Preprint version; final version available at http://ieeexplore.ieee.org/document/8119820
IEEE Transactions on Pattern Analysis and Machine Intelligence (2017)
DOI: 10.1109/TPAMI.2017.2777486

Learning Kinematic Structure Correspondences Using Multi-Order Similarities

Hyung Jin Chang, *Member, IEEE,* Tobias Fischer, *Student Member, IEEE,* Maxime Petit, *Member, IEEE,* Martina Zambelli, *Student Member, IEEE,* and Yiannis Demiris, *Senior Member, IEEE*

Abstract—In this paper, we present a novel framework for finding the kinematic structure correspondences between two articulated objects in videos via hypergraph matching. In contrast to 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. Thus our method allows matching the structure of objects which have similar topologies or motions, or a combination of the two. Our main contributions can be summarised as follows: (i) casting the kinematic structure correspondence problem into a hypergraph matching problem by incorporating multi-order similarities with normalising weights, (ii) introducing a structural topology similarity measure by aggregating topology constrained subgraph isomorphisms, (iii) measuring kinematic correlations between pairwise nodes, and (iv) proposing 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 synthetic and real data, outperforming various other state of the art methods. Our method is not limited to a specific application nor sensor, and can be used as building block in applications such as action recognition, human motion retargeting to robots, and articulated object manipulation amongst others.

Index Terms—Articulated kinematic structure correspondences, hypergraph matching, subgraph isomorphism aggregation, kinematic correlation, combinatorial local motion similarity, humanoid robotics.

1 Introduction

A kinematic structure represents motion properties as well as shape information of an object in a topological manner [2], [3]. The encoded relationship between rigid body parts connected by kinematic joints can be considered as a mid-level representation of general objects. In this paper, we focus on finding correspondences between two kinematic structures extracted from different objects' image sequences using a new hypergraph matching framework. Such accurate and efficient estimation of kinematic correspondences of heterogeneous objects is beneficial in the computer vision and robotics fields. Some application areas are learning by imitation [4], human motion retargeting to robots [5], [6], human action recognition from different sensors [7], behaviour discovery and alignment [8], affordance based object/tool categorisation [9], body scheme learning for robotic manipulators [10], and articulated object manipulation [11], [12]. Although our framework can be applied to generic objects, we mainly focus on sequences containing humans and various robots as they are involved in the aforementioned applications (see Fig. 1).

As building kinematic structures of articulated objects from visual input data is an active research topic both in computer vision and robotics, various approaches to generate accurate kinematic structures have been presented [2], [13], [14], [15]. However, there are few works utilising the generated kinematic structures for the aforementioned high level tasks. For instance, Fayad *et al.* [15] used the kinematic structure as a basis of 3D reconstruction. Sturm *et al.* [16] and Katz *et al.* [12] applied kinematic structures to robot manipulations, but their methods are restricted to relatively simple objects. We propose a new way of utilising the kinematic structure for matching kinematic correspondences between complex articulated objects belonging to different categories, such as human bodies, robots, and animals. We also extend our framework to applications with heterogeneous sensory data including RGB and RGB-D sequences. A key application lies in the robotics field, where the correspondence between a human teacher and a humanoid robot can be established, which allows imitation learning without a pre-defined model of the human.

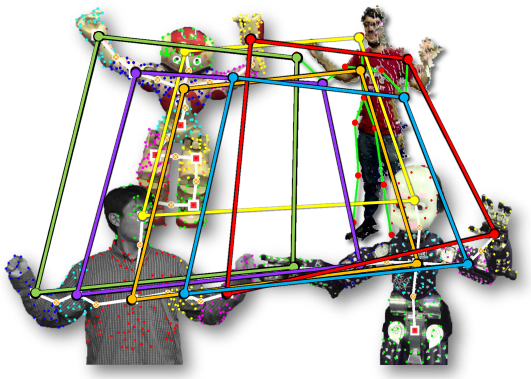


Fig. 1. The proposed framework reliably builds up kinematic structure correspondence matches across heterogeneous objects captured with different sensors. Our method can for example find correspondences between a upper-body dancing human in a 2D grey image sequence, the iCub and NAO humanoid robots in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

Most conventional correspondence finding methods in the

- *All authors are with the Department of Electrical and Electronic Engineering, Imperial College London, United Kingdom, SW7 2AZ. E-mail: {hj.chang, t.fischer, m.petit, m.zambelli13, y.demiris}@imperial.ac.uk*
- *Research presented in this paper is a continuation of Chang et al. [1] and includes results from [1].*
- *Acknowledgements: This work was supported in part by EU FP7 project WYSIWYD under Grant 612139. We thank Minsu Cho for fruitful discussions. The authors gratefully acknowledge the support from the members of the Personal Robotics Lab.*

computer vision area are restricted to two static images of the same category [17], [18], [19] or the same object with different poses or views [20], [21], [22], [23]. Local shape feature and graph matching based methods [24], [25], [26] have been researched actively for decades. However, most of these approaches are based on object appearance and local shape features applied to two static images, which do not include dynamic information.

Graph matching methods have been widely used for the correspondence matching problem [27], [28]. Hypergraph matching is an extension of traditional graph matching which is limited to pairwise relationships, whereas a hypergraph permits higher-order relationships (three or more nodes), achieving robust matches even under large variations [25], [26], [29], [30], [31]. Hypergraph matching has a wide potential for finding matches among various applications, but not many applications use this method because of the difficulty in designing the high-order similarity. Most conventional similarity measures for graph matching are based on informative and discriminative local descriptors and their geometric relations, which are not suitable for our framework as our image sequences can potentially have very different appearances while still having clearly defined kinematic correspondences.

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 assume that an articulated object is composed of a set of rigid segments and the structure represents the connections between segments. We propose new measures in order to simultaneously consider structural topology (first order), kinematic correlation (second order), and combinatorial motion (third order). We incorporate these measures into the hypergraph matching framework with weight normalisation. Our experiments show that the proposed method outperforms state-of-the-art appearance based and skeleton graph alignment based methods quantitatively and qualitatively on both synthetic and real data.

In Section 2, we discuss approaches to estimate articulated structures and other correspondence matching methods. In Section 3, we present a framework for estimating the kinematic structure correspondences via hypergraph matching. In Section 4, we provide a new dataset and compare the framework with other methods in various aspects. In the final section we discuss new research directions which can emerge from our work, with a focus on possible computer vision and robotics applications.

2 Related Work

Kinematic structure building: A kinematic structure looks similar to its corresponding skeleton (a framework of bones), but also incorporates kinematic properties between rigid body parts. A general approach for the kinematic structure estimation from visual data has been segmenting rigid body parts based on motion and subsequently building connections between the segments. For example, factorisation based articulated motion segmentation methods [32], [33] were presented showing that the rank of the feature trajectory matrix can indicate the kinematic joint type. However, these methods cannot deal with a high number of articulations because the factorization is generally sensitive to non-Gaussian noise, so that few tracking errors can harm the result [34]. Yan and Pollefeys [13] estimated a kinematic chain based on intersecting motion subspaces with kinematic joints placed at the intersections. However, such accurate matrix rank detection based approaches are highly vulnerable to noisy feature trajectories. Also, critical parameters need to be tuned at each step, for example the rank estimation parameter, the local sampling size, and the highest dimension size. Moreover, these approaches have shown poor structure estimation results on complex articulations [2].

Ross *et al.* [14] proposed a probabilistic graphical model learning method that adaptively learned the number of joints and their connections. However, as the number of articulations increases, the computation time grows exponentially. Applying this method to complex articulations is hard, as it is prone to local minima. An energy based multiple model fitting approach was proposed by Fayad *et al.* [15], which simultaneously performs segmentation and 3D reconstruction. The benefits of this approach are that neither assumptions about the skeleton structure of the object, nor the number of motion segments are required in advance. The set of 2D feature points is decomposed into overlapping rigid-bodies, and the kinematic structure joints are derived from the regions of the overlap. This method has shown impressive 3D articulated shape reconstruction performances even for complex structures. However, the focus is on full 3D reconstruction of articulated objects requiring overlaps between motion segments in order to resolve the depth scale ambiguity in 3D reconstruction. Recently, Chang and Demiris [3] presented a complex structure estimation method by combining motion and skeleton information, showing state-of-the-art estimation performance even for complex objects.

Structure correspondence matching: A path similarity based skeleton graph matching was developed by Bai and Latecki [35] and was applied for shape recognition based on object silhouettes [36]. Although performing well among clean silhouette images, the method requires noiseless skeletons as input, and cannot be applied to noisy images with background clutter.

The structure correspondence matching is similar to the graph alignment problem in the bioinformatics field for the alignment of protein-protein interactions networks [37]. 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. In our case, the kinematic structure graphs are much smaller and no characteristic information about nodes is available. Moreover, we are interested in one-to-one mapping for every node of one graph to exactly one node of the other. Thus the network alignment has to be global.

One of the earliest implementations of a global network alignment algorithms is called IsoRank [38]. It is not only based on topological similarities, as sequence similarities between nodes are also taken into account. Graph aligner [39] and its variants [40] are algorithms for global network alignment based only on the network topology, using a highly constrained measure of topological similarity between two networks called “graphlet degree vector” [41].

Recently, other global alignment methods that do not necessarily rely on node similarity information have been presented. NETAL [42] and MAGNA++ [43] are among the best existing global network alignment methods relying on topological alignment [44], and they have outperformed graph aligner [39] and its variants [40]. NETAL [42] constructs global alignments greedily, and can optionally take node similarity into account. It can also be used to focus solely on topological similarities for fast network alignment. MAGNA++ [43] is based on a genetic algorithm that simultaneously optimises node and edge conservation to improve the quality of the alignment, compared to conserving nodes or edges individually. It has been extended to multiMagna++ [45] to allow multiple instead of pairwise network alignments. However, these

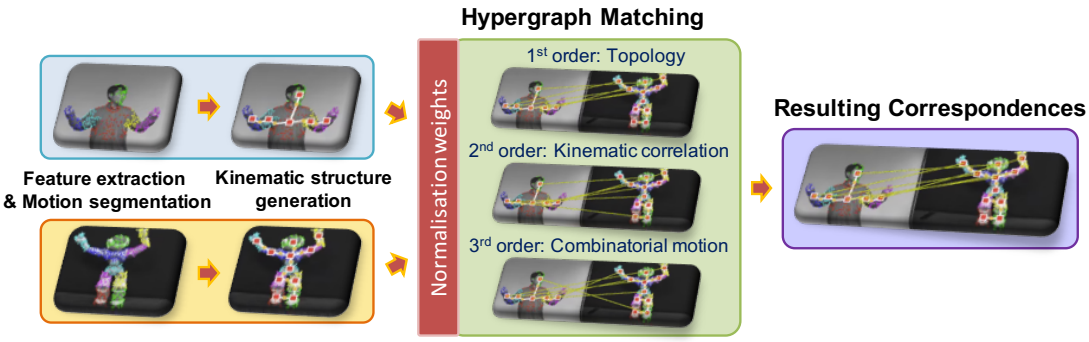


Fig. 2. Overall pipeline of the proposed method. First, features are extracted from the two image sequences, and the motion segments are estimated. Based on this, we generate the kinematic structures of the object within each image sequence. We use hypergraph matching with normalised weight terms to simultaneously consider structural topology, kinematic correlation and combinatorial motion. The resulting correspondences are robust to noise and outliers, and can be used in various computer vision and robotics applications.

methods failed to detect proper and consistent correspondences where graphs present symmetries (see Section 4).

Motion describing methods: Several approaches are concerned with describing characteristic motion patterns. Del Pero *et al.* [20] found correspondences between objects of the same class based purely on their characteristic motions. This is used for the discovery of behaviours such as walking and running, rather than matching of rigid segments as in our case. Jacquet *et al.* [46] 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.

A common issue when describing motions is to find representations which are translation, scale and rotation invariant. Hadfield *et al.* [47] extracted 3D motion descriptors for action recognition based on the scene flow for a set of interest points, and describe their motions using spherical histograms. They achieve scale invariance by normalising the histograms, and rotational invariance by applying principal components analysis to find the roll, pitch and yaw of the rotation.

A popular approach to achieve invariance in scale, translation and rotation is to use the geodesic distance on the Riemannian manifold. This is due to the fact that the geodesic distance on the Riemannian submanifold between two rotation matrices is more meaningful than other metrics from a geometric point of view, while being boundedly equivalent to other metrics [48]. Hartley *et al.* [49] analyse the problem of rotation averaging from the theoretical side, again emphasizing the importance of the geodesic distance. Schulz *et al.* introduced a framework which employs the geodesic distance to find deformations within 3D models of real objects [50]. In addition, the geodesic distance has been proven to be effective in practical applications, *e.g.* for video trajectory estimation and smoothing [51]. In contrast to our work, the authors find a rotation matrix which minimises the distance between the original (noisy) trajectory and the new trajectory, which is also smoothed at the same time. We go in the opposite direction and find the minimum and maximum rotation which can be caused by a joint. A method for separating objects based on their motions has been presented by Ochs and Brox [52]. They construct a hypergraph from triplets of trajectories, but argue that transferring the hypergraph to an ordinary graph is required to achieve scale and rotational invariance.

Hypergraph matching: Graph matching has been used in a variety of computer science applications, such as computer vision and machine learning, to find correspondences between two feature sets [53]. Classical first order methods consider node-wise unary correspondences. They are based, for example, on local descriptions, and are thus prone to ambiguities due to similar patterns or non-discriminative local appearance [30]. In second order graph matching, both nodes and edges are involved, allowing the enforcement of geometric consistency but, at the same time, making the matching problem essentially a quadratic assignment problem known to be NP hard [54]. Stochastic sampling was used in [55] to solve the graph matching problem, whereas a random walk view was proposed in [24]. Other second order methods include the work on matching between deformable graphs as proposed in [28].

Hypergraph matching is defined as taking higher order relationships between graphs into account, which allows incorporating more complex features and representations [29], [30], [56]. Zass and Shashua [29] addressed the problem in a probabilistic way and solved it using convex optimisation. Formulations of the hypergraph matching as a third order tensor optimisation problem were presented in [30], [56], which showed significant improvements over second order graph matching methods. Shashua *et al.* [57], Leordeanu and Sminchisescu [58], and Bulò and Pelillo [59] presented hypergraph clustering methods which can take the higher order relationships among sets of data points into account. The main limitation of the aforementioned methods is that they cannot model mixed higher order relationships, as each hyperedge should contain the same number of nodes [60]. Recently, a pure discrete method has been devised [61], accounting for both unary and higher order affinity terms. Lee *et al.* [25] present the reweighted random walk hypergraph matching method (RRWHM) which is an extension of the reweighted random walk [24] to the third order term. RRWHM generalises the hypergraph matching formulation to cover relations of features in arbitrary orders. Nguyen *et al.* [26] present one of the most recent methods and show that their tensor block coordinate ascent method is particularly robust to a large number of outliers.

3 Methodology

Our goal is to find corresponding 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 1) one target object exists in each scene, and 2) the features are extracted from every part of the object. To build the kinematic structure from 2D image sequences, we adopt a state-of-the-art kinematic structure generation method [2]. We

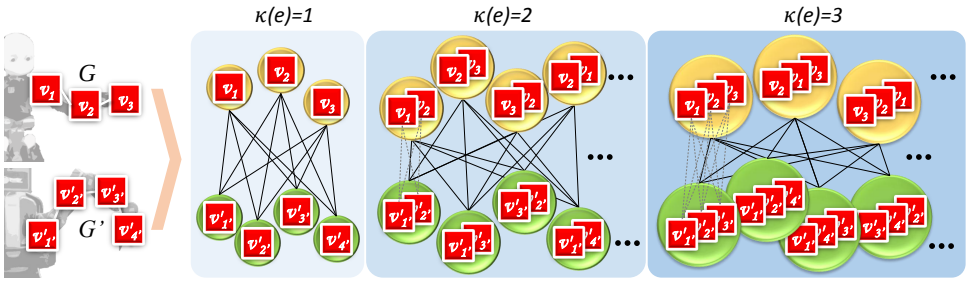


Fig. 3. Conceptual illustration of hypergraph edges showing the three different orders $\kappa(e) = 1$, $\kappa(e) = 2$ and $\kappa(e) = 3$.

also show an extension to the 3D case by employing the Kinect skeleton tracker.

In Section 3.1 we define the notations needed in the kinematic structure formulation. In Section 3.2, we outline the hypergraph matching and the similarity functions. Specifically, we introduce a new similarity term based on 2D kinematic structure topology in Section 3.2.1, followed by a pairwise term which measures kinematic correlation in Section 3.2.2, and we present a new third order combinatorial local motion term using the logarithm of rotation matrices in Section 3.2.3. In Section 3.3, we extend our framework to RGB-D sequences. An overview of our method is shown in Fig. 2.

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 $f \in \{1, ..., F\}$ as sequence index, where 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 i^{th} segment where $i \in \{1, ..., N\}$, and N as the total number of segments. We denote y_i^f as the centre position of segment S_i obtained by averaging its points at frame f.

In this work, we adopt the same way of representing the kinematic structure as [2], [13], *i.e.* we utilise 2D kinematic structures rather than 3D reconstructed structures. They use a non-cyclic graph model $G = (V, E)$ to indicate the topological connections between rigid body parts, which generally covers most articulated objects. Each node $v_i \in V$ 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. We also 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

Hypergraph matching is particularly suitable for our setting due to various reasons. Firstly, hypergraph matching does not require any pre-training, which is a requirement in our one-shot correspondence finding setting. Secondly, it allows merging heterogeneous similarity measures of varying orders, which leads to improved robustness and higher accuracy. Finally, the merging process is computationally efficient, such that the computational time is mainly dependent on the calculation of the individual terms rather than the merging.

A general hypergraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of nodes $v_i \in \mathcal{V}$, 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. A representative illustration of hyperedges is shown in Fig. 3. In this work, we generate the hypergraphs based on the two kinematic structures G and G'. We consider the hypergraph nodes $v_i \in \mathcal{V}$ as the kinematic structure nodes $v_i \in V$ (*i.e.* $V = \mathcal{V}$), so that each $e \in \mathcal{E}$ can represent any tuple of nodes (see Fig. 3).

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 (structural topology, $\mathcal{F}^1$), second (kinematic correlation, $\mathcal{F}^2$) and third order (combinatorial motion, $\mathcal{F}^3$) similarity terms, as follows:

$$\begin{aligned}\mathcal{F}(X) = w_1 \sum_{i}^{N} \sum_{i'}^{N'} \mathcal{F}^1_{(i,i')} X(i,i') \\ + w_2 \sum_{\substack{i,j \\ i \neq j}}^{N} \sum_{\substack{i',j' \\ i' \neq j'}}^{N'} \mathcal{F}^2_{(i,i')(j,j')} X(i,i') X(j,j') \\ + w_3 \sum_{\substack{i,j,k \\ i \neq j \\ j \neq k \\ i \neq k}}^{N} \sum_{\substack{i',j',k' \\ i' \neq j' \\ j' \neq k' \\ i' \neq k'}}^{N'} \mathcal{F}^3_{(i,i')(j,j')(k,k')} X(i,i') X(j,j') X(k,k').\end{aligned} \tag{1}$$

The normalising weights w_1, w_2, and w_3 play an important role in balancing the effects of the similarity terms. Previous hypergraph matching works [25], [26], [62] 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. It is important to note that in our framework each similarity term itself is already normalised to take values between 0 and 1 (see Eq.(5), Eq.(9) and Eq.(14)). As it can be seen from Eq.(1), the element set of each summation term is the same as the ordered arrangements in which no element occurs more than once, *i.e.* $\kappa(e)$-permutations of N, denoted as $P^N_{\kappa(e)}$:

$$P^N_{\kappa(e)} = \frac{N!}{(N - \kappa(e))!} \text{ for } \mathcal{G}; \; P^{N'}_{\kappa(e)} = \frac{N'!}{(N' - \kappa(e))!} \text{ for } \mathcal{G}', \tag{2}$$

where $\kappa(e) = 1$, $\kappa(e) = 2$ and $\kappa(e) = 3$ for the first, second and third order similarities, that is $e = (v_i)$, $e = (v_i, v_j)$, $e = (v_i, v_j, v_k)$, respectively. The total number of elements for each similarity term is the multiplication of $P^N_{\kappa(e)}$ and $P^{N'}_{\kappa(e)}$. Consequently, if not normalised, higher order similarity terms have more impact to the matching result than lower order terms.

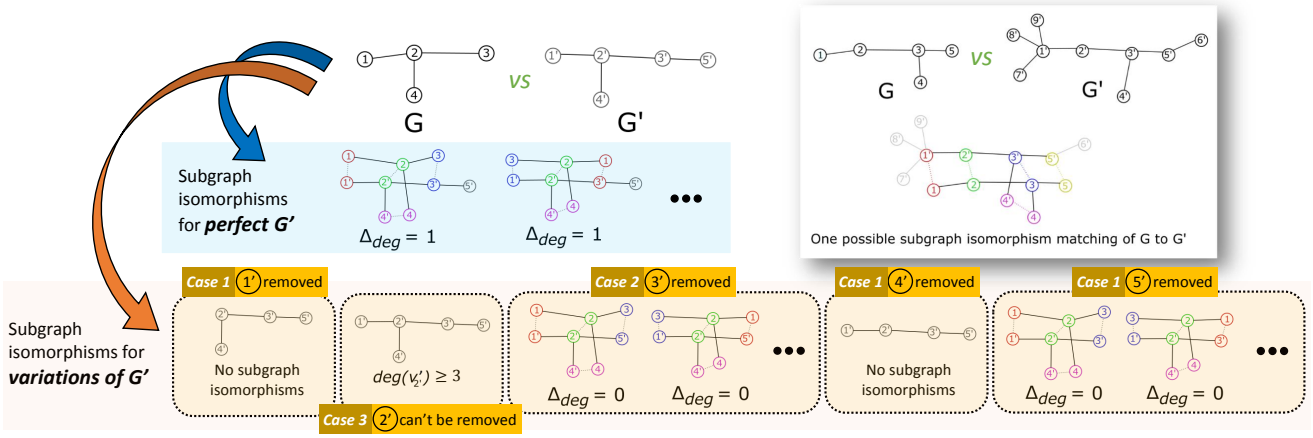


Fig. 4. Example of the consecutive subgraph isomorphisms of G into G'. The subgraph isomorphisms of G into a perfect (*i.e.* noise free, without node removal) G' are used to produce the correspondence matrix $\mathcal{M}$ (illustrated in a cyan background region). Using a node removal process, we generate the subset $\mathbb{G}'$ composed of all valid graph variations G'_k. Note that the variation $G' \setminus v'_{2'}$ is not added to $\mathbb{G}'$ as $deg(v'_{2'}) = 3$, and thus $v'_{2'}$ cannot be removed (illustrated in a yellow background region). **Top right**: Example of a possible subgraph isomorphism X_l of G into G'. Note that $\mathbf{A}^*_l(1,1') = |deg(v_1) - deg(v'_{1'})| = |1-4| = 3$ triggers the local constraint θ. The global constraint τ is checked against $\sum_{i,i'} \mathbf{A}^*_l(i,i') = 3+0+0+0+1 = 4$. The subgraph isomorphism is only kept if both the local and the global constraint are validated.

Therefore, we propose multi-order similarities with normalising weights $w_{\kappa(e)}$ which are inverse proportional to the number of elements, such that:

$$w_{\kappa(e)} = \frac{1}{P^N_{\kappa(e)}} \times \frac{1}{P^{N'}_{\kappa(e)}} = \frac{\big(N-\kappa(e)\big)!}{N!} \times \frac{\big(N'-\kappa(e)\big)!}{N'!}. \quad (3)$$

The well-balanced combination of the three similarity functions through the weights w_1, w_2, and w_3 allows us to effectively merge the three different pieces of information and obtain a more accurate and meaningful correspondence between kinematic structures.

3.2.1 First Order Similarity: Structure Topology

As the aim of our framework is to find correspondences between objects of different appearance, we cannot employ classical feature point descriptors such as SIFT [63]. Instead, the first order similarity is a measure of structural topology, *i.e.* the relations of nodes which form an object. This measure allows to find matches between objects which have similar properties in terms of their underlying structure, while being invariant to their appearance, shape and size. One way of representing kinematic structures to study their topological structure is by assimilating them as undirected graphs, where the centre of body-parts y_i are the nodes $v_i \in V$ of the network. The first order similarity can be reformulated as a graph alignment problem based solely on topological structure, *i.e.* how edges connect nodes, whereby the edges do not have any attribute or weight associated.

The general idea is as follows. We find substructures of a graph $G' = (V', E')$ which topographically match a graph $G = (V, E)$ which has less nodes than G', *i.e.* $N' \geq N$ without loss of generality (see 3.2.1.A). We then apply constraints on these matches, as we want to avoid matches which are topographically unlikely (see 3.2.1.B). More specifically, we discard matches including nodes with largely differing number of neighbours (called local constraint), and matches where the number of neighbouring nodes overall differs significantly between the graphs (called global constraint). In fact, higher weights are given to graph matches which only slightly differ in the overall number of neighbouring nodes (see 3.2.1.C). We also consider the case that the bigger graph G' contains more nodes than the actual kinematic structure represented by G', and thus vary this graph by systematically removing nodes and performing the same matching strategy on the varied graphs (see 3.2.1.D). The overall measure $\mathcal{F}^1$ is then found as the mean of the quality measures with and without graph variations (see 3.2.1.E).

A. Subgraph isomorphism: We define the subgraph isomorphism following Valiente [64]. Remember that $N \leq N'$ without loss of generality, that is G' contains more nodes than G.

Definition 3.1 (Subgraph isomorphism). *A subgraph isomorphism from a graph $G = (V, E)$ into a graph $G' = (V', E')$ is an injection $\phi : V \to V'$ such that if*

$$(v_i, v_j) \in E, \text{ then } \big(\phi(v_i), \phi(v_j)\big) \in E'.$$

The top-right part of Fig. 4 represents an example of a subgraph isomorphism of G to G'. To find the subgraph isomorphisms, we use the VF2 algorithm [65][1], which is one of the most recent methods and has shown a clear superiority in different scenarios [66]. The output of VF2 is a set of binary assignment matrices $\mathbf{X} = \{X_1, \cdots, X_L\}$, with each assignment matrix $X_l = \{0,1\}^{N \times N'}$ $(l = \{1, \cdots, L\})$ representing one subgraph isomorphism. Following the notation introduced in Section 3.2, the entry $X_l(i,i')$ equals 1 if v_i is matched with $v'_{i'}$ in the l^{th} subgraph isomorphism, and 0 otherwise.

Note that a subgraph of G' which is isomorphic to G may not have the same number of edges per node as G, but may contain more edges; in other words, a subgraph isomorphism might not be a perfect match.

B. Local and global constraints: We thus introduce additional constraints to eliminate subgraph isomorphisms which are likely to be a bad match. The local constraint is based on a threshold θ for the edge degree difference of individual node pairs, and the global constraint introduces a threshold τ over the sum of all edge degree differences. We introduce the edge degree difference matrix $\mathbf{A}$ with entries $\mathbf{A}(i,i') = |deg(v_i) - deg(v'_{i'})|$, where $deg(\cdot)$ denotes the number of edges of this node. The matrix $\mathbf{A}^*_l = \mathbf{A} \circ X_l$ (with $\circ$ denoting the Hadamard product) contains

1. Implemented in the R package '*igraph*': igraph.org

the edge degree differences for all node pairs which are matched in X_l. Then, we discard all subgraph isomorphisms X_l for which $\exists \mathbf{A}_l^*(i,i') > \theta$ (local constraint) or $\sum_{i,i'} \mathbf{A}_l^*(i,i') > \tau$ (global constraint). The threshold parameters θ and τ are experimentally determined, and their impact on the performance is discussed in Section 4.6.

C. Quality measure: To measure the quality of the subgraph isomorphisms, we introduce the set $\mathbf{X}^d = \{X_l \mid \sum_{i,i'} \mathbf{A}_l^*(i,i') = d\}$, which contains subgraph isomorphisms X_m^d with the same edge degree difference d. The smaller the difference d, the higher is the expected quality. We generate the probabilistic correspondence matrix $\mathcal{M}$ as a weighted sum, where subgraph isomorphisms with a low difference d are given a higher weight:

$$\mathcal{M} = \sum_{d=0}^{\tau} \frac{1}{d+1} \sum_{m=1}^{|\mathbf{X}^d|} \frac{1}{|\mathbf{X}^d|} X_m^d, \tag{4}$$

where $|\mathbf{X}^d|$ denotes the number of subgraph isomorphisms in $\mathbf{X}^d$. We normalise $\mathcal{M}$ such that the sum over all entries in one row equals 1. Therefore, the entry $\mathcal{M}(i,i')$ contains the probability that v_i matches $v'_{i'}$ considering the structural topology of G and G', originating from various subgraph isomorphisms after applying local and global constraints.

D. Node removal process: In order to cope with imperfections in the estimated kinematic structure, we also consider possible variations $G' \setminus v'_{i'}$ of G', which indicates that node $v'_{i'}$ is removed from G' along with the respective edges. The node removal process depends on the degree $deg(v'_{i'})$ as follows:

- **Case 1:** if $deg(v'_{i'}) = 1$, remove $v'_{i'}$.
- **Case 2:** if $deg(v'_{i'}) = 2$, where $v'_{i'}$ is linked with $v'_{j'}$ and $v'_{k'}$, an edge is created to link $v'_{j'}$ and $v'_{k'}$, and $v'_{i'}$ is removed.
- **Case 3:** if $deg(v'_{i'}) \geq 3$, this node was unlikely extracted incorrectly, as it has a large number of edges. Thus the node is not removed.

The valid variation set, denoted by $\mathbb{G}'$, contains all graph variations $G' \setminus v'_{i'}$ whose corresponding node $v'_{i'}$ falls in case 1 or 2. Examples of such successive graph variations and their corresponding subgraph isomorphisms can be found in Fig. 4. For each element $G'_k \in \mathbb{G}'$, we calculate the correspondence matrix $\mathcal{M}'_k$ using Eq.(4) by substituting G' with G'_k.

The proposed procedure is described in Algorithm 1.

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

Input : G, G'
Output : $\mathcal{F}^1$
$\mathbf{X} \leftarrow$ VF2(G, G')
$\mathcal{M} \leftarrow$ calculate by Eq.(4) based on $\mathbf{X}$
$\mathbb{G}' \leftarrow$ generated according to node removal process
for $k \in [1 : |\mathbb{G}'|]$ **do**
  $\mathbf{X}'_k \leftarrow \text{VF2}\left(G, G'_k\right)$
  $\mathcal{M}'_k \leftarrow$ calculate by Eq.(4) based on $\mathbf{X}'_k$
$\mathcal{F}^1 \leftarrow \frac{1}{|\mathbb{G}'|} \sum_{k=1}^{|\mathbb{G}'|} \frac{\mathcal{M}+\mathcal{M}'_k}{2}$

E. First order correspondence matrix: The first order similarity matrix $\mathcal{F}^1$ based on structural topology is then calculated as the mean between the probabilistic correspondence matrices with and without node removal as shown in Eq.(5):

$$\mathcal{F}^1_{(i,i')} = \frac{1}{|\mathbb{G}'|} \sum_{k=1}^{|\mathbb{G}'|} \left(\frac{\mathcal{M}(i,i') + \mathcal{M}'_k(i,i')}{2} \right). \tag{5}$$

3.2.2 Second Order Similarity: Kinematic Correlation

The motivation for the second order similarity is as follows. We hypothesise that node pairs $(v_i, v_j) \in G$ and $(v'_{i'}, v'_{j'}) \in G'$ should be matched together if the nodes within each pair move similarly and are located nearby. In order to do so, we first calculate the kinematic correlation proximities between in-body nodes for both G and G' separately, where in the following we use the calculations of G as example. Following [2], the kinematic proximity $\mathbf{P}_{(i,j)}$ between node v_i and node v_j (with segment centre points y_i and y_j) can be effectively measured by considering both relative moving velocity difference $\|(y_i^f - y_i^{f-1}) - (y_j^f - y_j^{f-1})\|$ and geodesic distance $\zeta(\cdot)$ along the skeleton:

$$\mathbf{P}_{(i,j)} = \underset{f \in F}{\text{median}} \left\{ \left\| (y_i^f - y_i^{f-1}) - (y_j^f - y_j^{f-1}) \right\| \cdot \zeta(y_i^f, y_j^f; \Psi^f) \right\}. \tag{6}$$

The median value over all frames F is taken in order to be robust to outliers. A large distance value of $\mathbf{P}_{(i,j)}$ implies that the pairwise nodes v_i and v_j are skeletally apart and move with different velocities. $\zeta(\cdot)$ is measured by the shortest path connecting the two node points within the skeleton distance map Ψ as was presented in [2][2]:

$$\zeta(y_i^f, y_j^f; \Psi^f) = \min_{\Gamma \in \mathcal{P}_{y_i^f, y_j^f}} \sum_{n=1}^{l(\Gamma)} \frac{1}{\Psi^f(p_n)}, \tag{7}$$

where Γ is a path connecting the two node points, $\mathcal{P}_{y_i, y_j}$ is the set of all possible paths, and $l(\Gamma)$ indicates the length of the connecting path. In [2] the skeleton distance map Ψ^f is generated within the object region Ω^f and its shape boundary $\delta\Omega^f$ (also known as a silhouette) for each frame f, and p_n is a point inside of the object region Ω^f and lying on the path Γ. The skeleton contains both, shape features and topological structures of the original objects. As a good representation of the skeleton, a distance transform [67] is defined as a function that returns the closest distance to the silhouette for each internal point p. The distance function $\Psi^f(p)$ of $\delta\Omega^f$ is defined as proposed in [68]:

$$\Psi^f(p) = \min_{q \in \delta\Omega^f} \big(dist(p,q) \big), \quad \forall p \in \Omega^f. \tag{8}$$

The distance metric is usually the Euclidean distance $dist(p,q) = \|p - q\|$.

Finally, based on our hypothesis that node pairs should be matched if the contained nodes have close kinematic proximities, the second order similarity function is calculated as:

$$\mathcal{F}^2_{(i,i')(j,j')} = \exp\big(- \|\mathbf{P}_{(i,j)} - \mathbf{P}_{(i',j')}\| \big). \tag{9}$$

3.2.3 Third Order Similarity: Combinatorial Motion

We consider characteristic combinatorial local motions, which are shared between different kinematic structures. As discussed in [25], [30], third order feature combination leads to geometric invariances (scale, rotation, translation) and better represents local information. Similarly, we consider the combinatorial kinematic rotation ranges

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

of three nodes and match node triplets with similar kinematic rotation ranges. As widely used in mechanical kinematics, we utilise joint limits to describe the kinematic rotation range of a joint [69], [70], *i.e.* the minimum and maximum rotation angle with respect to another joint.

To build the combinatorial motion range descriptor, we first need to find the joint positions[3]. 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 Fig. 5. For each frame f, we find the M nearest points of S_i to the segment centre y_j^f, and denote them as neighbouring points $\mathcal{N}_{i\to j}^f$. The joint position J_{i-j}^f is then defined by:

$$J_{i-j}^f = \frac{1}{2M}\sum_{m=1}^{M}\Big(\mathcal{N}_{i\to j}^f(m) + \mathcal{N}_{j\to i}^f(m)\Big). \tag{10}$$

Note that $\mathcal{N}_{i\to j}^f \neq \mathcal{N}_{j\to i}^f$, but $J_{i-j}^f = J_{j-i}^f$ 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 Fig. 5):

$$\alpha_{i-j}^f = \angle(\mathbf{v}_{i,i-j}^f, \mathbf{v}_{j,i-j}^f) = \arctan\left(\frac{\|\mathbf{v}_{i,i-j}^f \times \mathbf{v}_{j,i-j}^f{}^T\|}{\mathbf{v}_{i,i-j}^f \cdot \mathbf{v}_{j,i-j}^f{}^T}\right). \tag{11}$$

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 [48], [51]. It is invariant to the absolute rotation, as only the relative rotation between the body parts is captured [72]. It is also invariant to scale, as the translation between the body parts is not considered in the geodesic distance. A more detailed investigation of the scale, rotational, and translational invariance can be found in [73], and its application to surface alignment which is somewhat similar to our problem in [74]. Thus for combinatorial body parts (i, j, k), we define the geodesic distance of pairwise rotation matrices as:

$$d_{(i,j,k)}^f = \|\mathrm{logm}(R_{i-j}^f{}^T R_{j-k}^f)\|_F \tag{12}$$

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:

$$\begin{aligned}\Upsilon_{(i,j,k)} = [&\min_{f\in F} d_{(i,j,k)}^f, \max_{f\in F} d_{(i,j,k)}^f,\\ &\min_{f\in F} d_{(j,k,i)}^f, \max_{f\in F} d_{(j,k,i)}^f,\\ &\min_{f\in F} d_{(k,i,j)}^f, \max_{f\in F} d_{(k,i,j)}^f].\end{aligned} \tag{13}$$

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

$$\mathcal{F}^3_{(i,i')(j,j')(k,k')} = \exp\big(-\|\Upsilon_{(i,j,k)} - \Upsilon_{(i',j',k')}\|\big). \tag{14}$$

The presented metric is a super-symmetric tensor, whose entries remain unchanged under any permutation of its indices [75], and thus can be represented as a tensor product which is a requirement for applying the metric in the hypergraph matching framework [30].

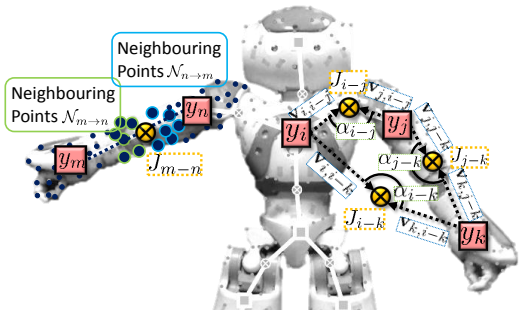


Fig. 5. Visualization of the joint estimates and motion range descriptor. The left half shows the segment centre positions y, and the neighbouring points $\mathcal{N}$ to find the joint positions J (Eq.(10)). The right half visualises the vectors $\mathbf{v}$ which are used to calculate the angles α (Eq.(11)). Best viewed in colour.

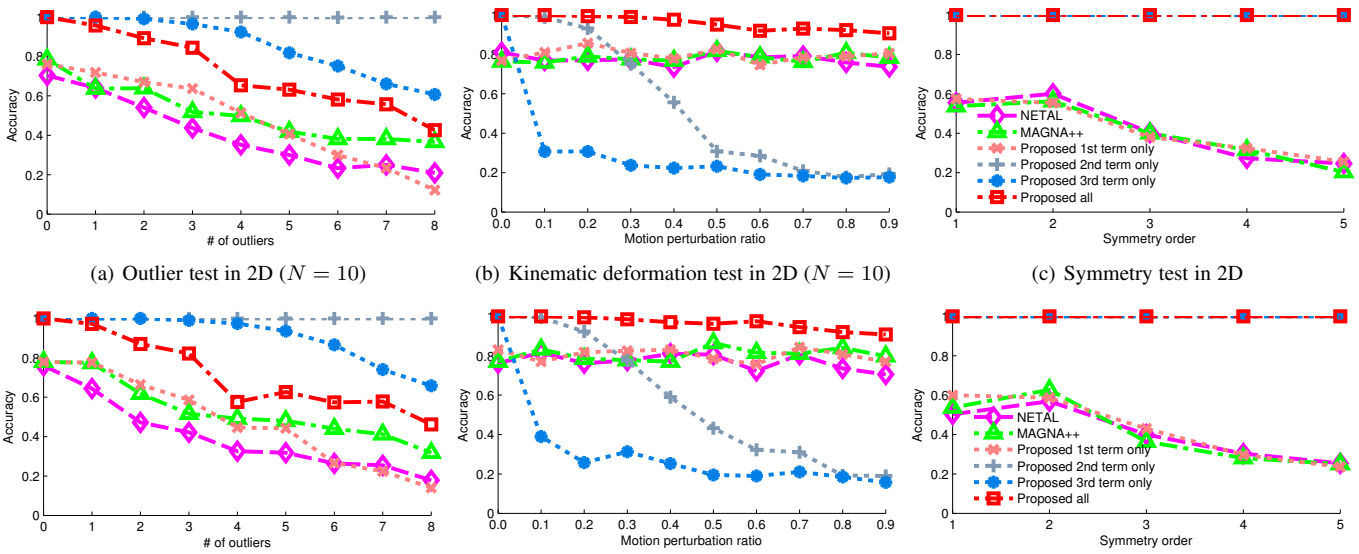
3.3 Correspondences using heterogeneous sensors

In order to compare 2D and 3D structures within one framework, we define the following properties for 3D structures. As the first order term describes the structural topology, it does not differ from the 2D case for 3D structures. The second order similarity term Eq.(6) in 3D space is calculated based on the joint's movement velocity in two consecutive frames. As geodesic distance measure Eq.(7) we use the length of the shortest path on the skeleton which connects the joints. For the third term calculation, we manually categorised the skeleton nodes found using the RGB-D camera into body part joints (*e.g.* head node) and revolute joints (*e.g.* shoulder) considering its location. If there is no body part node between two revolute joints (*e.g.* between elbow and shoulder), we introduce a body part node located in the centre of the line connecting the two revolute joints. Then, we define the directed angles and their corresponding rotational matrices in the same framework of the 2D structure, and the geodesic distance of pairwise rotation matrices is calculated by Eq.(12).

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 with their default parameters: NETAL [42] and MAGNA++ [43].

3. There have been many works trying to find joint positions as part of kinematic structure estimation [13], [14], [15], [71]. Yan and Pollefeys [13] estimate joint positions as the intersection of two motion subspaces, but their method is vulnerable to noise. The estimate in [14] is based on probabilistic inference which can find the number of joints and their connections adaptively, but the method is sensitive to the prior and has difficulty in recovering from a poor initialisation. In [71] a SVM based method which is conceptually similar to our proposal is presented, but ours is simpler than theirs. We tested these methods, and have found that their joint estimates are similar or worse compared to ours.



(a) Outlier test in 2D ($N = 10$) (b) Kinematic deformation test in 2D ($N = 10$) (c) Symmetry test in 2D

(d) Outlier test in 3D ($N = 10$) (e) Kinematic deformation test in 3D ($N = 10$) (f) Symmetry test in 3D

Fig. 6. 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.

For the appearance based approaches, we used Agglomerative Correspondence Clustering (ACC) [17], Reweighted Random Walks for Graph Matching (RRWM) [24], and Progressive Graph Matching (PGM) [18]. For all comparisons, the authors' original implementations were used. For our implementation, we set $\tau = 3$ and $\theta = 1$, and we validate this choice of parameters in Section 4.6. As the ground truth matches are known, the accuracy is measured by the ratio of correct node matches in graph G divided by N (N is the total number of nodes in G and we assume that each node of G has a matching node in G'). To encourage future work from other researchers on this newly proposed problem we release our code along with our new dataset[4]. All experiments were performed using a PC with an 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 [42], [43] trying to find correspondences between the two structures. Each quantitative result in these experiments is acquired by averaging 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 results, 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 Fig. 6(a) and 6(d), even in the presence of severe topology changes our proposed method finds matches more accurately than graph alignment methods, as our method benefits from the kinematic (second order) and motion (third order) terms. While the second and third order terms perform better than the overall proposed method in this outlier test, the experiments below show that including the first order term is important in other scenarios.

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 Fig. 6(b) and 6(e). Even though the second and third order 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 Fig. 6(c) and 6(f), both the graph matching methods and using only the first term resulted in low accuracies, but our method can find correct matches using the motion information.

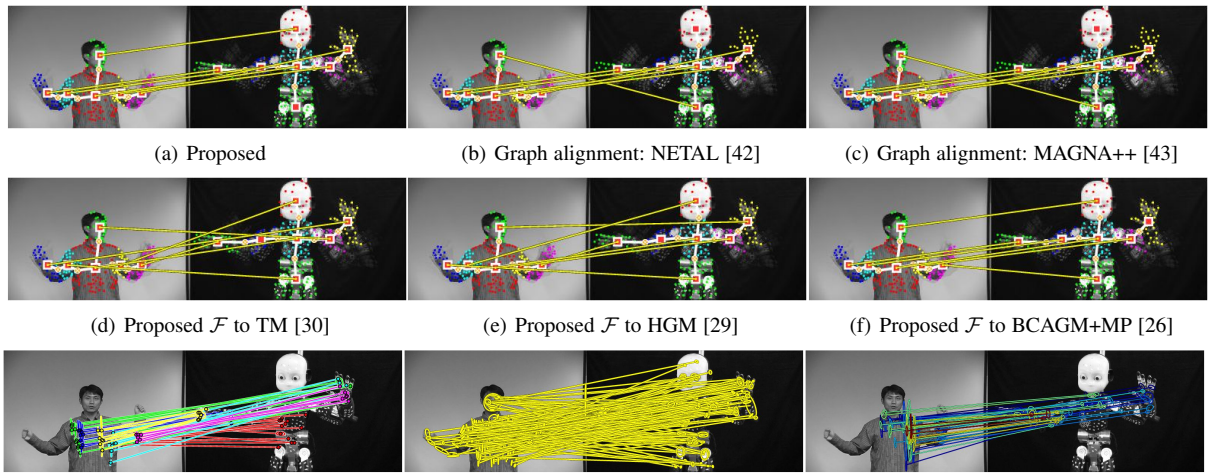
Through these experiments, we have validated that the structural topology as well as the kinematic and motion information are characteristic properties for dynamic structures. While each individual term suffers depending on the perturbation, the combination of all three terms results in increased accuracy and robustness to a range of perturbations.

4.2 Real Kinematic Structure Datasets

Imperial-PRL-KSC-Dataset: The Imperial-PRL-Dataset [2] has previously been used for testing the complex kinematic structure generation performance. It contains simple structures as well as 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 constructed more sequences using various humanoid robots[5]

4. Both the code and the dataset are available at: www.imperial.ac.uk/PersonalRobotics.

5. We utilised three robots: iCub (www.icub.org), NAO (www.softbankrobotics.com) and Baxter (www.rethinkrobotics.com)

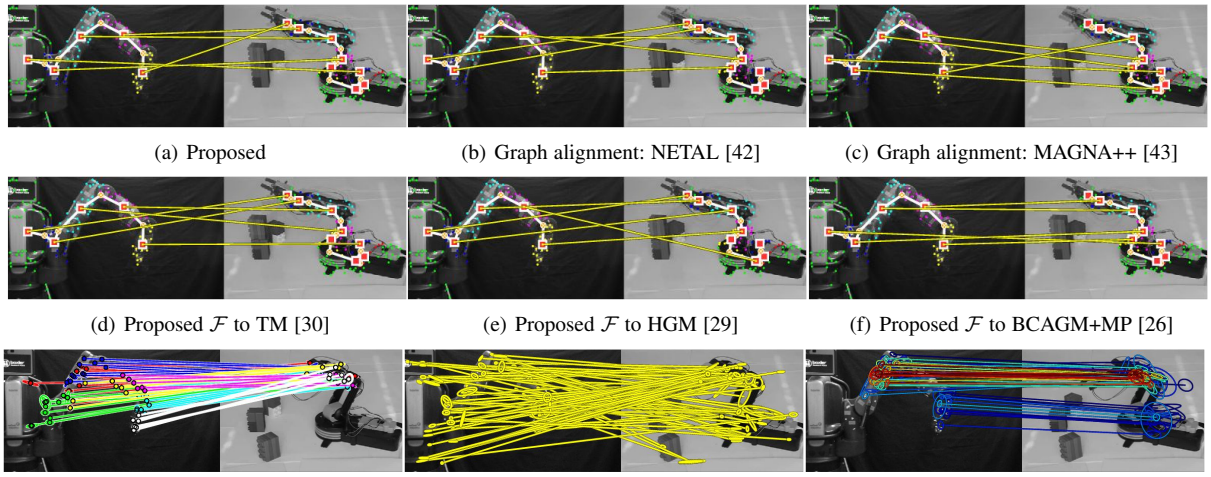


(a) Proposed (b) Graph alignment: NETAL [42] (c) Graph alignment: MAGNA++ [43]

(d) Proposed $\mathcal{F}$ to TM [30] (e) Proposed $\mathcal{F}$ to HGM [29] (f) Proposed $\mathcal{F}$ to BCAGM+MP [26]

(g) Appearance feature matching: ACC [17] (h) Appearance feature matching: RRWM [24] (i) Appearance feature matching: PGM [18]

Fig. 7. Experiments on real image datasets: dancing human *vs.* dancing iCub. Two static images are used for the appearance based methods (best viewed in colour).

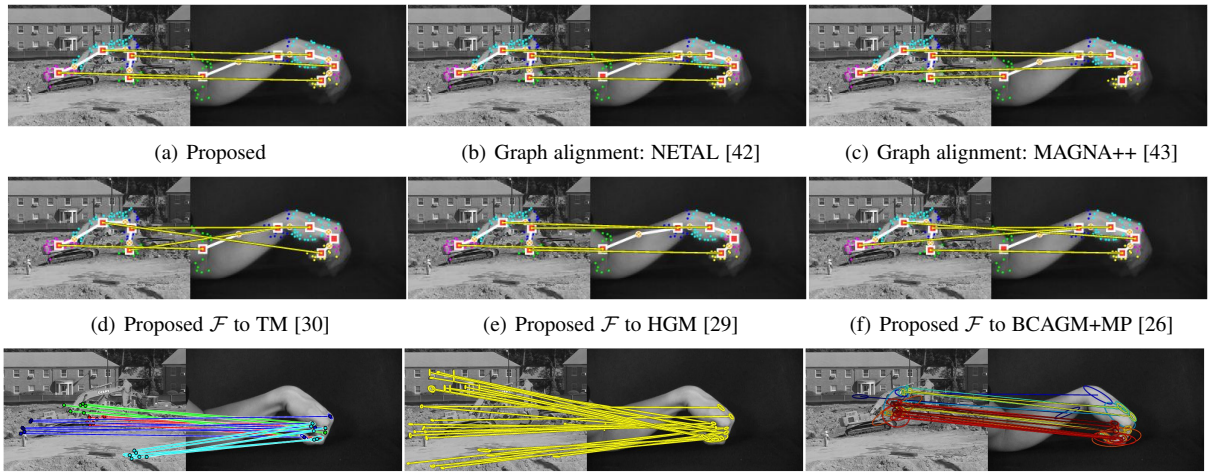


(a) Proposed (b) Graph alignment: NETAL [42] (c) Graph alignment: MAGNA++ [43]

(d) Proposed $\mathcal{F}$ to TM [30] (e) Proposed $\mathcal{F}$ to HGM [29] (f) Proposed $\mathcal{F}$ to BCAGM+MP [26]

(g) Appearance feature matching: ACC [17] (h) Appearance feature matching: RRWM [24] (i) Appearance feature matching: PGM [18]

Fig. 8. Experiments on real image datasets: Baxter *vs.* OWI-535 Robotic Arm Edge (best viewed in colour).



(a) Proposed (b) Graph alignment: NETAL [42] (c) Graph alignment: MAGNA++ [43]

(d) Proposed $\mathcal{F}$ to TM [30] (e) Proposed $\mathcal{F}$ to HGM [29] (f) Proposed $\mathcal{F}$ to BCAGM+MP [26]

(g) Appearance feature matching: ACC [17] (h) Appearance feature matching: RRWM [24] (i) Appearance feature matching: PGM [18]

Fig. 9. Experiments on real image datasets: yellow crane *vs.* digging arm (best viewed in colour).

and human motion, and have performed many tests with different sequence combinations. We generated the kinematic structures G and G' using [2] with manually set segment numbers, which 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 [76] and

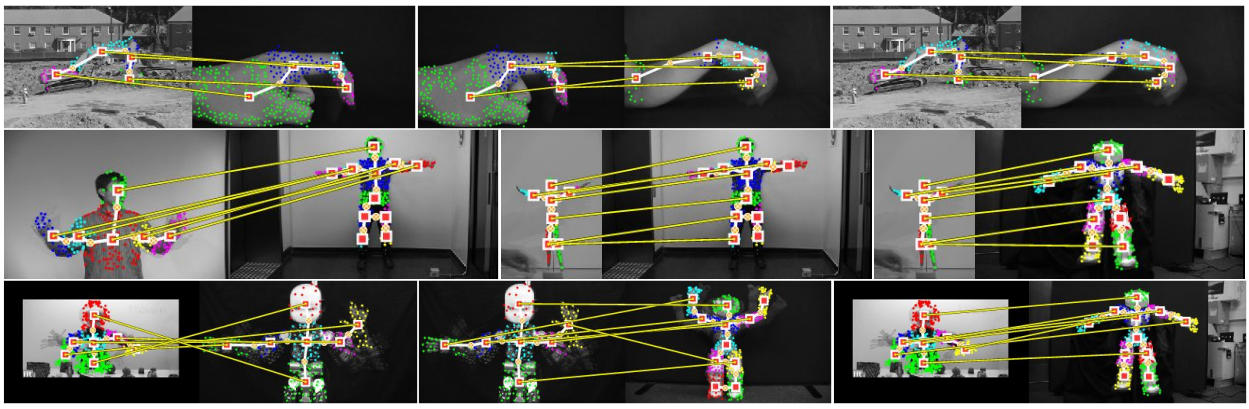


Fig. 10. Various kinematic structure correspondence matching results using the proposed method (best viewed in colour). The bottom-left correspondence is analysed in more detail in Fig. 12.

TABLE 1
Performance on the real kinematic structure dataset. The numbers are the mean and standard deviation (in parenthesis) respectively.

Methods	Accuracy (%)
NETAL [42]	67.72(±39.59)
MAGNA++ [43]	63.42(±33.66)
Proposed $\mathcal{F}$ to HGM [29]	35.06(±30.73)
Proposed $\mathcal{F}$ to TM [30]	29.76(±21.65)
Proposed $\mathcal{F}$ to BCAGM [26]	52.66(±25.02)
Proposed $\mathcal{F}$ to BCAGM+IPFP [26]	45.28(±15.50)
Proposed $\mathcal{F}$ to BCAGM+MP [26]	69.09(±20.65)
Proposed $\mathcal{F}$ first term only	85.19(±20.39)
Proposed $\mathcal{F}$ second term only	68.81(±31.58)
Proposed $\mathcal{F}$ third term only	67.81(±19.72)
Proposed $\mathcal{F}$ to RRWHM [25] without weight normalisation Eq.(3)	88.23(±13.07)
Proposed $\mathcal{F}$ to RRWHM [25] with weight normalisation Eq.(3)	**92.99 (±10.41)**

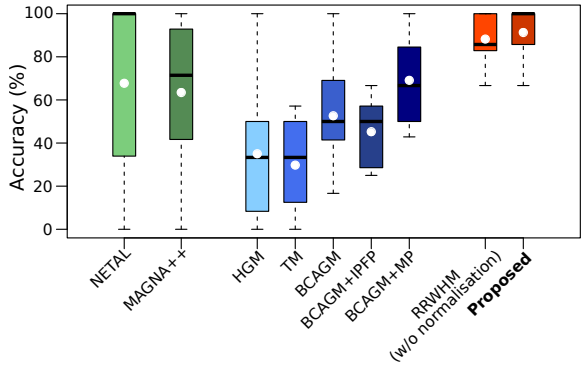


Fig. 11. Boxplots of the performance of state of the art methods on the real kinematic structure dataset. The bottom and top of the box are respectively the lower and upper quartile, the central bold band is the median, and the white dot is the mean, showing that our proposed method outperformed the others.

the SIFT descriptor [63] as was described in [18]. Then, ACC [17], RRWM [24] and PGM [18] are applied separately using the code provided by the authors.

As shown in Fig. 7, 8 and 9, our method clearly outperforms other methods for finding accurate kinematic correspondences. Especially the graph alignment approaches often failed to distinguish symmetric matches, and the appearance based correspondence matching approaches cannot 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.

Fig. 10 shows more correspondence matching results between heterogeneous objects. Even though there are many outlier nodes in G', our method finds accurate matches. The bottom left matching result is especially interesting, as the matching is symmetrically reversed. This is due to the motions executed by the robots, as one robot moves its arm upwards while the other robot moves its arm downwards. This is further illustrated in Fig. 12, where we present image sequences of the pairs in order to demonstrate their motion directions. From these results, it can be observed that the proposed combinatorial motion similarity term is able to imply motion directions.

In Table 1 and Fig. 11, we show the quantitative matching accuracy based on manually generated ground truth matchings. We compare our method to NETAL [42] and MAGNA++ [43]. Our proposed framework outperforms those methods, both in terms of higher accuracy as well as significantly lower variance ($p < 0.05$), which highlights the robustness of our method when applied to different sequences. We utilise different hypergraph matching algorithms which allow to merge different order similarity information in a structured way to obtain topological, kinematic and motion correspondences. The notion "Proposed $\mathcal{F}$ to [·]" in Table 1 denotes that we use the approximation algorithms of [25], [26], [29], [30] to find the matches based on $\mathcal{F}$ (Eq.(1)), *i.e.* we do not directly use the hypergraph methods for feature matching, but use their approximation algorithms instead.

Using the approximation algorithm of RRWHM [25] to solve Eq.(1) performs best because of the stochastic scheme in RRWHM which updates the correspondence matrix more robustly than other optimisation-based methods [77], especially when the graph structure is dynamic and largely deformed by motions. Furthermore, RRWHM has been shown to achieve better performance in the case of non-convex functions as in Eq.(1).

In Table 1, one can see that combining the three different terms leads to higher accuracy compared to using the individual terms only. We also show that the proposed term normalisation weights $w_{\kappa(e)}$ play an important role by balancing the structural topology and motion similarity. We have observed that most failure cases without the normalisation term were due to mismatched nodes having similar motions, which is expected as the motion term has

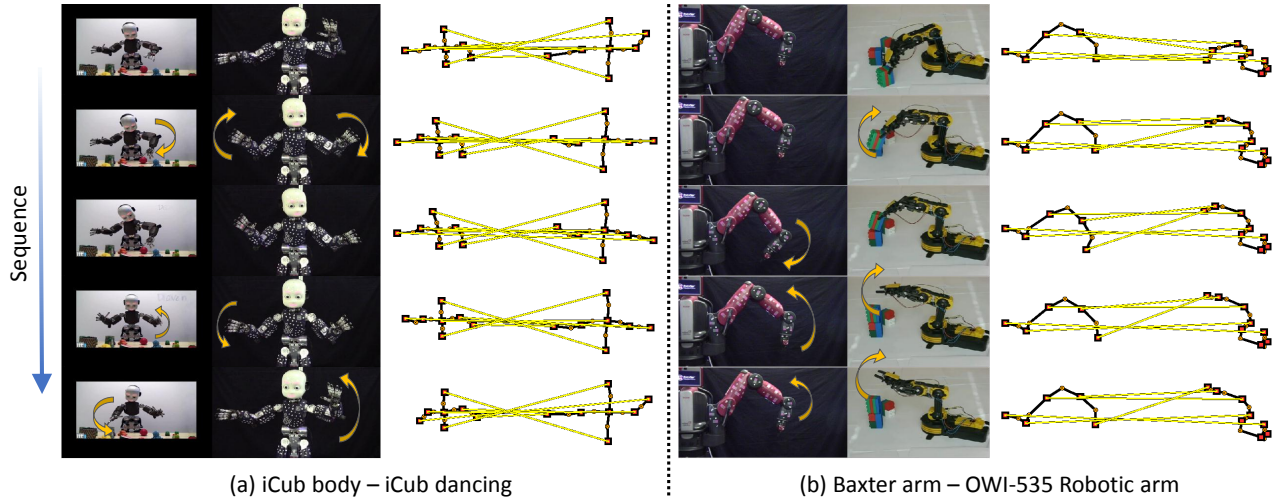
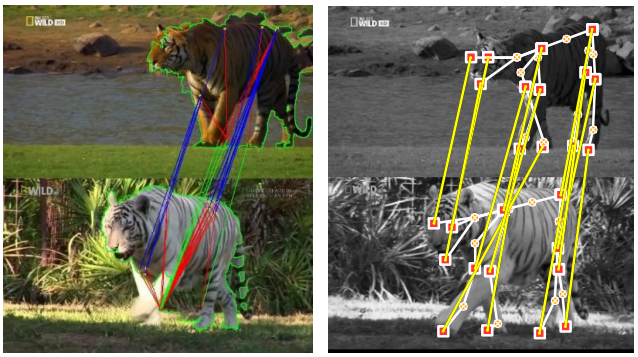


Fig. 12. This figure shows that the combinatorial motion term can distinguish motion directions (best viewed in colour). (a) Sequential matches of iCub body sequence and iCub dancing sequence. The matches are upside-down, as the left iCub is moving its hands downwards, whilst the right iCub is waving its hands upwards. (b) Sequential matches of Baxter arm and OWI-535 Robotic Arm Edge sequence. The matches are leftside-rightside.



(a) (b)

Fig. 13. (a) The method in [20] finds a corresponding motion cluster representing the swinging leg. The red lines show the 10 closest pairs of local feature points in the descriptor space. The blue lines represent corresponding anchor points, and the green lines connect swinging body parts. (b) The proposed method can correspond an overall kinematic structure correctly considering structure topology as well as motion (best viewed in colour).

the highest weight when the normalisation weights are not being used.

TigDog Dataset: We have applied the proposed method to another articulated motion pattern matching dataset, namely the TigDog Dataset [8], [20]. Del Pero *et al.* [8], [20] presented an unsupervised method discovering characteristic motion patterns in videos of highly articulated objects with the dataset. They represented articulated object motion using a collection of ordered pairs of local feature point trajectories (PoTs) [20]. This dataset contains video shots for three different classes: tigers, horses and dogs. The 2D locations of 19 landmarks are manually annotated in each frame (left eye, neck, front left ankle, *etc.*). We have constructed a kinematic structure using the given landmark locations and applied it to our proposed correspondence framework. As shown in Fig. 13, the method presented in [20] was able to find only a limited number of local correspondences with similar motions. In comparison, our method can correctly find correspondences between all body parts considering structural topology and motion characteristics.

4.3 Correspondences using Heterogeneous Sensors

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 Fig. 1, the proposed method is able to find correspondences between similar kinematic structure joints even if their appearances are visually very different, as well as in the presence of strong motion variations. In addition, as shown in Fig. 1 and Fig. 14, the proposed method can even find kinematic structure matches across different sensors, namely RGB images and the skeleton extracted by a RGB-D camera.

We also employed the proposed method for an advanced robotics application. It allows the robot to find correspondences between self-learned representations of its own body and body parts of a human [6]. More specifically, we use an iCub humanoid robot and perform motor babbling, *i.e.* issuing random motor commands. Then the robot can find kinematic structure correspondences between its own body parts, *e.g.* between its left and right arm as shown in Fig. 15(a). In a further step, our method can also find correspondences between the robot arm and the kinematic structure of a human extracted by a RGB-D camera, as demonstrated in Fig. 15(b). This could then be used for imitation learning, as the robot found correspondences between its own joints and the body parts of the human. We elaborate further on the use of our framework for imitation learning in our conclusions.

4.4 Matching Similar Motions

We evaluated our method on the MPI Dexter 1 hand dataset [78] in order to validate whether it can be used for matching similar motions. The dataset consists of seven sequences of hand motions of a single actor. Note that the graph representation of all the fingers except the thumb is the same. 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) and 31.11% (MAGNA++) which are only considering structural topology. This demonstrates the robustness of our newly designed third order similarity motion term, which is key to overcome the hand's ambiguous topological structure. Therefore our method can also be

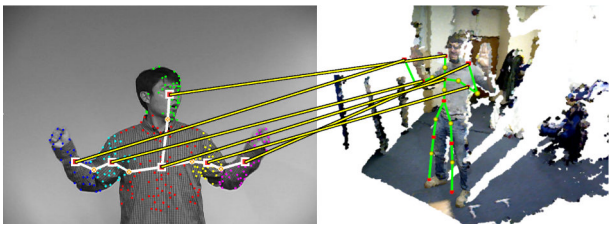


Fig. 14. The proposed method can establish kinematic correspondences between RGB video and depth video.

applied to matching similar motions without the need of adapting parameters, as long as the kinematic nodes and joint positions are provided or can be estimated.

4.5 Time complexity analysis

The time complexity for the calculation of the first term is $\mathcal{O}(N^2)$ in the best case and $\mathcal{O}(N!N)$ in the worst case [65], while the second term's time complexity is $\mathcal{O}(N^2)$, and $\mathcal{O}(N^3)$ for the third term. As kinematic structures typically have a relatively low number of nodes, the computation is quite fast. On our proposed real dataset (Imperial-PRL-KSC-Dataset), it takes 1.69 ± 1.09s, 0.11 ± 0.07s, and 2.20 ± 1.70s for the first, second and third order similarity term calculation respectively (unoptimised Matlab and R codes), as well as 0.02 ± 0.001s for the RRWHM hypergraph matching [25] (optimised C++ code).[6]

4.6 Validations on the parameters of the first order term

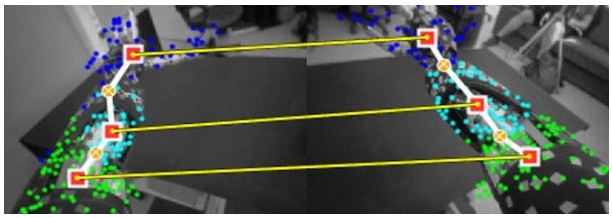
We validated the parameters of the first order topology term as detailed in Table 2. For the local constraint, we did not find a significant performance difference between $\theta = 1$ and $\theta = 2$. However, the performance decreases for $\theta \geq 3$, because in this case the constraint is rarely triggered. Values of $\theta \geq 3$ are not plausible either, as this would mean that the skeleton is highly distorted. For example, $\theta = 3$ would allow a node with just one neighbour to be matched with another node which has four neighbouring joints.

TABLE 2
Validations on the threshold parameters. The numbers are the mean and standard deviation (in parenthesis) respectively.

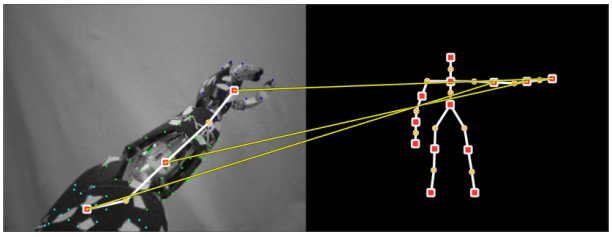
Parameter		Accuracy (%)
τ	θ	
3	1	92.99 (±10.41)
3	2	91.26 (±11.32)
3	3	88.23 (±19.83)
3	4	88.23 (±19.83)
2	1	81.30 (±29.32)
4	1	92.99 (±10.41)
5	1	92.99 (±10.41)
6	1	92.99 (±10.41)
2	2	85.19 (±28.62)
4	2	91.26 (±11.32)
5	2	91.26 (±11.32)
6	2	91.26 (±11.32)

The global constraint τ limits the allowed total edge difference between the two graphs. As shown in Table 2, a low value of $\tau \leq 2$ leads to decreased performance, as not enough noise is allowed. For $\tau \geq 3$, the impact of τ is low, as Eq.(4) weights subgraphs inverse proportionally to the total difference in degree of the graphs.

6. The other hypergraph matching methods [26], [29], [30] take roughly the same time for computation when compared to RRWHM.



(a) Corresponding iCub self-body (left arm) to self-body (right arm)



(b) Corresponding iCub self-body (left arm) to others (human left arm)

Fig. 15. The proposed method can be used (a) for an iCub robot to find correspondences between iCub's partial arm structure captured using the iCub's RGB camera to its own other body part, and even (b) to full body human structure captured using a RGB-D camera.

In other words, even if τ is increased, the newly found subgraph isomorphisms have relatively low weight. In case of graphs with a high number of nodes, one should adjust τ accordingly, *i.e.* τ should be increased for complex skeletons with a high number of articulations.

5 Conclusion and Future Work

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, as well as being robust against frequently encountered symmetries around the skeleton axis. 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 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 believe that our method opens exciting opportunities in the computer vision and robotics fields, with new research directions emerging in a variety of applications such as learning by demonstration, modeless object manipulation learning, articulated object categorisation, human motion retargeting to a humanoid and tool usage learning. 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. Employing our method would have the benefit that the structure of the robot does not have to be known in advance, which is important for generic

applicability across different robots, as well as robots which have to adapt to damage [79]. Similarly, our method would also allow the robot to adapt to users with different kinds of physical impairment by using personalised motion ranges [80].

REFERENCES

[1] H. J. Chang, T. Fischer, M. Petit, M. Zambelli, and Y. Demiris, "Kinematic structure correspondences via hypergraph matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 4216–4425.

[2] H. J. Chang and Y. Demiris, "Unsupervised Learning of Complex Articulated Kinematic Structures combining Motion and Skeleton Information," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3138–3146.

[3] ——, "Highly articulated kinematic structure estimation combining motion and skeleton information," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2017.

[4] Y. Wu and Y. Demiris, "Towards one shot learning by imitation for humanoid robots," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2010, pp. 2889–2894.

[5] 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*, 2014, pp. 1–10.

[6] M. Zambelli, T. Fischer, M. Petit, H. J. Chang, A. Cully, and Y. Demiris, "Towards anchoring self-learned representations to those of other agents," in *Proc. IEEE Int. Conf. Intell. Robots and Syst. Workshops*, 2016.

[7] A. A. Chaaraoui, J. R. Padilla-Lopez, and F. Florez-Revuelta, "Fusion of Skeletal and Silhouette-based Features for Human Action Recognition with RGB-D Devices," in *Proc. Int. Conf. Comput. Vis. Workshop*, 2013, pp. 91–97.

[8] L. Del Pero, S. Ricco, R. Sukthankar, and V. Ferrari, "Behavior discovery and alignment of articulated object classes from unstructured video," *Int. J. Comput. Vis.*, vol. 121, no. 2, 2017.

[9] Y. Wu and Y. Demiris, "Learning dynamical representations of tools for tool-use recognition," in *Proc. IEEE Int. Conf. Robot. Biomimetics*, 2011, pp. 2664–2669.

[10] J. Sturm, C. Plagemann, and W. Burgard, "Body schema learning for robotic manipulators from visual self-perception," *J. Physiol. Paris*, vol. 103, no. 3-5, pp. 220–231, 2009.

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

[12] D. Katz, M. Kazemi, J. A. Bagnell, and A. Stentz, "Interactive Segmentation, Tracking, and Kinematic Modeling of Unknown 3D Articulated Objects," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2013, pp. 5003–5010.

[13] J. Yan and M. Pollefeys, "A Factorization-Based Approach for Articulated Nonrigid Shape, Motion and Kinematic Chain Recovery From Video," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 5, pp. 865–877, 2008.

[14] D. Ross, D. Tarlow, and R. Zemel, "Learning Articulated Structure and Motion," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 214–237, 2010.

[15] J. Fayad, C. Russell, and L. Agapito, "Automated articulated structure and 3D shape recovery from point correspondences," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2011, pp. 431–438.

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

[17] M. Cho, J. Lee, and K. M. Lee, "Feature Correspondence and Deformable Object Matching via Agglomerative Correspondence Clustering," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2009, pp. 1280–1287.

[18] M. Cho and K. M. Lee, "Progressive graph matching: Making a move of graphs via probabilistic voting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 398–405.

[19] T. Taniai, S. N. Sinha, and Y. Sato, "Joint recovery of dense correspondence and cosegmentation in two images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 4246–4255.

[20] L. Del Pero, S. Ricco, R. Sukthankar, and V. Ferrari, "Articulated motion discovery using pairs of trajectories," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 2151–2160.

[21] ——, "Discovering the Physical Parts of an Articulated Object Class From Multiple Videos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 714–723.

[22] L. Wei, Q. Huang, D. Ceylan, E. Vouga, and H. Li, "Dense human body correspondences using convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1544–1553.

[23] Y. Zhou, E. Antonakos, J. Alabort-i Medina, A. Roussos, and S. Zafeiriou, "Estimating correspondences of deformable objects "in-the-wild"," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 5791–5801.

[24] M. Cho, J. Lee, and K. M. Lee, "Reweighted Random Walks for Graph Matching," in *Proc. European Conf. Comput. Vis.*, 2010, pp. 492–505.

[25] J. Lee, M. Cho, and K. M. Lee, "Hyper-graph Matching via Reweighted Random Walks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 1633–1640.

[26] Q. Nguyen, A. Gautier, and M. Hein, "A Flexible Tensor Block Coordinate Ascent Scheme for Hypergraph Matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 5270–5278.

[27] L. Torresani, V. Kolmogorov, and C. Rother, "Feature correspondence via graph matching: Models and global optimization," in *Proc. European Conf. Comput. Vis.*, 2008, pp. 596–609.

[28] F. Zhou and F. D. la Torre, "Deformable Graph Matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 2922–2929.

[29] R. Zass and A. Shashua, "Probabilistic Graph and Hypergraph Matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008.

[30] O. Duchenne, F. Bach, I.-S. Kweon, and J. Ponce, "A Tensor-Based Algorithm for High-Order Graph Matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2383–2395, 2011.

[31] M. Cho, J. Sun, O. Duchenne, and J. Ponce, "Finding matches in a haystack: A max-pooling strategy for graph matching in the presence of outliers," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 2083–2090.

[32] P. Tresadern and I. Reid, "Articulated Structure From Motion by Factorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 1110–1115.

[33] J. Yan and M. Pollefeys, "Automatic Kinematic Chain Building from Feature Trajectories of Articulated Objects," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2006, pp. 712–719.

[34] T. Brox and J. Malik, "Object segmentation by long term analysis of point trajectories," in *Proc. European Conf. Comput. Vis.*, 2010, pp. 282–295.

[35] X. Bai and L. J. Latecki, "Path Similarity Skeleton Graph Matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 7, pp. 1282–1292, 2008.

[36] W. Shen, Y. Wang, X. Bai, H. Wang, and L. J. Latecki, "Shape clustering: Common structure discovery," *Pattern Recognition*, vol. 46, no. 2, pp. 539–550, 2013.

[37] M. Koyutürk, Y. Kim, U. Topkara, S. Subramaniam, W. Szpankowski, and A. Grama, "Pairwise alignment of protein interaction networks," *J. Comput. Biol.*, vol. 13, no. 2, pp. 182–199, 2006.

[38] R. Singh, J. Xu, and B. Berger, "Global alignment of multiple protein interaction networks with application to functional orthology detection," *Proc. National Academy Sciences*, vol. 105, no. 35, pp. 12 763–12 768, 2008.

[39] O. Kuchaiev, T. Milenković, V. Memišević, W. Hayes, and N. Pržulj, "Topological network alignment uncovers biological function and phylogeny," *J. Royal Society Interface*, vol. 7, no. 50, pp. 1341–1354, 2010.

[40] O. Kuchaiev and N. Pržulj, "Integrative network alignment reveals large regions of global network similarity in yeast and human," *Bioinformatics*, vol. 27, no. 10, pp. 1390–1396, 2011.

[41] T. Milenkovic and N. Przulj, " Uncovering Biological Network Function via Graphlet Degree Signatures," *Cancer Informatics*, vol. 6, pp. 257–273, 2008.

[42] 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*, vol. 29, no. 13, pp. 1654–1662, 2013.

[43] V. Vijayan, V. Saraph, and T. Milenković, "MAGNA++: Maximizing Accuracy in Global Network Alignment via both node and edge conservation," *Bioinformatics*, vol. 31, no. 14, pp. 2409–2411, 2015.

[44] L. Meng, A. Striegel, and T. Milenkovic, "Local versus global biological network alignment," *Bioinformatics*, vol. 32, no. 20, pp. 3155–3164, 2016.

[45] V. Vijayan and T. Milenković, "Multiple network alignment via multiMAGNA++," in *Proc. Int. Workshop Data Mining in Bioinformatics at ACM Conf. Knowledge Discovery and Data Mining*, 2016.

[46] B. Jacquet, R. Angst, and M. Pollefeys, "Articulated and Restricted Motion Subspaces and Their Signatures," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 1506–1513.

[47] S. Hadfield, K. Lebeda, and R. Bowden, ""Natural Action Recognition Using Invariant 3D Motion Encoding"," in *Proc. European Conf. Comput. Vis.*, 2014, pp. 758–771.

[48] D. Q. Huynh, "Metrics for 3D Rotations: Comparison and Analysis," *J. Math. Imaging and Vis.*, vol. 35, no. 2, pp. 155–164, 2009.

[49] R. Hartley, J. Trumpf, Y. Dai, and H. Li, "Rotation Averaging," *Int. J. Comput. Vis.*, vol. 103, no. 3, pp. 267–305, 2013.

[50] 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.*, vol. 24, no. 2, pp. 539–560, 2015.

[51] C. Jia and B. Evans, "3D rotational video stabilization using manifold optimization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2013, pp. 2493–2497.

[52] P. Ochs and T. Brox, "Higher order motion models and spectral clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 614–621.

[53] D. Conte, P. Foggia, C. Sansone, and M. Vento, "Thirty years of Graph Matching in Pattern Recognition," *Int. J. Pattern Recognition Artificial Intell.*, vol. 18, no. 3, pp. 265–298, 2004.

[54] T. Cour, P. Srinivasan, and J. Shi, "Balanced Graph Matching," in *Proc. Advances Neural Inform. Process.*, vol. 19, 2007, pp. 313–320.

[55] J. Lee, M. Cho, and K. M. Lee, "A Graph Matching Algorithm Using Data-Driven Markov Chain Monte Carlo Sampling," in *Proc. Int. Conf. Pattern Recognit.*, 2010, pp. 2816–2819.

[56] M. Leordeanu, A. Zanfir, and C. Sminchisescu, "Semi-supervised learning and optimization for hypergraph matching," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2011, pp. 2274–2281.

[57] A. Shashua, R. Zass, and T. Hazan, "Multi-way clustering using super-symmetric non-negative tensor factorization," in *Proc. European Conf. Comput. Vis.*, 2006, pp. 595–608.

[58] M. Leordeanu and C. Sminchisescu, "Efficient hypergraph clustering," in *Int. Conf. AI and Statistic*, 2012, pp. 676–684.

[59] S. R. Bulò and M. Pelillo, "A game-theoretic approach to hypergraph clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1312–1327, 2013.

[60] M. Hein, S. Setzer, L. Jost, and S. S. Rangapuram, "The total variation on hypergraphs-learning on hypergraphs revisited," in *Advances in Neural Information Processing Systems*, 2013, pp. 2427–2435.

[61] J. Yan, C. Zhang, H. Zha, W. Liu, X. Yang, and S. M. Chu, "Discrete Hyper-Graph Matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1520–1528.

[62] Y. Zeng, C. Wang, Y. Wang, X. Gu, D. Samaras, and N. Paragios, "Dense Non-rigid Surface Registration Using High-Order Graph Matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 382–389.

[63] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE Int. Conf. Comput. Vis.*, 1999, pp. 1150–1157.

[64] G. Valiente, *Algorithms on Trees and Graphs*. Berlin: Springer, 2002.

[65] L. P. Cordella, P. Foggia, C. Sansone, and M. Vento, "A (Sub)Graph Isomorphism Algorithm for Matching Large Graphs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 10, pp. 1367–1372, 2004.

[66] H.-C. Ehrlich and M. Rarey, "Systematic benchmark of substructure search in molecular graphs - From Ullmann to VF2," *J. Cheminformatics*, vol. 4, no. 1, pp. 1–17, 2012.

[67] A. K. Jain, *Fundamentals of Digital Image Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1989.

[68] R. Strzodka and A. Telea, "Generalized distance transforms and skeletons in graphics hardware," in *Proc. IEEE TCVG Conf. on Visualization*, 2004, pp. 221–230.

[69] G. Nicolas, F. Multon, G. Berillon, and F. Marchal, "From bone to plausible bipedal locomotion using inverse kinematics," *J. Biomech.*, vol. 40, no. 5, pp. 1048–1057, 2007.

[70] W. Xu, J. Wang, K. Yin, K. Zhou, M. Van De Panne, F. Chen, and B. Guo, "Joint-aware manipulation of deformable models," *ACM Trans. Graph.*, vol. 28, no. 3, pp. 35:1–35:9, 2009.

[71] B. Rosman and S. Ramamoorthy, "Learning spatial relationships between objects," *Int. J. Robot. Research*, vol. 30, no. 11, pp. 1328–1342, 2011.

[72] M. Moakher, "Means and Averaging in the Group of Rotations," *SIAM J. Matrix Anal. Appl.*, vol. 24, no. 1, pp. 1–16, 2002.

[73] M. Hilaga, Y. Shinagawa, T. Kohmura, and T. L. Kunii, "Topology Matching for Fully Automatic Similarity Estimation of 3D Shapes," in *Proc. ACM Conf. Comput. Graph. Interactive Techn.*, 2001, pp. 203–212.

[74] T. Tung and T. Matsuyama, "Geodesic mapping for dynamic surface alignment," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 5, pp. 901–913, 2014.

[75] E. Kofidis and P. A. Regalia, "On the best rank-1 approximation of higher-order supersymmetric tensors," *SIAM J. Matrix Anal. Appl*, vol. 23, no. 3, pp. 863–884, 2002.

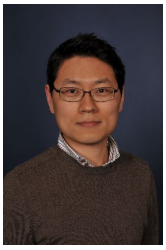
[76] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in *Proc. British Mach. Vis. Conf.*, 2002, pp. 36.1–36.10.

[77] F. Zhou and F. D. la Torre, "Factorized Graph Matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 127–134.

[78] S. Sridhar, A. Oulasvirta, and C. Theobalt, "Interactive Markerless Articulated Hand Motion Tracking Using RGB and Depth Data," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 2456–2463.

[79] A. Cully, J. Clune, D. Tarapore, and J.-B. Mouret, "Robots that can adapt like animals," *Nature*, vol. 521, no. 7553, pp. 503–507, 2015.

[80] Y. Gao, H. J. Chang, and Y. Demiris, "Iterative path optimisation for personalised dressing assistance using vision and force information," in *Proc. IEEE Int. Conf. Intell. Robots and Syst.*, 2016, pp. 4398–4403.



Hyung Jin Chang received his B.S. and Ph.D. degree from the School of Electrical Engineering and Computer Science, Seoul National University, Seoul, Republic of Korea. He is a post doctoral researcher with the Department of Electrical and Electronic Engineering at Imperial College London. His current research interests include articulated structure learning, human robot interaction, object tracking, human action understanding and user modelling.

Tobias Fischer received the B.Sc. degree from the Ilmenau University of Technology, Germany, in 2013, and the M.Sc. degree in Artificial Intelligence from the University of Edinburgh, U.K., in 2014. He is currently pursuing the Ph.D degree in the Personal Robotics Lab at Imperial College London. His research interests include both computer vision and human vision, visual attention, machine learning and computational neuroscience. Tobias is interested in applying this knowledge to developmental robotics.

Maxime Petit received the M.Sc. degree in computer science from the University of Paris-Sud, France, and an engineering degree in biosciences (bio-informatics and modelling) from the National Institute of Applied Sciences (INSA) Lyon, France, both in 2010. In 2014, he received a Ph.D. in Neurosciences from the National Institute of Science and Medical Research (INSERM), Lyon, within the Robot Cognition Laboratory. He was a Research Associate at the Personal Robotics Lab, Imperial College London, and is Research Associate at Ecole Centrale de Lyon (France) since 2017. His research interests include developmental robotics, memory and reasoning in robotics, especially linked to social interaction through spoken language with a human.

Martina Zambelli received her B.Sc. degree in Information Engineering in 2011 and her M.Sc. degree in Automation Engineering in 2013, both from the University of Padova, Italy, and developed her master thesis project in the Automatic Control department at KTH - Royal Institute of Technology, Stockholm, Sweden, in 2013. She is currently a Ph.D. student in the Personal Robotics Lab at the Electrical and Electronic Department, Imperial College London, since January 2014, under the supervision of Prof. Yiannis Demiris. Her research interests involve autonomous robotics, artificial cognitive systems, human-robot interaction, autonomous system, with particular focus on learning methods for humanoid robots.

Yiannis Demiris (SM03) received the B.Sc (Honours) and Ph.D. degrees from the Department of Artificial Intelligence, University of Edinburgh, Edinburgh, U.K. He is a Professor at the Department of Electrical and Electronic Engineering at Imperial College London, where he heads the Personal Robotics Laboratory. His current research interests include human robot interaction, machine learning, user modeling, and assistive robotics; he has published more than 150 journal and peer reviewed conference papers on these topics. Professor Demiris was the Chair of the IEEE International Conference on Development and Learning in 2007 and the Program Chair of the ACM/IEEE International Conference on Human Robot Interaction in 2008. He was a recipient of the Rector's Award for Teaching Excellence in 2012 and the FoE Award for Excellence in Engineering Education in 2012. He is a senior member of the IEEE, and a fellow of the IET, BCS and the Royal Statistical Society.