Harnessing Human Manipulation

NSF/ARL Workshop on Cloud Robotics: Challenges and Opportunities
February 27-28, 2013

Matthew T. Mason[1], Nancy Pollard[1], Alberto Rodriguez[1] and Ryan Kerwin[2]
[1]The Robotics Institute — Carnegie Mellon University
<matt.mason, nsp, albertor>@cs.cmu.edu
[2]Georgia Institute of Technology
ryankerwin@gatech.edu

***Abstract*— This paper describes a new project to analyze the vast corpus of human manipulation videos available on the web, using citizen scientists. Humans understand human manipulation well enough to segment videos appropriately, to assign intent, and to recognize actions even from poor imagery. A large corpus of analyzed human manipulation would add to our understanding of manipulation, support automated analysis, facilitate collaborative manipulation and generally accelerate robotic manipulation research. The paper presents our initial results with a web interface used by nine "citizen scientists" to analyze sandwich making.**

I. Introduction

Robotic manipulation research frequently draws on human manipulation for inspiration, but our appreciation of human manipulation is primarily anecdotal and introspective. A deeper, more principled understanding of human manipulation will help to set the direction for robotic manipulation research, inspire new designs, catalyze the development and integration of a broad range of robotic manipulation skills, provide a foundation for new planning and control techniques, and ultimately accelerate the contributions of robotic technologies in new applications as well as old.

The goal of our work is to achieve this deeper understanding and acceleration of robotic manipulation research by using citizen scientists to analyze the vast corpus of human manipulation available on the internet. We hope to emulate the success of LabelMe [58, 59] which provides a large corpus of labelled images in support of computer vision research. The present paper describes the very beginning of the project, including a web interface used by nine "citizen scientists" (actually members of our lab, and some of their family) to analyze a video of a human working in a kitchen.

II. Related Work

This section discusses related work in human and robot manipulation, segmentation, action units, and datasets.

This work was supported by National Science Foundation [NSF-IIS-0916557].

a) Human Manipulation: Our goal of analyzing human manipulation is preceded by work analyzing coordination patterns in human reaching, grasping, and lifting tasks. Researchers have attempted to identify distinguishing features of these tasks such as how finger grip force varies with friction [33] and how posture varies with task [8, 54, 57, 75, 12]. Researchers have examined coordination explicitly, including between the arm and hand [32, 38], within the joints of the hand [60, 44, 42], and in forces applied at the fingers [33, 40, 39]. Other groups have explored potential objective criteria for motor control, such as minimizing jerk [27], torque change [68] or other metrics of exertion or energy [16, 21, 15], maximizing end state comfort [26, 63, 56], and maximizing inertial symmetry of a grasp [67]. Grasp taxonomies have also been developed to structure the space of configurations taken by the hand, from basic taxonomies [49, 62] to taxonomies targeted at tasks of daily living [22, 35, 76] or for skilled machining tasks [19].

Drawing on studies of human manipulation, robotics researchers have proposed organizational strategies for robotic grasps [6, 31], including the concept of virtual fingers [31], the use of grasp synergies [17], and metrics for high quality grasps [4]. Human behavior imitation has been used for learning of robotic grasping tasks [37], and also for more dynamic tasks such as pole swing-up [61], air hockey [7], and box tumbling [53].

Most human manipulation research, however, has focused on the tasks of reaching, grasping, and lifting, with the assumption that the object remains fixed in place. Much less studied are strategies involving object movement. There are some examples, such as that of Rosenbaum and his colleagues [55] who present a method for evaluating costs of object-positioning and object-rotation. More often such studies focus on non-human manipulation, so that it seems possible that we know more about manipulation in apes than in humans! For example, Torigoe [66] identified over 500 different manipulation acts and compared their use across 74 different species, and Van Schaik and colleagues [69] identify and classify numerous instances of tool use in various animal species.

Our work employs humans to parse human manipulation behaviors. Along these lines, the research of Byrne [13] on great apes is intriguing. In the context of skill acquisition in great apes, he suggests that primitives may be identified even in novel observed actions and perhaps statistically related to their likely effects.

b) Segmentation: The present project uses humans to segment video images, but there remains the problem of combining the segmentations of different human subjects, and the closely related question of whether automated segmentation is feasible. Segmentation of human motion is of interest in a number of communities, including computer graphics, computer vision, and for human subjects studies, and there is a rich background to draw from. Both model-based approaches [28, 50, 2] and unsupervised approaches [5, 9, 73, 51] have been proposed for human motion tracking and for segmentation and labeling of discrete actions. Model-based approaches, however, rely on the presence of hand-annotated training data, along with a well defined set of actions, neither of which are available at this stage of our investigation. Unsupervised approaches to date tend to rely on rather abstract features of the motion such as ellipsoidal representations or pixel flow, which creates a challenge for capturing details such as contact configurations, changing grasps, sliding vs. tumbling etc. As such, we are steered towards crowd sourcing, where we hope to make the use of the ability of people to understand very much about an action from a relatively small number of pixels. Of particular note, however, is the research of Müller and his colleagues [46, 47], who propose tracking high level features such as "right hand moving forwards" that may be useful for our human annotators.

c) Compositional Units of Action: Researchers have documented for years evidence for the existence of compositional units used in the production of human behavior and motion. Aside from the biological implications of the validity of that claim, the identification and classification of compositional units has proved key in the development of the respective fields. First, they provide a model for behavior production; Second, compositional units constitute a language to reason, represent and quantify human behavior.

The most succesful example is the introduction of Facial Action Units by Ekman and Friesen [23, 25, 18]. In a higly referenced manual by the name of Facial Action Coding System (FACS) they provide a classification and anatomically-based quantitative description of facial motion. Facial movement, either gesture or expression, is explained as the composition and concatenation of 44+ individual action units. Each unit is described in terms of how to detect it, how to score its magnitude and how to produce it. FACS coding produces a sequence of units and intensities that have been used to measure emotion, pain, and psychological disorders [65, 24, 3, 74, 70, 1] as well as to synthesize facial expressions in animation [71, 52, 34, 72].

Compositional units are also important in the context of human speech production. Although there is no obvious division of speech when analyzed as an acoustic signal, it has been proposed that from the point of view of the control of the articulatory system there are compositional units of speech production [29, 10, 11]. The introduction of the atomic actions of speech has also proven useful to understand the emergence of speech errors [45, 30].

d) Datasets: The creation of large, comprehensive and community-shared databases plays a key role in research development. A prominent example in the robotics community is the labeled and segmented set of 800,000 images that LabelMe [58, 59] has produced over the years. The LabelMe dataset has had a large impact in the communities of automatic of object detection and recognition. Another relevant example is the Cohn-Kanade (CK) dataset [36, 41] for the automatic detection and recognition of facial expression. The CK dataset contains 2105 annotated image sequences from 182 subjects of varying ethnicity, showing multiple instances of most primary Facial Action Units (FACS) [23, 25]. It has had a large influence in computer vision, machine learning, behavioral sciences and human-computer interaction. In face detection and modeling, the CMU Pose, Illumination, and Expression (PIE) [64] and AR Face [43] databases have played a key role in computer vision and computer graphics.

III. First results

We conducted a pilot study to explore the use of crowd sourcing to analyze human manipulation behaviors. Specifically our pilot study addresses the following issues:

- Can ordinary untrained users effectively analyze human manipulation behaviors from ordinary video data alone?
- Can we develop a web-based interface that serves video at sufficient resolution in time and space, and allows users to segment and label clips effectively?
- Can a crowd of individuals working independently produce coherent analysis of a video?

The study uses a single video [14] from the Carnegie Mellon University Multimodal Activity (CMU-MMAC) Database [20]. The video shows a male subject gathering peanut butter and jelly from a refrigerator, gathering bread and utensils from cabinets and drawers, and making a peanut butter and jelly sandwich. Despite the presence of a variety of sensors on his head, body, arms, and the back of his hands, the motions seem natural. From our single view, the subject's hands are obscured when inside the refrigerator, cabinets, or drawers, and when inside the bread sack. Otherwise we have a good view of the action. The video is 1024x768, 30fps and about 4 minutes long, but was served to users at a resolution of 410x308.

We recruited a crowd of nine users for the task of segmenting and labeling video clips, using the web interface shown in Figure 1. The system maintains a pool of video clips on a server. On demand it randomly draws a clip from its pool, and presents it to the user. The user then selects and labels as many clips as he or she likes, possibly overlapping. When finished, the labelled clips are added to the pool and a new video clip is served. The users were not trained, except for some brief instructions on the web page. The experiment

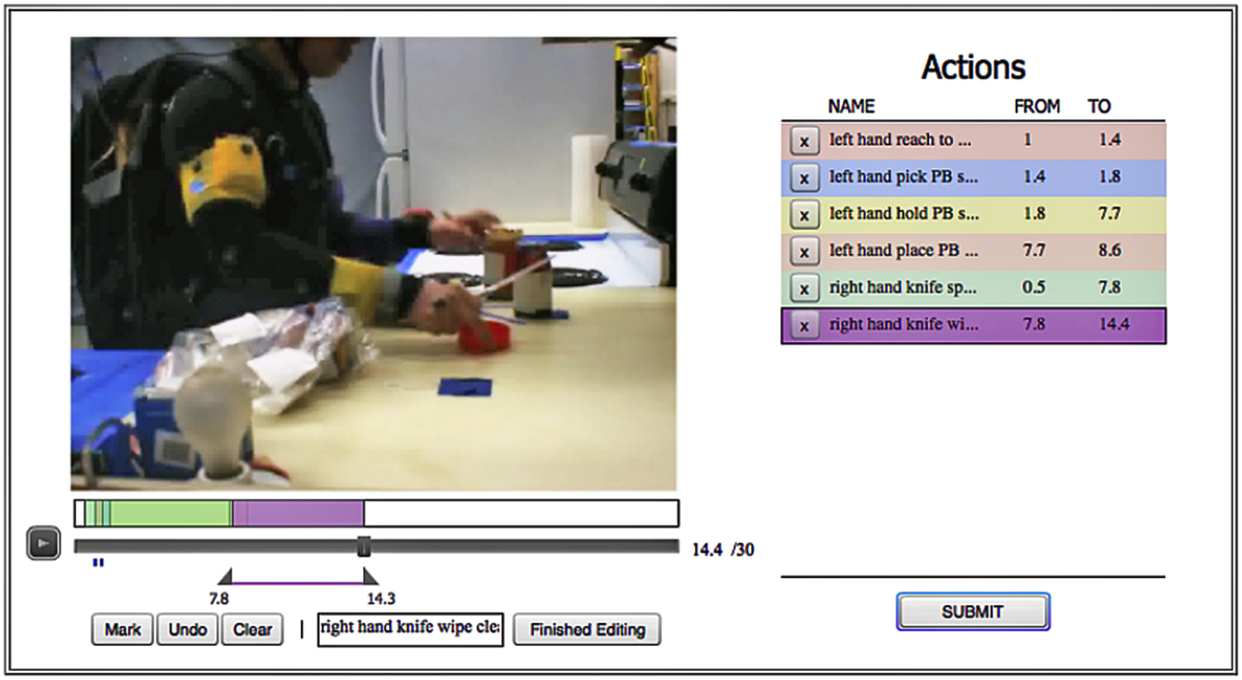

Fig. 1. Screen shot of the segmentation interface used in the pilot project. Note that at this point the subject is using what Napier [48] termed a "combined grip" to hold the knife in while grasping a lid with thumb and middle fingers.

lasted for a week over which the users could participate on demand.

To analyze the results, the authors carefully segmented and labelled the first 30 seconds of the video, which included getting the peanut butter and jelly jars from the refrigerator, and placing them on the counter. The result comprises 55 primitive actions, which, for the purpose of the study, we treat as ground truth. Table I shows a representative sequence.

t_{start}	t_{end}	Label
26.02	27.16	Hold both jars in line of sight.
27.17	28.08	Left hand, holding PB jar, reaches to top of door.
27.17	28.08	Left hand switches PB jar grasp to thumb and index cylinder.
27.28	31.12	Footstep right. Turn right. Three more steps to square shoulders with counter. Get close to counter.
28.08	28.08	Left hand grasps fridge door handle. Middle finger atop handle. Index and thumb fingers against side.
28.10	29.03	Left hand pushes door to start closing.
29.03	29.05	Left hand releases door.
29.05	29.20	Fridge door swings closed.
29.03	29.05	Left hand changes grasp of PB jar to cylinder grasp.

TABLE I
REPRESENTATIVE SEQUENCE OF GROUND TRUTH SEGMENTS AND LABELS IN THE PILOT PROJECT.

Our nine crowd members labelled 463 actions, 74 of them in the first 30 seconds. We manually compared each crowd label with our ground truth, and mapped each crowd label to a corresponding ground truth primitive, or sequence of primitives. Twenty of the crowd's actions mapped directly to ground truth primitives, three of them were more primitive than our ground truth primitives, two of them defied classification, and the remainder mapped to sequences of primitives. We then compared the start and stop times with ground truth. Figure 2 shows a histogram of time errors. There are some outliers, corresponding to ambiguities in matching crowd labels against ground truth. Otherwise we were surprised at the overall accuracy.

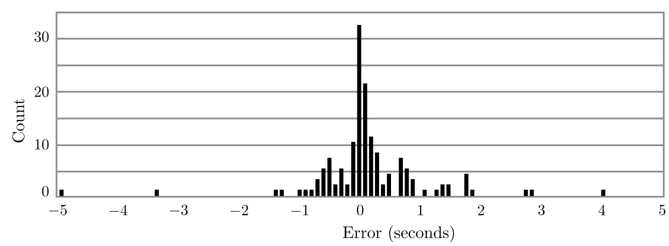

Fig. 2. Histogram of time errors.

IV. Conclusion

We conclude from the study:

- As we had hoped, the pilot project demonstrates that humans can analyze video, and the users were surprisingly consistent in their segmentation.
- It is interesting that people generally refer to the *effects* of actions, rather than the actual behavior of the hands or body, e.g. "Turned on the light" rather than "tapped the switch with the fingertips of the left hand".
- Even the small sample of clips and labels in Table I illustrates a few problems: Should we consider locomotion acts as part of manipulation? What about perception

acts? Should we label the motion of an object after release? How do we deal with obscured actions? What level of detail do we want in the label of a video segment?

- A large portion of the manipulation action in the video cannot be classified under the grasp–move–ungrasp paradigm. Even some of the grasps are novel. Some users labelled the grasp of the refrigerator handle a hook grasp, but that is not quite correct. The subject only inserts his fingertips, so that his hand can readily disengage and slide over the top of the door as it opens. It would be useful to have a more comprehensive list of manipulation primitives to better describe these videos.

The current results are far from sufficient to have a significant impact on our understanding of human manipulation, but we are encouraged in our hope that human analysis of human manipulation, using the techniques of citizen science, has the potential to provide a comprehensive corpus of analyzed human manipulation behaviors.

REFERENCES

[1] Z. Ambadar, J. Cohn, and L. Reed, "All smiles are not created equal: Morphology and timing of smiles perceived as amused, polite, and embarrassed/nervous," *Journal of Nonverbal Behavior*, vol. 33, no. 1, pp. 17–34, 2009.

[2] O. Arikan, D. A. Forsyth, and J. F. O'Brien, "Motion synthesis from annotations," *ACM Transactions on Graphics (SIGGRAPH)*, vol. 22, no. 3, pp. 402–408, 2003.

[3] A. Ashraf, S. Lucey, J. Cohn, T. Chen, Z. Ambadar, K. Prkachin, and P. Solomon, "The painful face-pain expression recognition using active appearance models," *Image and Vision Computing*, vol. 27, no. 12, pp. 1788–1796, 2009.

[4] R. Balasubramanian, L. Xu, P. D. Brook, J. R. Smith, and Y. Matsuoka, "Human-guided grasp measures improve grasp robustness on physical robot," in *IEEE International Conference on Robotics and Automation*, 2010.

[5] J. Barbic, A. Safonova, J.-Y. Pan, C. Faloutsos, J. K. Hodgins, and N. S. Pollard, "Segmenting motion capture data into distinct behaviors," in *Proceedings of Graphics Interface*, 2004.

[6] G. A. Bekey, H. Liu, R. Tomovic, and W. J. Karplus, "Knowledge-based control of grasping in robot hands using heuristics from human motor skills," *IEEE Transactions on Robotics and Automation*, vol. 6, no. 9, pp. 709–722, 1993.

[7] D. C. Bentivegna, A. Ude, C. G. Atkeson, and G. Cheng, "Learning to act from observation and practice," *International Journal of Humanoid Robotics*, vol. 1, no. 4, 2004.

[8] R. M. Bongers, C. F. Michaels, and A. W. Smitsman, "Variations of tool and task characteristics reveal that tool-use postures are anticipated," *Journal of Motor Behavior*, vol. 36, no. 3, pp. 305–315, 2004.

[9] M. E. Brand and V. Kettnaker, "Discovery and segmentation of activities in video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 844–851, 2000.

[10] C. Browman and L. Goldstein, "Towards an articulatory phonology," *Phonology Yearbook*, vol. 3, no. 21, pp. 9–252, 1986.

[11] ——, "Articulatory gestures as phonological units," *Phonology*, vol. 6, no. 2, pp. 201–251, 1989.

[12] R. Burgess-Limerick and B. Abernethy, "Qualitatively different modes of manual lifting," *Journal of Industrial Ergonomics*, vol. 19, no. 5, pp. 413–417, 1997.

[13] R. Byrne, "Imitation as behaviour parsing," *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, vol. 358, no. 1431, p. 529, 2003.

[14] Carnegie Mellon University Multimodal Activity (CMU-MMAC) Database, "Peanut butter sandwich video," http://kitchen.cs.cmu.edu/Main/S13_Sandwich_Video.zip.

[15] C. C. Chang, D. R. Brown, D. S. Bloswick, and S. M. Hsiang, "Biomechanical simulation of manual lifting using spacetime optimization," *Journal of Biomechanics*, vol. 34, no. 4, pp. 527–532, 2001.

[16] C. K. Chow and D. H. Jacobson, "Studies of human locomotion via optimal programming," *Mathematical Biosciences*, vol. 10, no. 3-4, pp. 239–306, 1971.

[17] M. Ciocarlie and P. K. Allen, "Hand posture subspaces for dexterous robotic grasping," *International Journal of Robotics Research*, vol. 28, pp. 851–867, 2009.

[18] J. Cohn and P. Ekman, "Measuring facial action by manual coding, facial EMG, and automatic facial image analysis," *Handbook of Nonverbal Behavior Research Methods in the Affective Sciences*, 2005.

[19] 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.

[20] F. De la Torre Frade, J. Hodgins, A. Bargteil, X. Martin Artal, J. Macey, A. Collado I Castells, and J. Beltran, "Guide to the Carnegie Mellon University multimodal activity (cmu-mmac) database," Robotics Institute, Pittsburgh, PA, Tech. Rep. CMU-RI-TR-08-22, April 2008.

[21] M. J. Dysart and J. C. Woldstad, "Posture prediction for static sagittal-plane lifting," *Journal of Biomechanics*, vol. 29, no. 10, pp. 1393–1397, 1996.

[22] S. J. Edwards, D. J. Buckland, and J. D. McCoy-Powlen, *Developmental and Functional Hand Grasps*. Slack Incorporated, Thorofare, New Jersey, 2002.

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

[24] P. Ekman and E. Rosenberg, *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS).* Oxford University Press, USA, 2005.

[25] P. Ekman, W. Friesen, and J. Hager, "Facial action coding system: The manual." *Network Research Information*, 2002.

[26] M. G. Fischman, "Constraints on grip-selection: minimizing awkwardness," *Perceptual and Motor Skills*, vol. 86, no. 1, pp. 328–330, 1998.

[27] T. Flash and N. Hogan, "The coordination of arm movements: an experimentally confirmed mathematical model," *Journal of Neuroscience*, vol. 5, no. 7, pp. 1688–1703, 1985.

[28] D. J. Fleet, M. J. Black, Y. Yacoob, and A. D. Jepson, "Design and use of linear models for image motion analysis," *International Journal of Computer Vision*, vol. 36, no. 3, pp. 171–193, 2000.

[29] C. Fowler, P. Rubin, R. Remez, and M. Turvey, "Implications for speech production of a general theory of action," *Language Production*, 1980.

[30] L. Goldstein, M. Pouplier, L. Chen, E. Saltzman, and D. Byrd, "Dynamic action units slip in speech production errors," *Cognition*, vol. 103, no. 3, pp. 386–412, 2007.

[31] T. Iberall, "Human prehension and dextrous robot hands," *International Journal of Robotics Research*, vol. 16, no. 3, 1997.

[32] M. Jeannerod, "Intersegmental coordination during reaching at natural visual objects," in *Attention and Performance IX (eds. J. Long and A. Baddeley).* Lawrence Erlbaum Associates, 1981, pp. 153–169.

[33] R. Johansson, "Sensory control of dexterous manipulation in humans," in *Hand and brain: The Neurophysiology and Psychology of Hand Movements (eds. A. Wing, P. Haggard, and J. J.R. Flangan).* Academic Press, New York, 1996, pp. 381–414.

[34] P. Kalra, A. Mangili, N. Thalmann, and D. Thalmann, "Simulation of facial muscle actions based on rational free form deformations," in *Computer Graphics Forum*, vol. 11. Wiley Online Library, 1992, pp. 59–69.

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

[36] T. Kanade, J. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 2000, pp. 46–53.

[37] S. B. Kang and K. Ikeuchi, "Toward automatic robot instruction from perception-mapping human grasps to manipulator grasps," *IEEE Journal of Robotics and Automation*, vol. 13, pp. 81–95, 1997.

[38] F. Lacquaniti and J. F. Soechting, "Coordination of arm and wrist motion during a reaching task," *Journal of Neuroscience*, vol. 2, no. 4, pp. 399–408, 1982.

[39] M. L. Latash, I. M. Gelfand, Z. M. Li, and V. M. Zatsiorsky, "Changes in the force-sharing pattern induced by modifications of visual feedback during force production by a set of fingers," *Experimental Brain Research*, vol. 123, no. 3, pp. 255–262, 1998.

[40] Z. M. Li, M. L. Latash, and V. M. Zatsiorsky, "Force sharing among fingers as a model of the redundancy problem," *Experimental Brain Research*, vol. 119, no. 3, pp. 276–286, 1998.

[41] P. Lucey, J. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2010, pp. 94–101.

[42] J. Lukos, C. Ansuini, and M. Santello, "Choice of contact points during multidigit grasping: effect of predictability of object center of mass location," *Journal of Neuroscience*, vol. 27, no. 14, pp. 3894–3903, 2007.

[43] A. Martinez, "The AR face database," *CVC Technical Report*, vol. 24, 1998.

[44] C. Mason, J. Gomez, and T. Ebner, "Hand synergies during reach-to-grasp," *Journal of Neurophysiology*, vol. 86, no. 6, pp. 2896–2910, 2001.

[45] R. Mowrey and I. MacKay, "Phonological primitives: Electromyographic speech error evidence." *Journal of the Acoustical Society of America*, 1990.

[46] M. Müller and T. Röder, "Motion templates for automatic classification and retrieval of motion capture data," in *ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA)*, 2006, pp. 137–146.

[47] M. Müller, A. Baak, and H.-P. Seidel, "Efficient and robust annotation of motion capture data," in *ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA)*, 2009, pp. 17–26.

[48] J. R. Napier, "The prehensile movements of the human hand," *J Bone Joint Surg Br*, vol. 38-B, no. 4, pp. 902–913, 1956. [Online]. Available: http://web.jbjs.org.uk/cgi/content/abstract/38-B/4/902

[49] J. Napier and R. Tuttle, *Hands*. Princeton University Press, 1993.

[50] N. Oliver, E. Horvitz, and A. Garg, "Layered representations for human activity recognition," in *IEEE International Conference on Multimodal Interfaces*, 2002, pp. 3–8.

[51] N. Peyrard and P. Bouthemy, "Content-based video segmentation using statistical motion models," in *Proceedings of the British Machine Vision Conference (BMVC)*, 2002, pp. 527–536.

[52] S. Platt and N. Badler, "Animating facial expressions," *ACM SIGGRAPH computer graphics*, vol. 15, no. 3, pp. 245–252, 1981.

[53] N. S. Pollard and J. K. Hodgins, "Generalizing demonstrated manipulation tasks," in *Workshop on the Algorithmic Foundations of Robotics*, Dec 2002.

[54] M. K. Rand and G. E. Stelmach, "Effect of orienting the finger opposition space in the control of reach-to-grasp movements," *Journal of Motor Behavior*, vol. 37, no. 1, pp. 65–78, 2005.

[55] D. A. Rosenbaum and M. J. Gaydos, "A method for obtaining psychophysical estimates of movement costs," *Journal of Motor Behavior*, vol. 40, no. 1, pp. 11–17, 2008.

[56] D. A. Rosenbaum, J. Vaughan, H. J. Barnes, and M. J. Jorgensen, "Time course of movement planning: selection of handgrips for object manipulation," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 18, no. 5, pp. 1058–1073, 1992.

[57] D. A. Rosenbaum, J. Vaughan, M. J. Jorgensen, H. J. Barnes, and E. Stewart, "Plans for object manipulation," in *Attention and Performance XIV - A silver jubilee: Synergies in experimental psychology, artificial intelligence and cognitive neuroscience (ed. D. E. Meyer and S. Kornblum)*. MIT Press, Bradford Books, Cambridge, 1993, pp. 803–820.

[58] B. Russell, A. Torralba, K. Murphy, and W. Freeman, "LabelMe: a database and web-based tool for image annotation," *International Journal of Computer Vision*, vol. 77, no. 1, pp. 157–173, 2008.

[59] B. Russell, A. Torralba, and W. T. Freeman, "LabelMe: The open annotation tool," http://labelme.csail.mit.edu/index.html.

[60] M. Santello, M. Flanders, and J. Soechting, "Postural hand synergies for tool use," *Journal of Neuroscience*, vol. 18, no. 23, pp. 10 105–10 115, 1998.

[61] S. Schaal and C. G. Atkeson, "Learning control in robotics," *IEEE Robotics and Automation Magazine*, vol. 17, pp. 20–29, 2010.

[62] G. Schlesinger, "Der Mechanische Aufbau der Kunstlichen Glieder. Ersatzglieder und Arbeitshilfen, part II," 1919.

[63] M. W. Short and J. H. Cauraugh, "Planning macroscopic aspects of manual control: end-state comfort and point-of-change effects," *Acta Psychologica*, vol. 96, no. 1-2, pp. 133–147, 1997.

[64] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1615–1618, 2003.

[65] Y. Tian, T. Kanade, and J. Cohn, "Recognizing action units for facial expression analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 97–115, 2001.

[66] T. Torigoe, "Comparison of object manipulation among 74 species of non-human primates," *Primates*, vol. 26, pp. 182–194, 1985.

[67] M. T. Turvey, K. Shockley, and C. Carello, "Affordance, proper function, and the physical basis of perceived heaviness," *Cognition*, vol. 73, no. 2, pp. B17–B26, 1999.

[68] Y. Uno, M. Kawato, and R. Suzuki, "Formation and control of optimal trajectory in human multijoint arm movement - minimum torque-change model," *Biological Cybernetics*, vol. 61, no. 2, pp. 89–101, 1989.

[69] C. Van Schaik, R. Deaner, and M. Merrill, "The conditions for tool use in primates: implications for the evolution of material culture," *Journal of Human Evolution*, vol. 36, pp. 719–741, 1999.

[70] P. Wang, F. Barrett, E. Martin, M. Milonova, R. Gur, R. Gur, C. Kohler, and R. Verma, "Automated video-based facial expression analysis of neuropsychiatric disorders," *Journal of neuroscience methods*, vol. 168, no. 1, pp. 224–238, 2008.

[71] K. Waters, "A muscle model for animation three-dimensional facial expression," *ACM SIGGRAPH Computer Graphics*, vol. 21, no. 4, pp. 17–24, 1987.

[72] A. Wojdeł and L. Rothkrantz, "Parametric generation of facial expressions based on FACS," in *Computer Graphics Forum*, vol. 24, 2005, pp. 743–757.

[73] L. Zelnik-Manor and M. Irani, "Event-based video analysis," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2001, pp. 123–130.

[74] Z. Zeng, M. Pantic, G. Roisman, and T. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, pp. 39–58, 2009.

[75] W. Zhang and D. A. Rosenbaum, "Planning for manual positioning: the end-state comfort effect for manual abduction-adduction," *Experimental Brain Research*, vol. 184, no. 3, pp. 383–389, 2008.

[76] J. Z. Zheng, S. D. L. Rosa, and A. M. Dollar, "An investigation of grasp type and frequency in daily household and machine shop tasks," in *IEEE International Conference on Robotics and Automation*, Shanghai, China, May 9–13 2011.