The Complexities of Grasping in the Wild

Yuzuko C. Nakamura[1], Daniel M. Troniak[1], Alberto Rodriguez[2], Matthew T. Mason[1], and Nancy S. Pollard[1]
[1] The Robotics Institute, Carnegie Mellon University
[2] Department of Mechanical Engineering, Massachusetts Institute of Technology
<yuzi, nsp, matt.mason>@cs.cmu.edu, troniak@gmail.com, albertor@mit.edu

***Abstract*— The recent ubiquity of high-framerate (120 fps and higher) handheld cameras creates the opportunity to study human grasping at a greater level of detail than normal speed cameras allow. We first collected 91 slow-motion interactions with objects in a convenience store setting. We then annotated the actions through the lenses of various existing manipulation taxonomies. We found manipulation, particularly the process of forming a grasp, is complicated and proceeds quickly. Our dataset shows that there are many ways that people deal with clutter in order to form a strong grasp of an object. It also reveals several errors and how people recover from them. Though annotating motions in detail is time-consuming, the annotation systems we used nevertheless leave out important aspects of understanding manipulation actions, such as how the environment is functioning as a "finger" of sorts, how different parts of the hand can be involved in different grasping tasks, and high-level intent.**

I. Introduction

For roboticists working on dexterous robots, observation of human manipulation continues to be an important way to understand the problem of grasping (e.g. [1, 2, 3]). One way of observing human grasping in detail is to use high-framerate video. Due to the growing ubiquity of high-framerate video cameras in phones, it is now feasible to capture a large number of grasping actions "in the wild" i.e. in everyday settings such as cluttered workspaces. The large number of actions and the everyday setting allows behaviors such as mistakes to be captured, and the high framerate reveals detailed finger movement and the making and breaking of contact.

This paper reports on one human subject taking items from store shelves, counters, and bins, and replacing them. Video was recorded using a single handheld camera at 120 frames per second. The researchers then analyzed that video using several classification systems, as well as ad hoc analyses that attempt to note high-level events in the recording not captured in the other taxonomies.

Unfortunately, RGB video is not amenable to automated analysis. Humans have to watch the video and record their observations. It is our hope that eventually this process can be partially automated using video analytics and behavior recognition, and that our annotations can function as ground truth data for future automated analytical tools.

The primary long range goal of this work is to develop an annotation system capable of describing manipulation behavior performed by one actor in such a way that the manipulation can be copied by another actor. Such an annotation system would need to be detailed and expressive enough to note all elements critical for duplicating the motion, but also flexible/abstract enough to be applicable across different robot hardware and hand morphologies.

This paper's contributions are the following: (1) Slow-motion video of interactions with a wide variety of objects; (2) analysis of this dataset and summary of major findings; and (3) application of those findings toward the development of an annotation system able to capture important elements of grasping.

The greatest surprise was the variety and complexity of behaviors we saw, even though the task domain is mostly picking and placing. Insights gained from this study include:

- The process of grasping in the presence of clutter can be complex, sometimes involving adjustment of a grasp or exploiting the environment, yet occurs quickly.
- Contact-guided placing is common.
- Collisions between effector and clutter or between object and clutter are commonplace. Error recovery is quick when it is necessary at all.
- Expected patterns of behavior based on grasp taxonomies and other prior work were observed but less frequently than we expected.

II. Related Work

To classify observations from video, we first look to existing taxonomies. Grasp pose taxonomies based on shape and function have existed for many years (e.g. [4, 5, 6, 7, 8]) and organize grasps based on aspects of power vs. precision, the shape of the object, and the shape of the hand. Researchers have also collected grasping data to refine these taxonomies (e.g. [9, 10]). Beyond static grasping, a number of taxonomies based on manipulation have also been developed (e.g. [11, 12, 13, 14, 15]).

Outside taxonomies centered around the hand and manipulation, there exist taxonomies for whole-body pose [16] and whole-body and facial movement [17, 18, 19]. Observations of great apes are also of interest [20]. For example, Byrne et al. [21] observe over 200 primitive actions, such as pick-out, pull-apart, and rotate-adjust, as necessary to describe feeding behaviors of mountain gorillas.

For our annotations, we choose the Elliott and Connolly taxonomy [11] for its description of the intrinsic movements of the hands and fingers, the Feix taxonomy [22] for its

This work was supported by National Science Foundation [NSF-IIS-1637853].

description of gravity-independent grasps, and the Bullock et al. taxonomy [12] for its description of changes in contact, motion, and prehension. We note that other researchers have annotated static grasps from video in the domains of cleaning and machine shop work [23] and contact and motion for some everyday tasks [12]. We contribute encodings for shelf picking and placing actions, choose a mix of taxonomies, and contribute observations of how these taxonomies succeed and fail to capture what we see.

One goal of our research is to understand manipulation primitives that may be useful in robotics. Manipulation primitives have been extensively explored. A tiny sample includes examples ranging from pushing [24], toppling [25], pivoting [26], opening doors [27], and moving objects out of the way [28] to making pancakes [29], making cookies [30], and folding towels [31]. An interesting outcome of our study is that we observe a number of evidently useful primitives that appear to have been less studied in robotics, such as levering an object up.

III. Dataset

The dataset analyzed consists of a collection of RGB videos of a single subject manipulating objects in a convenience store. The videos were captured by one of the researchers using the iSight camera on an iPhone 5S (120 frames per second, 1280x720 resolution).

Continuous video capture of the entire visit was infeasible due to limitations in disk space and battery; thus videos were captured discontinuously and subsequently trimmed and pieced together to form a single video. In total, 91 interactions between the subject and 60 convenience store objects were observed and analyzed. These interactions collectively took place over a period of 3 minutes and 9 seconds of discontinuous video.

The subject was given instruction on which items to manipulate as she moved about the store. On occasion, the subject was encouraged to increase the variety of manipulation actions when possible, such as to twirl a turnstile or regrasp an apple. When finished, the subject attempted to replace the items back in their original locations. The subject has identified herself as being right-handed.

Objects manipulated by the subject include beverage bottles, cans, cups and Tetra Paks; salad dressing, tea, salt and cream packets; dry condiment shakers; a refrigerator door; various packaged foods, such as ice cream, potato chips and candy bars; plastic knives, forks, and spoons; napkins; a plastic sign; a plastic bag; a turnstile; an apple; a pizza box; a wrapped hoagie; plastic salad boxes; a plastic sauce cup with lid; and steel tongs.

A compressed version of the dataset and annotations are available online at `http://www.cs.cmu.edu/~dtroniak/nsh_shop_120.webm` and `http://www.cs.cmu.edu/~ynakamur/projects/complexities/annotations.zip`.

IV. Methodology

The captured video was viewed and analyzed with the aim of noting any significant events or processes that would be helpful for instructing a robotic actor to be able to replicate the manipulation. The researchers manually labeled the dataset using several existing taxonomies, as well as through other lenses where a taxonomy does not exist:

- Static grasp pose taxonomy created by Feix et al. [22]. This taxonomy collects poses from previous taxonomies and separates hand shapes based on function (power, precision, or intermediate), thumb position (abducted or adducted) and which surfaces of the hand are used to secure the object (palm, finger pads, or sides of fingers).
- Intrinsic (within-hand) hand motion categories observed by Elliott and Connolly [11], which describe motions a hand uses to manipulate an object already in the hand.
- Bullock et al.'s manipulation taxonomy [12], which creates broad categories of manipulation based on the presence or absence of contact (C), prehension (P), motion (M), intrinsic hand motion (W), and motion at contact points (A). The taxonomy is high-level and doesn't assume any particular hand morphology.
- The lens of errors and recovery from errors.
- The lens of contacts and when they are important to execution of a motion, either aiding or constraining manipulation.

Annotating the video through these lenses often involved noting intention as well – why that choice of grasp; what is the purpose and end effect of a particular intrinsic hand motion; what was the hand attempting to do when the error occurred?

In a first pass, the entire video was annotated through each of the above lenses. We then focused on a small number of actions that contained examples of interesting recurring phenomena (e.g. levering up, regrasps, errors/error recovery), and ranged from the simplest actions (milk bottle place #1) to the most complicated actions (cutlery pick #2) observed. We cleaned up the annotations for these motions to make them consistent and to use more fine-resolution frame numbers instead of seconds, and then plotted their annotations in the form of a timeline. The eight motions selected were:

- Zone bar pick (0:11-0:22 – Fig. 3)
- Zone bar place (0:22-0:29 – Fig. 3)
- Mountain Dew pick (1:01-1:08 – Fig. 3)
- Milk bottle place #1 (3:41-3:46 – Fig. 3)
- Pepsi cup pick (4:09-4:15 – Fig. 3)
- Cutlery pick #2 (5:56-6:34 – Fig. 4)
- Lay's chips pick (8:25-8:36 – Fig. 5)
- Pizza box pick (9:35-9:46 – Fig. 5)

In the timelines, we used color to distinguish between annotations that fell within a taxonomy (grey blocks) and new ones not found in that taxonomy (green blocks). For annotations using the Bullock, Ma, and Dollar (BMD) taxonomy, the "new" annotations correspond to moments when multiple actions are being performed by different parts of the hand – for example, two fingers holding an object in a stable grasp while the rest form a grasp of a second object. While these moments could be annotated as a single BMD category (usually C [P,NP] M W A), we decided to annotate

the actions of the different units of the hand separately to be more descriptive of what is happening. The downside is that this way of annotation is more complicated.

Due to the general and comprehensive nature of the BMD taxonomy, an annotation was possible at every point in time during grasping except when the hand is off-screen or occluded. Gaps in the BMD timeline correspond to these situations.

When analyses had no annotations associated with them, their empty timelines were excluded from the figure. For example, there were no miscellaneous annotations in the Mountain Dew pick action and no intrinsic manipulation annotations in the milk bottle place action (see Fig. 3).

V. Results

Annotations for the selected clips are shown in Figs. 3-5. The accompanying video shows these motions. This section outlines insights obtained from these and other annotations.

A. The process of forming a grasp is complex.

The high framerate video reveals detailed grasping strategies that are hard to see in normal 30 fps video. The examples shown in the video indicate that the process of forming a grasp is as complex and worthy of notice as the final achieved grasp pose itself. While it is simple to pinch small items between two or more fingers and instantly form a grasp that way, many of the grasps observed featured some kind of hand pose adjustment between the time of making contact and forming the final grasp. Fig. 1a is an example of how and why adjustments occur between contact and final grasp: first, ulnar fingers use the rim of the box to lift one side, exposing the bottom surface (frame 1). Then a complicated sequential pattern of finger lifting and recontacting (frames 2-5) results in the final grasp (last frame). This final grasp involving the bottom surface of the box is much more secure, but not possible until the bottom surface has been lifted up enough for fingers to be placed underneath.

In general, we find that the process of forming a grasp has multiple phases:

1) Approach and preshaping: changing the pose of the arm or hand in anticipation of grasping
2) Contact: compliantly making contact with some part of the object
3) Dealing with clutter: maneuvering fingers into spaces, singulating an object, or pushing its surfaces away from nearby surfaces.
4) Taking weight: bracing or adjusting pose to take full weight of object
5) Lift: able to move object with full arm now that stable grasp has been formed
6) Grasp adjustment: to more comfortable grasp

For small, light, or unobstructed objects, some of these phases may not be necessary. Sometimes singulating the object and pulling it further into the hand to form a grasp happen simultaneously (see Fig. 1b).

Ungrasping involves similar phases but in reverse (for example, touching down and letting go of weight instead of taking weight and lifting). Similar to grasping, many ungrasping motions are not just opening the hand to break contact; they instead involve some kind of in-hand motion or grasp change before contact is broken. Approximately 25 of the 53 grasping examples (47%) feature post-contact grasp adjustments before a final grasp, and 16 of the 48 placing examples (33%) feature pre-release grasp adjustment (see Fig. 1c).[1]

B. Environment-aided grasping

Before prehension is achieved, the human hand is nevertheless able to manipulate an object (for example, lifting a corner or edge up, tilting an object out, singulating an object by pressing down, etc.). The way it does this is by using the environment as a "finger" of sorts, which provides an opposing surface that a hand can use to "grasp" an object securely enough to manipulate it. Being able to exploit these environmental contacts appears to be important for grasping objects when a normal pinch grasp is not feasible.

We also found that gaps in the environment are also exploited in order to aid grasping. Fingers can be inserted into gaps and extended in order to create more space, as in the case of the soymilk pick (1:29). The pizza box pick (Fig. 5) is an example of both exploiting a gap to contact the side of the box and then using that contact to form an environment-aided grasp.

C. Insights from the grasp taxonomy analysis

Fig. 2 shows new and in-between grasps found in the video. (a) The placement of the index finger is flexible and can be abducted away from other fingers, resulting in variations on existing grasps. (b) There exists a family of lateral grasps involving the side of fingers other than the index finger, possibly in conjunction with the index finger to strengthen the grasp. (c) Storage grasps involving the ulnar fingers or the crease between the thumb and index finger are specialized grasps that allow manipulation or a second grasp to be performed by unused fingers. (d) Deformable objects like potato chip bags resulted in unusual grasps that use a mix of side and pad opposition. (e) Some in-between grasps were found like an apple grasp in between the precision sphere and precision disk grasps, and a milk bottle grasp similar to a tripod grasp but stronger and more stable.

We also observed objects initially grasped with a weak/precision grasp being regrasped into a power grasp. Figs. 1a and 4 are examples of this.

Although we focused on stable grasp poses e.g. times when there is no motion occurring within the hand, the cutlery pick action (Fig. 4) was an exception. During this action, small motions within the hand (such as lifting the middle finger) can instantly change one grasp into another (e.g. from prismatic 2 finger to inferior pincer).

[1] A single action in the video could contain multiple grasping and placing examples, so the total number of grasps and places is greater than the number of actions captured.

Fig. 1: Examples of (a) regrasping into a stronger grasp, (b) simultaneous levering out and grasp formation, (c) contact-guided placing, and (d) error correction (pinky is withdrawn from bin while approaching).

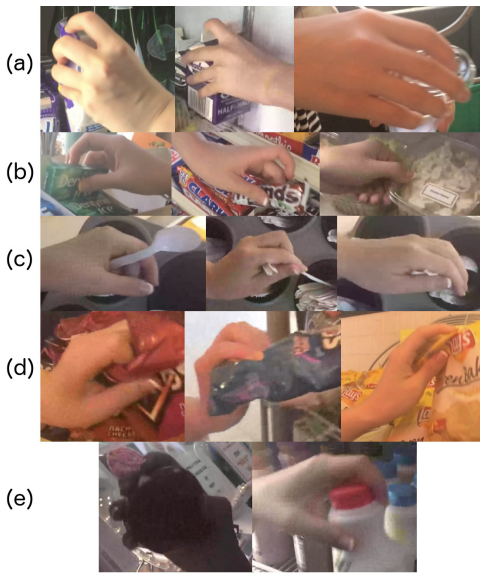

Fig. 2: New and in-between grasps observed in the video. (a) Variations with index finger extended or placed in a different area than other fingers. (b) Variations on lateral pinch grasp using middle and ring fingers. (c) Storage grasps allow manipulation and multiple grasps. (d) Deformable objects create a variety of opposition types (side and pad). (e) In-between grasps.

D. Insights from intrinsic manipulation analysis

We observed intrinsic movements like squeezing a bottle into the palm; interdigital steps to reorient stick-like objects in the hand; and rocking objects back or forth to help remove them from clutter. In particular, we noticed from the cutlery pick action (Fig. 4) that the interdigital step is a broad, high-level category that contains various smaller motions that people use to reorient objects in the hand.

Because the Elliott and Connolly taxonomy is designed for motions to manipulate objects already grasped, we did not analyze the intrinsic, environment-aided hand motions used before prehension to manipulate the object. An extension of this intrinsic manipulation taxonomy for non-prehensile manipulation could be useful.

E. Insights from Bullock, Ma, and Dollar (BMD) analysis

Throughout the picking/placing process, the hand is very rarely still, with either the whole arm, individual fingers, or both moving for the entire time in most examples. This analysis reveals that the human hand is very efficient when grasping, parallelizing work. For example, Fig. 1d shows approach to an object (whole-arm motion) occurring at the same time as error correction (within-hand motion to pull the pinky finger out of the way).

One limitation of the BMD taxonomy is that there is no way to annotate the common scenario when motion is occurring both outside the hand and within the hand simultaneously (i.e. a motion-within-hand (W) plus a motion-not-within-hand (NW) annotation), or when some contacts are changing while others are static (motion-at-contact (A) plus motion-not-at-contact (NA) annotation). In other words, within-hand and at-contact motion "mask" external motion or still contacts. The ability to indicate both are occurring simultaneously complicates the process of annotating motion, but may be important for the goal of being able to instruct robots to copy human grasping actions.

As the Pepsi cup pick (Fig. 3) indicates, full arm motion with a stable grasp pose can denote very different kinds of forces and motions. It can denote a smooth motion (pulling an object out of a space), or the shaking used during part of this motion. It is not able to distinguish between these two types of motion, which makes sense as the BMD taxonomy was designed to be augmented with other manipulation taxonomies. In particular, a taxonomy to describe motion direction and force type [32] may be a good choice to use here.

F. Errors and error recovery

The subject was instructed not to take any particular care when grasping. As a result, errors are observed from time to time, appearing in 13 of the 91 captured actions. Errors were corrected very quickly and the intended motion eventually succeeded with only one exception (tea packet push (4:53-5:04)). Fig. 1d shows an example of a quickly-corrected error, where a finger slips into a bin and is lifted without interrupting the grasping motion. Other errors we noticed included an edge of the object hitting other objects, pinches missing/failing to secure an object, and actions failing to insert an object into the intended location.

G. Insights from contact analysis

One analysis of the video focused on contacts and noted whenever contact was important to the motion. These motions fell into two categories: (1) contact was established purposefully in order to aid the motion (contact guidance), and (2) haptic feedback rather than visual feedback was possibly driving the action.

In the first case, contact was helpful for completing a motion. In 13 of 48 placing actions, an initial contact between a corner of the object was first established, and then the constraints created by that contact were used to guide the object into place. Fig. 1c is an example of such a movement where a contact is established.

The second case contained most examples of error correction as well as motions that were incidentally contact-heavy. For example, the Pepsi cup grab (Fig. 3) involved contacts that needed to be broken; this task was accomplished by shaking the cup.

VI. Discussion

Our video analysis reveals that the grasping process is surprisingly complex but fast. It takes advantage of environmental contacts and touch feedback. An initial non-prehensile "grasp" is often used to manipulate the object to make a final power grasp possible.

To create an annotation system sufficient to describe and prescribe manipulation, grasp poses, intrinsic manipulation, and generic manipulation taxonomies are useful. Grasp poses often reflect the end goal of the action e.g. a stable grasp of an object that is suitable for transporting and placing. However, pose taxonomies need to be extended to describe the flexible aspects of the grasp (for example, to instruct a robot that the index finger can be separated from other fingers and be used to tip the object out) and to include storage grasps. The intrinsic manipulation taxonomy is useful for describing manipulation of already-grasped objects and could be extended to include non-prehensile manipulation. The generic manipulation taxonomy lives up to its goals of being general enough to describe all manipulation without being tied to any one hand morphology, and is useful for segmenting motions into phases. However, its main limitation is the difficulty in describing multitasking/concurrent manipulation, which occurs regularly in human grasping.

In addition to those elements, a prescriptive annotation system needs to also describe the role of the environment in grasping and be able to convey intent. Force types and object motion may help fill this intention gap. Different force types can achieve different objectives, like smooth movement for transport vs. shaking to break contacts/suction force. Object motion hints at purpose as well. For example, some non-prehensile manipulation is done for the purpose of lifting a corner of the object up (changing the object's configuration). How this is achieved (e.g. by motion within the hand) is as important as knowing the purpose (e.g. to change the object configuration to expose the object's bottom side).

Our work has implications for robotic grasping. For example, compliant contact and contact guidance found in human grasping suggests that ongoing work on compliant control is important. In addition, dexterous in-hand manipulation seems important even for simple pick-and-place tasks. If human hands are any indication, work on the more difficult task of dexterous manipulation, including non-prehensile manipulation, will aid work on the forming of stable cage grasps.

The dataset has several limitations. First, the motions are primarily picking and placing motions that are performed by a single subject who is aware of being recorded. Because only one subject was recorded, some aspects of grasping may be idiosyncratic to her. Second, the objects are usually grasped without any intention of being used or placed in a different location.

VII. Conclusion

In this work we captured a dataset of slow-motion actions in a convenience store setting. We analyzed this video through the lenses of different manipulation taxonomies – a grasp pose taxonomy, an intrinsic manipulation action taxonomy, and a generic manipulation taxonomy – as well as through lenses focused on errors and contacts. We found that the process of grasping is complex and deserves more focus, particularly in situations with clutter or environmental constraints. Grasping is not only complex but also *quick* – with multiple goals being worked toward at the same time, such as one motion both singulating an object and drawing it into the hand – and *heavily reliant on touch* for corrections. The process of annotating elements of manipulation is time-consuming and at times reliant on high-level understanding of the video such as being able to infer the intention of a motion or series of motions. As such, there are many challenges to using human motion examples to inform robotic grasping. However, awareness of the complexity and strategy involved in grasping may help us design more robust and effective grasping processes.

References

[1] M. Khansari, E. Klingbeil, and O. Khatib, "Adaptive human-inspired compliant contact primitives to perform surface–surface contact under uncertainty," *The International Journal of Robotics Research*, vol. 35, no. 13, pp. 1651–1675, 2016.

[2] T. Feix, I. M. Bullock, and A. M. Dollar, "Analysis of human grasping behavior: Object characteristics and grasp type," *IEEE transactions on haptics*, vol. 7, no. 3, pp. 311–323, 2014.

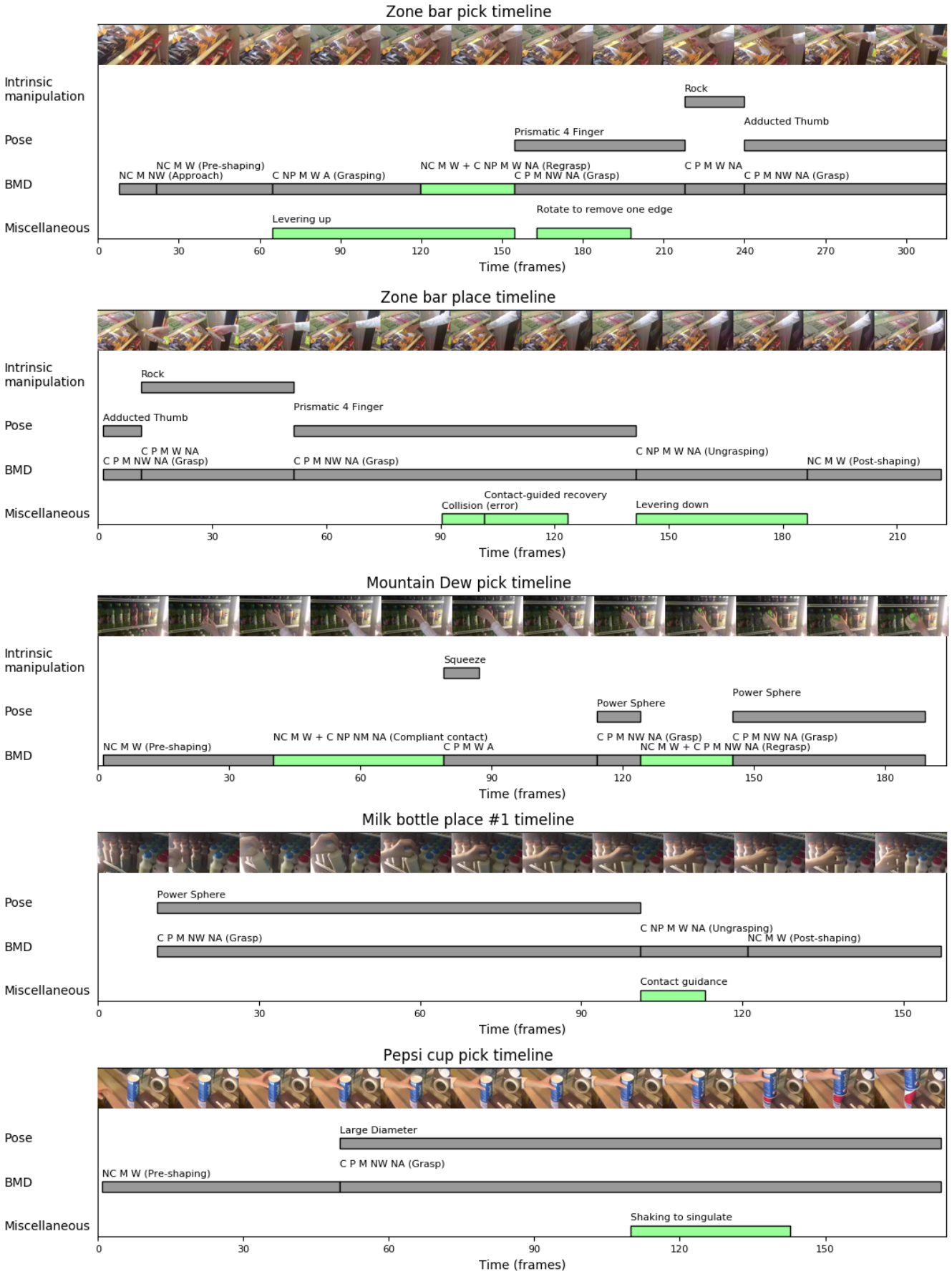

Fig. 3: Timelines for Zone bar pick, Zone bar place, Mountain Dew pick, milk place #1, and Pepsi cup pick.

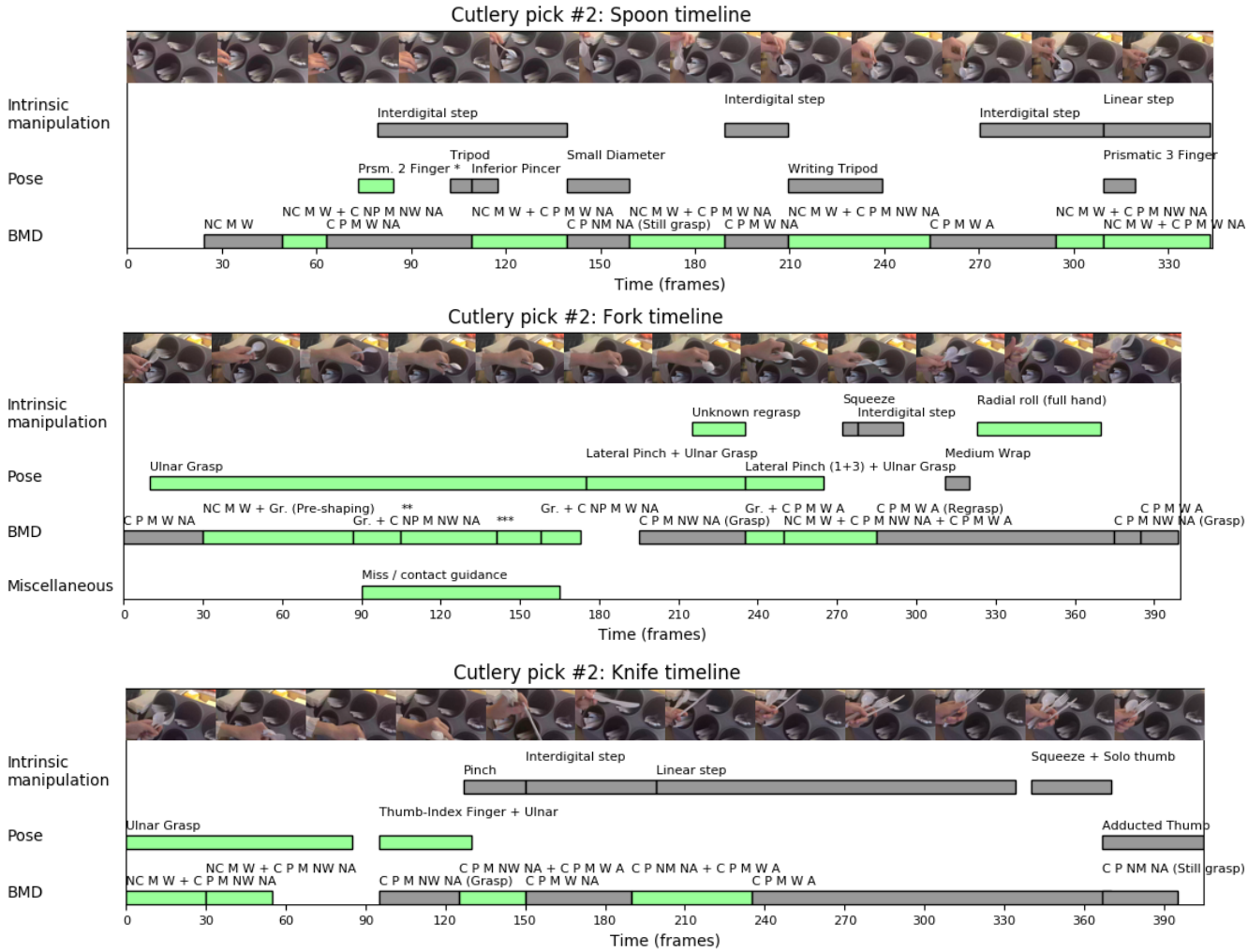

Fig. 4: Timelines for cutlery pick #2. *This grasp is capable of sliding the spoon out, but part of the spoon is supported by the environment, so this grasp is an environment-aided prehensile grasp. **NC M W + C P NM NA + C NP M W A (different fingers holding, preshaping, and manipulating). ***NC M W + C P M NW NA + C NP M NW A (same as previous but with full-arm motion)

[3] L. Y. Chang, S. S. Srinivasa, and N. S. Pollard, "Planning pre-grasp manipulation for transport tasks," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*. IEEE, 2010, pp. 2697–2704.

[4] I. G. Schlesinger, "Der mechanische aufbau der künstlichen glieder," in *Ersatzglieder und Arbeitshilfen*. Springer, 1919, pp. 321–661.

[5] J. R. Napier, "The prehensile movements of the human hand," *Journal of Bone and Joint Surgery*, vol. 38, no. 4, pp. 902–913, 1956.

[6] N. Kamakura, M. Matsuo, H. Ishii, F. Mitsuboshi, and Y. Miura, "Patterns of static prehension in normal hands," *The American Journal of Occupational Therapy: Official Publication of the American Occupational Therapy Association*, vol. 34, no. 7, pp. 437–445, 1980.

[7] T. Iberall, "The nature of human prehension: Three dextrous hands in one," in *Robotics and Automation. Proceedings. 1987 IEEE International Conference on*, vol. 4. IEEE, 1987, pp. 396–401.

[8] M. Cutkosky, "On grasp choice, grasp models, and the design of hands for manufacturing tasks," *IEEE Transactions on Robotics and Automation*, vol. 5, no. 3, pp. 269–279, 1989.

[9] B. Abbasi, E. Noohi, S. Parastegari, and M. Žefran, "Grasp taxonomy based on force distribution," in *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*. IEEE, 2016, pp. 1098–1103.

[10] H. Marino, M. Gabiccini, A. Leonardis, and A. Bicchi, "Data-driven human grasp movement analysis," in *ISR 2016: 47st International Symposium on Robotics; Proceedings of*. VDE, 2016, pp. 1–8.

[11] J. M. Elliott and K. Connolly, "A classification of manipulative hand movements," *Developmental Medicine & Child Neurology*, vol. 26, no. 3, pp. 283–296, 1984.

[12] I. M. Bullock, R. R. Ma, and A. M. Dollar, "A hand-centric classification of human and robot dexterous manipulation," *Haptics, IEEE Transactions on*, vol. 6, no. 2, pp. 129–144, 2013.

[13] L. Y. Chang and N. S. Pollard, "Video survey of pre-grasp interactions in natural hand activities," June 2009.

[14] F. Worgotter, E. E. Aksoy, N. Kruger, J. Piater, A. Ude, and M. Tamosiunaite, "A simple ontology of manipulation actions based on hand-object relations," *Autonomous Mental Development, IEEE Transactions on*, vol. 5, no. 2, pp. 117–134, 2013.

[15] D. Leidner, C. Borst, A. Dietrich, M. Beetz, and A. Albu-Schäffer, "Classifying compliant manipulation tasks for automated planning in robotics," in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*. IEEE, 2015, pp. 1769–1776.

[16] J. Borras and T. Asfour, "A whole-body pose taxonomy for loco-manipulation tasks," in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*. IEEE, 2015, pp. 1578–1585.

[17] N. Abe and J.-P. Laumond, "Dance notations and robot motion," in *Proceedings of the 1st Workshop of the Anthropomorphic Motion Factory at LAAS-CNRS '14*, Toulouse, France, 2014.

[18] P. Ekman and W. Friesen, *Facial action coding system: A technique for the measurement of facial movement*. Consulting Psychologists Press, Palo Alto, 1978.

[19] J. Cohn and P. Ekman, "Measuring facial action by manual coding,

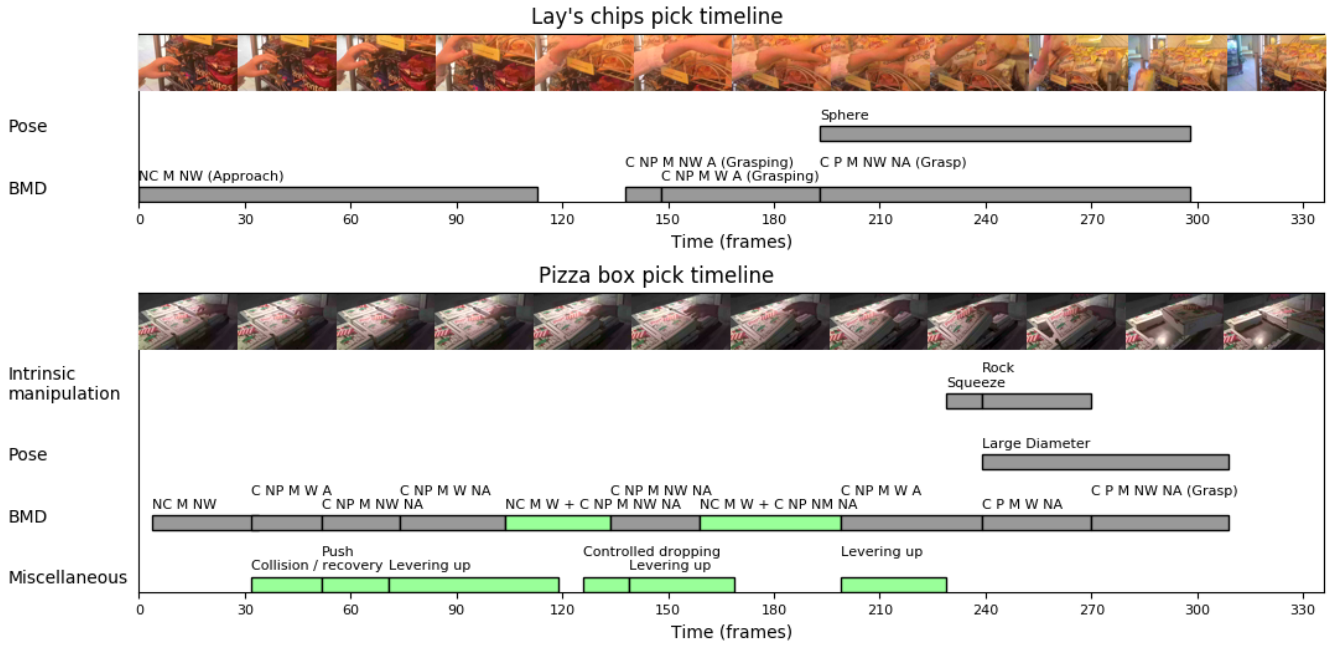

Fig. 5: Timelines for Lay's chips pick and pizza box pick

facial EMG, and automatic facial image analysis," *Handbook of Nonverbal Behavior Research Methods in the Affective Sciences*, 2005.

[20] T. Torigoe, "Comparison of object manipulation among 74 species of non-human primates," *Primates*, vol. 26, no. 2, pp. 182–194, 1985.

[21] R. W. Byrne, J. M. Byrne *et al.*, "Manual dexterity in the gorilla: bimanual and digit role differentiation in a natural task," *Animal Cognition*, vol. 4, no. 3-4, pp. 347–361, 2001.

[22] T. Feix, J. Romero, H.-B. Schmiedmayer, A. M. Dollar, and D. Kragic, "The GRASP taxonomy of human grasp types," *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 1, pp. 66–77, 2016.

[23] I. M. Bullock, T. Feix, and A. M. Dollar, "The Yale human grasping dataset: Grasp, object, and task data in household and machine shop environments," *The International Journal of Robotics Research*, p. 0278364914555720, 2014.

[24] M. T. Mason, "Mechanics and planning of manipulator pushing operations," *International Journal of Robotics Research*, vol. 5, no. 3, pp. 53–71, 1986.

[25] K. M. Lynch, "Toppling manipulation," in *IEEE International Conference on Robotics and Automation*, 1999.

[26] E. Yoshida, M. Poirier, J.-P. Laumond, O. Kanoun, F. Lamiraux, R. Alami, and K. Yokoi, "Pivoting based manipulation by a humanoid robot," *Autonomous Robots*, vol. 28, no. 1, pp. 77–88, 2010.

[27] E. Klingbeil, A. Saxena, and A. Y. Ng, "Learning to open new doors," in *IEEE/RSJ Conference on Intelligent Robots and Systems (IROS)*, 2010.

[28] M. Stilman and J. Kuffner, "Navigation among movable obstacles: Real-time reasoning in complex environments," *International Journal of Humanoid Robotics*, vol. 2, no. 4, pp. 479–503, 2005.

[29] M. Tenorth, U. Klank, D. Pangercic, and M. Beetz, "Web-enabled robots – robots that use the web as an information resource," *IEEE Robotics and Automation Magazine*, vol. 18, 2011.

[30] M. Bollini and D. Rus, "Cookies, anyone? http://web.mit.edu/newsoffice/2011/cookies-anyone.html."

[31] J. Maitin-Shepard, M. Cusumano-Towner, J. Lei, and P. Abbeel, "Cloth grasp point detection based on multiple-view geometric cues with application to robotic towel folding," in *IEEE International Conference on Robotics and Automation*, 2010.

[32] J. Liu, F. Feng, Y. C. Nakamura, and N. S. Pollard, "A taxonomy of everyday grasps in action," in *Humanoid Robots (Humanoids), 2014 14th IEEE-RAS International Conference on*. IEEE, 2014, pp. 573–580.