D Shape Collections. *ACM Trans. Graph.* 35, 6, Article 210 (Nov. 2016), 210:1–210:12 pages. <https://doi.org/10.1145/2980179.2980238>
- Li Yi, Hao Su, Xingwen Guo, and Leonidas J. Guibas. 2017b. SyncSpecCNN: Synchronized Spectral CNN for 3D Shape Segmentation. In *CVPR*. IEEE Computer Society, 6584–6592.
- Mehmet Ersin Yumer and Levent Burak Kara. 2012. Co-abstraction of Shape Collections. *ACM Trans. Graph.* 31, 6, Article 166 (Nov. 2012), 11 pages. <https://doi.org/10.1145/2366145.2366185>
- Christopher Zach, Manfred Klopschitz, and Marc Pollefeys. 2010. Disambiguating visual relations using loop constraints. In *CVPR*. IEEE Computer Society, San Francisco, CA, 1426–1433. <http://dblp.uni-trier.de/db/conf/cvpr/cvpr2010.html#ZachKP10>
- Hongyang Zhang, Vatsal Sharan, Moses Charikar, and Yingyu Liang. 2018. Recovery Guarantees for Quadratic Tensors with Limited Observations. *CoRR* abs/1811.00148 (2018).
- Zaiwei Zhang, Zhenxiao Liang, Lemeng Wu, XiaoWei Zhou, and Qixing Huang. 2019. Path-Invariant Map Networks. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Piscataway, NJ, USA.
- XiaoWei Zhou, Menglong Zhu, and Kostas Daniilidis. 2015. Multi-image Matching via Fast Alternating Minimization. In *ICCV*. IEEE Computer Society, Santiago, Chile, 4032–4040.

## A DETAILS ABOUT ALTERNATING OPTIMIZATION

In this section, we provide details on the alternating minimization procedure for optimizing (5), which is presented in the following general form:

$$\sum_{r=1}^N \sum_{s=1}^N \sum_{t=1}^n w_{rst} (R_{rst} - \sum_{l=1}^{\bar{m}} a_{rl} b_{sl} c_{tl})^2 \quad (15)$$

where  $w_{rst}$  encode the elements of  $C_{jik}^{in}$ .

The idea of alternating minimization to alternate the optimizations of  $A$ ,  $B$  and  $C$ . In this case, the objective function becomes quadratic in the variables, and the optimal solution at each iteration can be obtained by solving a linear system.

**Optimizing  $A$ .** When  $B$  and  $C$  are fixed, we can see that optimizations of different rows of  $A$  are decoupled:

$$\min_{a_{r1}, \dots, a_{r\bar{m}}} \sum_{s=1}^N \sum_{t=1}^n w_{rst} (R_{rst} - \sum_{l=1}^{\bar{m}} a_{rl} b_{sl} c_{tl})^2, \quad 1 \leq r \leq N \quad (16)$$

Define  $H^{A,r} \in \mathbb{R}^{\bar{m} \times \bar{m}}$  and  $g^{A,r} \in \mathbb{R}^{\bar{m}}$ , whose elements are given by

$$\begin{aligned} H_{ll'}^{A,r} &:= \sum_{s=1}^N \sum_{t=1}^n w_{rst} b_{sl} b_{s'l'} c_{tl} c_{t'l'}, \quad 1 \leq l, l' \leq \bar{m} \\ g_l^A &:= \sum_{s=1}^N \sum_{t=1}^n w_{rst} b_{sl} c_{tl} R_{rst}, \quad 1 \leq l \leq \bar{m} \end{aligned} \quad (17)$$

It is easy to see that the optimal solution for

$$(a_{r1}, \dots, a_{r\bar{m}})^T := (H^{A,r})^{-1} g^{A,r}. \quad (18)$$

Note that the complexities of forming matrix  $H^A$  and vector  $g^A$  are  $O(N^2 n \bar{m}^2)$  and  $O(N^2 n \bar{m})$ , respectively. Solving the linear system in (18) costs  $O(N \bar{m}^3)$ . So the total computational cost for this step is  $O(N^2 n \bar{m}^2)$ . It should be noted that for large-scale systems,  $w_{rst}$  are sparse, so the computational cost is further reduced.

**Optimizing  $B$  or  $C$**  is done in a similar fashion. Define matrices  $H^{B,s}, H^{C,t} \in \mathbb{R}^{\bar{m} \times \bar{m}}$  and vectors  $g^{B,s}, g^{C,t} \in \mathbb{R}^{\bar{m}}$ , whose elements are given by

$$\begin{aligned} H_{ll'}^{B,s} &:= \sum_{r=1}^N \sum_{t=1}^n w_{rst} a_{rl} a_{r'l'} c_{tl} c_{t'l'}, \quad 1 \leq l, l' \leq \bar{m} \\ g_l^B &:= \sum_{r=1}^N \sum_{t=1}^n w_{rst} a_{rl} c_{tl} R_{rst}, \quad 1 \leq l \leq \bar{m} \\ H_{ll'}^{C,t} &:= \sum_{r=1}^N \sum_{s=1}^N w_{rst} a_{rl} a_{r'l'} b_{sl} b_{s'l'}, \quad 1 \leq l, l' \leq \bar{m} \end{aligned} \quad (19)$$

$$g_l^{C,t} := \sum_{s=1}^N \sum_{r=1}^N w_{rst} a_{rl} b_{sl} R_{rst}, \quad 1 \leq l \leq \bar{m} \quad (20)$$

The optimal solutions are given by

$$(b_{s1}, \dots, b_{s\bar{m}})^T := H^{B,s}{}^{-1} g^{B,s} \quad (21)$$

and

$$(c_{t1}, \dots, c_{t\bar{m}})^T := H^{C,t}{}^{-1} g^{C,t}, \quad (22)$$

respectively. Again, the complexity of optimizing  $B$  or  $C$  is  $O(N^2 n \bar{m}^2)$ .

**Convergence rate.** ([Bezdek and Hathaway 2003]) shows that under the assumptions of (1) a local optimal  $\sum_{l=1}^{\bar{m}} \mathbf{a}_l^* \otimes \mathbf{b}_l^* \otimes \mathbf{c}_l^*$  to (15) is unique, (2) and the Hessian at this local optimal is positive definite, and (3) we start from the an initial solution that is close to this local optimal, then alternating minimization converges to this local optimal in a geometric rate. In our experiments, we found that our approach always converges to a stationary point  $\sum_{l=1}^{\bar{m}} \mathbf{a}_l^* \otimes \mathbf{b}_l^* \otimes \mathbf{c}_l^*$ .

Note that although  $\sum_{l=1}^{\bar{m}} \mathbf{a}_l^* \otimes \mathbf{b}_l^* \otimes \mathbf{c}_l^*$  may be unique, the values of  $A^*$ ,  $B^*$  and  $C^*$  is dependent on the initial solution.