

# K-Best Transformation Synchronization

Yifan Sun \* Jiacheng Zhuo\*  
UT Austin

Austin, TX 78712

{yifansun12, jzhuo}@utexas.edu

Arnav Mohan

Liberal Arts and Science Academy

Austin, TX 78745

arnavmohan@gmail.com

Qixing Huang

UT Austin

Austin, TX 78712

huangqx@cs.utexas.edu

## Abstract

*In this paper, we introduce the problem of  $K$ -best transformation synchronization for the purpose of multiple scan matching. Given noisy pair-wise transformations computed between a subset of depth scan pairs,  $K$ -best transformation synchronization seeks to output multiple consistent relative transformations. This problem naturally arises in many geometry reconstruction applications, where the underlying object possesses self-symmetry. For approximately symmetric or even non-symmetric objects,  $K$ -best solutions offer an intermediate presentation for recovering the underlying single-best solution. We introduce a simple yet robust iterative algorithm for  $K$ -best transformation synchronization, which alternates between transformation propagation and transformation clustering. We present theoretical guarantees on the robust and exact recoveries of our algorithm. Experimental results demonstrate the advantage of our approach against state-of-the-art transformation synchronization techniques on both synthetic and real datasets.*

## 1. Introduction

Computing consistent relative transformations across multiple RGB or RGBD images (or transformation synchronization), is a fundamental problem that enjoys rich computer vision applications, including multi-view structure-from-motion [20, 36, 18], scan registration for 3D reconstruction [26], and assembly of fractured objects [23], to name just a few. However, almost all existing approaches stand on the assumption that there is a single plausible solution. This assumption becomes inappropriate if there exist multiple potential solutions, e.g., when the underlying object is symmetric.

In this paper, we introduce the problem of  $K$ -best transformation synchronization. The same as the standard setting of transformation synchronization (e.g. [41, 6, 13, 2, 33, 9]),  $K$ -best transformation synchronization takes as input a collection of partial objects (e.g., an RGB image collection or an RGBD image collection) and their relative transfor-

mations computed using an off-the-shelf algorithm. The difference in  $K$ -best transformation synchronization is that besides noise and outliers, there are multiple modes in the input relative transformations due to the underlying symmetries. In light of this, the output of  $K$ -best transformation consists of, for each partial object,  $K$  transformations between its local coordinate system and a world coordinate system. We expect that these transformations can describe the different modes in the input relative transformations. Note that we do not specify the value of  $K$ , and it is automatically inferred from the input relative transformations.

$K$ -best transformation synchronization appears to be a lot harder than single-best transformation synchronization due to computational challenges in separating multiple-modal correct input transformations and the incorrect input transformations. For example, recent state-of-the-art transformation synchronization techniques (e.g. [41, 4, 33]) assume that when representing pairwise transformations using data matrices, the spectral-gap of the data matrix that encodes correct input transformations shall be much bigger than the spectral norm of the data matrix that encodes incorrect input transformations (c.f. [41]). However, this assumption may not hold anymore when applying existing transformation synchronization techniques to the  $K$ -best setting, where the spectral behavior of the correct data matrix may be close to a random matrix, and the desired spectral separation is not available anymore.

In this paper, we propose a novel approach for  $K$ -best transformation synchronization. Our approach alternates between transformation propagation and transformation clustering. Transformation propagation is motivated from the success of using the power method for spectral transformation synchronization [35, 32, 31, 34, 3]. In contrast to the averaging operator used in the standard power method, we perform transformation clustering to find dominant modes of the propagated transformations, which reflect the underlying symmetry. We describe a simple procedure to determine the optimal value for  $K$  by analyzing the transformation synchronization results under different  $K$ . From the theoretical point of view, we establish strong exact recovery conditions of our approach on the underlying symmetry and relative transformations. For transforma-

\*Both authors contributed equally

tion synchronization of approximately symmetric objects, we present a post-processing step for K-best transformation synchronization that jointly picks the best pose for each partial object among the candidate poses obtained through transformation propagation and transformation clustering. In this case, the K-best solutions serve as an intermediate representation for standard transformation synchronization.

We have tested our approach on synthetic examples from ShapeNetCore [11] and real examples from ScanNet [19]. The results show that our approach leads to considerable improvements over state-of-the-art transformation synchronization techniques. On subsets of instances that show strong self-symmetries (which we denote as ShapeNetCoreSym and ScanNetSym for the rest of this paper), our approach reduces the state-of-the-art mean angular rotation error from  $12.1^\circ$  to  $6.7^\circ$  on ShapeNetCoreSym and from  $31.6^\circ$  to  $19.6^\circ$  on ScanNetSym, respectively. Our approach also leads to considerable improvements on entire benchmark datasets. The mean angular rotation errors improve from  $13.7^\circ$  to  $9.7^\circ$  on ShapeNetCore and from  $36.3^\circ$  to  $21.3^\circ$  on ScanNet. Our approach also shows competing performance against state-of-the-art rotation synchronization techniques on large RGB image collections (e.g., Cornell Arts Quad [18] and San Francisco dataset [14]).

## 2. Related Works

**Early works in map synchronization** follow the general methodology of applying the cycle-consistency criterion to improve maps computed between pairs of shapes in isolation [26, 23, 45, 30, 39]. Despite the empirical success, a common limitation of these methods is their exact recovery conditions, i.e., under what conditions can the underlying maps be recovered, are unknown. In addition, it is difficult to apply them for K-best transformation synchronization. For example, one straight-forward approach is to apply these methods recursively by removing input relative transformations that are compatible with the current solutions. However, this simple extension does not work since a single pair-wise map may be shared by multiple consistent relative transformations.

**Convex optimization techniques.** Modern map synchronization techniques leverage matrix formulations of object maps. [24] establishes the equivalence between the cycle-consistency constraint and the fact that the matrix that stores pair-wise maps in blocks is positive semidefinite or low-rank. Based on this fact, convex optimization techniques formulate map synchronization as solving a constrained low-rank matrix recovery problem [24, 42, 4, 15, 13, 8, 29, 33, 10], where the input matrix encodes the input maps, and the recovered matrix describes the cycle-consistent maps. The key advantage of such formulations is that they admit tight exact recovery conditions, namely, under what conditions of the input maps, the underlying maps can be exactly recovered.

Despite the key theoretical advantage, convex optimiza-

tion formulations cannot be easily modified to output multiple synchronizations. The major bottleneck is that the exact recovery conditions require spectral separation between noise and the underlying synchronizations, which do not hold any more in the presence of multiple solutions.

**Spectral techniques.** One family of map synchronization techniques are based on spectral decomposition of data matrices that encode maps in blocks [39, 32, 34, 6, 2]. Existing techniques also assume the existence of the spectral gap, making them not suitable for K-best transformation synchronization. Our propagation-aggregation scheme is motivated from the power method, yet we modify the scheme so that it works for continuous transformations. The clustering scheme is conceptually similar to the sparsification operator in truncated power methods [27, 44]. However, the analysis of our approach is more challenging because the spectral gap of the data matrix is small for symmetric objects.

**Non-convex optimization techniques.** A recent line of transformation synchronization techniques focuses on non-convex optimizations, including alternating minimization [48], reweighted least squares [21, 12, 25], reweighted factorization [1], and continuous optimization [38, 28]. Despite the efficiency of non-convex techniques, their success requires choosing good initial solutions. While such initial solutions are easy to obtain in the single-best setting (e.g., random sampling or spectral initialization), they become difficult to obtain in the K-best setting (e.g., spectral initialization is not applicable anymore).

**M-best MRF inference.** In a relevant problem of computing M-best solutions for discrete MRF inference, [5] modifies the objective function to spin out multiple solutions in a recursive manner. In contrast, we focus on the specific application of rotation synchronization. We also propose a specific algorithm and analyze exact recovery conditions.

**Joint map and symmetry synchronization.** In a recent work [37], the authors proposed a method for optimizing dense correspondences between objects that possess self symmetries. The key idea is to leverage lifting to obtain a map representation which provides lossless encoding of symmetry groups. The method leverages low-rank matrix recovery to estimate the underlying symmetry and pairwise maps. In contrast, this paper focuses on synchronizing rigid transformations, and we propose a novel method that combines transformation propagation and clustering. Both the problem setting and the proposed algorithm are different from those of [37].

**Bayesian pose graph optimization.** Another recent work [7] studied transformation synchronization from the perspective of Bayesian inference. Under Bingham distributions of poses, it outputs a probability distribution of synchronization results. The inference approach leverages MCMC. However, due to the single-mode nature of the Bingham distribution, this approach is mostly suitable for modeling the uncertainties of single-best transformation synchronization.

### 3. Approach

In this section, we introduce the technical details of our approach. We begin with the problem statement and an overview of our approach in Section 3.1. We then elaborate different components of our approach from Section 3.2 to Section 3.4.

#### 3.1. Problem Statement and Approach Overview

**Problem statement.** We begin introducing a few notations. In this paper, we represent a transformation  $T \in SE(3)$  using a  $4 \times 4$  matrix. Unless otherwise noticed, we denote the rotation and translation components of  $T$  as  $R \in \mathbb{R}^{3 \times 3}$  and  $\mathbf{t} \in \mathbb{R}^3$ , respectively. Likewise, we denote the rotation and translation components of  $T_{ij}$  as  $R_{ij}$  and  $\mathbf{t}_{ij}$ , respectively.

K-best transformation synchronization considers the setting where we have  $n$  partial objects (e.g., RGB-D images)  $\mathcal{I} = \{I_1, \dots, I_n\}$  capturing an underlying symmetric object. With  $\Sigma_i$  we denote the local coordinate system associated with  $I_i$ . Let  $\Sigma$  be a world coordinate system, and we assume  $\Sigma = \Sigma_1$  in this paper. The input is given by relative rigid transformations computed along a connected graph  $G = (\mathcal{L}, \mathcal{E})$  using an off-the-shelf algorithm. Specifically, we have a relative transformation  $T_{ij}^{in} \in SE(3) : \Sigma_j \rightarrow \Sigma_i$  for each edge  $(i, j) \in \mathcal{E}$ . Following the convention, we assume  $\mathcal{E}$  is symmetric, i.e.,  $\forall (i, j) \in \mathcal{E}, (j, i) \in \mathcal{E}$  and  $T_{ji}^{in} = T_{ij}^{in-1}$ . K-best transformation synchronization assumes that  $\forall (i, j) \in \mathcal{E}, \exists 1 \leq k_{ij}, k_{ji} \leq K$ ,

$$T_{ij}^{in} = T_{jk_{ji}}^* T_{ik_{ij}}^{*-1} + E_{ij},$$

where  $\mathcal{T}_i^* = \{T_{i1}^*, \dots, T_{iK}^*\}$ ,  $1 \leq i \leq n$  collects  $K$  underlying transformations from  $\Sigma_i$  to  $\Sigma$  associated with  $I_i$ , and  $E_{ij}$  is the measurement noise. For symmetric objects, it is expected that  $\mathcal{T}_i^*$  is consistent with the underlying symmetry group.

The goal of K-best transformation is to determine the size  $K$  and compute for each input object  $I_i$  a set of rigid transformations  $\mathcal{T}_i = \{T_{ik}, 1 \leq k \leq K\}$  to recover  $\mathcal{T}_i^*$ .

**Approach overview.** The key idea of our approach is motivated from the fact that when looking at composite transformations between a pair of images along paths in  $G$ , the correct ones (there may be multiple ones due to the underlying symmetry) shall be realized by multiple such paths (c.f.[22]). In light of this, we propose to alternate between propagating transformations and clustering transformations. The propagation operation explores all potential relative transformations from  $S_1$  to each  $S_i$ , while the clustering operation ensures that the number of propagated relative transformations does not explode. As highlighted below, our approach consists of a pre-processing step, a core propagation-and-clustering procedure, a procedure for determining  $K$ , and a post-processing step.

The pre-processing step of our approach computes a dense set of samples  $\mathcal{S} \subset SE(3)$  for rigid transformations.

$\mathcal{S}$  allows us to convert a hard continuous optimization problem into a more manageable discrete optimization problem. As we will see later, such a discretization also significantly boosts the performance of transformation clustering. Moreover, our sampling strategy possesses an effective indexing data structure, such that most samples are not visited during the execution of our algorithm.

The core propagation-and-clustering procedure assumes a given  $K$  and computes  $\mathcal{T}_i^*$ ,  $1 \leq i \leq n$  in an iterative manner. Given the current transformation sets  $\mathcal{T}_i^{(l)}$ ,  $1 \leq i \leq n$  at iteration  $l$ . The propagation operation computes  $\overline{\mathcal{T}}_i^{(l+1)}$  by propagating signals from neighboring vertices of  $I_i$ . The clustering step computes  $\mathcal{T}_i^{(l+1)}$  by picking  $K$  most repeating elements in  $\overline{\mathcal{T}}_i^{(l+1)}$ . This propagation-and-clustering procedure terminates when  $\mathcal{T}_i^{(l)}$  are fixed or the maximum number of iterations is reached.

Our approach determines the optimal value of  $K$  by running the propagation-and-clustering procedure on different values of  $K$ . We then introduce a scoring function that measures a description efficiency score of a solution and pick  $K$  that maximizes this description efficiency.

The post processing step jointly picks the best solution among the candidate poses computed during the propagation-and-clustering procedure. We solve this problem by combing leading eigenspace computation and rounding.

#### 3.2. Preprocessing: Transformation Sampling

In this section, we introduce a sampling strategy for  $SE(3)$ . In particular, this sampling strategy can be indexed using a sparse indexing data structure, and most of the samples are not revisited during each run of our algorithm. This allows us to use very dense samples to marginalize the discretization error.

We perform sampling for rotations and translations separately. Specifically, we use a regular grid to sample the space of translations. Let  $d_{\max}$  be the maximal translation along each axis, we use a grid of resolution  $N^3$  for possible translations, meaning the size of each cell is  $\frac{d_{\max}}{N}$ . We expect that the sampling density is larger than the measurement noise, and we set  $N = 128$  in our experiments. In addition, we set  $d_{\max} = 2L$ , where  $L$  is the averaged size of each input object. With  $\mathcal{S}_t \subset \mathbb{R}^3$  we denote the resulting translation sample set. Sampling the space of rotations  $SO(3)$  is a bit more involved, due to the complicated structure of  $SO(3)$ . We propose to decompose  $SO(3) = (S^2/\{1, -1\} \times S^1)$ , where  $S^i$  denotes the unit sphere in  $\mathbb{R}^{i+1}$ . Using this decomposition, we simply concatenate samples for  $S^1$  and  $S^2/\{1, -1\}$  to obtain samples for  $SO(3)$ . To sample  $S^1$ , we place uniform samples along  $S^1$ . To sample  $S^2/\{1, -1\}$ , we employ the HEALPix grid [43]. In our experiments, we set the grid resolutions to be  $2^\circ$  for both  $S^1$  and  $S^2/\{1, -1\}$ . With  $\mathcal{S}_R$  we denote the resulting rotation sample set. Finally, we define the entire sample set  $\mathcal{S}$  for

$SE(3)$  by concatenating the rotation and translation sample sets, i.e.,  $\mathcal{S} = \mathcal{S}_R \times \mathcal{S}_t$ .

Our approach frequently utilize a projection operator  $\mathcal{P}_{\mathcal{S}} : SE(3) \rightarrow \mathcal{S}$  that snaps any rigid transformation onto the transformation sample set  $\mathcal{S}$ :

$$\mathcal{P}_{\mathcal{S}}(T) = [R_{\mathcal{S},T} | \mathbf{t}_{\mathcal{S},T}],$$

where  $\forall T \in SE(3)$ ,

$$R_{\mathcal{S},T} = \operatorname{argmin}_{R' \in \mathcal{S}_R} \|R - R'\|_{\mathcal{F}}, \quad \mathbf{t}_{\mathcal{S},T} = \operatorname{argmin}_{\mathbf{t}' \in \mathcal{S}_t} \|\mathbf{t} - \mathbf{t}'\|.$$

Both projection operators can be easily implemented using the sparse indexing structure.

### 3.3. Transformation Propagation and Clustering

The core of the proposed K-best transformation synchronization approach is an iterative procedure for transformation propagation and clustering. This procedure assumes a given  $K$  and outputs a transformation set  $\mathcal{T}_i \subset \mathcal{S}$ ,  $|\mathcal{T}_i| = K$  for each partial object  $I_i$ . The input to this procedure is given by relative transformations  $T_{ij}^{in}, (i, j) \in \mathcal{E}$  computed using an off-the-shelf pairwise matching technique. Let  $\mathcal{T}_i^{(l)}, 1 \leq i \leq n$  denote the transformation sets at iteration  $l$ . Initially, we set

$$\mathcal{T}_1^{(0)} = \{I_4\}, \quad \mathcal{T}_i^{(0)} = \emptyset, \quad 2 \leq i \leq n.$$

At each iteration  $l$  of the iterative procedure, we first propagate  $\mathcal{T}_i^{(l)}$  to neighboring objects to generate candidate transformation sets  $\overline{\mathcal{T}}_i^{(l+1)} \subset \mathcal{S}$ ,  $1 \leq i \leq n$ :

$$\overline{\mathcal{T}}_i^{(l+1)} := \bigcup_{j \in \mathcal{N}(i)} \{\mathcal{P}_{\mathcal{S}}(T_{ji}^{in} \cdot T_j) | T_j \in \mathcal{T}_j^{(l)}\}. \quad (1)$$

where  $\mathcal{N}(i) \subset \{1, \dots, n\}$  denotes the indices of adjacent objects of  $I_i$  in  $\mathcal{G}$ . Given each candidate transformation set  $\overline{\mathcal{T}}_i^{(l+1)}$ , we compute its corresponding transformation set  $\mathcal{T}_i^{(l+1)}$  at iteration  $l+1$  by extracting  $K$  dominant modes:

$$\mathcal{T}_i^{(l+1)} \leftarrow \text{K-Mode}(\overline{\mathcal{T}}_i^{(l+1)}). \quad (2)$$

As  $\overline{\mathcal{T}}_i^{(l+1)} \subset \mathcal{S}$ . We simply set  $\mathcal{T}_i^{(l+1)}$  as the  $K$  elements that are most frequent in  $\overline{\mathcal{T}}_i^{(l+1)}$ . Note that if  $|\overline{\mathcal{T}}_i^{(l+1)}| < K$ , then we simply set  $\mathcal{T}_i^{(l+1)} = \overline{\mathcal{T}}_i^{(l+1)}$ .

The iterative procedure of propagation and clustering terminates when

$$\mathcal{T}_i^{(l+1)} = \mathcal{T}_i^{(l)}, \quad \forall 1 \leq i \leq n,$$

or the maximum number of iterations is reached. With  $\mathcal{T}_i^*, 1 \leq i \leq n$  we denote the output of this iterative procedure.

The iterative procedure described above is dependent on the root object, which we set as  $I_1$ . Our experiments show that the final results are insensitive to the choice of the root object.

### 3.4. Estimation of $K$

In this section, we define a scoring function to evaluate the output of the propagation and clustering procedure described above. To this end, we augment the notation of the out as  $\mathcal{T}^{out}(K) := \{\mathcal{T}_i^*(K), 1 \leq i \leq n\}$  for a fixed  $K$  with respect to the set of input relative transformations  $\mathcal{T}_{\text{pair}}^{in} := \{T_{ij}^{in}, (i, j) \in \mathcal{E}\}$ . We then define a consistency score between one input rigid transformation  $T_{ij}^{in}$  and two corresponding transformation sets  $\mathcal{T}_i$  and  $\mathcal{T}_j$ :

$$s(\mathcal{T}_i, \mathcal{T}_j, T_{ij}^{in}) = \begin{cases} 1 & \mathcal{P}_{\mathcal{S}}(T_{ij}^{in} T_i) = T_j, \exists T_i \in \mathcal{T}_i, T_j \in \mathcal{T}_j \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

We proceed to define a score of  $\mathcal{T}^{out}(K)$  with respect to  $\mathcal{T}_{\text{pair}}^{in}$  as follows:

$$\text{score}(\mathcal{T}^{out}(K), \mathcal{T}_{\text{pair}}^{in}) := \frac{1}{K} \sum_{(i,j) \in \mathcal{E}} s(\mathcal{T}_i^*(K), \mathcal{T}_j^*(K), T_{ij}^{in}). \quad (4)$$

Intuitively,  $\text{score}(\mathcal{T}^{out}(K), \mathcal{T}_{\text{pair}}^{in})$  measures a description efficiency score, i.e., the ratio between the number of input instances that are compatible with the solution set and the size of the solution set. It is thus expected that we have a drop in  $\text{score}(\mathcal{T}^{out}(K^*+1), \mathcal{T}_{\text{pair}}^{in})$  right after the underlying ground-truth value  $K^*$ , which is the size of the symmetry group that the underlying object possesses. In light of this, we set

$$K := \operatorname{argmin}_{1 \leq K \leq K_{\max} - 1} \frac{\text{score}(\mathcal{T}^{out}(K+1), \mathcal{T}_{\text{pair}}^{in})}{\text{score}(\mathcal{T}^{out}(K), \mathcal{T}_{\text{pair}}^{in})} \quad (5)$$

for a sufficiently large  $K_{\max}$ .

### 3.5. Postprocessing: Infer the Best Solution

When the notion of the best solution makes sense, e.g., the underlying symmetry is only approximate, we perform a post-processing step to jointly pick one candidate pose  $T_i^* \in \mathcal{T}_i^*, 1 \leq i \leq n$  as the optimal solution to transformation synchronization. We formulate this step as MRF inference and adopt a spectral relaxation [17]. Specifically, we form a data matrix  $X \in \mathbb{R}^{(n-1)K+1 \times (n-1)K+1}$ , so that  $\forall 2 \leq i, j \leq n, 1 \leq k, k' \leq K$ ,

$$X_{(i-2)K+k+1, (j-2)K+k'+1} = \begin{cases} \delta(\mathcal{P}_{\mathcal{S}}(T_{ij}^{in} T_{ik}) = T_{jk'}) & (i, j) \in \mathcal{E} \\ \delta(\mathcal{P}_{\mathcal{S}}(T_{i1}^{in} T_{id}) = T_{i1} \wedge k' = K) & (i, 1) \in \mathcal{E} \\ \delta(\mathcal{P}_{\mathcal{S}}(T_{1j}^{in}) = T_{1j} T_{id} \wedge k = K) & (1, j) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases}$$

To perform the spectral initialization, we first compute the leading  $K$  eigen-vectors  $U \in \mathbb{R}^{((n-1)K+1) \times K}$  of  $X$ . We then project the first canonical basis  $e_1$  into  $\mathbb{R}^{(n-1)K+1}$  (i.e., the first element is 1 and the remaining elements are 0) into the column space of  $U$ , obtaining  $\mathbf{v} = \mathbf{U}\mathbf{U}^T e_1$ . We finally set the single-best pose for  $I_i$  as

$$T_i^* = T_{ik_i}, \quad k_i = \operatorname{argmax}_{1 \leq k \leq K} |\mathbf{v}_{(i-2)K+k}|.$$

## 4. Analysis

In this section, we present an analysis of our algorithm by showing exact recovery conditions of K-best solutions as well as the single-best solution. The technical proofs are deferred to the supplemental material.

**Noise model.** The noise model we analyze is inspired from the noise models introduced in the literature for analyzing convex optimization techniques for map synchronization (e.g., [41, 15, 34]). The key advantage of our model is that we do not assume the noisy input maps (i.e., those deviate from the underlying ground-truth and those induced by the underlying symmetry group) are random. Instead, we consider adversarial noise for these noisy input maps. Moreover, we assume correct measurements take different modes independently with approximately equal probability, which is a viable assumption for pairwise matching that outputs one solution among multiple competing solutions. On the other hand, we assume each edge of the observation graph  $\mathcal{G}$  is added with probability  $p_e$ , following the Erdos-Renyi random graph model. This enables us to establish simple exact recovery conditions. Please refer to the supplemental material for an exact recovery condition where  $\mathcal{G}$  is fixed.

Specifically, we consider the following model for generating the input data. The model is parameterized by the number of objects  $n$ , the edge probability  $p_e$ , the good measurement probability  $p_g$ , and an offset  $\delta$  for modeling approximate symmetries:

- Independently and identically for any object pair  $(i, j)$ ,  $\Pr[(i, j) \in \mathcal{E}] = p_e$ . In other words, the expected number of neighboring vertices for each vertex of  $\mathcal{G}$  is  $np_e$ .
- Independently and identically for any edge  $(i, j) \in \mathcal{E}$ , with probability  $\frac{p_g}{K^*} + \delta$ ,  $\mathcal{P}_S(T_{ij}^{in} T_{i1}^{gt}) = \mathcal{P}_S(T_{j1}^{gt})$ , and with probability  $\frac{p_g}{K^*} - \frac{\delta}{K^*-1}$ ,  $\mathcal{P}_S(T_{ij}^{in} T_{i1}^{gt}) = \mathcal{P}_S(T_{jk}^{gt}), \forall 2 \leq k \leq K^*$ ;
- Independently and identically for any edge  $(i, j) \in \mathcal{E}$ , with probability  $1 - p_g$ ,  $T_{ij}$  can be an arbitrary transformation in  $SE(3)$ . In particular, we do not place any assumption on the noise of the initial maps.

Intuitively, the above rules implies that, each of the  $K^*$  ground-truth transformations exist in the input with approximately equal probability (controlled by  $\delta$ ), so that it is possible to recover them all.

**Exact recovery conditions.** We then present a recovery condition on the parameters  $p_e$  and  $p_g$  of such graph families as follows:

**Theorem 4.1. (Main theorem)** *If the input data follows the noisy model described above. Suppose  $p_e = n^{\frac{1}{d}-\epsilon} - 1$  for constants  $d > 3$  and  $0 < \epsilon < 1$ , and moreover,  $p_g$  satisfies*

$$p_g > c^{\frac{1}{d}}$$

for some absolute constant  $c > 0.75$ . Then with probability at least  $1 - O(\frac{1}{n^3})$ , 1) our algorithm recovers the underlying ground-truth  $K^*$ , and 2) after  $l = O(d)$  iterations,  $\mathcal{T}_i^{(l)}$  recovers all the  $K^*$  underlying ground-truth transformations of partial object  $I_i$  for every  $2 \leq i \leq n$ .

**Theorem 4.2.** *Under the assumption of Theorem 4.1, the post processing step recovers the unique global optimal solution with probability at least  $1 - O(\frac{1}{n^3})$ , whenever  $\delta \geq O(\sqrt{\frac{\log(n)}{\sqrt{np_e}}})$ .*

Besides the recovery condition for computing the K-best solution, Theorem 4.1 and Theorem 4.2 also indicate the advantage of using the K-best solution for computing the single-best solution. Note that under adversarial noise, exact recovery conditions of state-of-the-art techniques (e.g. [42, 24]) can only handle  $\leq 50\%$  incorrect pairwise input. Using the K-best representation, we can extend the upper bound to  $\frac{K^* - O(1)}{K^*} - o(1)$ , showing a clear advantage when  $K^*$  is large.

## 5. Experimental Evaluation

### 5.1. Experimental Setup

**Datasets.** We consider three datasets for the evaluation. The first dataset collects 100 models sampled from ShapeNetCore [11]. Among these 100 models, 30 of them have strong self-symmetries, and they comprise ShapeNetCoreSym. We place 30 simulated scans from each model. To simulate scanning noise, we add i.i.d. Gaussian perturbation along the viewing directions (c.f. [26]). The second dataset is ScanNet [19], which collects range scans and their 3D reconstructions from 706 rooms. We uniformly sample 100 scenes and uniformly sample 30 scans per scene for experimental evaluation. The ground-truth is given by their poses with the 3D reconstruction. Among these rooms, we select 80 scenes that possess noticeable self-symmetries. These rooms comprise ScanNetSym. Please refer to the supplemental material for details about the selected models from ShapeNet and the subsets ShapeNetCoreSym and ScanNetSym. The third dataset collects three large-scale image collections Notre-Dame, Cornell ArtQuad, and San Francisco, which are widely used for the purpose of evaluating multi-view structure from motion.

To generate the initial rigid transformations for ShapeNetCore and ScanNet, we employed the Open3D implementation [46] of fast global registration [47], which is a state-of-the-art approach for pairwise alignment.

For all the results of our approach reported in this section, we set  $K_{\max} = 10$  and the maximum number of iterations as 100.

**Baseline comparison.** We compare our approach to three state-of-the-art techniques for transformation synchronization: RobustAlign [16], MRF-SFM [18], and IRLS [12, 25].

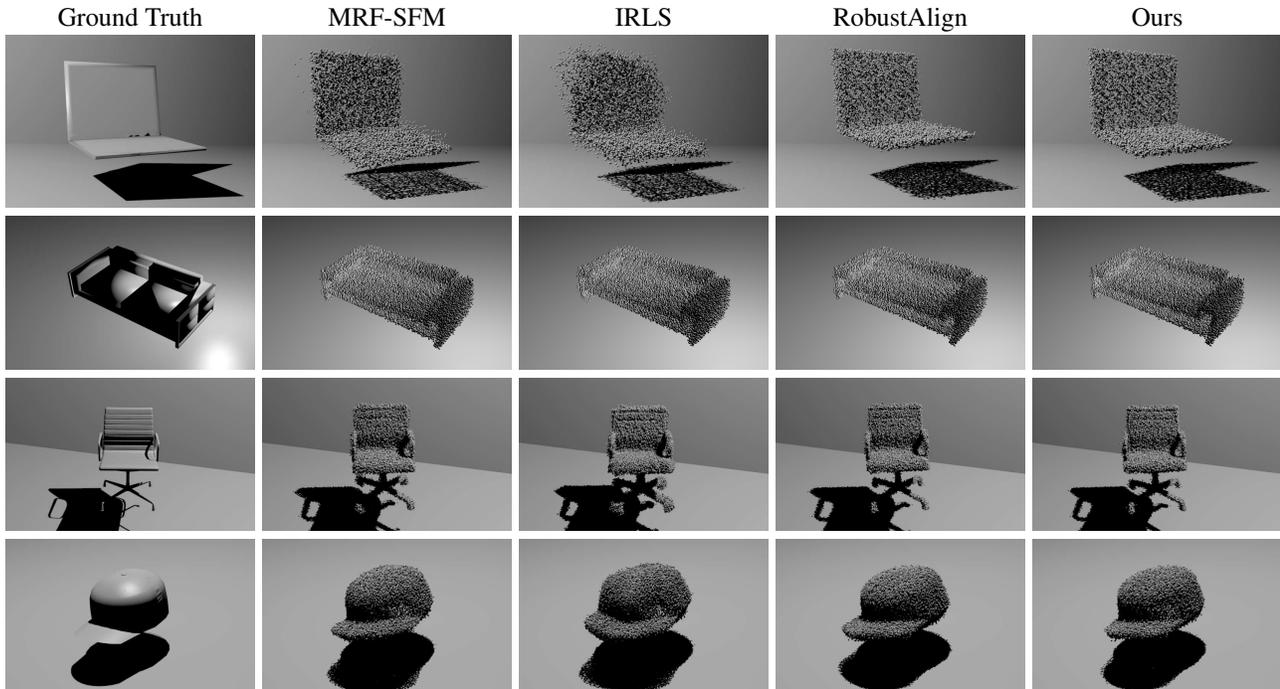


Figure 1: Qualitative comparisons between our approach (single-best) and baseline approaches on randomly sampled ShapeNetCore objects [11]. We use the rotation and translation computed by the algorithms to reconstruct the objects in point-cloud-form, and render the point clouds. Our methods drastically outperform IRLS and MRF-SFM. If look at the edge of the laptop, the gully on the back of the chair and the shadow of the hat, one can still tell that our approach is better than the RobustAlign method.

Specifically, RobustAlign applies robust global optimization to perform transformation synchronization. MRF-SFM performs rotation synchronization and translation synchronization in order. Each synchronization problem is formulated as solving an MRF by placing samples. Note that unlike our approach, MRF-SFM only outputs a single-best solution. IRLS combines a state-of-the-art rotation synchronization technique [12] and a state-of-the-art translation synchronization technique [25]. Both of them are based on reweighted least squares. Similar to the sequential procedure of MRF-SFM, we first apply [12] to estimate the rotations. We then apply [25] to estimate the translations.

We also compare our approach to additional state-of-the-art techniques, including Torsello et al. [40], Arrigoni et al. [2] and Rosen et al. [33], which are based on spectral and convex optimization, and Birdal et al. [7], which is based on probability inference. We found that their performance is at most similar to the top performing approach among RobustAlign, MRF-SFM and IRLS. Thus, we defer the technical details to the supplemental material.

**Evaluation protocol.** We perform experimental evaluation under two settings: one for symmetric objects and another for general objects. Both settings use a primitive evaluation protocol, which evaluates a predicted relative transformation  $(R, \mathbf{t})$  with respect to its ground-truth relative transformation  $(R^*, \mathbf{t}^*)$  by measuring the angular rotation error  $err_a(R^*, R) := \|\log(R^* R^T)\|/\sqrt{2}$  [18] and the transla-

tion error  $err_t(\mathbf{t}, \mathbf{t}^*) := \|\mathbf{t} - \mathbf{t}^*\|$ . Note that  $err_t(\mathbf{t}, \mathbf{t}^*)$  is dependent on the coordinate system associated with each scan. In this paper, we always place the origin of each coordinate system at the center of each scan. We also normalize each translation error with respect to the diagonal length of the bounding box of the underlying model. With this setup, the first setting consider objects with self-symmetries (i.e., models in ShapeNetCoreSym and ScanNetSym). Instead of evaluating with respect to the single underlying ground-truth, we define the angular rotation error or translation error associated with each pair of scans as the smallest one with respect to poses induced from the underlying symmetry group. Since our approach outputs multiple solutions per pair, we take the maximum rotation error or translation error among all pairwise solutions derived from our approach. Note that this metric implicitly favors baseline approaches, since they only output one relative pose per scan pair. As shown in Table 1, we report the percentage of pairs (across all models in each benchmark dataset) whose angular rotation errors or translation errors are smaller than a varying threshold. The second setting is a special case of the first setting, where we simply evaluate with respect to the underlying ground-truth. This setting applies to all models in ShapeNetCore and ScanNet. We apply the same protocol to report the rotation and translation errors.

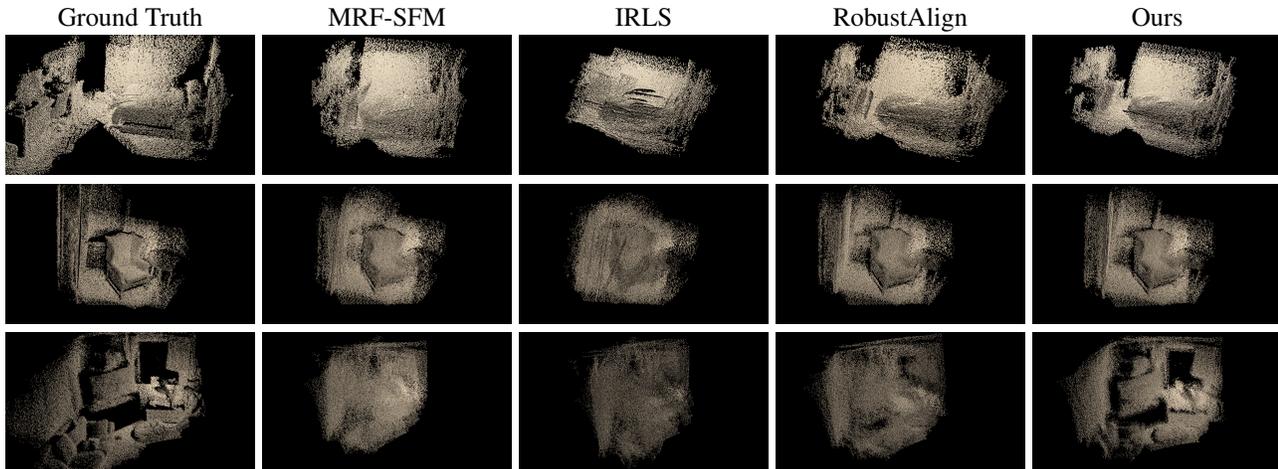


Figure 2: Qualitative comparisons between our approach (single-best) and baseline approaches on randomly sampled ScanNet scenes [19]. We use the rotation and translation computed by the algorithms to reconstruct the scenes in point-cloud-form, and render the point clouds. It is visually obvious that our approach outperforms the three methods that we are comparing with.

## 5.2. Results on ShapeNetCore

Figure 1 and Table 1 show the qualitative and quantitative results of our approach and baseline approaches, respectively. Overall, our approach delivers accurate results. The mean rotation/translation errors of our approach are  $6.7^\circ/0.17$  and  $9.7^\circ/0.24$  on ShapeNetCoreSym and ShapeNetCore, respectively. In contrast, the top performing baseline approaches only achieved  $12.1^\circ/0.19$  and  $13.7^\circ/0.39$ , respectively.

On ShapeNetCoreSym, the top-performing baseline is MRF-SFM. We can understand this from the fact that most incorrect input pairwise transformations are caused by the underlying symmetries. Such inputs favor MRF-SFM, as in the limit, if 1) all input transformations are accurate when factoring out the underlying symmetry, and 2) MRF-SFM places accurate samples, then MRF-SFM shall output one consistent solution. In this experimental setup, our approach still outperforms MRF-SFM due to the fact that we can sample significantly more densely than MRF-SFM. Moreover, baseline approaches that employ continuous optimization result in large errors on ShapeNetCoreSym. One explanation is that in the presence of multiple competing modes in the input data, existing continuous optimization techniques (even under robust norms) tend to find averages of existing solutions rather than identifying a single consistent mode.

On ShapeNetCore, the behavior of the baseline approaches changes. Most ShapeNetCore models possess strong consistent matches. As a result, RobustAlign and IRLS yield better results than MRF-SFM, which is restricted by the relatively low sampling density. Still, our approach exhibits noticeable performance gains from all baseline approaches. This shows the advantage of using the K-best solution as an intermediate representation for approximately symmetric or even non-symmetric objects.

The running time of our approach on ShapeNetCore is 1.2s on average on a desktop with 32G memory and a 3.2G Hz CPU. In contrast, the fastest baseline approach, RobustAlign, takes 0.6s on average on the same platform.

## 5.3. Results on ScanNet

As illustrated in Table 1 and Figure 2, our approach outperforms all baseline approaches on ScanNet. Similar to the results on ShapeCoreSym, all baseline approaches exhibit large errors on ScanNetSym. The results again show that when applying continuous optimization techniques (even with robust norms) to input data that exhibit multiple modes, existing approaches return the average of different modes rather than finding one mode. Hence these approaches exhibit large error under the K-best protocol.

The performance gaps between our approach and baseline approaches on ScanNet are larger than those on ShapeNetCore as well. We can understand this from the fact that the outlier ratio of the input is high on ScanNet, making it extremely difficult to handle for both continuous optimization based techniques and MRF-SFM. Our approach alleviates this issue by following a flexible two-step procedure, namely, we first detect K-dominant modes, from which we can then compute the best solution.

Compared to the running time on ShapeNetCore, our approach took longer to converge on ScanNet. The averaged running time was 4.6s. In contrast, it was 1.2s on ShapeNetCore. This increase in running time is again largely due to the high outlier ratio. In contrast, RobustAlign took 0.8s on average on this dataset.

## 5.4. Results on Multi-View Structure-from-Motion

We also tested our approach on three large-scale structure-from-motion examples for the purpose of rotation synchronization: Notre-dame [12], Cornell-Artsquad [18]

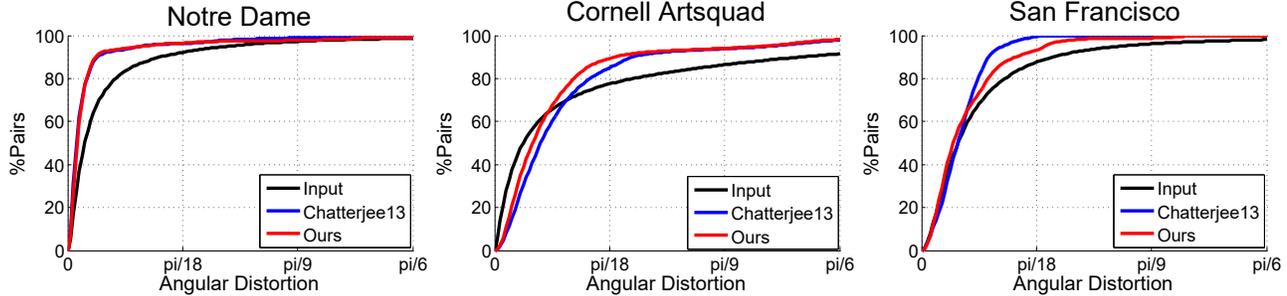


Figure 3: Quantitative evaluations of our method on Notre-dame [12], Cornell-Artsquad and San-Francisco [18]. Our approach is similar to [12] in terms of overall performance.

	ShapeNetCoreSym (K-Best)						ScanNetSym (K-Best)						ShapeNetCore (Single-Best)						ScanNet (Single-Best)					
	Rotation			Trans.			Rotation			Trans.			Rotation			Trans.			Rotation			Trans.		
	3°	30°	Mean	0.05	0.25	Mean	3°	30°	Mean	0.05	0.25	Mean	3°	30°	Mean	0.05	0.25	Mean	3°	30°	Mean	0.05	0.25	Mean
Input[46]	41.4	76.1	24.5	34.1	61.2	0.36	32.5	51.1	41.2	31.9	51.1	0.69	42.1	62.1	39.1	24.8	52.8	0.65	19.8	38.0	69.9	12.5	29.3	1.35
MRF-SFM[18]	51.5	84.6	12.1	45.1	78.1	0.19	32.1	62.3	31.6	36.5	64.3	0.42	47.8	81.1	13.7	31.2	55.9	0.48	33.6	55.9	36.3	22.7	35.1	0.81
IRLS[12, 25]	42.5	78.3	25.2	32.3	65.2	0.40	34.2	53.1	35.5	33.2	55.1	0.65	51.1	79.1	20.5	31.0	61.2	0.45	31.1	51.0	45.4	24.5	36.7	0.78
RobustAlign[16]	42.7	79.2	23.2	33.4	67.3	0.37	34.4	52.2	34.5	34.2	57.2	0.64	<b>52.3</b>	78.5	22.5	31.2	<b>63.3</b>	0.39	33.8	50.5	43.4	24.7	38.4	0.70
Our approach	<b>67.2</b>	<b>91.3</b>	<b>6.7</b>	<b>49.3</b>	<b>80.1</b>	<b>0.17</b>	<b>41.1</b>	<b>84.4</b>	<b>19.6</b>	<b>38.5</b>	<b>67.3</b>	<b>0.29</b>	51.8	<b>87.1</b>	<b>9.7</b>	<b>36.2</b>	58.9	<b>0.24</b>	<b>39.6</b>	<b>63.9</b>	<b>21.3</b>	<b>26.7</b>	<b>39.4</b>	<b>0.56</b>
$K^{gt}$	69.9	95.1	5.8	53.3	86.1	0.15	44.5	87.3	16.9	40.2	70.1	0.26	55.3	89.8	15.2	39.4	61.9	0.21	40.6	66.1	19.1	28.7	42.1	0.52
3×binsize	59.2	88.3	8.9	42.5	78.4	0.21	35.1	80.1	22.6	32.7	65.4	0.32	47.9	84.9	11.5	33.4	56.4	0.27	35.3	60.8	22.9	25.2	38.1	0.59
2×binsize	65.1	90.2	6.5	48.1	79.3	0.18	39.6	83.3	20.1	36.8	65.7	0.30	50.7	86.2	9.9	34.7	57.3	0.25	38.1	62.7	22.9	24.9	38.1	0.57
0.5×binsize	70.2	90.4	6.5	48.8	79.2	0.16	44.3	82.3	18.9	39.1	66.4	0.30	52.4	86.7	9.9	38.1	58.2	0.24	40.1	64.9	20.5	28.6	40.4	0.58
Vary-root	±4.7	±3.1	±0.3	±4.5	±2.7	±0.02	±3.2	±2.5	±0.5	±5.1	±2.8	±0.02	±3.8	±3.1	±0.3	±4.1	±2.7	±0.01	±3.3	±2.9	±0.4	±3.5	±2.2	±0.02

Table 1: Benchmark evaluation on our approach and baseline approaches. The columns labeled as 3°, 30°, 0.05 and 0.25 tell the portion of the algorithm output that is within these error bracket. The columns labeled as 'mean' tell the mean rotation error or the translation error. Our approach outperforms the three methods that we are comparing with.

and San-Francisco [18]. To this end, we modify our approach by only performing propagation and clustering on rotation samples. As these datasets possess a single-best underlying ground-truth, we only evaluate the single-best solution of our approach. Overall, our approach is comparable to the state-of-the-art rotation synchronization approach [12] on each instance (See Figure 3). In particular, our approach even possesses a slight performance gain on Cornell Artsquad. This is quite encouraging as our approach uses an iterative procedure on discrete samples, while [12] is based on continuous optimization. Computationally, our approach is also comparable to [12]. On the same computing platform described above, our approach took 34.2s, 122.1s, and 212.3s on Norte Dame, Cornell Artsquad and Scan Francisco, respectively, while [12] took 23.1s, 95.1s, and 172.1s, respectively.

### 5.5. Ablation Study

**How good is the predicted  $K$ ?** Our approach predicts the underlying  $K^*$  accurately, i.e., the prediction accuracies on ShapeNet and ScanNet are 98.2% and 97.4%, respectively. We also run our approach with the ground-truth  $K^*$ . As shown in Table 1, the rotation and translation errors improve slightly. However, the performance gaps are not salient.

**How sensitive is the sampling density?** Our approach is insensitive to the bin size we use for generating the samples for SE(3). As shown in Table 1, we vary the bin size by 3×, 2×, 0.5× the default value. As shown we see that with 3× and 2× bin sizes, the rotation and translation errors increase slightly. With 0.5× bin size, the percentages

with respect to 3° increase slightly while those with respect to 30° drop slightly. However, the overall accuracies remain approximately the same when varying the sampling density.

**How sensitive is the root node?** We also tested our approach by varying the root object for each dataset. Experimental results show that the variance in rotation errors and translation errors appears to be small. As shown in Table 1, the variance in mean rotation error is less than 0.5°, and the variance in mean translation error is less than 0.02, both of which are small when compared to the absolute values.

## 6. Conclusions

In this paper, we have described an approach for synchronizing multiple consistent transformations among a collection of objects. Our approach takes as input a set of pair-wise transformations and outputs K-best solutions, each of which specifies a consistent match across the input objects. We introduce an iterative clustering method, which alternates between transformation propagation and transformation clustering. These K-best solutions can be subsequently used to derive a single-best solution via computing the leading eigenvector of a data matrix that encodes the input pairwise transformations. We provide strong exact recovery conditions of our approach. Experimental results demonstrate the usefulness of our approach on both symmetric objects and non-symmetric objects.

**Acknowledgement.** Qixing Huang would like to acknowledge support from NSF DMS-1700234, a Gift from Snap Research, and a hardware Donation from NVIDIA.

## References

- [1] Federica Arrigoni, Andrea Fusiello, Beatrice Rossi, and Pasqualina Fragneto. Robust rotation synchronization via low-rank and sparse matrix decomposition. *CoRR*, abs/1505.06079, 2015. [2](#)
- [2] Federica Arrigoni, Beatrice Rossi, and Andrea Fusiello. Spectral synchronization of multiple views in se (3). *SIAM Journal on Imaging Sciences*, 9(4):1963–1990, 2016. [1](#), [2](#), [6](#)
- [3] Chandrajit Bajaj, Tingran Gao, Zihang He, Qixing Huang, and Zhenxiao Liang. Smac: simultaneous mapping and clustering using spectral decompositions. In *International Conference on Machine Learning*, pages 334–343, 2018. [1](#)
- [4] Afonso S Bandeira, Moses Charikar, Amit Singer, and Andy Zhu. Multireference alignment using semidefinite programming. In *Proceedings of the 5th conference on Innovations in theoretical computer science*, pages 459–470. ACM, 2014. [1](#), [2](#)
- [5] Dhruv Batra, Payman Yadollahpour, Abner Guzman-Rivera, and Gregory Shakhnarovich. Diverse m-best solutions in markov random fields. In *European Conference on Computer Vision*, pages 1–16. Springer, 2012. [2](#)
- [6] Florian Bernard, Johan Thunberg, Peter Gemmar, Frank Hertel, Andreas Husch, and Jorge Goncalves. A solution for multi-alignment by transformation synchronisation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2161–2169, 2015. [1](#), [2](#)
- [7] Tolga Birdal, Umut Simsekli, Mustafa Onur Eken, and Slobodan Ilic. Bayesian pose graph optimization via bingham distributions and tempered geodesic MCMC. In *NeurIPS*, pages 306–317, 2018. [2](#), [6](#)
- [8] Jesus Briales and Javier Gonzalez-Jimenez. Initialization of 3d pose graph optimization using lagrangian duality. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5134–5139. IEEE, 2017. [2](#)
- [9] Luca Carlone and Giuseppe C. Calafiore. Convex relaxations for pose graph optimization with outliers. *IEEE Robotics and Automation Letters*, 3(2):1160–1167, 2018. [1](#)
- [10] Luca Carlone and Giuseppe C Calafiore. Convex relaxations for pose graph optimization with outliers. *IEEE Robotics and Automation Letters*, 3(2):1160–1167, 2018. [2](#)
- [11] 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. Shapenet: An information-rich 3d model repository. *CoRR*, abs/1512.03012, 2015. [2](#), [5](#), [6](#)
- [12] Avishek Chatterjee and Venu Madhav Govindu. Efficient and robust large-scale rotation averaging. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 521–528, 2013. [2](#), [5](#), [6](#), [7](#), [8](#)
- [13] Kunal N. Chaudhury, Yuehaw Khoo, and Amit Singer. Global registration of multiple point clouds using semidefinite programming. *SIAM Journal on Optimization*, 25(1):468–501, 2015. [1](#), [2](#)
- [14] David M Chen, Georges Baatz, Kevin Köser, Sam S Tsai, Ramakrishna Vedantham, Timo Pyhäläinen, Kimmo Roimela, Xin Chen, Jeff Bach, Marc Pollefeys, et al. City-scale landmark identification on mobile devices. In *CVPR 2011*, pages 737–744. IEEE, 2011. [2](#)
- [15] Yuxin Chen, Leonidas Guibas, and Qixing Huang. Near-optimal joint object matching via convex relaxation. In *Proceedings of the 31st International Conference on International Conference on Machine Learning-Volume 32*, pages II–100. JMLR. org, 2014. [2](#), [5](#)
- [16] Sungjoon Choi, Qian-Yi Zhou, and Vladlen Koltun. Robust reconstruction of indoor scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5556–5565, 2015. [5](#), [8](#)
- [17] Timothee Cour and Jianbo Shi. Solving markov random fields with spectral relaxation. In *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics*, pages 75–82, 2007. [4](#)
- [18] David Crandall, Andrew Owens, Noah Snavely, and Dan Huttenlocher. Discrete-continuous optimization for large-scale structure from motion. In *CVPR 2011*, pages 3001–3008. IEEE, 2011. [1](#), [2](#), [5](#), [6](#), [7](#), [8](#)
- [19] Angela Dai, Angel X. Chang, Manolis Savva, Maciej Halber, Thomas A. Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. *CoRR*, abs/1702.04405, 2017. [2](#), [5](#), [7](#)
- [20] Venu Madhav Govindu. Lie-algebraic averaging for globally consistent motion estimation. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 1, pages I–I. IEEE, 2004. [1](#)
- [21] Richard I. Hartley, Khurram Aftab, and Jochen Trumpf. L1 rotation averaging using the weiszfeld algorithm. In *CVPR*, pages 3041–3048. IEEE Computer Society, 2011. [2](#)
- [22] Qi-Xing Huang, Guo-Xin Zhang, Lin Gao, Shi-Min Hu, Adrian Butscher, and Leonidas J. Guibas. An optimization approach for extracting and encoding consistent maps in a shape collection. *ACM Trans. Graph.*, 31(6):167:1–167:11, 2012. [3](#)
- [23] Qi-Xing Huang, Simon Flöry, Natasha Gelfand, Michael Hofer, and Helmut Pottmann. Reassembling fractured objects by geometric matching. *ACM Transactions on Graphics (TOG)*, 25(3):569–578, 2006. [1](#), [2](#)
- [24] Qi-Xing Huang and Leonidas Guibas. Consistent shape maps via semidefinite programming. In *Proceedings of the Eleventh Eurographics/ACMSIGGRAPH Symposium on Geometry Processing*, pages 177–186. Eurographics Association, 2013. [2](#), [5](#)
- [25] Xiangru Huang, Zhenxiao Liang, Chandrajit Bajaj, and Qixing Huang. Translation synchronization via truncated least squares. In *Advances in neural information processing systems*, pages 1459–1468, 2017. [2](#), [5](#), [6](#), [8](#)
- [26] Daniel F Huber. *Automatic three-dimensional modeling from reality*. PhD thesis, Robotics Institute, Carnegie Mellon University, 2002. [1](#), [2](#), [5](#)
- [27] Michel Journée, Yurii Nesterov, Peter Richtárik, and Rodolphe Sepulchre. Generalized power method for sparse principal component analysis. *Journal of Machine Learning Research*, 11(Feb):517–553, 2010. [2](#)
- [28] Gim Hee Lee, Friedrich Fraundorfer, and Marc Pollefeys. Robust pose-graph loop-closures with expectation-maximization. In *IROS*, pages 556–563. IEEE, 2013. [2](#)
- [29] Spyridon Leonardos, Xiaowei Zhou, and Kostas Daniilidis. Distributed consistent data association via permutation synchronization. In *ICRA*, pages 2645–2652. IEEE, 2017. [2](#)

- [30] A. Nguyen, M. Ben-Chen, K. Welnicka, Y. Ye, and L. Guibas. An optimization approach to improving collections of shape maps. In *Eurographics Symposium on Geometry Processing (SGP)*, pages 1481–1491, 2011. [2](#)
- [31] Deepti Pachauri, Risi Kondor, Gautam Sargur, and Vikas Singh. Permutation diffusion maps (pdm) with application to the image association problem in computer vision. In *Advances in Neural Information Processing Systems*, pages 541–549, 2014. [1](#)
- [32] Deepti Pachauri, Risi Kondor, and Vikas Singh. Solving the multi-way matching problem by permutation synchronization. In *Advances in neural information processing systems*, pages 1860–1868, 2013. [1](#), [2](#)
- [33] David M Rosen, Luca Carlone, Afonso S Bandeira, and John J Leonard. Se-sync: A certifiably correct algorithm for synchronization over the special euclidean group. *The International Journal of Robotics Research*, 38(2-3):95–125, 2019. [1](#), [2](#), [6](#)
- [34] Yanyao Shen, Qixing Huang, Nati Srebro, and Sujay Sanghavi. Normalized spectral map synchronization. In *Advances in Neural Information Processing Systems*, pages 4925–4933, 2016. [1](#), [2](#), [5](#)
- [35] Amit Singer and H-T Wu. Vector diffusion maps and the connection laplacian. *Communications on pure and applied mathematics*, 65(8):1067–1144, 2012. [1](#)
- [36] Noah Snavely, Steven M Seitz, and Richard Szeliski. Photo tourism: exploring photo collections in 3d. In *ACM transactions on graphics (TOG)*, volume 25, pages 835–846. ACM, 2006. [1](#)
- [37] Yifan Sun, Zhenxiao Liang, Xiangru Huang, and Qixing Huang. Joint map and symmetry synchronization. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 251–264, 2018. [2](#)
- [38] Niko Sünderhauf and Peter Protzel. Towards a robust backend for pose graph slam. In *2012 IEEE International Conference on Robotics and Automation*, pages 1254–1261. IEEE, 2012. [2](#)
- [39] Andrea Torsello, Emanuele Rodola, and Andrea Albarelli. Multiview registration via graph diffusion of dual quaternions. In *CVPR 2011*, pages 2441–2448. IEEE, 2011. [2](#)
- [40] Andrea Torsello, Emanuele Rodolà, and Andrea Albarelli. Multiview registration via graph diffusion of dual quaternions. In *The 24th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011, Colorado Springs, CO, USA, 20-25 June 2011*, pages 2441–2448, 2011. [6](#)
- [41] Lanhui Wang and Amit Singer. Exact and stable recovery of rotations for robust synchronization. *Information and Inference: A Journal of the IMA*, 2:145193, December 2013. [1](#), [5](#)
- [42] Lanhui Wang and Amit Singer. Exact and stable recovery of rotations for robust synchronization. *Information and Inference: A Journal of the IMA*, 2(2):145–193, 2013. [2](#), [5](#)
- [43] Anna Yershova, Swati Jain, Steven M Lavalley, and Julie C Mitchell. Generating uniform incremental grids on  $so(3)$  using the hopf fibration. *The International journal of robotics research*, 29(7):801–812, 2010. [3](#)
- [44] Xiao-Tong Yuan and Tong Zhang. Truncated power method for sparse eigenvalue problems. *Journal of Machine Learning Research*, 14(1):899–925, 2013. [2](#)
- [45] Christopher Zach, Manfred Klopschitz, and Marc Pollefeys. Disambiguating visual relations using loop constraints. In *CVPR*, pages 1426–1433. IEEE Computer Society, 2010. [2](#)
- [46] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Open3d: A modern library for 3d data processing. *CoRR*, abs/1801.09847, 2018. [5](#), [8](#)
- [47] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Fast global registration. In *European Conference on Computer Vision*, pages 766–782. Springer, 2016. [5](#)
- [48] Xiaowei Zhou, Menglong Zhu, and Kostas Daniilidis. Multi-image matching via fast alternating minimization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4032–4040, 2015. [2](#)