Towards Understanding Complex Human Dexterous Manipulation Strategies: Kinematics of Gaiting-based Object Rotations

Jimin Hong and Aaron M. Dollar, *Senior Member, IEEE*

***Abstract*— This paper presents a novel method for tracking gaiting-based (changing contacts, reciprocal, cyclical) within-hand manipulation strategies of a human hand. We present a kinematic model that relies on data collected from 6-DOF magnetic sensors attached to 7 external sites on the hand. The sensors are calibrated by three procedures—sensor-to-fingertip, constrained fingertip workspace limits, and flat hand configuration. Subjects rotated two cubes of different sizes around the 3 object-centric axes, while a synchronized camera recorded the object motion. Hand motions were segmented and then averaged using dynamic time warping (DTW) to yield a representative time-series motion primitive for the given task. The hand movements of two subjects during cube rotation tasks were reconstructed using a 22-degree of freedom (DOF) hand kinematic model. Based on a qualitative evaluation of the joint movements, intrasubject correlations of joint angles were found.**

***Clinical Relevance*— This work lays out a methodology and preliminary qualitative results for the novel choice of studying the within-hand manipulation strategies instead of other commonly used criteria such as the execution speed for accomplishing various complex tasks. This insight would better characterize hand motions for stroke survivors undergoing rehabilitation and for improved functionality in artificial hand designs.**

I. Background

The importance of the human hand's ability to interface with the physical world cannot be understated. Endowing artificial end effectors with the level of manipulative dexterity that human hands exhibit is the Holy Grail, and as such, human hands provide a high-performance platform to study manipulation for clinical and robotic purposes. In the field of rehabilitation, hand functionality is usually tested with a small set of tasks and given scores that measure the speed of task completion, e.g. [1]–[4]. None of these methods, however, observe the detailed strategies used in completing the required tasks.

Humans frequently use within-hand manipulation (WIHM), where "within-hand" is defined as "parts of the hand moving with respect to the hand", in order to more dexterously interact with objects in the environment [5]. However, little work has been done to examine these manipulative movements, which can be quite complex, with multiple fingers moving independently and each making and breaking contact. Although manipulation is possible without within-hand dexterity, WIHM offers advantages such as improved positioning accuracy, efficiency, and object manipulability in

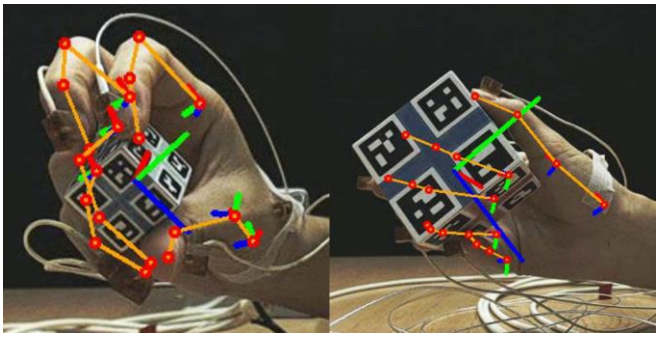

Fig. 1. Kinematic hand model superimposed on camera images

cluttered, unstructured environment with reduced upper arm mobility [5].

Some of the previous works have categorized manipulation primitives for robot and human hands alike. [6] constructs a taxonomy of grasps in action (which naturally includes manipulative instances unlike conventional static grasps) based on force types, motion directions, and stiffness. [7] divides the hand's intrinsic movements (as opposed to extrinsic movements of the wrist and the upper arm) into those based on whether the fingers move simultaneously or sequentially, whether the fingers move as a unit or in opposite directions, and the thumb movements. [8] categorizes hand-object interactions into translations between fingertips and the palm, linear shift, and simple and complex rotations, and [9] improves upon this system mainly by dividing shift into simple and complex shifts. [5] provides 5 types of robotic manipulations: regrasping, in-grasp manipulation, finger gaiting, finger pivoting/tracking, rolling, and sliding. [10] constructs a tree that categorizes manipulations based on conditions such as whether the grasp is prehensile or whether there is motion at contact. The node of most interest, the within-hand prehensile manipulation, is further divided into rotation and translation about and along the three hand-centric axes. [11] performed a dimensional reduction of the hand's grasping motion by identifying its postural synergies.

These works provide useful descriptive language for manipulations, but there are only a few, if any, quantitative studies of the human hand's dexterous capabilities in vivo during complex manipulation tasks due to two major challenges. One is the difficulty in accurately modeling the human hand, which is a complex biomechanical device with 22+ degrees of freedom. [12] assesses the validity of common assumptions made in various kinematic hand models in the literature, such as orthogonal and intersecting rotation axes at thumb joints. Another challenge is that tracking the hand in

*This work was supported in part by the US Army Medical Research and Materiel Command, under contract W81XWH-15-C-0125.

J. Hong and A. M. Dollar are with the Mechanical and Engineering and Materials Science Department, Yale University, New Haven, CT 06511 USA (e-mail: {jimin.hong, aaron.dollar}@yale.edu).

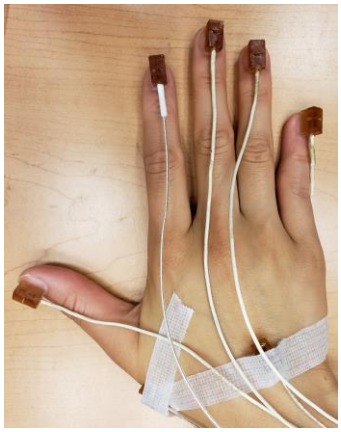

Fig. 2. Sensors are attached to the fingernails, thumb metacarpal, and middle metacarpal (reference).

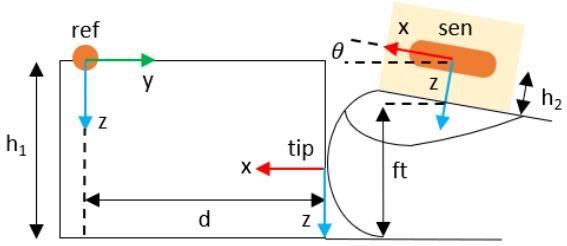

Fig. 3. Calibration from the sensor on the fingernail to fingertip

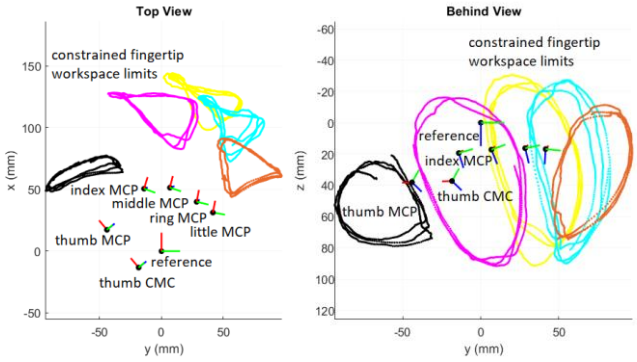

Fig. 4. Top (left) and behind (right) views of subject 2's hand

vivo during manipulation often requires an apparatus that may interfere with the naturalness of the manipulation motion. Optical motion tracking is subject to occlusion and requires at least 3 markers per rigid body (without any kinematic constraints) to reconstruct its 6-DOF pose. Wearables, such as the CyberGlove, interfere with manipulation especially at the contact interfaces [13]. Magnetic Resonance Imaging (MRI) can provide accurate bone structures, but it can be expensive [14]. Finally, vision-based hand pose estimation techniques using deep learning, such as OpenPose [15] and DeepLabCut [16], although sensor-free, do not have an underlying kinematic structure.

This work addresses the aforementioned challenges by developing a kinematic hand model that utilizes 7 magnetic sensors, the minimum number of sensors required to reconstruct the 22-DOF hand, attached to the select non-interfering external sites on the hand. This model is then used to study the hand motion during our experimental cube rotation tasks (Fig. 1).

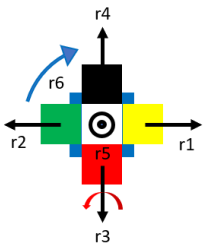

Fig. 5. A flattened view of the cube and rotations about the 3 object-centric axes in opposite directions

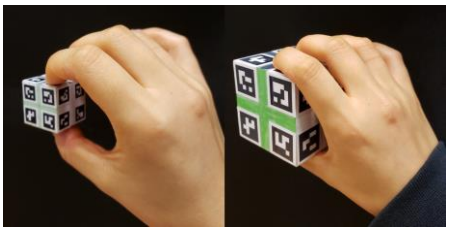

Fig. 6. Initial hand-object configuration for small and large cubes. The green side faces the radial direction while the white faces the palm.

II. Methods

A. Data Acquisition

Hand motions were recorded using a webcam synchronized with trakSTAR™ (Ascension Technology Corporation, VT), a 6-DOF magnetic motion tracking system with a medium range transmitter (workspace volume: x: 20-51 cm, y: ±23 cm, z: ±15 cm) and 7 Model 180 2 mm diameter sensors (sampling rate: 80 Hz, positional accuracy: 1.4 mm RMS, angular accuracy: 0.5 deg RMS). Each sensor was embedded in a Urethane rubber mold (durometer 80) to reduce unintended rotation about the long axis of the sensor. Five sensors were rigidly attached to the fingernails, one along the thumb metacarpal (MC) link, and another on the back of the hand along the middle MC for reference (Fig. 2). This study was approved by the Institutional Review Board (IRB#2000024214).

B. Data Calibration

Three types of calibration were conducted: sensor-to-fingertip (Fig. 3), constrained fingertip workspace limits (hereafter referred to as "constrained workspace"), and flat hand configuration (Fig. 2). For the constrained workspace, the thumb was constrained at the interphalangeal (IP) joint, and each finger at the distal interphalangeal (DIP) and proximal interphalangeal (PIP) joints, with a 3D-printed plank taped on the palmar side such that the digit remains straight. The subject then rotated the digit to its extremities, forcing the metacarpophalangeal (MCP) joint to reach its limits. The resultant workspace limits and the base joints for subject 2 are shown in Fig. 4.

C. Task Protocol

Two subjects (s1, s2) with healthy dominant right hands were recruited to rotate two birchwood cubes of different sizes (c1, c2) about the 3 object-centric axes normal to the cube faces in both clockwise and counterclockwise directions (r1—r6) as shown in Fig. 5. They were instructed to continuously rotate the cube in the given direction starting from the initial configuration (Fig. 6) for approximately 20 seconds (n = 500

frames, nonuniformly sampled). Because the motion is cyclical, the subjects conducted each trial once. The smaller cube (c1) has a side length of 2.5 cm corresponding to those used in box and block test, while the bigger cube (c2) has a side length of 5.7 cm equivalent to a standard Rubik's cube.

III. Kinematic Hand Model

The hand is modeled with the following assumptions:

- bones modeled as perfect rigid bodies
- straight fingers at zero flexion/extension
- intersecting, orthogonal joint axes
- DIP, PIP, IP joints modeled as ideal revolute joints
- MCP joints modeled as universal joints
- CMC joint modeled as a spherical joint
- joint axes fixed relative to their associated links
- parallel flexion and extension joint axes in each digit
- Minimal skin displacement at the sensor attachments

A 22-DOF hand is reconstructed by mapping the 6-DOF sensor data from the fingertips, thumb MC, and hand reference to joint angles. The frames at each joint are affixed to the link proximal to the joint and are defined such that the x-axis points distally along the link, y-axis in the ulnar direction, and z-axis toward the pad. Rotation at the multi-DOF joints modeled as universal and spherical joints (MCP and CMC, respectively) follow the Euler convention, where abduction/adduction (a/a) coincides with azimuth, flexion/extension (f/e) with elevation, and axial twist with roll. The full hand model is shown in Fig. 7. where solid lines indicate a complete fixture between the connected entities (fingertip-sensor-DIP joint); dotted lines a fixed target base frame with respect to local reference (finger MCP joints with respect to hand reference, and thumb MCP joint with respect to thumb MC); dashed lines the digit linkages; and half-dotted line the partial fixture between the CMC joint frame and the hand reference, where only its orientation, but not position, is fixed with respect to the reference.

A. Base Joint Pose Determination

The positions of the MCP joints for the thumb and the fingers were determined from the constrained workspace by treating each constrained digit as a single rigid body and locating its center of rotation according to [17]. Finger MCP joints are located with respect to hand reference, while thumb MCP joint is located with respect to thumb MC.

The constrained workspaces for the four fingers resemble an almond shape as shown in Fig. 4, consistent with the fact that MCP abduction/adduction limit decreases with increasing magnitude of flexion/extension. The base frames at the finger MCP joints were oriented based on the principal axes of the constrained workspace and the flat hand configuration such that zero angle deflection occurs when the hand is flat and that there would be no axial twist in the fingers.

The thumb's constrained workspace is circular rather than almond-like. The base frame at the thumb MCP joint was

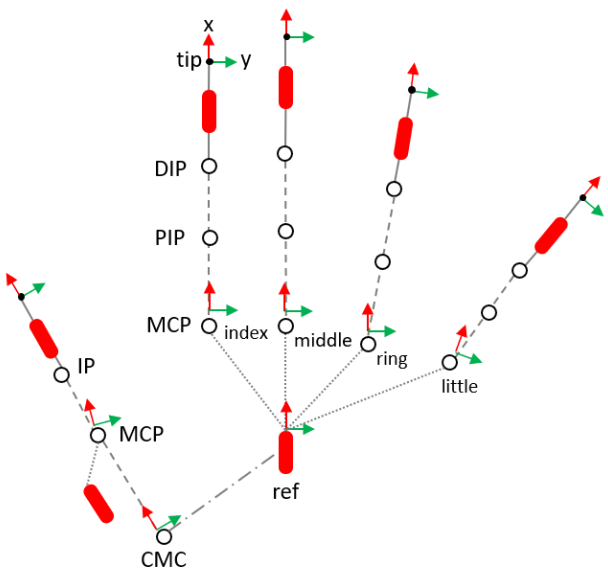

Fig. 7. Complete hand model: circles are joints; red ovals are sensors; red arrows are x-axes; green arrows are y-axes; solid lines indicate a complete fixture between the connected entities; dotted lines indicate fixed target base frame with respect to local reference sensor; dashed lines indicate digit linkages; and the half-dotted line indicates a partial fixture between the frame at the CMC joint and the hand reference, where only its orientation, but not position, is fixed.

oriented based on the average tip orientation of the constrained workspace and the flat hand configuration.

The thumb CMC joint's target position is backed out from the MCP joint in the flat hand configuration by displacing the MC link length in the -x direction with respect to the MCP frame. Its orientation is set to be equal to that of the MCP frame in the flat hand configuration and fixed with respect to the reference.

B. Link Length Determination

Thumb distal and MC links, as well as the finger distal and intermediate links, were measured using a caliper. The lengths for the remaining links—the proximal links for all digits (note that the thumb proximal link is between its distal and MC links)—were derived by calculating the finger length as the average distance between the calculated base joint position and the measured fingertip locations from the constrained workspace calibration.

C. Digit Reconstruction

Fingertip pose is fixed to the distal link for all digits, so the locations of the DIP joints for the fingers and of the IP joint for the thumb are locked with respect to the sensor readings. For the thumb, the IP joint angle is calculated by minimizing the distance between the MCP joint from the calibration step and the end of the 2-bar linkage extending from the fingertip sensor data. Similarly, the DIP and PIP joint angles for each finger are calculated by minimizing the distance between the MCP joint from the calibration step and the end of the 3-bar linkage extending from the fingertip while obeying the joint limits specified in (1). The actual MCP joints are hence unhinged with respect to the reference due to the position error.

$$max/min\ dip = 0.75 * pip \pm 24 \quad (1)$$

TABLE I. JOINT CHOICES FOR TASK SEGMENTATIONS

	Subject 1		Subject 2	
	Cube 1	**Cube 2**	**Cube 1**	**Cube 2**
Rot 1	max t IP	max t IP	max t IP	max t IP
Rot 2	max t IP	max t IP	max t IP	max t IP
Rot 3	min r PIP	min r PIP	min r PIP	min r PIP
Rot 4	min r PIP	min r PIP	min r PIP	min r PIP
Rot 5	min r MCP f/e	min r MCP f/e	min r MCP f/e	min l MCP a/a
Rot 6	min t MCP f/e	min t MCP f/e	min t MCP f/e	min t MCP f/e

IV. DATA ANALYSIS

For the purposes of this analysis, the following 22 joint angles were considered: thumb IP, thumb MCP (f/e and a/a), thumb CMC (f/e, a/a and twist), finger DIP, finger PIP, and finger MCP (f/e and a/a).

A. Motion Segmentation

The joint that seemed to delineate the discrete cycles in each task in the most distinct manner was identified based on the video recordings. The same joint choice between related tasks (e.g. between rotations in opposite directions) was preferred to establish a common ground for comparison such that the segmented cycles are not out of phase with their counterparts. Table I enumerates joint choices used to segment the time-series data. Except for the case of s2c2r5, in which little MCP a/a instead of ring MCP f/e was used, all rows (across subjects and cube sizes) have the same joint choice. And except for the rotations 5 and 6, the joint choices for the opposite rotations (1 and 2, and 3 and 4) are the same. A different joint choice for segmenting s2c2r5 was used because while ring MCP f/e data provided effective points for segmentation for the other 3 tasks under rotation 5, the ring finger in s2c2r5 did not behave consistently across the different cycles, sometimes moving with the index and the middle fingers and other times with the little finger.

Fig. 8 is an example of motion segmentation using ring MCP's f/e local minima. Although more local minima are present in the plot, only the minima that segment the data along identifiable cycles were preserved. The video recordings of the motions between adjacent segmentation points were reviewed to eliminate portions that do not contribute to the cyclical motion, which include slipping or regrasping that occur every so often. In the case of s1c2r5, the third segment just before and after 6 seconds was removed.

B. Dynamic Time Warping (DTW) Methods

The similarities among the motion segments must be quantified. Generally, calculating a divergence between two time-series data requires that the data or the corresponding feature vectors be of equal length. Uniform resampling to achieve this would lose information about the original data. It would also linearly and globally stretch or compress the data, which is not ideal in this case since we desire characteristic epochs to be aligned so that varying manipulation speed would not be a factor in distinguishing different strategies. Dynamic time warping (DTW) [20] calculates the divergence between two time-series with variable lengths by making them uniform in length while aligning phase-shifted epochs. DTW operates

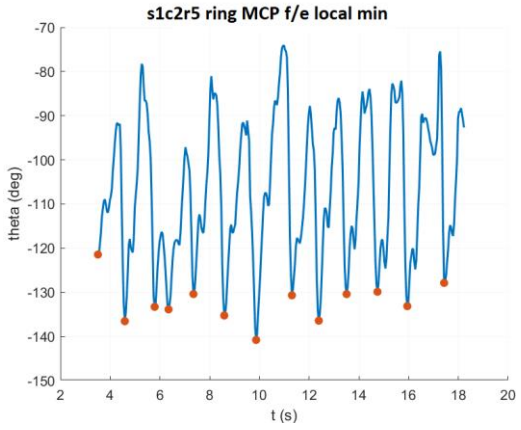

Fig. 8. Motion segmentation of the s1c2r5 task using ring MCP joint's f/e local minima. The third segment was removed due to its relative brevity and dissimilarity from the other segments.

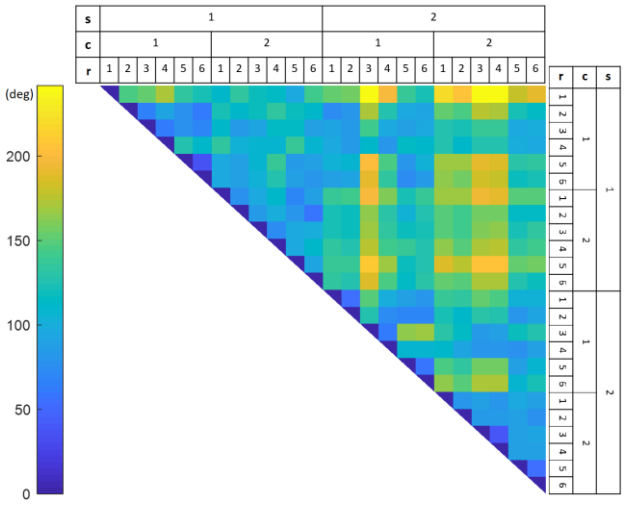

Fig. 9. Normalized correlation matrix between motion averages within each task category

by replicating the frames such that the sum of squared Euclidean distances is minimized according to

$$D(i,j) = min\begin{Bmatrix} D(i-1,j) + d(i,j) \\ D(i-1,j-1) + d(i,j) \\ D(i,j-1) + d(i,j) \end{Bmatrix}, D(1,1) = d(1,1), \quad (2)$$

where *d(i,j)* corresponds to the Euclidean distance between the frame *i* of one time-series and the frame *j* of the other. The optimal path is calculated through matrix *D(i,j)* by starting at the last frames of both time-series (bottom right corner of the matrix) and moving backwards through the smallest distance values (moving left or up) towards the first frames (upper left corner of the matrix).

DTW, however, can only handle two time-series at once. Multiple segments per task category are thus aligned and converted to uniform length using the DTW-based batch synchronization method [18]. This method is asymmetric and needs a reference time-series segment, which, in this case, was set to be the average segment obtained through barycenter averaging based on DTW (DBA) [19]. The longest sequence in the batch was used as the initial value for the batch average. These DTW-based algorithms are prone to over-stretching a segment to minimize its distance to the other segment, which

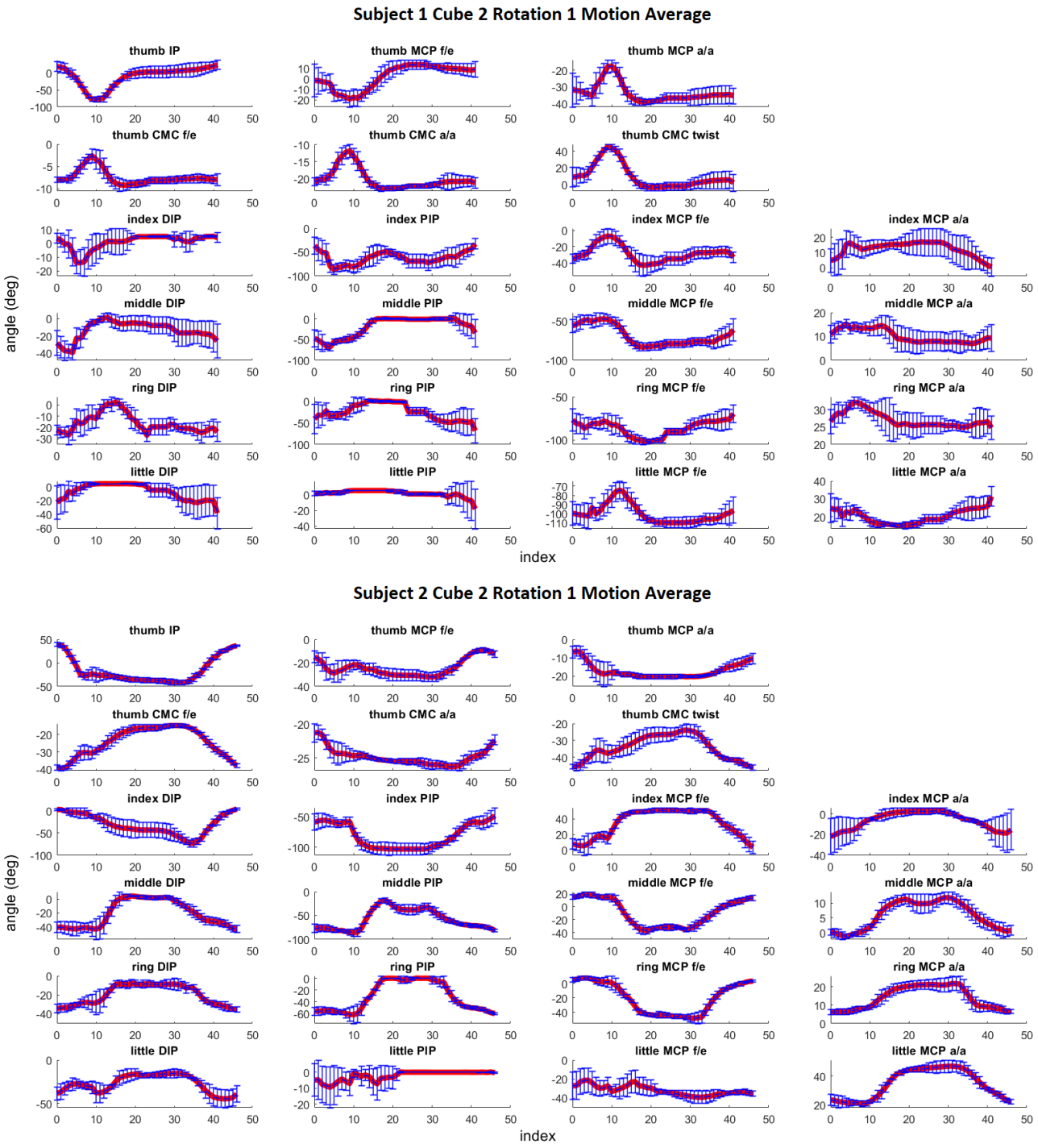

Fig. 10. Comparison of the 22-DOF time-series for the cube 2 rotation 1 task motion primitives between subject 1 (top) and subject 2 (bottom)

may result in accentuated amplitudes in the presence of local maxima. To prevent this, the maximum number of points that can be warped through replication was set to be the difference in length between the shortest and the longest segments among the segments being averaged.

V. Results and Discussion

The 22-DOF motion primitives for each of the task categories were calculated using DBA, and their normalized pairwise divergences calculated using DTW were visualized in a correlation matrix in Fig. 9. The matrix shows that the upper right and the lower left quadrants are generally warmer than the upper left and the lower right quadrants, suggesting that there is less similarity between subjects regardless of tasks than across tasks for a given subject. The matrix also shows that s1c1r1 is the most distinct strategy with greatest distances from other task averages. The darkest cells indicating small distances, and hence similar strategies, appear between s1c1r5 and s1c1r6 and between s2c2r3 and

s2c2r4, each pair being rotations in opposite directions. In general, every other off-diagonal cell corresponding to rotations in opposite directions tend to be dark, suggesting that the strategies for cube rotations in opposite directions tend to be more similar than when compared to other manipulation tasks.

Fig. 10 compares the joint angle time-series for the cube 2 rotation 1 (approximately about the hand's ulnar axis) across the two subjects. The motion averages are displayed as red solid lines with blue error bars one standard deviation below and above the average. Note that the y-axis scales are not equal among the joints to show variance at highest resolution. Both sets of plots were segmented at thumb IP's local maximum extension. Subject 1's thumb joints are characterized by more distinguished maxima compared to subject 2. Similar trajectories between different joints indicating strong couplings can be observed, such as between middle and ring fingers (4^{th} and 5^{th} rows, respectively) for subject 2. Coincidentally, the index joints (3^{rd} row) for subject 2 exhibit a vertically inverted trend from the middle and the ring joints, showing a reciprocal motion between the index and the middle-ring unit. Subject 2 employs slightly more consistent strategies compared to subject 1 as shown by smaller error bars. There is very little, if at all, correlation between the two subjects for this particular task, but this remains to be proven quantitatively. The video recordings suggest that the difference in the strategies employed by the two subjects for this task may be attributed to the different orientations of the grasp with respect to gravity, where gravity aids manipulation for subject 1 while subject 2 rotates the cube against the gravity. Although only 2 of the 24 sets of plots are displayed here, a qualitative evaluation of these plots reveals that 1. in general, there is a greater intrasubject joint signal correlation (indicated by the size of the error bars) than intersubject correlation, and 2. the different cube sizes seem to affect the range of the joint motion but not the finger coordination patterns.

VI. Future Work

This work is novel in obtaining in vivo joint angles during dexterous within-hand manipulation with minimal interference—despite a slight strain from the sensor wires, the interface between the object and skin remains undisrupted. With the methodology in place, the study should be expanded to include more subjects, objects, and manipulation tasks. With a richer dataset, the 22-DOF joint space can be reduced to a lower-dimensional manifold using principal component analysis to quantify the couplings and synergies in the hand posture space. The motion segments for various tasks can also be clustered and the cluster inertias analyzed to quantify the intersubject and intrasubject similarities in WIHM strategies. The 6-DOF pose of the object obtained through marker tracking can be used to better segment the motion and can be correlated with hand motions to reveal strategies that yield the desired object motion. Also, the object's motion deviance from instructed directions can be analyzed to study the natural axes of manipulation motion relative to the hand. For robotic applications, we seek to explore the mapping between the object motion and the finger motion, although such relationship will not be one-to-one.

References

[1] C. W. Wu, H.-J. Seo, and L. G. Cohen, "Influence of Electric Somatosensory Stimulation on Paretic-Hand Function in Chronic Stroke," *Arch. Phys. Med. Rehabil.*, vol. 87, no. 3, pp. 351–357, Mar. 2006, doi: 10.1016/j.apmr.2005.11.019.

[2] T. Stamm, M. Mathis, D. Aletaha, M. Kloppenburg, K. Machold, and J. Smolen, "Mapping hand functioning in hand osteoarthritis: Comparing self-report instruments with a comprehensive hand function test," *Arthritis Care Res.*, vol. 57, no. 7, pp. 1230–1237, 2007, doi: 10.1002/art.22989.

[3] C. M. Light, P. H. Chappell, and P. J. Kyberd, "Establishing a standardized clinical assessment tool of pathologic and prosthetic hand function: Normative data, reliability, and validity," *Arch. Phys. Med. Rehabil.*, vol. 83, no. 6, pp. 776–783, Jun. 2002, doi: 10.1053/apmr.2002.32737.

[4] A. Kapandji, "[Clinical test of apposition and counter-apposition of the thumb].," *Ann. Chir. Main Organe Off. Soc. Chir. Main*, vol. 5, no. 1, pp. 67–73, 1986, doi: 10.1016/S0753-9053(86)80053-9.

[5] R. R. Ma and A. M. Dollar, "On dexterity and dexterous manipulation," in *2011 15th International Conference on Advanced Robotics (ICAR)*, 2011, pp. 1–7, doi: 10.1109/ICAR.2011.6088576.

[6] J. Liu, F. Feng, Y. C. Nakamura, and N. S. Pollard, "Annotating Everyday Grasps in Action," in *Dance Notations and Robot Motion*, J.-P. Laumond and N. Abe, Eds. Cham: Springer International Publishing, 2016, pp. 263–282.

[7] J. M. Elliott and K. J. Connolly, "A Classification of Manipulative Hand Movements," *Dev. Med. Child Neurol.*, vol. 26, no. 3, pp. 283–296, 1984, doi: 10.1111/j.1469-8749.1984.tb04445.x.

[8] C. E. Exner, "In-Hand Manipulation Skills," in *Development of hand skills in children*, Rockville, MD: American Occupational Therapy Association, 1992, pp. 33–45.

[9] K. Pont, M. Wallen, and A. Bundy, "Conceptualising a modified system for classification of in-hand manipulation," *Aust. Occup. Ther. J.*, vol. 56, no. 1, pp. 2–15, 2009, doi: 10.1111/j.1440-1630.2008.00774.x.

[10] I. M. Bullock, R. R. Ma, and A. M. Dollar, "A Hand-Centric Classification of Human and Robot Dexterous Manipulation," *IEEE Trans. Haptics*, vol. 6, no. 2, pp. 129–144, Apr. 2013, doi: 10.1109/TOH.2012.53.

[11] M. Santello, M. Flanders, and J. F. Soechting, "Postural Hand Synergies for Tool Use," *J. Neurosci.*, vol. 18, no. 23, pp. 10105–10115, Dec. 1998, doi: 10.1523/JNEUROSCI.18-23-10105.1998.

[12] I. M. Bullock, J. Borràs, and A. M. Dollar, "Assessing assumptions in kinematic hand models: A review," in *2012 4th IEEE RAS EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, 2012, pp. 139–146, doi: 10.1109/BioRob.2012.6290879.

[13] G. D. Kessler, L. F. Hodges, and N. Walker, "Evaluation of the CyberGlove As a Whole-hand Input Device," *ACM Trans Comput-Hum Interact*, vol. 2, no. 4, pp. 263–283, Dec. 1995, doi: 10.1145/212430.212431.

[14] G. Stillfried, U. Hillenbrand, M. Settles, and P. van der Smagt, "MRI-Based Skeletal Hand Movement Model," in *The Human Hand as an Inspiration for Robot Hand Development*, R. Balasubramanian and V. J. Santos, Eds. Cham: Springer International Publishing, 2014, pp. 49–75.

[15] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," *ArXiv181208008 Cs*, May 2019.

[16] A. Mathis *et al.*, "DeepLabCut: markerless pose estimation of user-defined body parts with deep learning," *Nat. Neurosci.*, vol. 21, no. 9, pp. 1281–1289, Sep. 2018, doi: 10.1038/s41593-018-0209-y.

[17] S. S. H. U. Gamage and J. Lasenby, "New least squares solutions for estimating the average centre of rotation and the axis of rotation," *J. Biomech.*, vol. 35, no. 1, pp. 87–93, Jan. 2002, doi: 10.1016/S0021-9290(01)00160-9.

[18] A. Kassidas, J. F. MacGregor, and P. A. Taylor, "Synchronization of batch trajectories using dynamic time warping," *AIChE J.*, vol. 44, no. 4, pp. 864–875, 1998, doi: 10.1002/aic.690440412.

[19] F. Petitjean, A. Ketterlin, and P. Gançarski, "A global averaging method for dynamic time warping, with applications to clustering," *Pattern Recognit.*, vol. 44, no. 3, pp. 678–693, Mar. 2011, doi: 10.1016/j.patcog.2010.09.013.

[20] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 26, no. 1, pp. 43–49, Feb. 1978, doi: 10.1109/TASSP.1978.1163055.