Unsupervised Learning of Complex Articulated Kinematic Structures combining Motion and Skeleton Information

Imperial College London
imperial.ac.uk/PersonalRobotics

Hyung Jin Chang and Yiannis Demiris
Personal Robotics Laboratory, EEE Department
Imperial College London, United Kingdom

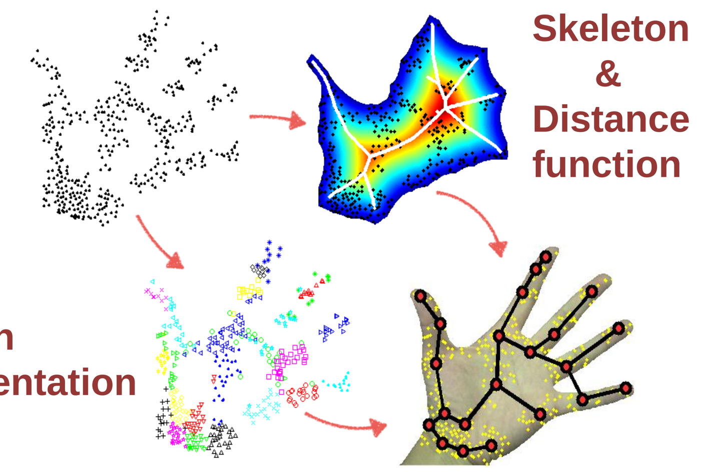
IEEE 2015 Conference on Computer Vision and Pattern Recognition
CVPR 2015 BOSTON JUNE 7-12

RESEARCH GOAL

Generate an ***accurate and plausible articulated kinematic structure***
via ***motion*** + ***skeleton*** information only using ***2D feature point trajectories***

KINEMATIC STRUCTURE LEARNING

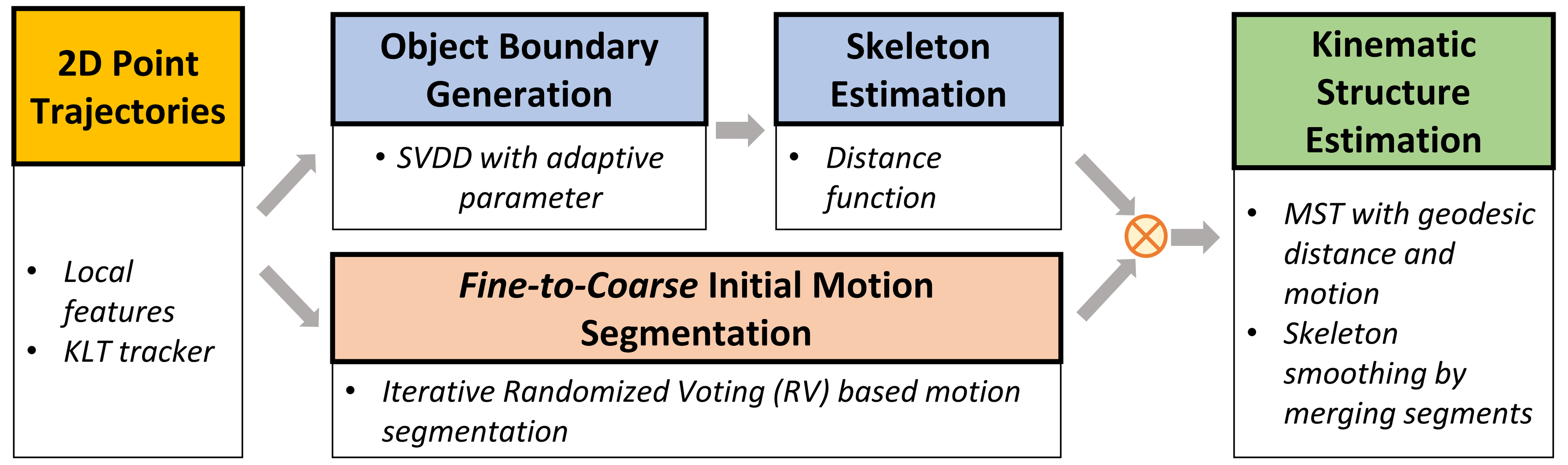
- ➢ Object kinematic recognition
- ➢ Human action recognition
- ➢ Robot body scheme learning
- ➢ Robot articulated object manipulation



CONTRIBUTIONS

i. Unsupervised articulated kinematic structure learning combining **motion** and **skeleton** information
ii. Iterative fine-to-coarse merging strategy for **adaptive motion segmentation** and **structure smoothing**
iii. Skeleton estimation from sparse feature points
iv. A new highly articulated object dataset

OVERALL FRAMEWORK



LIMITATIONS OF CURRENT METHODS

- ✓ Relatively simply articulations only ⇨ *Extreme case: human hand*
- ✓ Motion information only ⇨ *No topological constraints*
- ✓ Shape information only ⇨ *No kinematic constraints*
- ✓ Slow & sensitive to parameters ⇨ *1.8 times faster & no parameter tuning required*

CONTACT

- **Video result**, **source code**, and **the new Imperial-PRL-dataset**.are available: www.imperial.ac.uk/PersonalRobotics
- Dr. Hyung Jin Chang (hj.chang@imperial.ac.uk) and Dr. Yiannis Demiris (y.demiris@imperial.ac.uk)
- This work was supported in part by the EU FP7 project WYSIWYD under Grant 612139

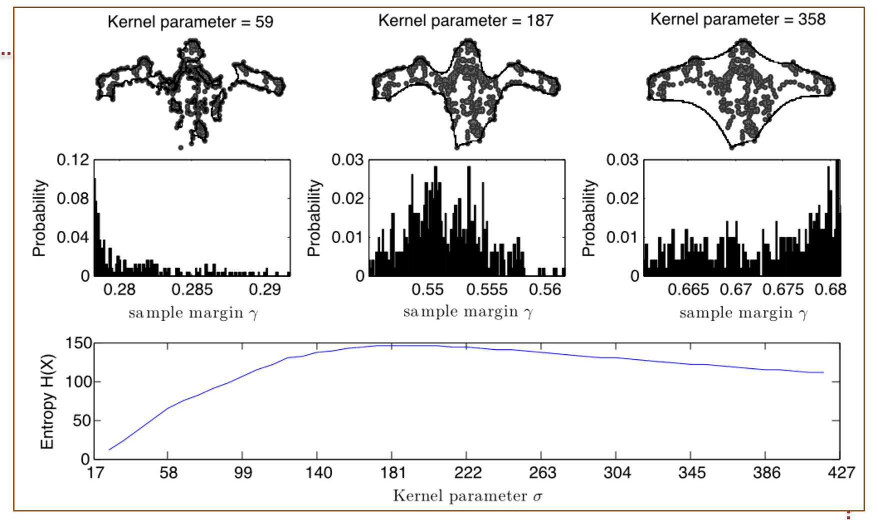
METHODOLOGY

FINE-TO-COARSE MOTION SEGMENTATION

- Difficult to ***estimate the precise number of motion segments***
- ***Iterative fine-to-coarse inference***
  - ❑ ***adaptively estimate*** an upper-bound number of initial motion segmentation
  - ❑ ***Randomised voting based motion segmentation (RV)***
    at least 8 points should be assigned to each segment to run
  - ❑ Perform RV algorithm iteratively by decreasing segment number until converge

Algorithm 1 Fine-to-coarse Motion Segmentation
Input: $x_i, i = 1, ..., N$ ▷ Point trajectories
Output: $S_k, k = 1, ..., \hat{c}$
$t \leftarrow 1$
$\hat{c}^t \leftarrow \lfloor N/8 \rfloor$ ▷ Initialise the number of segments
repeat
  $S_k^t \leftarrow$ RV motion segmentation$(\{x_i\}_{i=1}^N, \hat{c}^t)$
  $c_{<8} \leftarrow 0$
  for $k = 1, ..., \hat{c}^t$ **do**
    if $|S_k^t| < 8$ **then**
      $c_{<8} \leftarrow c_{<8} + 1$
  $\hat{c}^{t+1} \leftarrow \hat{c}^t - c_{<8}$
  $t \leftarrow t + 1$
until $c_{<8} = 0$

SKELETON FROM SPARSE FEATURE POINTS



❖ **Adaptive object boundary generation**

- ❑ Formulate a tight boundary

$$F(\mathbf{R}, \mathbf{a}) = \mathbf{R}^2 + C\sum_i \xi_i$$

subject to $\|x_i - \mathbf{a}\|^2 \le \mathbf{R}^2 + \xi_i, \quad \xi_i \ge 0 \quad \forall i$

- ❑ Kernel function for non-linear boundary: $K(x_i, x_j) = exp(-\|x_i - x_j\|/2\sigma^2)$
- ❑ A novel ***optimal kernel parameter selection*** by calculating the **entropy** of the ***sample margin*** distribution

$$\gamma(x_i) = \frac{\mathbf{a}\cdot\Phi(x_i)}{\|\mathbf{a}\|}$$

$$\hat{\sigma}^f = \arg\max_\sigma H(\gamma(X^f))$$

❖ **Skeleton distance function generation**

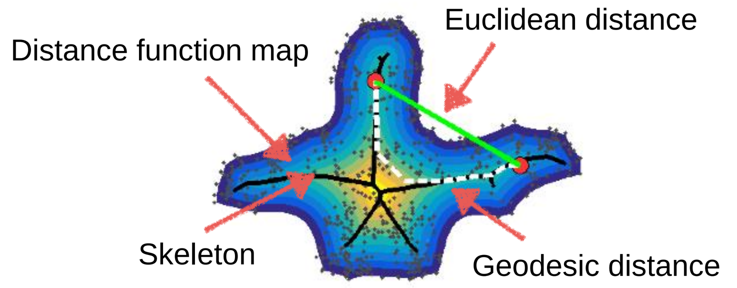
- ❑ ***Skeleton***: a set of all centre points of maximal circles inside the object boundary
- ❑ ***Skeleton distance function***: the closest distance to the boundary $\Psi(p) = \min_{q\in\delta\Omega}(dist(p,q))$

KINEMATIC STRUCTURE ESTIMATION

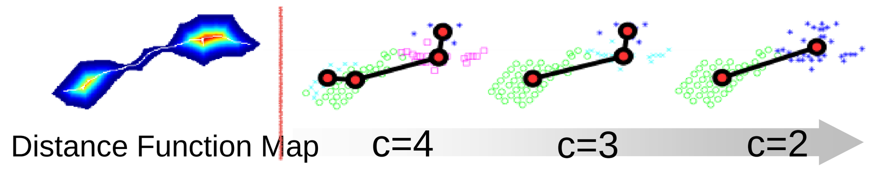
- ❑ Kinematic structure = Graphical model $G = (V, E)$
  - ▪ Nodes ***V***: motions segment centre $y_1, ..., y_{\hat{c}}$
  - ▪ Node proximity ***E***:

$$E(y_k, y_l) = \operatorname{median}_{f\in F}\{(\zeta(y_k^f - y_l^f; \Psi^f)) \times \|\dot{y}_k^f - \dot{y}_l^f\|\}$$

Geodesic distance *Motion difference*



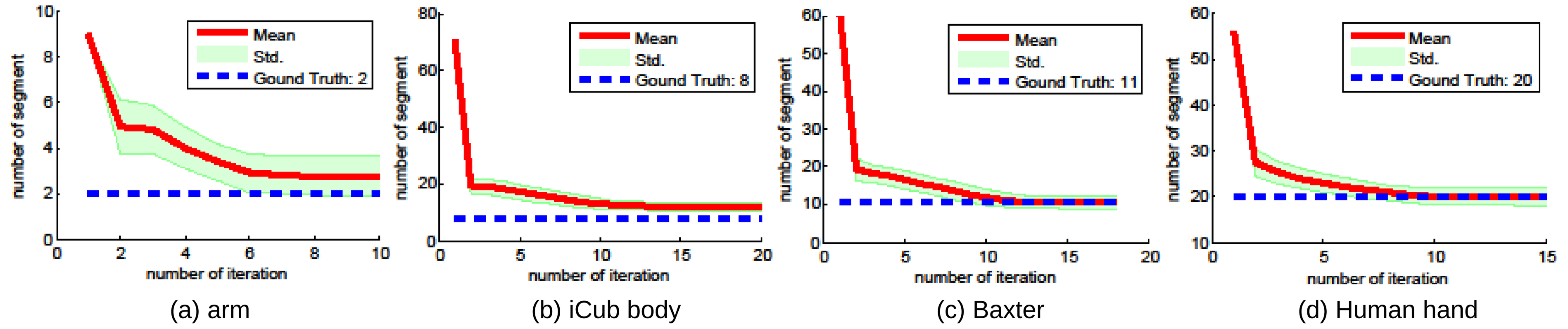
- ❑ Contorted structure smoothing by ***iterative merging guided by skeleton distance function***



EXPERIMENTAL RESULTS

▪ **Quantitative Results**

The motion segments number converges to the GT by the proposed iterative merging process.

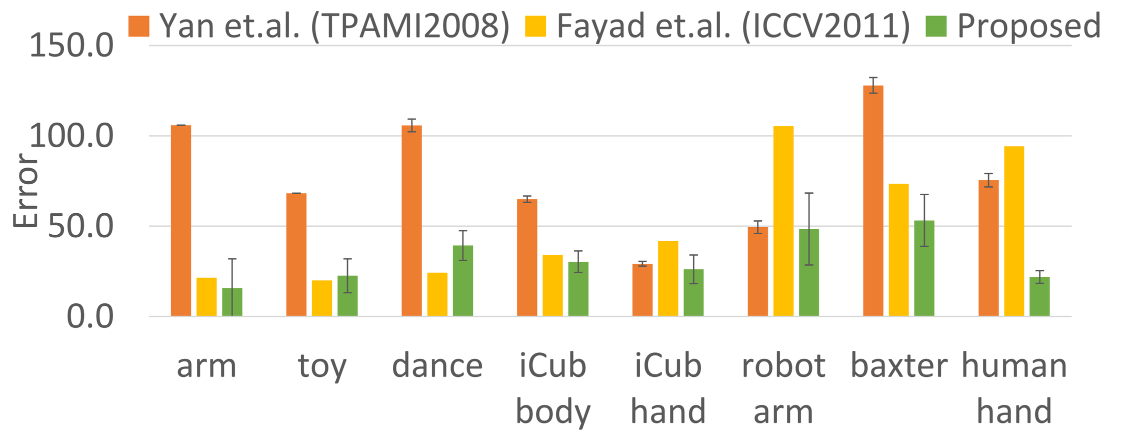


(a) arm (b) iCub body (c) Baxter (d) Human hand

Spatial deviation of each segment

$$error = \frac{1}{\hat{c}\cdot F}\sum_{k=1, f=1}^{\hat{c}, F}\left(\min_{g=1\ldots c_{GT}}\|y_k^f - y_g^f\|\right) \times \left(1 + \frac{|\hat{c} - c_{GT}|}{c_{GT}}\right)$$

Structural complexity

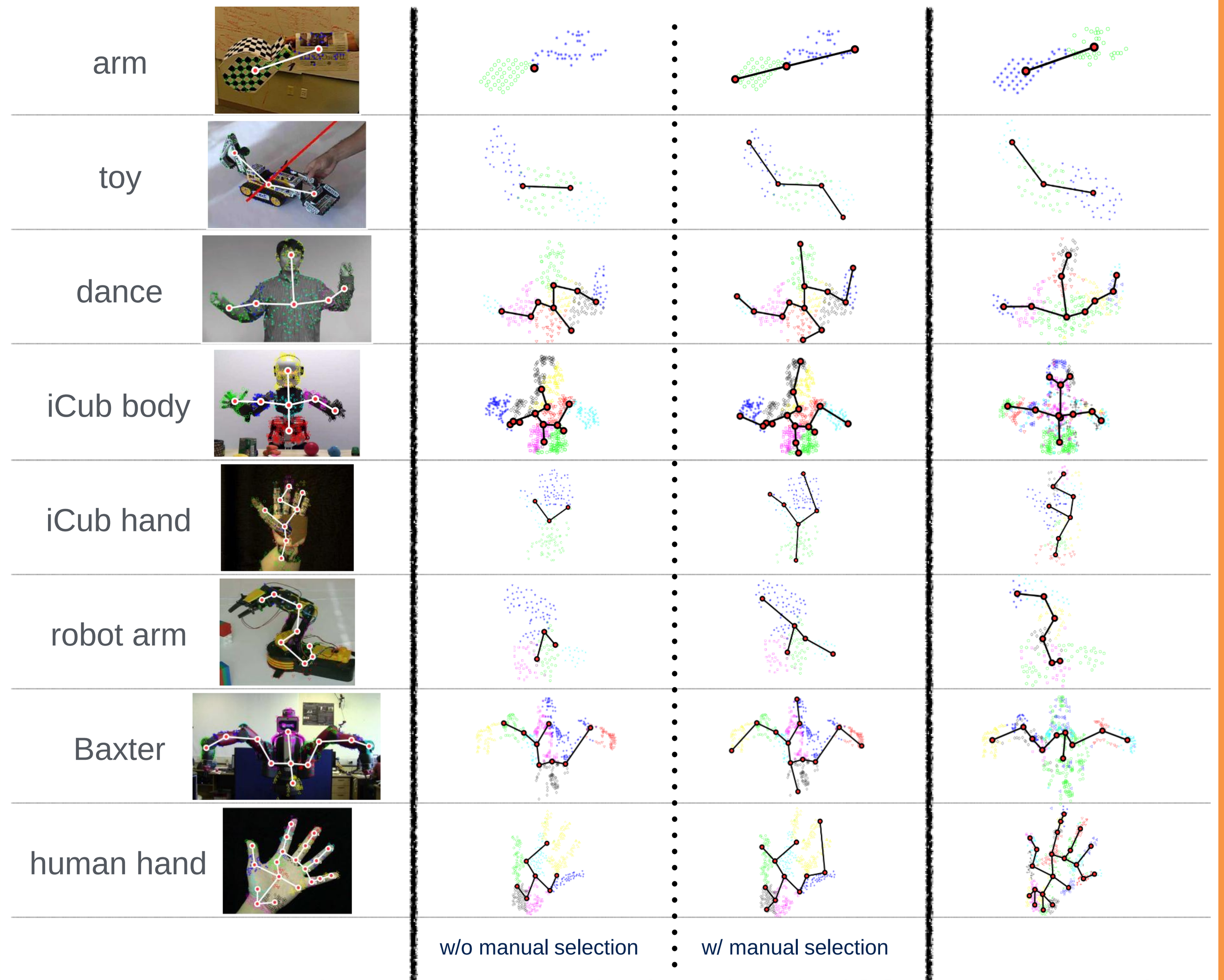


NEW IMPERIAL-PRL-DATASET

- Multiple motion segments
- Multiple articulations
- Concurrent motions
- Diverse complex motions
  - *Rotation / affine / scaling*
- Diverse dynamic objects
  - *Robots (iCub & Baxter) / human*

Dataset	# of seg.	# of points	# of frames	motion concur.
arm	2	77	30	no
toy	3	93	69	no
dance	6	236	60	yes
robot arm	8	144	737	yes
iCub body	7	573	250	yes
iCub hand	8	154	280	yes
Baxter	11	484	454	yes
human hand	20	450	634	yes

▪ **Qualitative Results**



Ground Truth Cost function based method [5] Proposed

[5] J. Fayad, C. Russell, and L. Agapito. Automated articulated structure and 3D shape recovery from point correspondences. In ICCV, 2011
[34] J. Yan and M. Pollefeys. Articulated motion segmentation using RANSAC with priors. In Lecture Notes in Computer Science, 2007
[35] J. Yan and M. Pollefeys. A factorization-based approach for articulated nonrigid shape, motion and kinematic chain recovery from video. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(5):865–877, May 2008.