REVIEW

Recent Data Sets on Object Manipulation: A Survey

Yongqiang Huang,[1] Matteo Bianchi,[2] Minas Liarokapis,[3] and Yu Sun[1,*]

Abstract

Data sets are crucial not only for model learning and evaluation but also to advance knowledge on human behavior, thus fostering mutual inspiration between neuroscience and robotics. However, choosing the right data set to use or creating a new data set is not an easy task, because of the variety of data that can be found in the related literature. The first step to tackle this issue is to collect and organize those that are available. In this work, we take a significant step forward by reviewing data sets that were published in the past 10 years and that are directly related to object manipulation and grasping. We report on modalities, activities, and annotations for each individual data set and we discuss our view on its use for object manipulation. We also compare the data sets and summarize them. Finally, we conclude the survey by providing suggestions and discussing the best practices for the creation of new data sets.

Keywords: robotics; robotic manipulation; robotic grasping; learning; data collection; data set

Introduction

Big and organized data sets are valuable in various scientific fields, primarily because they are crucial for revealing hidden patterns, testing hypotheses, and evaluating algorithms. The demand for new data sets follows the advancement of a multidisciplinary field or the evolution of particular problems. In robotics grasping and manipulation, many data sets have recently been created by a number of groups for different research purposes and have been shared with the robotics community. It is possible to find human motion data sets,[1] instrumental activities of daily living (IADLs) data sets,[2] object and model data sets,[3] object geometry and motion data sets,[4] haptic interaction data sets,[5] among others. The data sets are not only crucial for evaluating and comparing the performances of novel methods[5] but they are also extremely valuable for motion/path planning (see Ref.[6] for a review), robotic learning and training,[7] and investigation of human behavior. The goal is to achieve a mutual inspiration between neuroscience and robotics, thus leading to the definition of effective design and control guidelines for artificial systems.[8]

The role of data sets should be not only to inform the control and development of robotic devices but also to verify or deny the correctness and effectiveness of an algorithm or system design, and expose the flaws or exemplify the strength of the algorithm or the design itself. However, to properly choose a *good* data set, we first need to know which data sets are already available, what information they include, and how they differ. Then we can decide on whether any data set would be useful and which data set would best serve the research purpose. We may also decide that none of the data sets suits the purpose, and the reason on which that particular decision is made can be used to improve on the existing data sets and prepare new data sets.

To help one with choosing the right data set(s) or deciding on creating new data sets, we contribute a review of data sets that we consider useful for research on object manipulation and grasping. The data sets were created no earlier than 2009 because earlier data sets are usually not supported or accessible.

Object manipulation is the process of changing in a controlled manner the position and orientation of an object to execute a specific task. In contrast to a gross motion such as waving and stretching, an object manipulation motion is a fine motion, and the body parts involved cover a much smaller physical space.

[1]*Department of Computer Science and Engineering, University of South Florida, Tampa, Florida.*
[2]*Research Center "E. Piaggio," Dipartimento di Ingegneria dell'Informazione, University of Pisa, Pisa, Italy.*
[3]*School of Engineering and Applied Science, Yale University, New Haven, Connecticut.*

*Address correspondence to: *Yu Sun, Department of Computer Science and Engineering, University of South Florida, 4202 East Fowler Avenue, ENB 118, Tampa, FL 33620,* E-mail: yusun@mail.usf.edu

In this survey, we mainly focus on data sets that contain object manipulation motions. Gross motions may be present in certain reviewed data sets, but they do not play a dominant role. Under this regard, it is worth to mention another type of data sets, which are specifically designed for whole-body motions or limb motions, and hence are not considered in this review. They are particularly important to humanoid research, behavioral science, rehabilitation, neuroscience, and human–computer interaction. One example is the KIT Whole-Body Human Motion Database[*,1] which focuses on human and object motions, which are annotated through motion description tags. It contains not only motion captured data in a raw format (e.g., marker motions of the capture system) but also information on the subject anthropometric measurements, the objects used, and the environment along with a data interface to transfer motion to different kinematic and dynamic models of humans and robots. Other large-scale motion databases available to the scientific community are reviewed in Ref.[1]

Furthermore, given the importance of grasping as one of the key topics in robotics research and the motivation for the development of effective robotic manipulators,[9] we also discuss data sets on human grasping. Indeed, grasping not only facilitates manipulation but also determines the arm motions in many object manipulations as indicated in Refs.,[10,11] and hence can be considered as premanipulation. As previously mentioned, human grasp data sets can offer useful insights not only to better understand the human behavior but also to shape the design and control of artificial systems.[12] Under this regard, it is worth to mention the concept of *hand postural synergies*,[13] that is, broadly, goal-directed kinematic patterns of covariation observed between the human hand joints. The underlying concept for a general geometrical interpretation of synergies is the dimensionality reduction, that is, the number of degrees of freedom (DOFs) of the human hand that can be controlled in an independent manner is actually smaller than the total number of available DOFs.[14] This idea has been successfully applied in robotics (1) to relax the control effort using less control variables[15] and (2) to reduce the dimensionality of the problem in robotic grasp planning with dexterous hands,[16] facilitating an on-line grasp synthesis.[17] At the same time, such dimensionality reduction has inspired the design of underactuated robotic hands.[18,19]

*https://motion-database.humanoids.kit.edu

Table 1. Publication year of data sets (including data sets of HandCorpus)

Year	*2009*	*2010*	*2011*	*2012*	*2013*	*2014*	*2015*	*2016*
Data sets	27	28	25	21	29	30	24	31
	32	33	13b	21+	34	35	36	37
	38		39	22	40	41	42c	43
	44		45	23	26	46	47	48
				49	42a	42b	50	51
				13a		52		

Ref.[21] includes two different data sets which are denoted as 21 and 21+, respectively.

In this article, we divide the data sets into three categories and present them separately: those that include mostly cooking activities, in the Data Sets of Cooking Activity section, those that include more general activities of daily living (ADLs), in the Data Sets of ADLs section, and the data sets on kinematics of human grasping of real or imaginary objects, in the Data Sets of Grasping section. Following a common classification of human hand-pose reconstruction systems,[20] we decided to organize the review of grasping and manipulation data sets considering pure vision-based acquisitions and wearable-based acquisitions as independent entries, as discussed in the Data Sets of Grasping section. It is worth noting that marker-based recordings of grasping and manipulation activities are also described in the Data Sets of Grasping section, because a certain level of wearability is still required because of the usage of optical markers placed on the dorsum of human hand. All data sets are summarized in Table 1 that classifies the data sets according to the year that they were published. In Table 2, we list the number of instances provided in each data set. When a data set contains sequences, we report the number of sequences, otherwise, we report the number of data samples.

In the first two categories, we present the data sets in ascending chronological order. For each data set, we report on the modalities, the activities performed, and annotations, and we give our view on how each data set relates to object manipulation. After reporting on the data sets one-by-one, we summarize them on the availability of modalities, object identifiability in annotated activities, and the forms of temporal segmentation of annotated activities. We also provide the lists of shared annotated activities for the ADL and cooking data sets, respectively.

The data sets reviewed in category 3 are hosted in the HandCorpus initiative website,[53] an open access repository for sharing data about human and robot hands,

Table 2. Number of instances provided by each data set

Cooking			ADLs			HandCorpus		
Data	*Size*	*Type*	*Data*	*Size*	*Type*	*Data*	*Size*	*Type*
27	20	ms	38	20	ms	13a	286	rg, g
32	218	ms	44	150	v	31	8739	rg
45	28	v	28	24	ms	33	19	rg, g
21	17	ms	25	60	ms	36	825	rg, g
21+	30	ms	26	120	ms	13b	285	g
22	44	v	49	20	v	39	114	g
23	256	v	41	979	ms	42a	3694	rg
24	273	v	46	18,210	g	42b	1	k
29	50	ms	43	~59,000	ms	42c	300	g
34	35	ms	37	~650,000	g			
40	88	v	48	>1000	ms			
30	67	ms	47	~12,000	g			
35	77 hour	ms	50	13	ms			
			51	193	ms			
			52	4	ms			

The type "ms" or "multimodal sequence" refers to sequences that contain multiple modalities.

ADLs, activities of daily living; g, grasp; k, kinematic model; rg, reach and grasp; v, RGB video.

Table 3. Links to data sets

References	*Links*
27	http://openlab.ncl.ac.uk/publicweb/publicweb/AmbientKitchen/KitchenData/Slice & Dice_dataset
32	http://kitchen.cs.cmu.edu
45	Images: http://ai.stanford.edu/~alireza/GTEA/ and the rest: www.dropbox.com/sh/q4s6nocyhpnauic/AAAvCTfVPCo1u0vTCOsHGwA_a?dl=0
21+	http://ai.stanford.edu/~alireza/GTEA_Gaze_Website
22	www.mpi-inf.mpg.de/departments/computer-vision-and-multimodal-computing/research/human-activity-recognition/mpii-cooking-activities-dataset
23	www.mpi-inf.mpg.de/departments/computer-vision-and-multimodal-computing/research/human-activity-recognition/mpii-cooking-composite-activities
24	www.mpi-inf.mpg.de/departments/computer-vision-and-multimodal-computing/research/human-activity-recognition/mpii-cooking-2-dataset
29	http://cvip.computing.dundee.ac.uk/datasets/foodpreparation/50salads
34	www.murase.m.is.nagoya-u.ac.jp/KSCGR
40	http://web.eecs.umich.edu/~jjcorso/r/youcook
30	http://robocoffee.org/datasets
35	http://serre-lab.clps.brown.edu/resource/breakfast-actions-dataset
38	https://ias.in.tum.de/software/kitchen-activity-data
44	www.cs.rochester.edu/~rmessing/uradl
28	UCI repository: https://archive.ics.uci.edu/ml/datasets/OPPORTUNITY+Activity+Recognition#, Challenge: www.opportunity-project.eu/challengeDataset
25,26	http://pr.cs.cornell.edu/humanactivities/data.php
49	www.csee.umbc.edu/~hpirsiav/papers/ADLdataset
41	https://archive.ics.uci.edu/ml/datasets/Dataset+for+ADL+Recognition+with+Wrist-worn+Accelerometer#
46	www.eng.yale.edu/grablab/humangrasping
52	http://wildhog.ics.uci.edu:9090/#EGOCENTRIC0%20Intel/Creative
13a	www.handcorpus.org/?p=97
31	www.handcorpus.org/?p=1596
33	www.handcorpus.org/?p=100
36	www.handcorpus.org/?p=1507
13b	www.handcorpus.org/?p=91
39	www.handcorpus.org/?p=103
42a	www.handcorpus.org/?p=1156
42b	www.handcorpus.org/?p=1298
42c	www.handcorpus.org/?p=1578
43	https://sites.google.com/site/brainrobotdata/home/push-dataset
37	https://sites.google.com/site/brainrobotdata/home/grasping-dataset
48	http://rpal.cse.usf.edu/imd
51	https://github.com/jrl-umi3218/ManipulationKinodynamics
47	www.gregrogez.net/research/egovision4health/gun-71
50	www.hci.iis.u-tokyo.ac.jp/~cai-mj/utgrasp_dataset.html

Ref.[21+] refers to both Gaze and Gaze+.

with the objective of advancing the state of the art of the analysis of both the biological and the artificial side. The HandCorpus goal, under an engineering point of view, is to devise design guidelines from biology observations for the development of effective robotic devices (see Ref.[19]) and grasp planning algorithms.[54] Among all the data sets in HandCorpus, we select and review nine data sets that have collections of human hand kinematics recorded in grasping and manipulation tasks.

For those who want to further examine the data sets covered in this work, we provide the links to all data sets in Table 3.

Data Sets of Cooking Activity

In this section, we present 13 data sets of cooking activities. The interest in studying cooking activities is motivated by the large number of interactions with the objects and the external environment that human hands and body usually undergo. The data sets include common visual-based acquisition modalities such as Red Green Blue (RGB) vision and depth (D) vision, as well as modalities that are less common such as skin temperature and body heat. RGB vision is used by all data sets. We first present each data set individually, describing the different characteristics: data type and size, modalities, equipment, annotations, etc. Then, we compare the data sets on their different descriptive fields and discuss their suitability and applicability for learning from demonstration (LfD),[55] also known as programming by demonstration, or imitation learning.

Slice&Dice

Slice&Dice[27] features four instrumented utensils that include three knives of different sizes and a spoon. Each utensil embeds in its handle a three-axis accelerometer. Twenty subjects participated and each subject prepared a salad or a sandwich freely using the ingredients provided by the experimenter. The acceleration data are accompanied by RGB videos. We consider embedding accelerometers inside objects a merit as, unlike vision-based sensors, they provide acceleration data that belong to a certain object alone, and are readily usable without running object recognition first.

CMU-MMAC

The CMU-MMAC data set[32] contains multimodal cooking activities of five recipes: brownie, eggs, pizza, salad, and sandwich. The modalities include RGB videos from static and wearable cameras, multichannel audios, motion capture, inertial measurement units (IMUs), and Radio Frequency Identification (RFID). We are not positive on the number of subjects who were involved, but we infer that it is between 39 and 45. Each subject prepared all the recipes. The data set also specifically recorded anomalous accidental events that occurred while cooking. Certain modalities are incomplete for certain recipes prepared by certain subjects. Annotations exist for 16 subjects while preparing brownies and correspond to the videos captured by the wearable camera. The annotations apply the structure of "verb + objectOne + preposition + objectTwo," whose components are assembled using grammar.

Except RFID tagging that merely reports the involvement of certain objects, all modalities are on human, which is contrary to the Slice&Dice data set.[27] The data set is rich in data of upper arm motions because of the combined use of motion capture and IMUs, and therefore is suitable for 3D manipulation motion analysis.

GTEA data set

The GTEA data set[45] includes egocentric videos of four subjects performing seven food/beverage preparing activities. The videos consist of 31,222 RGB images. Annotations consist of simple verbs (put, take, pour, etc.) and names of objects (cup, sugar, etc.). Object recognition or manually drawn bounding boxes on objects are required before analysis of the object motion.

Gaze and Gaze+

The Gaze data set[21] contains RGB egocentric videos of 14 subjects preparing meals using provided ingredients on a table. The videos were captured using an eye-tracking camera and, therefore, are accompanied by gaze data. The Gaze+ data set[21] (later referred to as Ref.[21+]) is an upgrade to Gaze and provides the two modalities in Gaze plus audio. The videos have higher resolution than Gaze and were captured in an instrumented kitchen instead of on a simple table. Ten subjects were involved and each one of them prepared a set of seven dishes. Actions and objects were annotated in the same way as in Gaze. Compared with static images, egocentric images have much larger proportions of the image showing object manipulation specifically and contain more detail, which we consider a merit. Analyzing object motion, however, would assume that object tracking has been done.

Data sets Gaze and Gaze+ share a reference. To distinguish them, in Table 1, 2, 3, 4, 8 and 9, we use Ref. 21+ to refer to Gaze+.

MPII Cooking, Cooking Composite, and Cooking 2

MPII sequentially created three data sets related to cooking: the MPII Cooking data set[22] that focuses on fine-grained activity, the MPII Cooking Composite data set[23] that focuses on composite activities composed of basic-level activities, and the MPII Cooking 2 data set[24] that unifies and is an upgrade of both.[22,23]

The MPII Cooking data set involved 12 subjects each preparing 1–6 out of 14 dishes and contains 45 RGB high-definition (HD) videos with a total length of more than 8 hours or 881,755 frames. The annotations include 65 activities, and 5609 instances were identified.

The MPII Cooking Composite data set included all the videos from the MPII Cooking data set and added 212 newly recorded videos. Eighteen more subjects than in the MPII Cooking data set participated. Different from the MPII Cooking data set, the MPII Cooking Composite data set annotations include four categories: activities (e.g., verbs), ingredients, tools, and containers, which combined are referred to as "attributes." There exist 218 attributes in the data set, among which 78 are activities. A total of 49,258 attribute instances have been identified that belong to 12,642 annotated temporal segments.

As a refined superset of Refs.,[22,23] the MPII Cooking 2 data set contains 273 videos involving 30 subjects. The data set contains 59 dishes, which consist of 14 diverse and complex dishes from Ref.,[22] and 45 shorter and simpler composite dishes from Ref.[23] A total of 222 attributes exist, among which 87 are activities. A total of 54,774 attribute instances have been identified that belong to 14,105 temporal segments. For the mentioned MPII data sets, the subjects were only told which dish to prepare, which led to natural activities with much variability.

Of all the data sets we include in this work, the MPII data sets altogether have the largest number of HD videos and annotation instances. Objects and fine actions are annotated in great detail, and 2D poses of upper body are also provided. For vision-based 2D object manipulation analysis, the amount of data and action variability of the MPII data sets can only be rivaled by the Brown breakfast data set,[35] if not unmatched.

50 Salad

The 50 Salad data set[29] extends Slice&Dice[27] by using accelerometers on more utensils and by including depth videos in addition to RGB videos. Twenty-five subjects participated in the study and each one of them prepared a mixed salad twice, following a specific sequence of tasks in each run. The sequences were produced by a statistical activity diagram, which would theoretically enable the same number of samples for each task sequence.

The annotation includes three high-level activities: prepare dressing, cut and mix ingredients, and serve salad. Each high-level activity is annotated with several low-level activities that include pre-phases, core-phases, and post-phases. The 50 Salad inherits the merit of Slice&Dice,[27] involves more subjects, enables 3D analysis with depth videos, and has finer annotations. In that regard, we recommend 50 Salad over Slice&Dice.

Actions for cooking eggs

The actions for cooking eggs data set[34] contains RGB-D videos of cooking activities for five egg menus, all of which were cooked by each of seven subjects. The labels contain only verbs: break, mix, bake, turn, cut, boil, season, and peel. We include this data set because it provides fine object manipulation motion, but because objects are not identified in any way, using the data set would rely on human and object tracking more heavily than other data sets.

YouCook

The YouCook data set[40] consists of 88 RGB cooking videos downloaded from YouTube. All the videos have a third person point of view. Although only 7 actions labels are used, as many as 48 object labels spanning 7 object categories exist, and object tracks are provided. We consider the richness of object labels and the availability of the objects tracks as the merits of the data set, of which the latter facilitates analysis of fine motion in 2D.

Actions for making cereal

In Ref.[30] the data of eight subjects are included while preparing cereal. The data set includes multiple modalities, including RGB-D videos, audios, estimated six DOFs object-pose trajectories, and object mesh models. We consider the object-pose trajectories as the merit of the data set. No other data sets that we include provide such modality, and using the trajectories alone suffices to conduct analysis on 3D object manipulation.

Brown breakfast

The Brown breakfast data set[35] contains roughly 78 hours of RGB videos involving 52 subjects captured at up to 18 distinct kitchens. In total, 10 recipes were prepared and each subject was reported to have prepared all 10 recipes, but available data for different subjects vary. Forty-eight coarse activity annotations exist and 11,267 annotation instances were identified. The statistics of the data set makes it a possible rival of the MPII data sets. It has the largest number of video frames (non-HD) among the data sets we include, more than the MPII data sets by 50%. The number of coarse annotation instances is not much lower than the MPII data sets, but the detail and richness of the annotation could not compete with MPII. The data set does include fine activity annotations, but the statistics and the description of the formation of such annotations are not yet available. Compared with MPII, the data set lacks 2D upper body-pose annotations.

Summary

Table 4 lays out the different modalities included in all the data sets in this category, and Figure 1 shows in descending order the count of data sets for each modality.

We can easily notice in Table 4 that Ref.[32] includes the highest number of acquisition modalities, most of which cannot be found in the other data sets. This is because the goal of Ref.[32] is to make the data set multimodal.

In Table 4, we can notice that RGB vision is used in all 13 data sets and is the sole acquisition modality of 5 data sets. The equipment required for recording RGB images is generally minimal and is easy to set up. Apart from evaluating certain vision-based algorithms, the recorded RGB video can also be used to verify that the data collection scene is properly set up, to spot any mistakes during the data collection process, and to segment the collected data. RGB images are matrices and carry much more information than other scalar modalities (such as acceleration) captured at a comparable frequency. The pose of the object or the human estimated from RGB images has a lower accuracy than if it is directly measured by a motion capture system, but it usually suffices for action recognition.

Only three data sets include depth images, only two data sets provide information on 3D acceleration on object, and only one data set provides sequences of estimated object poses. Despite the high accuracy it provides, a motion capture system is used only in Ref.,[32]

Table 4. Modalities cooking

Modalities	*References*												
	27	*32*	*45*	*21*	*21+*	*22*	*23*	*24*	*29*	*34*	*40*	*30*	*35*
RGB video (image sequences)	■	■	■	■	■	■	■	■	■	■	■	■	■
Depth image sequence									■	■		■	
Audio		■			■							■	
RFID		■											
Motion capture		■											
3D acceleration on human		■											
3D rotational velocity on human		■											
2D acceleration on human		■											
body heat		■											
3D magnetic field on human		■											
Skin temperature		■											
Skin conductance		■											
Gaze				■	■								
3D acceleration on object	■								■				
6D object-pose track												■	
Object models												■	
Air temperature		■											
Light		■											
Moving camera		■	■	■	■						■		
No. of subjects	20	>39	4	14	10	12	30	30	25	7	n/a	8	52

n/a: not applicable; RFID: Radio Frequency Identification.

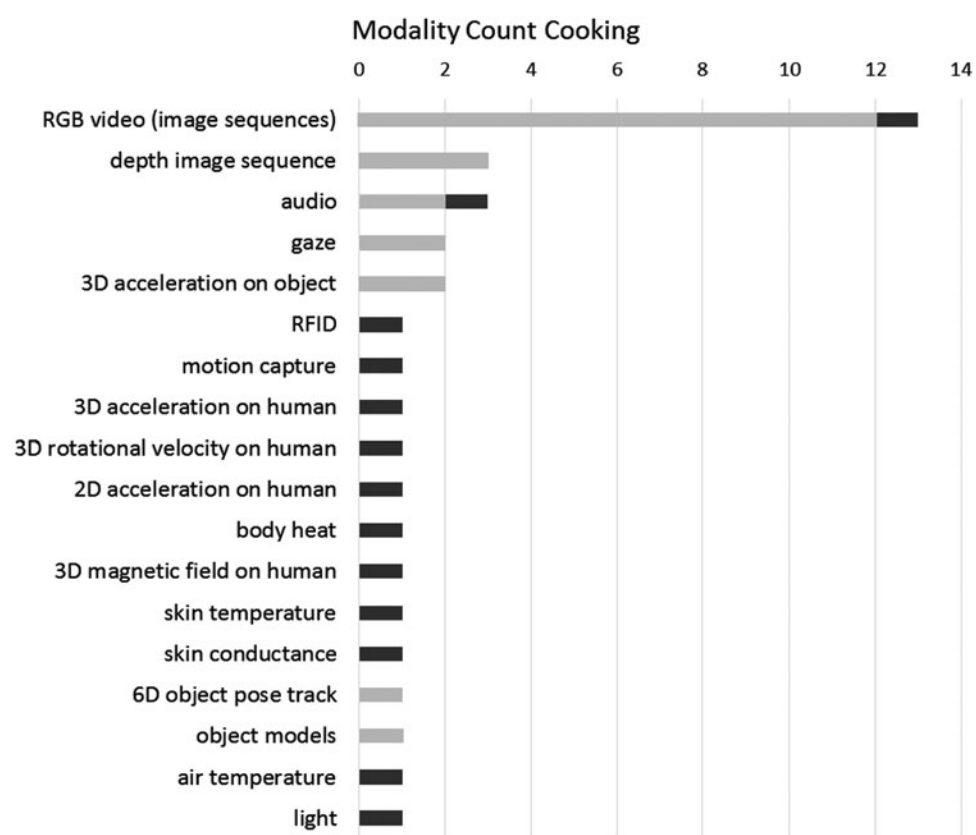

FIG. 1. Count of data sets for each modality. Black denotes the modalities of CMU-MMAC,[32] gray denotes the modalities of the other data sets.

possibly because of its cost, the lack of portability, and the effort required for the system setup. It is worth to mention that one of the most envisioned applications of these data sets in robotics is for LfD. What is commonly done is to use movement sequences of objects as input for training while testing is performed through physical manipulation of real objects by the robot (see Ref.[55]). The particular class of applications motivated us to consider and evaluate which characteristics are important to create a data set that is designed for LfD. Ideally, we would like to have readily available sequences of 3D object poses, which include positions and orientations as in Ref.[30] Object 3D acceleration data can be also converted to 3D position as in Refs.,[27,29] and RGB-D images can be used to estimate 3D object poses as in Ref.[34]

Another important aspect to take into account in data collection for LfD is the environment, either real or laboratory based. For example, among the data sets described in this article, Refs.[21+,35,40] were collected in real kitchens, and the other data sets were collected either on a tabletop or in a laboratory kitchen. An important difference between a real and a laboratory kitchen lies in the amount of clutter in the background: real kitchens generally have more clutter, which increases the difficulty of object recognition, which may lower the accuracy of object recognition. As object-pose estimations are fed into LfD as input, a possibly lower accuracy of object recognition is undesirable.

Activity annotations can be useful for various purposes. For example, if the annotations are short sentences describing a video, natural language processing can be combined with vision to provide higher accuracy on action/object recognition, or generate more annotations.[56] Annotations in the form of words can be used to represent motion activity classes. Paulius et al.[57] labeled 65 cooking videos including 798 labeled motion instances and 1229 labeled objects. The knowledge in the cooking videos is represented by a network of motions and objects, which is called functional object-oriented network (FOON). The FOON is continuously updated and maintained at this website.[†] We identified the annotated activities that are shared by multiple cooking data sets and list those data sets in Table 8. We combine similar annotations and specify each in the cells.

Data Sets of ADLs

In this section, we present 10 data sets of ADLs, 3 data sets of grasps acquired using camera, and 2 data sets of robot motion. The interest of studying ADLs is motivated by the extensive variety of the objects that human hands interacted with daily, and the variety of the environments where these interactions take place. Compared with Data Sets of Cooking Activity section, this section introduces additional modalities such as 3D kinematics of objects, force and torque on objects and on joints of robotic arms, and sequences of estimated human skeleton. Apart from action recognition, the application fields of the data sets include hand-pose recognition for human–machine interaction, grasp analysis, and deep learning, among others. Following the format in Data Sets of Cooking Activity section, we first review each data set individually, and then we discuss the use of motion capture and we provide more details on data set suitability for LfD.

TUM kitchen

The TUM kitchen data set[38] contains multimodal data of set-a-table activities. The modalities include RGB and raw Bayer pattern videos, motion capture, RFID, and reed sensor. Four subjects each transported certain objects from the cupboard, the counter, and the drawer to a table, and then laid them out in a specified way. The subjects transported the objects one by one as a robot would do, and also several objects at a time as naturally done by a human. The data set also includes repetitive activities of picking up and putting down objects. The annotations cover the entire duration of the set-a-table activity, which starts with *Reaching* through *ReleaseGraspOfSomething*. The actions of the left hand, the right hand, and the trunk were annotated, respectively.

Similarly to CMU-MMAC,[32] the data set identifies the objects involved during motion execution, and the availability of motion capture makes it a good candidate for 3D analysis on pick-and-place motion.

Rochester ADLs

The Rochester ADLs data set[44] contains RGB videos of five subjects performing certain ADLs and IADLs that can be summarized as using phone, writing, drinking and eating, and preparing food. Each video records one activity. Similar to the MPII data sets[22–24] and the Brown breakfast data set,[35] the Rochester ADLs data set would rely on human and object recognition to be useful for 2D fine motion analysis.

Opportunity

The opportunity data set[28] contains multimodal data of five morning ADLs runs and one drill run for each of four subjects. Motion sensors were densely deployed on

[†]www.foonets.com

the human body, on the objects, and in the environment. The modalities on the human body include IMUs, 3D accelerometers, and 3D localizers. The modalities on the objects include 3D accelerometers and 2D rotational velocity sensors. The annotations consist of five "tracks": locomotion, high-level activities, mid-level gestures, low-level actions, and objects for the left and the right hand, respectively.

The data set distinguishes itself from others that we include by using accelerometers and rotational velocity sensors on *both* the hand and the objects. As object manipulation analysis focuses on the interaction between hand and objects, data that include the motion of both the hand and the objects are desired. The data set is comparable with 50 Salad,[29] CMU-MMAC,[32] and TUM kitchen[38] in modality availability, although the last three data sets target cooking scenarios. For the objects, the data set includes 2D rotational velocity, which is unavailable in 50 Salad. For the human body, the data set lacks motion capture, which is available in CMU-MMAC and TUM kitchen, but alternatively provides 3D acceleration and 3D rotational velocity.

Cornell CAD-60 and CAD-120

The CAD-60[25] and the CAD-120 data sets[26] are both RGB-D video data sets. CAD-60 includes video sequences of 4 subjects performing 12 ADLs in 5 different indoor environments. Each sequence corresponds to one instance of a certain activity. The CAD-120 data set recorded 4 subjects each performing 10 high-level activities. Each subject performed every high-level activity multiple times with different objects. The annotations include 10 low-level activities and 12 object affordances.

CAD-60 and CAD-120 feature skeleton data, which include tracks of 3D position of all 15 joints and 3D orientation of 11 joints. The skeleton data in these data sets were generated using the NITE library that complements the PrimeSense sensors and were, therefore, estimated data. By comparison, the skeleton data collected using a motion capture system are actual physical measurements and, therefore, can be regarded as ground truth. Thus, the accuracy of the skeleton data in CAD-60 and 120 is lower than the accuracy of those collected with a motion capture system. Nevertheless, the skeleton data are directly usable for 3D fine motion analysis, a characteristic we consider as an advantage of these data sets.

First person ADLs

The first person ADLs data set by Pirsiavash and Ramanan[49] contains RGB videos captured using a GoPro camera. It recorded 20 subjects performing 18 ADLs. Forty-two objects were annotated by annotators with bounding boxes, tracks, and the status as to whether the object is being interacted with. Similar to Gaze(+),[21] with first person images, the working area of the hands is emphasized. However, as the data set includes a single modality, using it for analysis on 2D fine motion would rely on object tracking.

Wrist-worn accelerometer

The wrist-worn accelerometer data set[41] contains accelerometer data of 16 subjects performing a total of 14 ADLs. The accelerometers were attached to the right wrists of the subjects and the data were recorded at the subjects' home. The data set contains 979 trials. For fine motion analysis, wrist acceleration may be less ideal than hand acceleration, but it remains a readily usable modality.

UCI-EGO

The UCI-EGO or general-HANDS data set[52] includes four sequences of object manipulation activities. Each sequence includes 1000 RGB-D frames captured using an egocentric camera. Various objects were involved and manipulated, but because the data set focuses on hand detection and pose estimation, the manipulation tasks performed with each object are relatively short. As other vision-oriented data sets, the use of UCI-EGO data set for object manipulation analysis relies on object tracking.

Yale human grasping

The Yale human grasping data set[46] contains 27.7 hours of RGB wide-angle videos of profession-related manipulation motion. Two machinists and two housekeepers participated. The data set is intended for grasping analysis. The annotations were done on two levels. On the first level, the grasp type was annotated along with the corresponding task name and object name. The second level provided the properties of the object and the task. A total of 18,210 grasp instances have been annotated. The data set includes prolonged videos of manipulation motion of machining and housekeeping alone, two categories that are not to be found in other data sets that we include.

UT Grasp

The UT grasp data set[50] contains data of four subjects who were asked to grasp a set of objects in a controlled environment (placed on a desktop) after a brief demonstration of how to perform each type of grasps. A subset of

17 grasp types from Feix's taxonomy were selected that are commonly used in everyday activities.[58] The videos were recorded using a HD head-mounted camera (GoPro Hero2) at 30 fps while subjects performed each grasp type with various hand poses. Annotations were also provided. UT grasp differs from Refs.[46,47] in that it consists of data captured in a controlled environment (in front of a desk) in contrast with the Refs.[46,47] for which data were collected in different parts of a house.

GUN-71

The GUN-71 data set[47] contains roughly 12,000 RGB-D images of grasps, each annotated with 1 of the 71 grasp classes in Ref.[59] The images were captured using a chest-mounted camera. A total of 28 objects per grasp were recorded, resulting in 1988 different hand–object configurations. An important difference between GUN-71 in Ref.[47] and the data sets in Refs.[46,50] is that in Refs.[46,50], images were captured during daily activities and the annotations follow the distribution of everyday object manipulations, that is, common grasp classes are much more represented than rare grasps. In contrast, care was taken in GUN-71 to ensure a balanced distribution of grasps and variability of data. To that end, three to four different objects were used for each grasp class, five to six views of the manipulation scene were considered for each hand–object configuration, eight subjects participated in the experiments (four males and four females), and five different houses were involved.

Google Push and Grasping

To facilitate deep learning in robotics, Google Brain publicly shares two data sets of movements of robotic arms: Push[43] and Grasping.[37]

The Push data set contains about 59,000 sequences of multimodal data of robotic arms pushing objects. A bin that contained different objects was placed in front of a seven DOFs robotic arm, and the arm repeatedly pushed the objects in one out of two ways: either pushing randomly or starting randomly from somewhere on the border of the bin and sweeping toward the middle. A camera was mounted behind the arm facing the bin. The bin contained 10–20 objects at a time, and the objects were swapped out for new objects after roughly 4000 pushes. Ten robotic arms were used.

The data include RGB images, recorded gripper pose (x, y, z, yaw, pitch), commanded gripper pose, robot joint position, and external torques. The data set provides 2 test sets each including 1500 sequences. One test set contains two different subsets of objects from the training set, and the other test set includes two sets of objects absent from the training set.

The Grasping data set is collected using a similar setup to that of Push. The data set contains about 650,000 sequences of multimodal data of robotic arms grasping objects. The modalities include RGB-D images, recorded and commanded gripper pose (position in x, y, z and orientation in quaternions), joint positions, joint velocities, joint external torques, and joint commanded torques.

Using Push or Grasping that involves robots only, one aims at learning to finish a task rather than learning to finish a task like a human. The absence of the retargeting problem[60] is an inherent convenience if the learned motion is to be executed by the same robot.

Manipulation kinodynamics

The manipulation kinodynamics data set[51] includes 3.2 hours of kinematics and dynamics information of objects grasped and manipulated by humans using five fingers. More specifically, the data of the object include mass, inertia, linear and angular acceleration, angular velocity, and orientation. For each of the five fingers, the collected data include friction, force, contact point position, and the axes of a right-handed local coordinate frame $(\mathbf{x}, \mathbf{y}, \mathbf{z})$, where axes $\mathbf{x}$ and $\mathbf{y}$ define the contact surface and axis $\mathbf{z}$ points toward the object. The data set does not include images or videos. The objects are custom made and can vary in mass distribution, friction, and shape. The performed motions vary in speed, direction, and task (e.g., emulating pouring). In total, 193 different combinations were recorded.

Pham et al.[51] provide a full suite of kinematics and dynamics data. It was created for investigating the mapping relationship between the kinematics features (velocity, acceleration, etc.) of a manipulated object and the underlying manipulating force, which is something similar to a Newtonian physical law. Both the cause of manipulation (the force) and the corresponding result (the kinematics) were measured and both were *of the object*, and no extra processing or estimation is needed. Therefore, we consider the data set as invaluable for manipulation research, although including RGB-D images would have made the data set more approachable to the computer vision community.

RPAL tool manipulation

The data set of Ref.[48] features tool manipulation by human and is still in the process of being created. The data set contains multimodal sequential data of

Table 5. Modalities activity of daily living

Modalities	*References*														
	38	*44*	*28*	*25*	*26*	*49*	*41*	*46*	*43*	*37*	*48*	*47*	*50*	*51*	*52*
RGB video (image sequences)	■	■		■	■	■		■	■	■	■	■	■		■
Depth image sequence				■	■					■	■	■			■
RFID		■													
Motion capture	■										■				
Bayer pattern video	■														
3D acceleration on human			■				■								
3D rotational velocity on human			■												
3D orientation on human			■												
3D location on human			■												
3D magnetic field on human			■												
Finger flexure											■				
Skeleton				■	■										
Joint angles				■	■										
3D acceleration on object			■											■	
3D rotational velocity and acceleration on object														■	
2D rotational velocity on object			■												
3D acceleration on furniture			■												
Reed switch on furniture	■		■												
Force on fingertip														■	
In-tool force											■				
In-tool torque											■				
Gripper pose									■	■					
Commanded gripper pose									■	■					
Robot joint position									■	■					
Robot joint velocity										■					
Robot joint commanded torque										■					
Robot joint external torque									■	■					
Moving camera						■		■				■	■		■
No. of subjects	4	5	4	4	4	20	16	4	—	—	n/a	8	4	—	2

subjects using different tools. The tool consists of four components from front to back: a swappable tooltip, a six DOFs force-and-torque (FT) sensor, a universal handle, and a six DOFs position-and-orientation (PO, x, y, z, yaw, pitch roll) tracker. When possible, another PO tracker is mounted on the object that interacts with the tool. Modalities recorded besides FT and PO data are top view RGB videos and depth sequences of the scene, and finger flexure. Currently available data are hosted at this website.[‡] As FT and PO data are of the tool, they can be used directly for manipulation learning, without the need of feature extraction, which is necessary for images.

Summary

Similar to what we do for the cooking data sets, here we lay out the modalities in all the data sets in Table 5, and we show in Figure 2 the count of data sets for each modality in descending order.

We can see from the figure that RGB vision is the most commonly used modality and is provided in 12 data sets excluding only those in Refs.[28,41,51] In fact, Ref.[28] did collect RGB videos but did not publish them. Motion capture data are very accurate and can be found in Ref.[38] that uses a markerless system, and in Ref.[48] that uses both an optical marker-based system and an electromagnetic (EM) system.

When an object is being manipulated, its orientation may change significantly (e.g., when a spatula is used to flip a bread), challenging the reliability of an optical marker-based system. Moreover, objects vary in shape and can be small, which limits the maximum number of markers that can be used. Also, during the execution of a task, a manipulated object generally has certain contact with another object or certain material (such as water), which makes the contact surface unavailable for mounting markers and the available mounting surface even smaller.

The mentioned reasons drove Ref.[48] to switch from an optical marker-based system to an EM alternative. The EM motion capture system consists of at least one source that defines the world frame and acts as the origin and one tracker that senses its position and orientation with respect to the source. The source

[‡]http://rpal.cse.usf.edu/imd

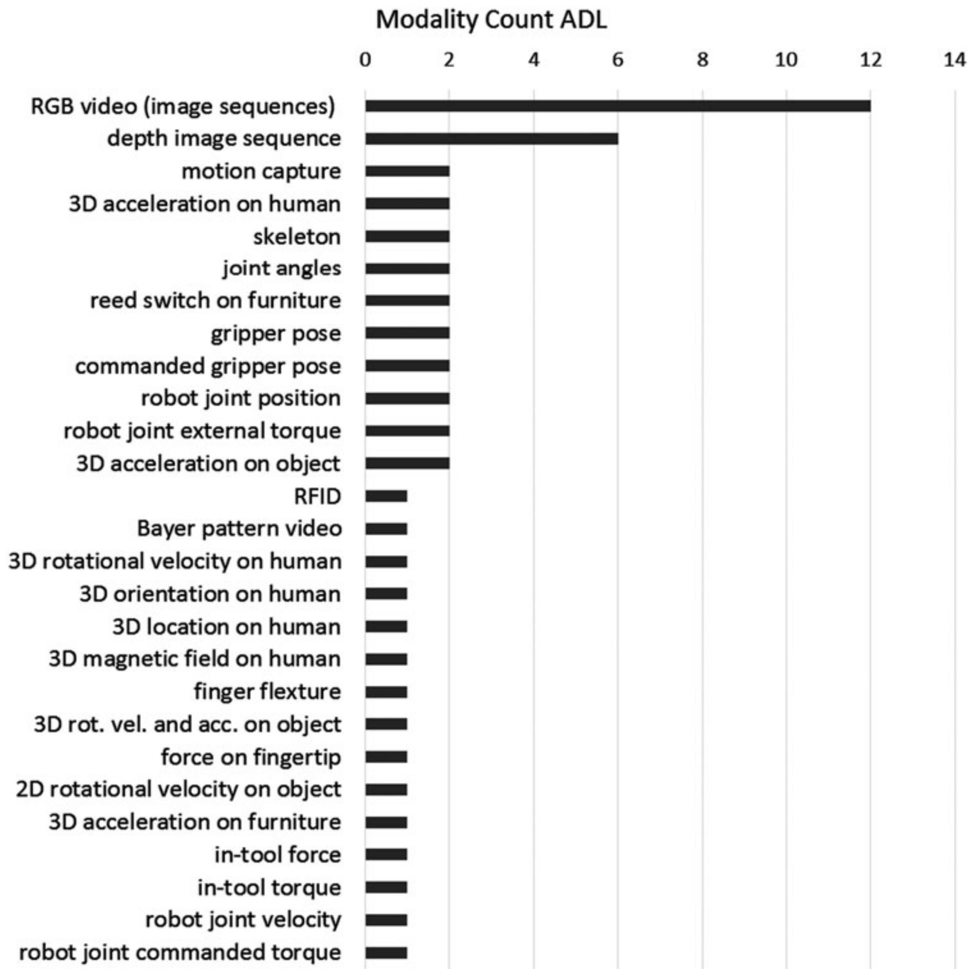

FIG. 2. Count of data sets for each modality.

and tracker are both connected to a processing station with cables. As the EM system uses cables, it does not require an unconstrained line-of-sight as in an optical marker-based system, and, therefore, a significant object-pose change causes occlusions and does not affect measurement accuracy. However, continuously rotating motion such as using a screwdriver has a possibility of finally putting stress on the cable and, therefore, requires extra attention.

In this section, we also include RGB vision-based grasp data sets[46,47,50] that are quite different from the wearable sensors-based data sets presented in Data Sets of Grasping section. Vision-based algorithms and methods can be used to perform both hand-pose estimation and classification using simple images of executed grasps or postures. However, the limited estimation accuracy and the lack of directly measured joint angles or fingertips positions limit the applicability of vision-based methods in the experimental analysis, design optimization, and control of actual robot artifacts (e.g., robot grippers and hands). The vision-based methods are also known to be prone to errors caused by changes in the environmental conditions (e.g., lighting conditions), and their efficiency can also be significantly affected by possible occlusions during the data collection process.

Unique to this section, Refs.[37,43,48,51] introduced the provision of force and torque. The data sets in Refs.[48,51] focus on the object while the data sets in Refs.[37,43] focus on the joints of the robotic system. Including force and torque enables modeling feedback, which makes the learning of object manipulation more physically realistic and helps with performing a learned task with a real object.

The data set described in Ref.[51] is intended for learning the relationship between kinematic features and

Table 6. Shared annotated activities of daily living

Activities	*References* 38	44	28	25	26	49	41
Use phone		Answer phone, dial on a phone		Talk on the phone		■	■
Write on whiteboard		■		■			
Drink		■	Sip	■	■	■	■
Eat		■			■		■
chop/cut		Chop	Cut	Chop			
Reach	■		■		■		
Release	Release grasp		■				
Comb hair						■	■
Brush teeth				■		■	■
Use computer				■		■	
Move					■	Dishes	
Stir			■	■			
Pour					■		■
Open	Door, drawer		■		■		
Close	Door, drawer		■		■		

We only consider low-level annotations for Ref.[28]

manipulating force during a manipulation motion *in general*, and not for a particular manipulation task. In simpler words, the data set focuses on *manipulation* rather than on *task*. As a consequence, the data set falls short of the requirement for LfD,[55] which focuses on manipulation *tasks*. As Pham et al.[51] used 3D printed objects, modifying the same[51] to make it suitable for LfD would require to change the current 3D object models to enable interaction with other objects while keeping the kinodynamics sensors from interfering with the manipulation tasks, task that may be nontrivial. In comparison, Ref.[48] focuses on recording data of tasks and is suitable for LfD, although it provides less fine-grained dynamics data than Ref.[51]

As for the cooking data sets, we identified the annotated activities that are shared by multiple ADLs data sets, and we list those data sets in Table 6. We combine similar annotations and specify each in the cells. For example, on the first row of Table 6, the annotated activity is summarized as “use phone,” whereas Ref.[44] specifically uses “answer phone” and “dial on a phone,” and Ref.[25] specifically uses “talk on the phone.”

Data Sets of Grasping

In this section, we introduce HandCorpus[§,53] and then review nine data sets about human hand kinematics recorded during grasping and manipulation tasks that are collected by the HandCorpus community and are stored in the HandCorpus repository. Unlike data sets,[46,47,50] which focus on video data, in this section we report on data sets about kinematic recording of human hand pose, in terms of sensor readings (marker positions and raw sensor data from a glove-based hand-pose reconstruction system) and joint angles of human hand during reach to grasp, manipulation, or grasps of real or imaginary objects.

The HandCorpus initiative

HandCorpus is an open access (no login or membership required) initiative/repository for sharing data sets, tools, and experimental results about human and robotic hands. The repository provides an accurate and coherent record for citing data sets, giving due credit to authors. Data sets are hierarchically indexed and can be easily retrieved using keywords and advanced search operations.

The motivation for HandCorpus is to provide the multidisciplinary “hand community” with a tool for benchmarking, to reuse results, and to foster collaborations between scientists in the fields of neuroscience and robotics. Of course, we study hands to understand the language of the human embodiment but also to try to reproduce this incredible language under a technological point of view. Under these considerations, the importance to have a common scientific framework is crucial.

The HandCorpus was originally created in 2011, within the European Project “The Hand Embodied (THE),” with the objective of making data collections and analyses about human hand publicly available. Since then, the HandCorpus website and structure have been ameliorated; presently (July 2016), the HandCorpus repository contains nine data sets about human hand kinematics recorded in grasping and manipulation tasks, with real or imaginary objects and two descriptions of kinematic models of human hands.

Regarding the latter point, it is possible to find (1) the description of a kinematic human hand model[61] devised from magnetic resonance imaging of a female hand, which provides axis locations in the form of transformation matrices, endowed with a visualization tool for the hand skeleton developed in Opensim,** (a freely available, user extensible software system to develop models of musculoskeletal structures and create dynamic simulations of movement); (2) the 3D coordinates of a static pose of a human hand using a motion

§www.handcorpus.org

**http://opensimulator.org

capture system and the kinematic model described in Ref.[42] The coordinates are expressed in C3D format[††] that represents a 3D biomechanics data standard.

In addition, HandCorpus contains five data sets about robotic hands. There are three entries on the description of device architecture and specs, as well as links to schematics and files to *in house* build the robotic devices. The OpenBionics underactuated, compliant, and modular robot hands[62] and the OpenBionics light-weight, affordable, anthropomorphic prosthetic hand[63] have been developed by the OpenBionics initiative and all the files required for their replication are available at the OpenBionics website.[‡‡] OpenBionics is an open-source initiative for the development of affordable, light-weight, modular robotic, and bionic devices founded in 2013.

The Pisa/IIT SH[19] is an affordable anthropomorphic robot hand, which embeds within its design the concept of kinematic synergies[13] to move according to the first human principal grasping pattern in free motion, and adaptability, which enables the hand to deform with the external environment to grasp a large variety of objects. The hand has 19 joints, but only uses 1 actuator, and it is very soft and safe, yet powerful and extremely robust.

The schematics of the SH can be retrieved from a direct link to Natural Machine Motion Initiative (NMMI).[§§] The NMMI is a modular open platform aiming to provide the scientific community with tools for fast and easy prototyping of soft robots, such as variable stiffness actuators, soft grippers, a pool of application-specific add-ons, an open mechanical standardized interconnection system, and common open electronics and software infrastructure to enable system integration.

The Pisa/IIT SoftHand has also served as the starting point for the development of affordable and easy-to-use prostheses, whose realization is actually pursued within the EU-H2020-funded grant SoftPro—the latter is also one of the sponsors of HandCorpus together with other 6 European grants (see Acknowledgments section), and the 22 international research groups across Europe, Asia, and the United States of America forming the HandCorpus community.

HandCorpus also contains kinematic recordings from two robotic hands: (1) the fingertip positions of RBO Hand 2[64] while enacting the Feix grasps[65] with the list of 3D coordinates of each of the five fingers and (2) the joint angles of a robot hand (schunk DLR HIT 4 fingers) while being teleoperated by a human hand wearing a Cyberglove.[66] The intention was to capture the workspace of the robotic hand, while avoiding possible collisions so as for the workspace to be modeled with the concept of principal motion directions, thus providing a reduced space for motion planning.

Table 7. HandCorpus statistics

	References								
Attributes	*13a*	*31*	*33*	*36*	*13b*	*39*	*42a*	*42b*	*42c*
Human postures	■	■	■	■	■	■	■	■	■
Joint angles	■	■	■		■	■	■		■
Marker coordinates								■	
Joint sensor raw data				■					
Static grasps	■		■	■	■	■			
Reach and grasp	■	■	■	■			■		
Free space							■	■	
Haptic exploration									■
Active marker motion capture system						■	■	■	■
Passive marker motion capture system	■								
Cyberglove		■	■	■	■				
Objects type	R	R	R	R	I	I	I	N	R
No. of DOFs	20	16	20	23	15	15	24	24	26
No. of subjects	7	30	1	5	1	1	1	1	1
Year	2012	2016	2010	2015	2011	2011	2013	2014	2015

DOFs, degrees of freedom; I, imaginary; N, no object; R, real.

Finally, HandCorpus contains tools for the analysis, visualization, and study of human and robot hands, including psychophysical investigation, tactile sensing, and biomechanical modeling. As previously mentioned, HandCorpus is also a hub to other open-access initiatives about robotic and human hands. A blog, a newsletter, a publication repository, and HandCorpus profiles in all major social networks are also provided. For further information, please visit the HandCorpus website, which is cross-platform, cross-browser, and easily accessible through mobile devices, such as Internet-enabled smartphones and tablets.

In Table 7, we report an overview of the data sets about human hand kinematics included in the HandCorpus, with a description of the different labels used to characterize data. Focusing on human hand kinematics, the following nine data sets refer to grasping and ADLs, such as haptic exploration.

DLR data set

The DLR data set (May 2012)[13] contains the kinematics of the human hand—joint angles (captured with a passive marker-based motion capture system; Vicon), while executing the grasps reported in Ref.[13] The results are different from those of Ref.,[13] because the objects grasped are real and the contact forces between

[††]www.c3d.org
[‡‡]www.openbionics.org
[§§]www.naturalmachinemotioninitiative.com

Table 8. Shared annotated cooking activities

Activity	*References*										
	27	*32*	*21*	*21+*	*24*	*29*	*34*	*40*	*30*	*35*	*45*
Chop/cut	Chop, slice, dice			Cut	Chop, cut, cut apart, cut dice, cut off ends, cut off inside, cut stripes, slice	Cut	Cut			Cut	
Peel/shave	Peel, shave			Peel	Peel	Peel	Peel			Peel	
Stir/mix	Stir	Stir		Mix	Mix, stir	Mix	Mix	Stir		Stir	Stir
Pour		■	■	■	■			■	Milk, cereal	■	■
Put/place		Put		Put	Put in, put on	Place		Put down		Put	Put
Take		■	■	■	Take lid, take out					■	■
Spread/smear	Spread		Spread	Spread	Spread					Smear	Spread
Eat/taste	Eat				Taste						
Scoop/spoon	Scoop		Scoop							Spoon	Scoop
Season/spice					Spice		Season	Season			
Turn/flip				Flip	Turn over		Turn	Flip			
Open/close food (container)		Open	■	■	■				■		■
Open/close drawer		Open			■						
Open/close dishwasher/oven				Oven	■						
Open/close cupboard/ fridge/microwave		■		Fridge	■						
Crack/break		Egg		■	Open egg		■			■	
Beat/whip		Beat egg			Whip						
Add					■	■				Teabag, salt and pepper, topping	
Squeeze				■	■					■	
Turn on/off				■	■						
Wash				■	■						
Dry				■	■						
Fill				■	■						

As Ref.[24] supercedes Refs.,[22,23] we only include Ref.[24] in the table.

the fingers and the object surface induce certain deformations to the hand postures. Data of 7 subjects were included and 23 different objects were grasped.

HUST data set

The HUST data set (March 2016)[31] reports the joint angles of the human hand while executing the grasping tasks of the Feix taxonomy.[65] During the experiments, the subjects (5) were seated and they had their right arm fixed on the table surface in a comfortable posture. The subjects were instructed to perform 33 types of tasks of the Feix taxonomy,[65] using a large number of objects. The human hand motion was captured with a data glove (Cyberglove system).

NTUA data set

The NTUA data set (May 2010)[33] investigates the role of hand synergies during reach to grasp. A subject was seated on a chair, while his trunk was restrained to the chair and his hand was placed on the table with the palm facing downwards. Objects of varying shape and size were placed at a higher point than the starting hand position. The user was instructed to move his arm to reach and grasp the object. For each trial, the starting hand position and the object position were kept the same. The human hand kinematics was described in terms of joint angles and captured with a Cyberglove system.

TU Berlin data set

The TU Berlin data set 1 (June 2015)[36] contains sensor raw data of 5 participants enacting the grasps of the Feix taxonomy (33 grasps).[65] During the resting periods, the subjects were asked to place their hand on a table surface. The human motion was captured with a Cyberglove.

UNIPI-ASU data set

The UNIPI-ASU data set (May 2011)[13] reports the joint angles of the human hand while grasping 57 imaginary objects according to the procedure proposed by Santello et al.[13] Human hand motion was once again captured with a Cyberglove.

UNIPI data sets

The UNIPI data set (October 2011)[39] contains the joints angles of 57 grasps of imaginary objects,[13] captured with an optical motion capture system (Phase Space). A single subject (male, 26) participated in the experiments.

The UNIPI data set 2 (September 2013)[42] contains the joint angles of the human hand of a female right handed subject described according to the model reported in Ref.[42] while grasping several imaginary objects, and recorded through a marker-based motion capture system (Phase Space). The subject was comfortably seated with the flat hand on the leg and was asked to move the hand so as to grasp an imaginary object for tool use, and then return to the rest position.

The UNIPI data set 3 (June 2014)[42] contains marker coordinates of a human hand of a single subject acquired through the Phase Space system while executing the Kapandji movement.[67] The kinematic model under investigation was described in Ref.[42]

The UNIPI data set 4 (September 2015)[42] contains data of a single subject who was blindfolded and was asked to haptically identify some common objects. Before each trial, the subject placed the dominant hand on a table and one of the objects was placed in random order about 30 cm in front of the hand. On a go signal, the subject reached out, explored, and identified the object. In addition, the subject was also asked to explore the surface curvature, edges, and textures of each object to prolong the exploration time. The human hand kinematics was captured with a Phase Space system and data represent joint angles.

Certain data sets share the same reference, and we distinguish those data sets using suffixes in Table 1, 2, 3, 7, and 9. Specifically, 13a refers to the DLR data set, 13b refers to the UNIPI-ASU data set, 42a refers to the UNIPI data set 2, 42b refers to the UNIPI data set 3, and 42c refers to the UNIPI data set 4.

Summary

In Table 7 we can observe that human hand kinematics can be acquired through different acquisition systems (active marker-based motion capture system, passive marker motion capture system [Vicon], and data glove) and using different descriptors (raw sensor data, joint angles, and marker 3D coordinates).

The most common acquisition modalities are (1) active marker motion capture system and (2) data glove. The main reason for this relies on the high accuracy (the amount of static marker jitter is inferior than 0.5 mm, usually 0.1 mm) and the ease in handling marker IDs for (1), whereas wearability and the fact that there is no need for implementing filtering techniques to reconstruct joint angles from marker measurement represent the main motivations for using (2). Regarding data descriptors, joint angles represent the most common type of data. This is intuitive, as this information is crucial for grasp planning,[54] to drive the design and control of robotic hands[12,19] and to improve the performance of hand-pose reconstruction systems, see Refs.[68,69]

Data format also varies (mainly .txt, .dat, and .mat, but also .C3D and .csv), although .mat and .txt are the most used. Main motivations for this are simplicity (.txt) and the fact that joint angles are usually obtained after a postprocessing phase, which is commonly performed in Matlab or Mathematica (.mat).

Finally, regarding the type of actions, reach to grasp and grasp of real objects are the most represented within HandCorpus, although grasping of imaginary objects, haptic exploration, and free hand motion can be also found.

In conclusion, a standardized procedure for data acquisition, formatting and organization is still lacking, and data reuse should be facilitated (e.g., using formats like the .C3D biomechanics standard). However, there is a clear trend in favor of the employment of motion capture and glove-based systems, joint angles as type of data, and .mat and .txt for data format. In this regard, it would be important to increase the information available on the kinematic model in use, with schemes and visual representations provided together with data sets, thus enabling a correct and simple reusage and interpretation of data. This is already partly done within HandCorpus, thanks to the usage of accompanying *Read me* file for the data sets, but it could be further improved through the adoption of common and unique data descriptions.

Discussion

Table 9 lays out different modalities included in the data sets of all three categories. We can clearly see that the different focuses require different modalities. A total of 25 out of 28 data sets in cooking/ADLs tasks have RGB videos, but none of the grasping data sets has any video except for the UNIPI data set 3, in which a video of the rendered skeleton of the hand while performing grasping actions is also provided and is freely available in the dedicated YouTube channel of the HandCorpus initiative.*** Such a visualization is important as it practically demonstrates the actions under investigation and increases the data comprehension.

***https://youtu.be/wTZdAFGjHpI

Table 9. Modalities

Modalities	Cooking													ADLs															HandCorpus								
	27	32	45	21	21+	22	23	24	29	34	40	30	35	38	44	28	25	26	49	41	46	43	37	48	47	50	51	52	13a	31	33	36	13b	39	42a	42b	42c
RGB video (image sequences)	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■		■	■	■		■	■	■	■	■	■		■									
Depth image sequence									■	■		■					■	■					■	■	■			■									
Audio		■			■							■																									
RFID		■												■																							
Motion capture		■												■										■								■	■	■	■	■	■
Glove-based system		■												■										■					■	■	■						
Bayer pattern video														■																							
3D acceleration on human		■														■				■																	
3D rotational velocity on human		■														■																					
3D orientation on human																■																				■	
3D location on human																■																				■	
2D acceleration on human		■																																			
Skeleton																	■	■																			
Joint angles																	■	■											■	■	■	■	■	■	■		■
body heat		■																																			
3D magnetic field on human		■														■																					
Skin temperature		■																																			
Skin conductance		■																																			
gaze				■	■																																
3D acceleration on object	■								■							■											■										
3D rotational velocity and acceleration on object																											■										
2D rotational velocity on object																■																					
6D object-pose track												■												■													
object models												■																									
3D acceleration on furniture																■																					
Reed switch on furniture														■		■																					
Force on fingertip																											■										
In-tool force and torque																								■													
Gripper pose																						■	■														
Commanded gripper pose																						■	■														
Robot joint position																						■	■														
Robot joint velocity																							■														
Robot joint commanded torque																							■														
Robot joint external torque																						■	■														
Air temperature		■																																			
Light		■																																			
Moving camera		■	■	■	■						■								■		■				■	■		■									
Activity type	C	C	C	C	C	C	C	C	C	C	C	C	C	A	A	A	A	A	A	A	A	P	G	A	G	G	M	M	G	G	G	G	G	G	G	M	H
No. of subjects	20	>39	4	14	10	12	30	30	25	7	n/a	8	52	4	5	4	4	4	20	16	4	—	—	n/a	8	4	—	2	7	30	1	5	1	1	1	1	1
Year	2009	2009	2011	2012	2012	2012	2012	2015	2013	2013	2013	2014	2014	2009	2009	2010	2011	2013	2012	2014	2014	2016	2016	2016	2015	2015	2016	2014	2012	2016	2010	2015	2011	2011	2013	2014	2015

A, ADL; C, cooking; G, grasping; H, haptic exploration; M, manipulation; P, pushing.

All nine data sets in the grasping category contain hand tracking data in contrast to the cooking and ADLs categories that contain only one or two data sets with hand tracking data combined. Seven out of 9 data sets in grasping have also joint/skeleton angles, comparing with 2 out of 15 data sets in ADLs tasks, and 0 out of 13 data sets in cooking tasks.

Research in object manipulation might find 3D object poses very useful. Explicit or readily usable recordings of object poses are available in Ref.[48] Poses of the robot end effector are provided in Refs.[37,43] Pieropan et al.[30] provide *estimated* object-pose trajectories. Object poses may be computed using acceleration and rotational velocity, and object motions that are simpler than poses can be obtained if a sensor actively takes samples and is attached to an object. Data sets with such a setup include:

(1) Refs.[27,29] in which objects were equipped with accelerometers.
(2) Ref.[28] in which objects were equipped with accelerometers and rotational velocity sensors. Furniture and appliances were equipped with reed switches and accelerometers.
(3) Ref.[38] in which doors were equipped with reed switches.

The shared activities demonstrate a consensus among different authors on what activities should be performed and annotated. For example, certain grasping taxonomies are often adopted and such directions can be helpful for those who try to create a new data set. However, not being a commonly shared activity does not necessarily mean an activity is not important. Therefore, we also provide the complete list of annotated activities at this website,[†††] for cooking and ADLs, respectively. The shared activities can also help with using more than one data set. If we want to study a certain shared activity, we could use several data sets that include this activity to access more modalities and higher variability. Objects that are involved in an activity may also be helpful for activity analysis. For all data sets except those in Ref.,[34] objects are identifiable in the annotated activities through

(1) being separately annotated,[23,24,26,40,45,46,49]
(2) being part of the annotation phrases,[21,21+,22,25,27,29,30,32,35,38,41,44,45] and
(3) being equipped with sensors
    (a) accelerometers,[27,28,29]
    (b) rotational velocity sensors,[28]
    (c) reed switches,[28,38] and
    (d) RFID.[32,38]

Temporal segmentation of annotated activities is also important for activity analysis. For Refs.,[46,47,50] temporal segmentation does not apply because they focus on grasp instances. All other data sets include temporal segmentation in the following forms:

(1) video subtitle,[27,30]
(2) explicit video time,[21+,49]
(3) frame number,[21–24,26,32,34,35,38,40,45]
(4) timestamp,[28,29] and
(5) implicit.[25,41,44]

As previously mentioned in the Introduction for the case of the whole-body motion data sets, we are aware of the existence of other related data sets; however, to keep this work focused, we do not include them. Examples of the excluded data sets are[‡‡‡]

(1) Refs.,[71,74] which are data sets that do not include object manipulation motions, or if they do, the object manipulation motions are sparse.
(2) Refs.,[3,75,76] which are data set of objects that are typically involved in manipulation, rather than data sets of motion.

Most data sets are intended for action recognition. However, researchers who work on LfD[55] intend to reproduce human actions rather than recognizing them. Thus, we suggest, in addition to choosing from the modalities we have reviewed, that a more ideal data set for LfD should also aim to provide readily usable data that are more closely related to dynamic and kinematic motion execution. Examples of suggested modalities include trajectories of object poses, joint poses of human upper body, hand posture, torque, and force between hand and object.

Finally, an important specification for creating useful data sets that can be used in robotics applications is to facilitate benchmarking. One interesting example is provided in Ref.,[3] in which the objects used for manipulation were chosen to cover different aspects of the

[†††]http://rpal.cse.usf.edu/motiondatasetreview/index.htm

[‡‡‡]HandCorpus represents an interesting example of data sets whose role can bridge the gap between neuroscience and robotics. However, in the literature, it is also possible to find data sets specifically built to address purely neuroscientific questions, which are currently out of the scope of this review. It is the case of the WAY-EEG-GAL,[70] which was designed to allow critical tests of techniques to decode sensation, intention, and action from scalp EEG recordings in humans who perform a grasp-and-lift task. Twelve participants performed lifting series (a total of 3936 trials) in which the object's weight, surface friction, or both, were changed unpredictably between trials. EEG, EMG, the 3D position of both the hand and object (through the Pholemus passive-marker magnetic motion capture system) as well as force/torque at both contact plates were recorded.

manipulation problem and object characteristics; RGB-D object scans, physical properties, and geometric models are also provided together with protocol examples and physical object delivery.

Conclusions

We reviewed 28 data sets on object manipulation and 9 data sets on grasping. We reported the characteristics and modalities of each data set individually, we gave our view on the relationship between each data set and object manipulation, and we compared and summarized all of them together.

The data sets were created to serve their own purposes and many of them are unique. Therefore, different modalities were used. The modalities range from popular video recording to rarely used air temperature and light. Many data sets were collected with numerous subjects, whereas some were collected with only one subject. Several data sets provide motion annotations. Twenty-three different cooking-related motions and 15 different ADL motions are annotated in the examined data sets. The survey provides a "map" for researchers in choosing the right existing data set(s) for their own research purposes. If the right data sets are not found, the researchers may decide on creating new data sets that will supplement the exiting data sets.

Observing the diversity of the data sets, we understand that trying to get a unique standard for the different types of data sets is clearly a daunting and challenging task. However, moving toward a common standardization that defines common data formats for common working areas as well as acquisition protocols would enable efficient data reusage and sharing, fostering collaborations, and creating large data sets that allow big data-driven approaches such as deep learning. It has been discussed recently in many conferences and workshops of the robotics community as one of several important initiatives. HandCorpus as one example is exploring a central depository approach, by collecting and processing existing data sets to provide consistent data formats and guarantee data quality.

This survey does not include data sets that, although are introduced in publications, are not openly available. Many of them were presented in the Workshop on Grasping and Manipulation Datasets that was organized under the International Conference on Robotics and Automation in May 2016. The workshop's report[77] provides a survey of those works and data sets.

Acknowledgments

This work was partly supported by the National Science Foundation under Grant Nos. 1421418 and 1560761, by the European Research Council under the Advanced Grant SoftHands (No. ERC-291166), by the EU H2020 projects SoftPro (No. 688857) and SOMA (No. 645599), and by the EU FP7 project WEARHAP (No. 601165), PACMAN (No. 600918), WALKMAN (No. 611832), and THE (No. 248587).

Author Disclosure Statement

No competing financial interests exist.

References

1. Mandery C, Terlemez Ö, Do M, et al. The kit whole-body human motion database. In: 2015 International Conference on Advanced Robotics (ICAR), IEEE, 2015. pp. 329–336.
2. Sun Y, Lin Y, Huang Y. Robotic grasping for instrument manipulations. In: International Conference on Ubiquitous Robots and Ambient Intelligence, 2016. pp. 1–3.
3. Calli B, Singh A, Walsman A, et al. The ycb object and model set: Towards common benchmarks for manipulation research. In: Proceedings of the 2015 IEEE International Conference on Advanced Robotics (ICAR), 2015.
4. Kopicki M, Detry R, Schmidt F, et al. Learning dexterous grasps that generalise to novel objects by combining hand and contact models. In: 2014 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2014. pp. 5358–5365.
5. Bekiroglu Y, Kragic D, Kyrki V. Learning grasp stability based on tactile data and hmms. In: 19th International Symposium in Robot and Human Interactive Communication, IEEE, 2010. pp. 132–137.
6. Bohg J, Morales A, Asfour T, Kragic D. Data-driven grasp synthesis—A survey. IEEE Trans Robot. 2014;30:289–309.
7. Mahler J, Pokorny FT, Hou B, et al. Dex-net 1.0: A cloud-based network of 3d objects for robust grasp planning using a multi-armed bandit model with correlated rewards.
8. Bianchi M, Moscatelli A. Human and robot hands.
9. Prattichizzo D, Trinkle JC. Grasping. In: Springer Handbook of Robotics, Springer, 2008. pp. 671–700.
10. Lin Y, Sun Y. Task-based grasp quality measures for grasp synthesis. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2015. pp. 485–490.
11. Lin Y, Sun Y. Grasp planning to maximize task coverage. Int J Robot Res. 2015;34:1195–1210.
12. Santello M, Bianchi M, Gabiccini M, et al. Hand synergies: Integration of robotics and neuroscience for understanding the control of biological and artificial hands, Phys Life Rev. 2016.
13. Santello M, Flanders M, Soechting JF. Postural hand synergies for tool use. J Neurosc. 1998;18:10105–110115.
14. Soechting JF, Flanders M. Flexibility and repeatability of finger movements during typing: Analysis of multiple degrees of freedom, J Comput Neurosci. 1997;4:29–46.
15. Ficuciello F, Palli G, Melchiorri C, Siciliano B. Experimental evaluation of postural synergies during reach to grasp with the ub hand iv. In: 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, 2011. pp. 1775–1780.
16. Ciocarlie M, Goldfeder C, Allen P. Dimensionality reduction for hand-independent dexterous robotic grasping. In: 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, 2007. pp. 3270–3275.
17. Ciocarlie MT, Allen PK. Hand posture subspaces for dexterous robotic grasping. Int J Robot Res. 2009;28:851–867.
18. Brown CY, Asada HH. Inter-finger coordination and postural synergies in robot hands via mechanical implementation of principal components analysis. In: 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, 2007. pp. 2877–2882.

19. Catalano MG, Grioli G, Farnioli E, et al. Adaptive synergies for the design and control of the pisa/iit softhand. Int J Robot Res. 2014;33:768–782.
20. Ciotti S, Battaglia E, Carbonaro N, et al. A synergy-based optimally designed sensing glove for functional grasp recognition. Sensors. 2016;16:811.
21. Fathi A, Li Y, Rehg JM. Learning to recognize daily actions using gaze. In: Proceedings of the 12th European Conference on Computer Vision—Volume Part I, ser. ECCV'12, 2012. pp.314–327.
22. Rohrbach M, Amin S, Andriluka M, Schiele B. A database for fine grained activity detection of cooking activities. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012.
23. Rohrbach M, Regneri M, Andriluka M, et al. Script data for attribute-based recognition of composite activities. In: European Conference on Computer Vision (ECCV), October 2012.
24. Rohrbach M, Rohrbach A, Regneri M, et al. Recognizing fine-grained and composite activities using hand-centric features and script data. 2015.
25. Sung J, Ponce C, Selman B, Saxena A. Human activity detection from rgbd images. In: AAAI Workshop on Pattern, Activity and Intent Recognition (PAIR), 2011.
26. Koppula HS, Gupta R, Saxena A. Learning human activities and object affordances from rgb-d videos. Int J Robot Res. 2013;32:951–970.
27. Pham C, Olivier P. Slice&Dice: Recognizing food preparation activities using embedded accelerometers. Springer. 2009.
28. Roggen D, Calatroni A, Rossi M, et al. Collecting complex activity datasets in highly rich networked sensor environments. In: 2010 Seventh International Conference on Networked Sensing Systems (INSS), June 2010. pp. 233–240.
29. Stein S, McKenna SJ. Combining embedded accelerometers with computer vision for recognizing food preparation activities. In: Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing, 2013. pp. 729–738.
30. Pieropan A, Salvi G, Pauwels K, Kjellstrom H. Audio-visual classification and detection of human manipulation actions. In: 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2014), September 2014. pp. 3045–3052.
31. Liu MJ, Xiong CH, Xiong L, Huang XL. Biomechanical characteristics of hand coordination in grasping activities of daily living. PLoS One. 2016;11:e0146193.
32. de la Torre F, Hodgins J, Bargteil A, et al. Guide to the carnegie mellon university multimodal activity (cmu-mmac) database. Robotics Institute, Carnegie Mellon University, Technical Report. CMU-RI-TR-08-22, July 2009.
33. Liarokapis MV, Artemiadis PK, Kyriakopoulos KJ. Telemanipulation with the dlr/hit ii robot hand using a dataglove and a low cost force feedback device. In: 2013 21st Mediterranean Conference on Control Automation (MED), June 2013, pp. 431–436.
34. Shimada A, Kondo K, Deguchi D, et al. Kitchen scene context based gesture recognition: A contest in ICPR2012. In: Advances in Depth Image Analysis and Applications, Lecture Notes in Computer Science, Vol. 7854, 2013. pp. 168–185.
35. Kuehne H, Arslan A, Serre T. The language of actions: Recovering the syntax and semantics of goal-directed human activities. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.
36. Deimel R, Brock O. A novel type of compliant and underactuated robotic hand for dexterous grasping. Int J Robot Res. 2015. Available online at http://ijr.sagepub.com/content/early/2015/08/13/0278364915592961.abstract
37. Levine S, Pastor P, Krizhevsky A, Quillen D. Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. CoRR. 2016;abs/1603.02199. Available online at http://arxiv.org/abs/1603.02199
38. Tenorth M, Bandouch J, Beetz M. The tum kitchen data set of everyday manipulation activities for motion tracking and action recognition. In: 2009 IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops), September 2009. pp. 1089–1096.
39. Bianchi M, Salaris P, Bicchi A. Synergy-based hand pose sensing: Reconstruction enhancement. Int J Robot Res. 2013. Available online at http://ijr.sagepub.com/content/early/2013/02/19/0278364912474078.abstract
40. Das P, Xu C, Doell R, Corso J. A thousand frames in just a few words: Lingual description of videos through latent topics and sparse object stitching. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013.
41. Bruno B, Mastrogiovanni F, Sgorbissa A. A public domain dataset for adl recognition using wrist-placed accelerometers. In: 2014 RO-MAN: The 23rd IEEE International Symposium on Robot and Human Interactive Communication, 2014. pp. 738–743.
42. Gabiccini M, Stillfried G, Marino H, Bianchi M. A data-driven kinematic model of the human hand with soft-tissue artifact compensation mechanism for grasp synergy analysis. In: 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, November 2013. pp. 3738–3745.
43. Finn C, Goodfellow IJ, Levine S. Unsupervised learning for physical interaction through video prediction. CoRR. 2016;abs/1605.07157. Available online at http://arxiv.org/abs/1605.07157
44. Messing R, Pal C, Kautz H. Activity recognition using the velocity histories of tracked keypoints. In: ICCV, 2009.
45. Fathi A, Ren X, Rehg JM. Learning to recognize objects in egocentric activities. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2011. pp. 3281–3288.
46. Bullock IM, Feix T, Dollar AM. The Yale human grasping data set: Grasp, object, and task data in household and machine shop environments. Int J Robot Res. 2014;34:251–255.
47. Rogez G, Supancic JS, Ramanan D. Understanding everyday hands in action from rgb-d images. In: Proceedings of the IEEE International Conference on Computer Vision, 2015. pp. 3889–3897.
48. Available online at http://rpal.cse.usf.edu/imd
49. Pirsiavash H, Ramanan D. Detecting activities of daily living in first-person camera views. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2012. pp. 2847–2854.
50. Cai M, Kitani KM, Sato Y. A scalable approach for discovering the visual structures of hand grasps. May 2015.
51. Pham TH, Kyriazis N, Argyros AA, Kheddar A. Hand-Object Contact Force Estimation From Markerless Visual Tracking. Submitted to IEEE Transactions on Pattern Analysis and Machine Intelligence. August 2016. Available online at https://hal.archives-ouvertes.fr/hal-01356138
52. Rogez G, Khademi M, Supančič III JS, et al. 3d hand pose detection in egocentric RGB-D images. CoRR. 2014;abs/1412.0065. Available online at http://arxiv.org/abs/1412.0065
53. Bianchi M, Liarokapis MV. HandCorpus, a new open-access repository for sharing experimental data and results on human and artificial hands. In: IEEE World Haptics Conference (WHC), April 2013.
54. Marino H, Ferrati M, Settimi A, et al. On the problem of moving objects with autonomous robots: A unifying high-level planning approach. IEEE Robot Autom Lett. 2016;1:469–476.
55. Billard A, Calinon S, Dillmann R, Schaal S. Robot programming by demonstration. In: Springer Handbook of Robotics, Berlin Heidelberg: Springer, 2008.
56. Donahue J, Hendricks LA, Guadarrama S, et al. Long-term recurrent convolutional networks for visual recognition and description. CoRR. 2014;abs/1411.4389. Available online at http://arxiv.org/abs/1411.4389
57. Paulius D, Milton R, Huang Y, et al. Functional object-oriented network for manipulation learning. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2016. pp. 1–8.
58. Bullock IM, Feix T, Dollar AM. Finding small, versatile sets of human grasps to span common objects. In: 2013 IEEE International Conference on Robotics and Automation (ICRA), May 2013. pp. 1068–1075.
59. Liu J, Feng F, Nakamura YC, Pollard NS. A taxonomy of everyday grasps in action. In: 2014 IEEE-RAS International Conference on Humanoid Robots, IEEE, 2014. pp. 573–580.
60. Gleicher M. Retargetting motion to new characters. In: Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques, ser. SIGGRAPH'98, New York, NY: ACM, 1998. pp. 33–42. Available online at http://doi.acm.org/10.1145/280814.280820
61. Stillfried G, Hillenbrand U, Settles M, van der Smagt P. Mri-based skeletal hand movement model. In: The Human Hand as an Inspiration for Robot Hand Development, Springer International Publishing, 2014. pp. 49–75.
62. Zisimatos AG, Liarokapis MV, Mavrogiannis CI, Kyriakopoulos KJ. Open-source, affordable, modular, light-weight, underactuated robot hands. In: 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2014. pp. 3207–3212.

63. Kontoudis GP, Liarokapis MV, Zisimatos AG, et al. Open-source, anthropomorphic, underactuated robot hands with a selectively lockable differential mechanism: Towards affordable prostheses. In: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), September 2015. pp. 5857–5862.
64. Deimel R, Brock O. A novel type of compliant and underactuated robotic hand for dexterous grasping. Int J Robot Res. 2015. [Epub ahead of print]; DOI: 10.1177/0278364915592961.
65. Feix T, Pawlik T, Schmiedmayer HB, et al. A comprehensive grasp taxonomy. In: Robotics, Science and Systems: Workshop on Understanding the Human Hand for Advancing Robotic Manipulation, 2009.
66. Rosell J, Suárez R, Rosales C, Pérez A. Autonomous motion planning of a hand-arm robotic system based on captured human-like hand postures. Auton Robots. 2011;31:87–102.
67. Kapandji A. [Clinical test of apposition and counter-apposition of the thumb]. Ann Chir Main. 1986;5:67–73.
68. Bianchi M, Salaris P, Bicchi A. Synergy-based hand pose sensing: Reconstruction enhancement. Int J Robot Res. 2013;32:396–406.
69. Bianchi M, Salaris P, Bicchi A. Synergy-based hand pose sensing: Optimal glove design. Int J Robot Res. 2013;32:407–424.
70. Luciw MD, Jarocka E, Edin BB. Multi-channel eeg recordings during 3,936 grasp and lift trials with varying weight and friction. Sci Data. 2014;1:140047.
71. Chaquet JM, Carmona EJ, Fernández-Caballero A. A survey of video datasets for human action and activity recognition. Comput Vis Image Underst. 2013;117:633–659.
72. Soomro K, Zamir AR, Shah M. UCF101: A dataset of 101 human action classes from videos in the wild. CRCV-TR-12-01, Technical Report, 2012.
73. Huynh T, Fritz M, Schiele B. Discovery of activity patterns using topic models. In: Proceedings of the 10th International Conference on Ubiquitous Computing, ser. UbiComp'08, 2008. pp. 10–19.
74. Ohn-Bar E, Trivedi M. The power is in your hands: 3d analysis of hand gestures in naturalistic video. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2013. pp. 912–917.
75. Lai K, Bo L, Ren X, Fox D. A large-scale hierarchical multi-view rgb-d object dataset. In: 2011 IEEE International Conference on Robotics and Automation (ICRA), May 2011. pp. 1817–1824.
76. Singh A, Sha J, Narayan K, et al. Bigbird: A large-scale 3d database of object instances. In: 2014 IEEE International Conference on Robotics and Automation (ICRA), May 2014. pp. 509–516.
77. Bianchi M, Bohg J, Sun Y. Latest datasets and technologies presented in the workshop on grasping and manipulation datasets. arXiv. 2016.

Cite this article as: Huang Y, Bianchi M, Liarokapis M, Sun Y (2016) Recent data sets on object manipulation: a survey. *Big Data* 4:4, 197–216, DOI: 10.1089/big.2016.0042.

Abbreviations Used

ADLs = activities of daily living
DOFs = degrees of freedom
EM = electromagnetic
FOON = functional object-oriented network
FT = force-and-torque
HD = high definition
IADLs = instrumental activities of daily living
IMU = inertial measurement unit
LfD = learning from demonstration
NMMI = Natural Machine Motion Initiative
PO = position-and-orientation
THE = The Hand Embodied