Kinematic Structure Correspondences via Hypergraph Matching

Hyung Jin Chang   Tobias Fischer   Maxime Petit   Martina Zambelli   Yiannis Demiris
Personal Robotics Laboratory, Department of Electrical and Electronic Engineering
Imperial College London, United Kingdom
{hj.chang, t.fischer, m.petit, m.zambelli13, y.demiris}@imperial.ac.uk

Abstract

In this paper, we present a novel framework for finding the kinematic structure correspondence between two objects in videos via hypergraph matching. In contrast to prior appearance and graph alignment based matching methods which have been applied among two similar static images, the proposed method finds correspondences between two dynamic kinematic structures of heterogeneous objects in videos. Our main contributions can be summarised as follows: (i) casting the kinematic structure correspondence problem into a hypergraph matching problem, incorporating multi-order similarities with normalising weights, (ii) a structural topology similarity measure by a new topology constrained subgraph isomorphism aggregation, (iii) a kinematic correlation measure between pairwise nodes, and (iv) a combinatorial local motion similarity measure using geodesic distance on the Riemannian manifold. We demonstrate the robustness and accuracy of our method through a number of experiments on complex articulated synthetic and real data.

1. Introduction

Building kinematic structures of articulated objects from visual input data is an active research topic in computer vision and robotics [27, 11, 17, 3]. The accurately estimated kinematic structure represents motion properties as well as shape information of an object in a topological manner, and it encodes relationships between rigid body parts connected by kinematic joints. Thus, it can be considered as a mid-level representation of general objects captured from different sensors such as RGB camera, depth camera, and robot encoders.

Accurate and efficient estimation of kinematic correspondences between heterogeneous objects is beneficial to many high level tasks such as learning by imitation [38], viewpoint invariant human action recognition by 3D skeletons [36], affordance based object/tool categorisation [39], articulated object manipulation [13, 17], and human motion retargeting to robots [33]. Therefore, in this paper we focus on finding correspondences between two articulated kinematic
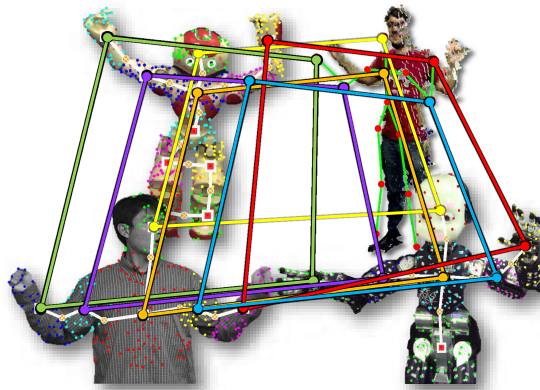

Figure 1. The proposed framework reliably builds up kinematic structure correspondence matches across heterogeneous objects captured with different sensors. Our method can for example find correspondences between a upper-body dancing human in a 2D grey image sequence, the iCub and NAO humanoid robots in 2D RGB videos, and a dancing human in depth image sequences.

structures extracted from different objects' image sequences using a new hypergraph matching framework.

There have been various approaches to generate accurate kinematic structures from visual data [41, 27, 11, 3]. However, there isn't much work on utilising the generated kinematic structures for higher level purposes. Fayad *et al.* [11] used the kinematic structure as a basis of 3D reconstruction. Sturm *et al.* [31] and Katz *et al.* [17] applied it to robot manipulations, but with relatively simple objects. We propose a new way of utilising the kinematic structure for matching kinematic correspondences between complex objects belonging to different categories, which can be extended to applications with heterogeneous sensory data as well.

Most of the conventional correspondence finding methods in the computer vision area are restricted to two static images [4, 6]. Skeleton corresponding points matching methods between two objects' silhouette images have been presented by Bai and Latecki [1]. Local shape feature and graph matching based methods [5, 20, 26] have been researched actively for decades. However, all these approaches are based on object

appearance and local shape features which do not include dynamic information.

Graph matching methods have been widely used for the correspondence matching problem [32, 45]. By introducing a high order tensor product scheme for the third or even higher order similarity terms, hypergraph matching methods can achieve robust matches even under large variations [43, 10, 20, 26]. Hypergraph matching has a wide potential for finding matches among various applications, but because of the difficulty in designing the high-order similarity, there are not many applications using this method.

In this paper, we present a novel hypergraph matching method capable of finding correspondences between articulated kinematic structures estimated from two different image sequences. We propose new similarity measures in order to consider structural topology (first order), kinematic correlation (second order) and combinatorial motion (third order) similarities simultaneously and incorporate them into the hypergraph matching framework with weight normalisation. We introduce a new topologically constrained subgraph isomorphism measure for the structural similarity and a restricted combinatorial motion descriptor for the motion similarity. Our experiments show that the proposed method outperforms appearance based and skeleton graph alignment based methods quantitatively and qualitatively on both, synthetic and real data.

2. Related Work

Kinematic structure building: The kinematic structure represents kinematic properties between rigid body parts, whilst the skeleton is generally a framework of bones. Motion based rigid body part segmentation and building connections between the segments have been a general approach for the kinematic structure estimation from visual data. Factorisation based articulated motion segmentation methods [34, 40] were presented showing that the rank of the feature trajectory matrix can indicate the kinematic joint type. Yan and Pollefeys [41] estimated a kinematic chain based on intersecting motion subspaces. Ross *et al.* [27] proposed a probabilistic graphical model learning method which learned the number of joints and their connections adaptively. However, as the number of articulation increases, the computation time grows exponentially. Also, the method is prone to local minima. An energy based multiple model fitting was proposed by Fayad *et al.* [11], which performs well for simple structures by balancing overall model complexity and local motion errors. However, it finds a moderate structure rather than an actual detailed structure as the segments are enforced to be overlapped. Recently, Chang and Demiris [3] presented a complex structure estimation by combining motion and skeleton information. They showed state-of-the-art estimation performance even for complex objects, but its randomisation of motion segmentation affects structure

consistency.

Structure correspondence matching: A path similarity based skeleton graph matching was developed by Bai and Latecki [1] and was applied for shape recognition based on object silhouettes [29]. Although performing well among clean silhouette images, the method requires noiseless skeletons as input, and can not be readily applied to images of actual moving objects. The structure correspondence matching is similar to the graph alignment problem in the bioinformatics field for the alignment of protein-protein interactions networks [18]. However, the general graph alignment data are quite different, as very large graphs (thousands of nodes) are used in conjunction with node similarity based on chemical properties. Recently, graph structure information based alignment methods such as NETAL [25] and MAGNA++ [37] have been presented, which do not rely on node similarities.

Motion description methods: Jacquet *et al.* [15] presented a relative transformation analysis method based on linear subspaces, but they focused on detecting the type of articulated motion between two restricted motion parts. Various metrics for 3D rotations, which can be used to describe motions, have been reviewed in [14]. Most metrics were found to be boundedly equivalent, and it was shown that the geodesic distance on the Riemannian submanifold between two rotation matrices has more meaning than the other metrics from a geometric point of view. Hartley *et al.* [12] analyse the problem of rotation averaging from the theoretical side, again emphasizing the importance of the geodesic distance. In addition, the geodesic distance has been proved to be effective in practical applications, *e.g.* for video trajectory estimation and smoothing [16]. Schulz *et al.* introduced a framework which employs the geodesic distance to find deformations within 3D models of real objects [28].

Hypergraph matching: Early second order graph matching approaches [21, 9] were based on the best rank-1 approximation of an affinity matrix and energy minimisation [32]. Stochastic sampling was used in [19] to solve the graph matching problem, whereas a random walk view was proposed in [5]. Other second-order methods include the work on deformable graphs matching, proposed in [45]. Higher-order relationships between graphs have been explored to derive hypergraph matching, which can incorporate more complex features and representations [43]. A formulation of the hypergraph matching as a third order tensor optimisation problem was presented in [10], which has shown significant improvements over second order methods. An extension to the third order setting of the reweighted random walk [5] was proposed by Lee *et al.* [20]. Recently, a pure discrete method has been devised [42], accounting for both unary and higher-order affinity terms and in line with the linear approximate framework, and a tensor block coordinate ascent methods was proposed for hypergraph matching [26].

3. Methodology

Our goal is to find corresponding kinematic joint matches between two articulated kinematic structures via hypergraph matching, whilst being accurate and plausible under appearance and motion variations. To this end, we mainly use 2D feature trajectories from two image sequences assuming that one target subject exists in each scene and the features are extracted from every part of the object. To build the kinematic structure from 2D image sequences, we adopt the state-of-the-art kinematic structure generation method [3].

3.1. Kinematic Structure Formulation

The 2D feature point trajectories of each image sequence are represented as x_p^f, with p as feature point index and $f = 1, ..., F$ as sequence index. F indicates the number of frames for each sequence. To indicate motion segments, we use S_i for the disjoint set of points belonging to the ith segment where $i = 1, ..., N$, and N as the total number of segments , and y_i denotes the centre position of segment S_i obtained by averaging its points.

In this work, we adopt the same way of representing the kinematic structure as [41, 3] which use a non cyclic graph model $G = (V, E)$ to indicate the topological connections between rigid body parts. Each node[1] $v_i \in V$ is assigned to the respective motion segment centre y_i, and the edge E_{ij} represents a connection between nodes v_i and v_j. In this work, we assume that all kinematic joints are revolute joints, as prismatic joints are less common and spherical joints can be easily decomposed into orthogonal revolute joints.

3.2. Hypergraph Matching for Kinematic Structure Correspondence

In order to estimate the correspondences between two kinematic structures, we use hypergraph matching methods. A general hypergraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of nodes $v \in \mathcal{V}$, and hyperedges $e \in \mathcal{E}$. Unlike usual graph edges, hyperedges enclose a subset of nodes from $\mathcal{V}$ with size $\kappa(e)$, referred to as the order of each hyperedge. In this work, we generate the hypergraphs based on the two kinematic structures G and G'. We consider the hypergraph nodes $v \in \mathcal{V}$ as the kinematic structure nodes $v \in V$ (*i.e.* $V = \mathcal{V}$), so that each $e \in \mathcal{E}$ can represent any tuple of nodes.

The hypergraph matching problem is to find mappings between nodes of two hypergraphs. Given two hypergraphs $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$, the goal is to find a subset in the set of correspondences $\mathcal{V} \times \mathcal{V}'$. Without loss of generality, we assume that $N \leq N'$ where $N = |\mathcal{V}|$ and $N' = |\mathcal{V}'|$. The subset of correspondences can be represented by the one-to-one binary assignment matrix $X \in \{0, 1\}^{N \times N'}$, where $X(i, i') = 1$ if $v_i \in \mathcal{V}$ matches $v_i' \in \mathcal{V}'$ and $X(i, i') = 0$ otherwise. We then define the similarity function $\mathcal{F}$ of a

matching subset as the weighted sum of the first, second and third order similarity terms, as follows:

$$\mathcal{F}(X) = w_1 \sum_i^N \sum_{i'}^{N'} \mathcal{F}^1_{(i,i')} X_i \qquad (1)$$

$$+ w_2 \sum_{\substack{i,j \\ i \neq j}}^N \sum_{\substack{i',j' \\ i' \neq j'}}^{N'} \mathcal{F}^2_{(i,i')(j,j')} X_i X_j$$

$$+ w_3 \sum_{\substack{i,j,k \\ i \neq j \neq k}}^N \sum_{\substack{i',j',k' \\ i' \neq j' \neq k'}}^{N'} \mathcal{F}^3_{(i,i')(j,j')(k,k')} X_i X_j X_k$$

where $X_i = X(i, i'), X_j = X(j, j')$ and $X_k = X(k, k')$. In particular, we define three similarity functions which are defined in the following sections and are related to the structure topology (first order $\mathcal{F}^1$), the kinematic correlation (second order $\mathcal{F}^2$), and the combinatorial motion (third order $\mathcal{F}^3$).

The normalising weights w_1, w_2, w_3 play an important role on balancing the effects of the similarity terms. Previous hypergraph matching works [44, 26] did not consider weights for the summation of different order terms. However, the similarity terms are not well-balanced, as the corresponding number of summation elements increases exponentially with the order. Consequently, higher order similarity terms have more impact to the matching result than lower order terms. Therefore, we propose multi-order similarities with normalising weights $w_{\kappa(e)}$ which are inverse proportional to the number of elements, such that:

$$w_{\kappa(e)} = \frac{(N - \kappa(e))!}{N!} \times \frac{(N' - \kappa(e))!}{N'!} \qquad (2)$$

where $\kappa(e) = 1$ (that is $e = \{v_i\}$), $\kappa(e) = 2$ (that is $e = \{v_i, v_j\}$) and $\kappa(e) = 3$ (that is $e = \{v_i, v_j, v_k\}$) for the first, second and third order correspondences, respectively. The well-balanced combination of all the three similarity functions through the weights w_1, w_2, w_3 allows us to effectively merge the three different pieces of information and obtain a more accurate and meaningful correspondence between kinematic structures.

In our framework, we leverage benefits of the reweighted random walk hypergraph matching (RRWHM) method [20] which allows to merge different order similarity information nicely to obtain topological, kinematic and motion correspondences. RRWHM shows comparable performances to other state-of-the-art methods [26] and in particular, it has been shown to achieve better performance in the case of non-convex functions as in Eq.(1). Thus, we use RRWHM [20] for solving Eq.(1) in our proposed framework. In the experiment section, we show the performance of other hypergraph matching methods applied to Eq.(1).

---

[1]The terms node and vertex are used interchangeably.

3.2.1 First Order Similarity: Structure Topology

Each vertex v_i of the kinematic structure graph $G = (V, E)$ is the centre of a body part which is reduced to a single point y_i. Thus, we cannot use classical point feature descriptors (such as SIFT [22]) to measure a similarity between two body parts (*i.e.* vertices) of two kinematic structures. Instead, we propose a new algorithm based on topologically constrained subgraph isomorphisms to find the first order node similarity function $\mathcal{F}^1$, which purely relies on the structural topology.

We find all possible subgraphs of G' which might be matched to the graph G under local and global constraints. The local constraint eliminates subgraphs where pairs of vertices are unlikely to be matched by considering the number of connected edges (degree). The global constraint is a threshold on the sum of the difference in degrees for all vertex pairs. This also serves as quality measurement, as a lower difference in degree indicates a higher similarity. We also variate G' by removing vertices one by one in order to cope with noise. We define the subgraph isomorphism following Valiente [35]:

Definition 3.1. *A subgraph isomorphism of a graph $G=(V, E)$ into a graph $G'=(V', E')$ is an injection $\mathcal{I} \subseteq V \times V'$ such that, for every pair of vertices $v_i, v_j \in V$ and $v'_i, v'_j \in V'$ with $(v_i, v'_i) \in \mathcal{I}$ and $(v_j, v'_j) \in \mathcal{I}$, then $(v_i, v'_j) \in E'$ if $(v_i, v_j) \in E$.*

Importantly, it is guaranteed that a subgraph isomorphism of G' has at least as many nodes and edges per node as G. However, the subgraph isomorphism by itself does not indicate a perfect match to G, as the subgraph isomorphism might not only contain more nodes than G, but matched nodes between the graphs may also have different edges associated.

We employ the commonly used VF2 algorithm [7, 8][2] to first determine the L subgraph isomorphisms $\mathbb{I}_l \in \{0, 1\}^{N \times N'}$ ($l = 1, ..., L$) of G into G'. The entry $\mathbb{I}_l(i, j)$ in the i-th row and j-th column of $\mathbb{I}_l$ equals 1 if v_i is matched with v'_j in the l-th subgraph isomorphism, and 0 otherwise.

Then, we apply the local topological constraint. For each matching pair of vertices, we extract the degree deg by counting the number of edges linked to each vertex, as well as the absolute difference of degrees: $\delta_{deg_{i,j}}(l) = |deg(v_i) - deg(v'_j)|$, where $\mathbb{I}_l(i, j) = 1$. We discard the subgraph isomorphism candidate $\mathbb{I}_l$, if there is a matching pair (v_i, v'_j) for which $\delta_{deg_{i,j}}(l) > \theta$. The global constraint is applied by calculating the total edge difference $\Delta_{deg}(l) = \sum \delta_{deg_{i,j}}(l)$ for the subgraph isomorphism $\mathbb{I}_l$ and discarding the candidate $\mathbb{I}_l$ if $\Delta_{deg}(l) > \tau$. The threshold parameters τ and θ are experimentally determined, and their impact on the performance is discussed in Section 4.2.

To measure the quality of the subgraph isomorphism $\mathbb{I}_l$, we introduce the subset Ω^d which contains all subgraph

---

Algorithm 1: Generating the first order similarity function using structural topology

> **Input** : G, G'
> **Output**: $\mathcal{F}^1$
> $\mathbb{I} \leftarrow \text{VF2}(G, G')$
> $\mathcal{M} \leftarrow$ calculate by Eq.(3) based on $\mathbb{I}$
> **for** $j \in [1 : N']$ **do**
> > $\mathbb{I}' \leftarrow \text{VF2}(G, \{G' \setminus v_j\})$
> > $\mathcal{M}'_j \leftarrow$ calculate by Eq.(3) based on $\mathbb{I}'$
>
> $\mathcal{F}^1 \leftarrow \frac{1}{N'} \sum_{j=1}^{N'} \frac{\mathcal{M} + \mathcal{M}'_j}{2}$

isomorphisms (Ω^d_ω) of $\mathbb{I}_l$, where $\Delta_{deg}(l) = d$. Thus, Ω^d contains subgraph isomorphisms with similar quality, and a lower difference d indicates a higher similarity. We generate a probabilistic correspondence matrix $\mathcal{M}$ as follows:

$$\mathcal{M} = \sum_{d=0}^{\tau} \frac{1}{d+1} \sum_{\omega=1}^{|\Omega^d|} \frac{\Omega^d_\omega}{|\Omega^d|}. \tag{3}$$

We normalise $\mathcal{M}$ such that the sum over all entries in one row equals 1, as each row i contains the probability that v_i matches vertex v'_j.

In order to cope with imperfections in the estimated kinematic structure, we also consider possible variations $\mathcal{M}'_j$ of G', where j indicates that node v'_j is removed along with the respective edges. The set $\mathcal{M}' = \{\mathcal{M}^{G' \setminus v'_1}, ..., \mathcal{M}^{G' \setminus v'_N}\}$ contains all subgraphs which can be found while sequentially removing one vertex.

The first order similarity matrix $\mathcal{F}^1$ based on structural topology is then found as the mean between the probabilistic correspondence matrix with and without node removal:

$$\mathcal{F}^1_{(i,i')} = \frac{1}{N'} \sum_{j=1}^{N'} \left(\frac{\mathcal{M}(i, i') + \mathcal{M}'_j(i, i')}{2} \right). \tag{4}$$

The proposed procedure is described in Algorithm 1.

3.2.2 Second Order Similarity: Kinematic Correlation

We measure the kinematic correlation from a pairwise motion and skeletal distance between two nodes. As was presented in [3], the kinematic distance $\mathbf{P}$ within a kinematic structure G can be effectively measured by considering both relative moving velocity difference and geodesic distance along the skeleton. Thus, the distance $\mathbf{P}_{(i,j)}$ between v_i and v_j is calculated based on the segment centre points y_i and y_j as follows:

$$\mathbf{P}_{(i,j)} = \underset{f \in F}{\text{median}} \{\|(y_i^f - y_i^{f-1}) - (y_j^f - y_j^{f-1})\| \times \zeta(y_i^f, y_j^f; \Psi^f)\}. \tag{5}$$

We take the median value over all frames F in order to be robust to outliers. The Ψ indicates the skeleton distance map generated by [3], and the skeletally geodesic distance ζ is

measured by a shortest path connecting the two node points within Ψ [3]. A large distance value implies that the pairwise nodes are skeletally apart and move with different velocity. Thus, the second-order similarity function is calculated as:

$$\mathcal{F}^2_{(i,i')(j,j')} = \exp(-\|\mathbf{P}_{(i,j)} - \mathbf{P}_{(i',j')}\|). \qquad (6)$$

3.2.3 Third Order Similarity: Combinatorial Motion

We consider characteristic combinatorial local motions, which are shared between different kinematic structures. As was discussed in [10, 20], third order feature combination leads to geometric invariances and better represents local information. Similarly, we consider the combinatorial kinematic rotation ranges of three nodes which capture local motion characteristics whilst being invariant to scale-rotation-translation and movement direction. As widely used in mechanical kinematics, we utilise the joint limits to describe its dynamic ranges. To characterise the motions which are executed by a revolute kinematic joint, we extract its motion range, *i.e.* the minimum and maximum rotation angle in respect to another joint.

To build the combinatorial motion range descriptor, we first need to find the joint positions. The location of a joint between two motion segments can be approximated as the point where two segments S_i and S_j encounter each other, as shown in the left half of Figure 2. For each frame f, we find the M nearest points of S_i to the segment centre v_j and vice versa, and denote them as neighbouring points $\mathcal{N}_{i \to j}$ (and $\mathcal{N}_{j \to i}$ respectively). The joint position J_{i-j} is then defined by $J_{i-j} = \frac{1}{2M} \sum_{m=1}^{M} (\mathcal{N}_{i \to j}(m) + \mathcal{N}_{j \to i}(m))$. Note that $\mathcal{N}_{i \to j} \neq \mathcal{N}_{j \to i}$, but $J_{i-j} = J_{j-i}$ for $i \neq j$.

Our novel combinatorial motion range descriptor is then built as follows. We consider three body parts (i, j, k) at frame f with segment centres y_i^f, y_j^f, y_k^f and their respective revolute joint positions J_{i-j}^f, J_{j-k}^f and J_{i-k}^f. We find the vectors between all joints and their respective centre points, *i.e.* $\mathbf{v}_{i,i-j}^f = y_i^f - J_{i-j}^f$, $\mathbf{v}_{j,i-j}^f = y_j^f - J_{i-j}^f$. Then, we find the directed angle α_{i-j}^f between the two vectors (see also right half of Figure 2):

$$\alpha_{i-j}^f = \angle(\mathbf{v}_{i,i-j}^f, \mathbf{v}_{j,i-j}^f) = \arctan\left(\frac{\|\mathbf{v}_{i,i-j}^f \times \mathbf{v}_{j,i-j}^{f \, T}\|}{\mathbf{v}_{i,i-j}^f \cdot \mathbf{v}_{j,i-j}^{f \, T}}\right).$$
$$(7)$$

Finally, we build the rotation matrix R_{i-j}^f based on α_{i-j}^f.

In the next step, we calculate the geodesic distance on the Riemannian manifold between all pairwise rotation matrices which describe combinatorial movements of three body parts. The geodesic distance measure has been shown to be a highly meaningful metric to describe 3D rotations [14, 16]. For combinatorial body parts (i, j, k), we define the geodesic distance vector of pairwise rotation matrices as follows:
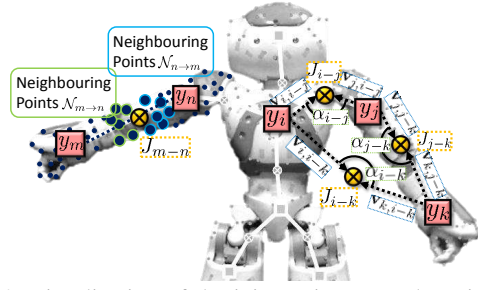
---

[3]We use the code available at http://www.imperial.ac.uk/PersonalRobotics to calculate each distance term.

Figure 2. Visualization of the joint estimates and motion range descriptor. The left half shows the segment centre positions y, and the neighbouring points $\mathcal{N}$ to find the joint positions J. The right half visualises the vectors $\mathbf{v}$ which are used to calculate the angles α (Eq.7). Best viewed in colour.

$$d_{(i,j,k)}^f = \|\mathrm{logm}(R_{i-j}^{f \, T} R_{j-k}^f)\|_F \qquad (8)$$

where logm is the principal matrix logarithm, and $\|\cdot\|_F$ is the Frobenius norm.

We then merge the sequence of geodesic distances $d_{(i,j,k)}^f$ and form a 6 dimensional feature vector $\Upsilon_{(i,j,k)}$. The entries of the vector are the minimum and maximum distances found over all frames for body parts (i, j, k). This corresponds to the minimum and maximum range of all possible combinatorial motions of body parts i, j and k. Thus, the feature vector to describe the combinatorial local motion is built as:

$$\Upsilon_{(i,j,k)} = [\min_{f \in F} d_{(i,j,k)}^f, \max_{f \in F} d_{(i,j,k)}^f,$$
$$\min_{f \in F} d_{(j,k,i)}^f, \max_{f \in F} d_{(j,k,i)}^f,$$
$$\min_{f \in F} d_{(k,i,j)}^f, \max_{f \in F} d_{(k,i,j)}^f]. \qquad (9)$$

To compare two feature vectors $\Upsilon_{(i,j,k)}$ and $\Upsilon_{(i',j',k')}$, we define the third order similarity measure as:

$$\mathcal{F}^3_{(i,i')(j,j')(k,k')} = \exp(-\|\Upsilon_{(i,j,k)} - \Upsilon_{(i',j',k')}\|). \quad (10)$$

We want to highlight that the presented metric is super-symmetric, and thus can be represented as a tensor product.

The local combinatorial motion metric has various benefits over other metrics. Using the geodesic distance on the Riemannian manifold has been found to be geometrically meaningful to describe 3D rotations [14]. The geodesic distance is invariant to the absolute rotation, as only the relative rotation between the body parts is captured [24]. Furthermore, it is invariant to scale, as the translation between the body parts is not considered in the geodesic distance.

4. Experiments

We evaluated our method on synthetic benchmarks and real image sequences generated from various objects by comparing it with state-of-the-art structure alignment methods and appearance based correspondence matching methods. In particular, we used the following graphical structure alignment approaches: NETAL [25] and MAGNA++ [37]. For
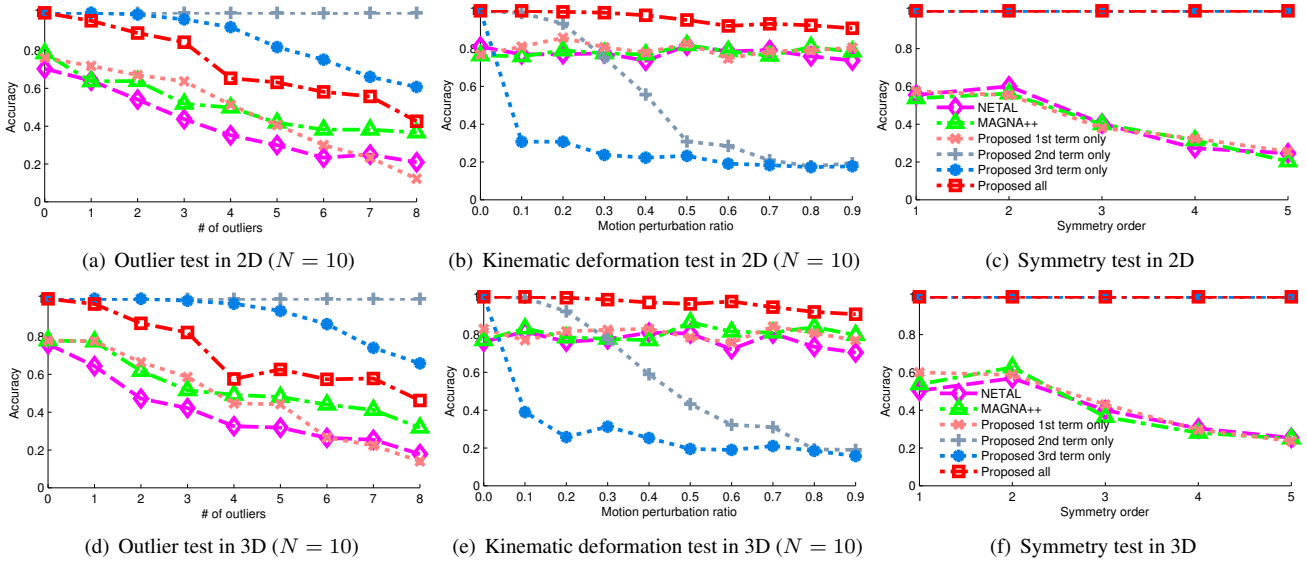
Figure 3. Performance according to outlier tests (left), kinematic deformation ratio (centre) and symmetry order changes (right). The kinematic structures in the top row are generated in 2D and the bottom row's results are from 3D structures. It can be observed that our method achieves the best performance over all other algorithms in most cases.

the appearance based approaches, we used Agglomerative Correspondence Clustering (ACC) [4], Reweighted Random Walks for Graph Matching (RRWM) [5], and Progressive Graph Matching (PGM) [6]. For all comparisons, the authors' original implementations were used. Also, we set $\tau = 3$ and $\theta = 1$. To encourage more following researches on this newly proposed problem we release our code along with our new dataset. All experiments were performed using a PC with a Intel Core i7-4770 CPU @ 3.40GHz (x8) and 32GB of RAM.

4.1. Synthetic Dataset

In this experiment we performed various comparative evaluations on synthetically generated kinematic structures in both 2D and 3D. For each trial, we randomly constructed a kinematic structure graph G and created a second graph G' by perturbing G. Then, we compared state-of-the-art graph alignment algorithms [25, 37] trying to find correspondences between the two structures. Each quantitative result in these experiments is acquired from averages of 100 random trials. For the first structure G, we randomly generated node positions, their movements and kinematic correlations, and the kinematic structure is built based on them. The kinematic range of each node is assigned by a uniform random distribution $\mathcal{U}(0, \sigma_m)$ and used to simulate dynamic movements. The kinematic joints are located in the middle of the connected nodes following the structure. The number of frames and the parameter σ_m do not affect the result, and are set to $F = 100$ and $\sigma_m = 50$.

First, we increased the number of outlier nodes in G' and randomly set their kinematics while preserving the kinematics of all other nodes. As shown in Figures 3(a) and 3(d),
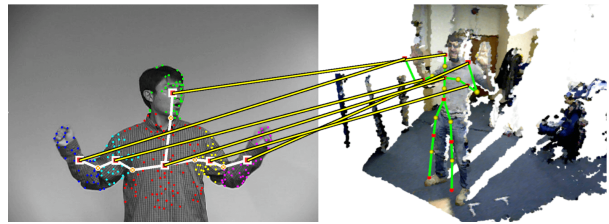

Figure 4. The proposed method can establish kinematic correspondences between RGB video and depth video.

even though there are severe topology changes, our proposed method finds matches more accurately than graph alignment methods, as our method benefits from the kinematic and motion terms.

Second, we increased each node's kinematic range in G' while not changing the structure topology. To generate G', we perturb G by adding newly generated random motions from $\mathcal{U}(0, \rho * \sigma_m)$, where ρ induces a motion perturbation ratio. The results are shown in Figures 3(b) and 3(e). Even though the motion terms are deteriorated by the motion perturbation, the first term helps to establish correct matches robustly, resulting in an overall high accuracy.

Third, we tested the robustness to structures having symmetric topology but non-symmetric motions. The symmetry order indicates a number of possible symmetric axes. As we can see in Figures 3(c) and 3(f), the graph alignment methods were easily confused by structure symmetry, but our method can find correct matches using the motion information. Through this test we have validated that the kinematic and motion information are characteristic properties for dynamic structures as much as structural topology and the proposed kinematic correlation term and combinatorial motion term are effectively designed.

4.2. Real Kinematic Structure Dataset

The Imperial-PRL-Dataset [3] has been used for testing the complex kinematic structure generation performance. It contains from simple structures to highly complex structures such as a human upper body and a human hand. However, the kinematic structures are not diverse enough to validate the correspondence matching performance. Thus, we newly constructed more sequences using various humanoid robots[4] and human motion, and have performed many tests with various combinations[5]. We generated the kinematic structures G and G' using [3] with manually set segment numbers, and they are used as input pairs for the proposed method and the graph alignment methods.

For the object appearance based matching methods, we used the first image of each sequence as input. Initial candidate feature correspondences were generated using the MSER detector [23] and the SIFT descriptor [22] as was described in [6]. Then, ACC [4], RRWM [5] and PGM [6] are applied separately using the codes provided by the authors.

As shown in Figure 5, our method clearly outperforms other methods for finding accurate kinematic correspondence. Especially the graph alignment approaches often failed to distinguish symmetric matches. It is clear that the appearance based correspondence matching approaches can not be applied to heterogeneous objects. These results on real image data are in line with the tests on synthetic data. Our method is able to establish similar kinematic structure matches even between visually totally different appearances and in the presence of strong motion variations.

Figure 6 shows more correspondence matching results between heterogeneous objects. Even though there are many outlier nodes in G', it finds accurate matches. Especially, the bottom left matching result is interesting. The matches are upside-down, as the left iCub is moving its hands downwards, whilst the right iCub is waving its hands upwards. This shows that the combinatorial motion term can distinguish motion directions.

In Table 1, we show the quantitative matching accuracy based on manually generated ground truth matchings by applying the proposed similarity function (Eq.(1)) to various hypergraph matching methods. It can be seen that our method outperforms other methods when used in conjunction with RRWHM [20]. This is because the stochastic scheme of RRWHM can update the correspondence matrix more robustly than other optimisation-based methods, especially when the graph structure is dynamic and largely deformed by motions. Furthermore, Table 1 shows that the proposed term normalisation weights $w_{\kappa(e)}$ play an important role by making a balance between the structural topology and mo-

---
[4] We utilised three robots: iCub (www.icub.org), NAO (www.aldebaran.com) and Baxter (www.rethinkrobotics.com)

[5] The extended dataset is available at www.imperial.ac.uk/PersonalRobotics.
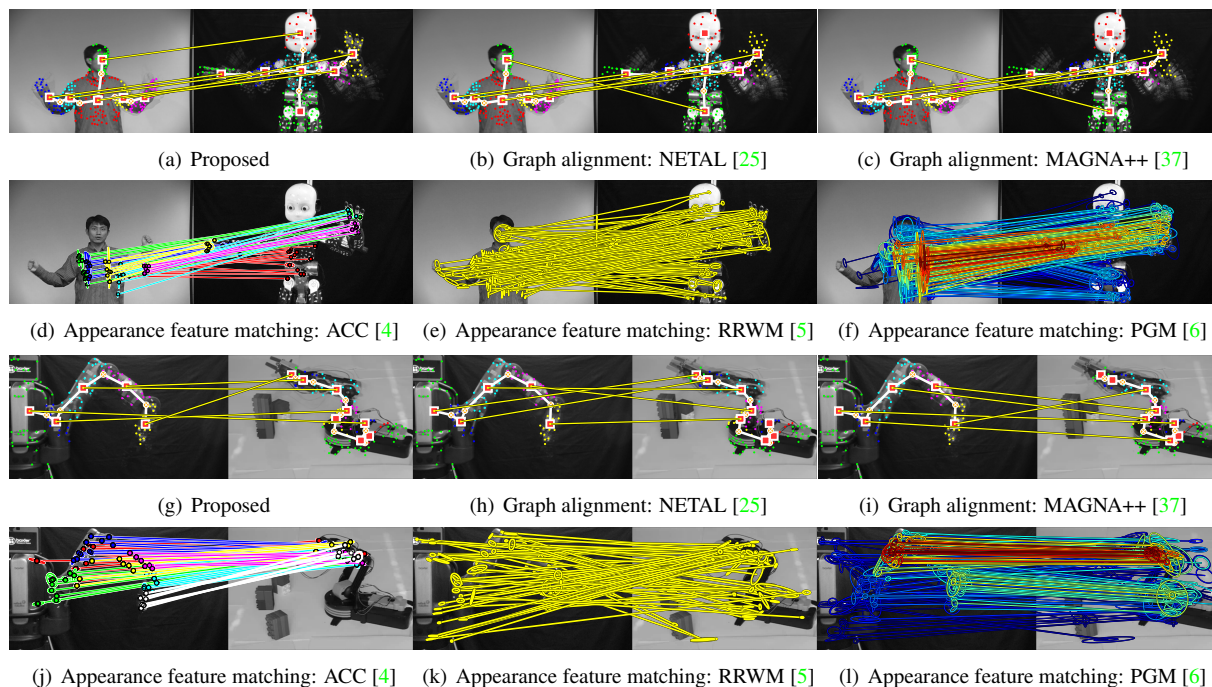
Table 1. Performance on the real kinematic structure dataset.

Methods	Accuracy (%)
Proposed $\mathcal{F}$ to HGM [43]	35.06($\pm$30.73)
Proposed $\mathcal{F}$ to TM [10]	29.76($\pm$21.65)
Proposed $\mathcal{F}$ to BCAGM+MP [26]	69.09($\pm$20.65)
NETAL [25]	67.72($\pm$39.59)
MAGNA++ [37]	63.42($\pm$33.66)
Proposed $\mathcal{F}$ to RRWHM [20] without weight normalisation Eq.(2)	88.23($\pm$13.07)
Proposed $\mathcal{F}$ to RRWHM [20] with weight normalisation Eq.(2)	**92.99 ($\pm$ 10.41)**

tion similarity. We have observed that most failure cases of RRWHM without the normalisation term were due to mismatched nodes having similar motions.

Furthermore, we consider the kinematic structure as a mid-level representation, so the proposed method can be applied to any kind of input device as long as the kinematic structure can be produced. As shown in Figures 1 and 4, the proposed method can even find kinematic structure matches across different sensors.

We evaluated our method on the MPI Dexter 1 hand dataset [30] in order to validate whether it can be used for matching similar actions. The dataset consists of 7 sequences of hand motions of a single actor. Because all sequences are from one actor's motion, we measured correct hand joints matching accuracy between similar motion pairs: 'fingercount'-'fingerwave', 'pinch'-'tigergrasp' *etc*. The kinematic nodes and joint positions are estimated from the provided fingertip positions and depth images. We achieved an accuracy of 86.67% (no parameter tuning), compared to 25.56% (NETAL [25]) and 31.11% (MAGNA++ [37]). This demonstrates the robustness of our third order motion term, as the dataset's hand topological structure is ambiguous.

4.3. Experimental Analysis

Time complexity analysis: On our proposed real dataset, it takes 1.69 ± 1.09s, 0.11 ± 0.07s, and 2.20 ± 1.70s for 1st, 2nd and 3rd order respectively (non-optimized Matlab implementation), as well as 0.02 ± 0.001s for the RRWHM hypergraph matching [20] (C++ implementation). As most kinematic structures contain only few nodes (in our real dataset from 4 to 11), VF2 is fast enough in our application. Furthermore, the time complexity of VF2 is similar to other state-of-the-art subgraph isomorphism algorithms [2].

Validations on the parameters of VF2: If the local constraint $\theta \geq 3$, the performance decreases as the constraint is rarely triggered. For the global constraint τ, a low value of $\tau \leq 2$ decreases performance as it does not allow enough noise. For $\tau \geq 3$, the impact of τ is low, as Eq.(3) weights subgraphs inverse proportionally to the total difference in degree of the graphs. Detailed experimental results are provided as a supplementary document.

(a) Proposed   (b) Graph alignment: NETAL [25]   (c) Graph alignment: MAGNA++ [37]

(d) Appearance feature matching: ACC [4]   (e) Appearance feature matching: RRWM [5]   (f) Appearance feature matching: PGM [6]

(g) Proposed   (h) Graph alignment: NETAL [25]   (i) Graph alignment: MAGNA++ [37]

(j) Appearance feature matching: ACC [4]   (k) Appearance feature matching: RRWM [5]   (l) Appearance feature matching: PGM [6]

Figure 5. Experiments on real image datasets. Two static images are used for the appearance based methods (best viewed in colour).
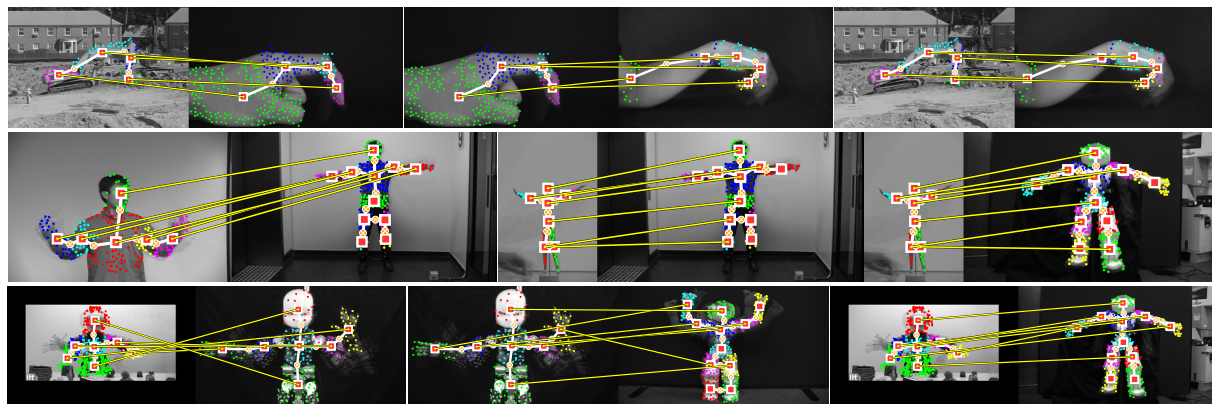

Figure 6. Various kinematic structure correspondence matching results using the proposed method (best viewed in colour).

5. Conclusion and Future Works

We have presented a novel approach to find kinematic structure correspondences between heterogeneous objects via the hypergraph matching method. Our method establishes both structural topology correspondences and their kinematics-based matches, effectively eliminating outliers from structure and motion variations. To find the structural topology similarity, we proposed a topologically constrained subgraph isomorphism aggregation which is robust to noise in the kinematic structures. For motion similarity, we employed the logarithm of the rotations between motion segments, which results in the geodesic distance between the kinematic joints. This has the advantage of being invariant to scale, orientation, and translation; leading to an universal combinatorial motion descriptor. Under variations of the structure topology, kinematics and symmetry, it can be observed that our method achieves the best performance over all other algorithms in most cases. As this is the first work on kinematic structure correspondence, we provide a challenging dataset containing the ground truth correspondences of various kinematic structures. Using this dataset, we have shown that our proposed method outperforms other existing approaches which are based on either solely structural topology, or object appearance. We plan to apply our method to learning by imitation tasks in robotics. There, finding correspondences between the kinematic structures of the human and the robot is one of the main problems, which can be effectively addressed using our method.

References

[1] X. Bai and L. J. Latecki. Path Similarity Skeleton Graph Matching for 3D Objects. *IEEE Trans. on PAMI*, 30(7):1282–1292, 2008. 1, 2

[2] V. Carletti, P. Foggia, and M. Vento. Performance comparison of five exact graph matching algorithms on biological databases. In *ICIAP*, pages 409–417, 2013. 7

[3] H. J. Chang and Y. Demiris. Unsupervised Learning of Complex Articulated Kinematic Structures combining Motion and Skeleton Information. In *CVPR*, pages 3138–3146, 2015. 1, 2, 3, 4, 7

[4] M. Cho, J. Lee, and K. M. Lee. Feature Correspondence and Deformable Object Matching via Agglomerative Correspondence Clustering. In *ICCV*, pages 1280–1287, 2009. 1, 6, 7, 8

[5] M. Cho, J. Lee, and K. M. Lee. Reweighted Random Walks for Graph Matching. In *ECCV*, pages 492–505, 2010. 1, 2, 6, 7, 8

[6] M. Cho and K. M. Lee. Progressive graph matching: Making a move of graphs via probabilistic voting. In *CVPR*, pages 398–405, 2012. 1, 6, 7, 8

[7] L. P. Cordella, P. Foggia, C. Sansone, and M. Vento. An Improved Algorithm for Matching Large Graphs. In *Proc. IAPR Workshop on Graph-Based Representations in Pattern Recognition*, pages 149–159, 2001. 4

[8] L. P. Cordella, P. Foggia, C. Sansone, and M. Vento. A (Sub)Graph Isomorphism Algorithm for Matching Large Graphs. *IEEE Trans. on PAMI*, 26(10):1367–1372, 2004. 4

[9] T. Cour, P. Srinivasan, and J. Shi. Balanced Graph Matching. *NIPS*, 19:313–320, 2007. 2

[10] O. Duchenne, F. Bach, I.-S. Kweon, and J. Ponce. A Tensor-Based Algorithm for High-Order Graph Matching. *IEEE Trans. on PAMI*, 33(12):2383–2395, Dec 2011. 2, 5, 7

[11] J. Fayad, C. Russell, and L. Agapito. Automated articulated structure and 3D shape recovery from point correspondences. In *ICCV*, pages 431–438, 2011. 1, 2

[12] R. Hartley, J. Trumpf, Y. Dai, and H. Li. Rotation Averaging. *IJCV*, 103(3):267–305, 2013. 2

[13] X. Huang, I. Walker, and S. Birchfield. Occlusion-aware multi-view reconstruction of articulated objects for manipulation. *Robot. Auton. Syst.*, 62:497–505, 2014. 1

[14] D. Q. Huynh. Metrics for 3D Rotations: Comparison and Analysis. *J. Math. Imaging and Vis.*, 35(2):155–164, 2009. 2, 5

[15] B. Jacquet, R. Angst, and M. Pollefeys. Articulated and Restricted Motion Subspaces and Their Signatures. In *CVPR*, pages 1506–1513, 2013. 2

[16] C. Jia and B. Evans. 3D rotational video stabilization using manifold optimization. In *ICASSP*, pages 2493–2497, 2013. 2, 5

[17] D. Katz, M. Kazemi, J. A. Bagnell, and A. Stentz. Interactive Segmentation, Tracking, and Kinematic Modeling of Unknown 3D Articulated Objects. In *ICRA*, pages 5003–5010, 2013. 1

[18] M. Koyutürk, Y. Kim, U. Topkara, S. Subramaniam, W. Szpankowski, and A. Grama. Pairwise alignment of protein interaction networks. *J. Comput. Biol.*, 13(2):182–199, 2006. 2

[19] J. Lee, M. Cho, and K. M. Lee. A Graph Matching Algorithm Using Data-Driven Markov Chain Monte Carlo Sampling. In *ICPR*, pages 2816–2819, 2010. 2

[20] J. Lee, M. Cho, and K. M. Lee. Hyper-graph Matching via Reweighted Random Walks. In *CVPR*, pages 1633–1640, 2011. 1, 2, 3, 5, 7

[21] M. Leordeanu and M. Hebert. A Spectral Technique for Correspondence Problems Using Pairwise Constraints. In *ICCV*, pages 1482–1489, 2005. 2

[22] D. G. Lowe. Object recognition from local scale-invariant features. In *ICCV*, pages 1150–1157, 1999. 4, 7

[23] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *BMVC*, pages 36.1–36.10, 2002. 7

[24] M. Moakher. Means and Averaging in the Group of Rotations. *SIAM J. Matrix Anal. Appl.*, 24(1):1–16, Jan. 2002. 5

[25] B. Neyshabur, A. Khadem, S. Hashemifar, and S. S. Arab. NETAL: A new graph-based method for global alignment of protein-protein interaction networks. *Bioinformatics*, 29(13):1654–1662, 2013. 2, 5, 6, 7, 8

[26] Q. Nguyen, A. Gautier, and M. Hein. A Flexible Tensor Block Coordinate Ascent Scheme for Hypergraph Matching. In *CVPR*, pages 5270–5278, 2015. 1, 2, 3, 7

[27] D. Ross, D. Tarlow, and R. Zemel. Learning Articulated Structure and Motion. *IJCV*, 88(2):214–237, 2010. 1, 2

[28] J. Schulz, S. Jung, S. Huckemann, M. Pierrynowski, J. S. Marron, and S. M. Pizer. Analysis of Rotational Deformations From Directional Data. *J. Comp. Graph. Stat.*, 24(2):539–560, 2015. 2

[29] W. Shen, Y. Wang, X. Bai, H. Wang, and L. J. Latecki. Shape clustering: Common structure discovery. *Pattern Recognition*, 46(2):539–550, 2013. 2

[30] S. Sridhar, A. Oulasvirta, and C. Theobalt. Interactive Markerless Articulated Hand Motion Tracking Using RGB and Depth Data. In *ICCV*, pages 2456–2463, 2013. 7

[31] J. Sturm, C. Stachniss, and W. Burgard. A Probabilistic Framework for Learning Kinematic Models of Articulated Objects. *J. Artif. Intell. Res.*, 41:477–626, 2011. 1

[32] L. Torresani, V. Kolmogorov, and C. Rother. Feature correspondence via graph matching: Models and global optimization. In *ECCV*, pages 596–609, 2008. 2

[33] T. Tosun, R. Mead, and R. Stengel. A General Method for Kinematic Retargeting: Adapting Poses Between Humans and Robots. In *Proc. ASME Int. Mech. Eng. Congress and Exposition*, pages 1–10, 2014. 1

[34] P. Tresadern and I. Reid. Articulated Structure From Motion by Factorization. In *CVPR*, pages 1110–1115, 2005. 2

[35] G. Valiente. *Algorithms on Trees and Graphs*. Springer, Berlin, 2002. 4

[36] R. Vemulapalli, F. Arrate, and R. Chellappa. Human Action Recognition by Representing 3D Skeletons as Points in a Lie Group. In *CVPR*, pages 588–595, 2014. 1

[37] V. Vijayan, V. Saraph, and T. Milenković. MAGNA++: Maximizing Accuracy in Global Network Alignment via both node and edge conservation. *Bioinformatics*, 31(14):2409–2411, 2015. 2, 5, 6, 7, 8

[38] Y. Wu and Y. Demiris. Towards one shot learning by imitation for humanoid robots. In *ICRA*, pages 2889–2894, 2010. 1

[39] Y. Wu and Y. Demiris. Learning dynamical representations of tools for tool-use recognition. In *ROBIO*, pages 2664–2669, 2011. 1

[40] J. Yan and M. Pollefeys. Automatic Kinematic Chain Building from Feature Trajectories of Articulated Objects. In *CVPR*, pages 712–719, 2006. 2

[41] J. Yan and M. Pollefeys. A Factorization-Based Approach for Articulated Nonrigid Shape, Motion and Kinematic Chain Recovery From Video. *IEEE Trans. on PAMI*, 30(5):865–877, May 2008. 1, 2, 3

[42] J. Yan, C. Zhang, H. Zha, W. Liu, X. Yang, and S. M. Chu. Discrete Hyper-Graph Matching. In *CVPR*, pages 1520–1528, 2015. 2

[43] R. Zass and A. Shashua. Probabilistic Graph and Hypergraph Matching. In *CVPR*, pages 1–8, 2008. 2, 7

[44] Y. Zeng, C. Wang, Y. Wang, X. Gu, D. Samaras, and N. Paragios. Dense Non-rigid Surface Registration Using High-Order Graph Matching. In *CVPR*, pages 382–389, 2010. 3

[45] F. Zhou and F. D. la Torre. Deformable Graph Matching. In *CVPR*, pages 2922–2929, 2013. 2