Bibliography

Adolph, K. E., Eppler, M. A., and Gibson, E. J. (1993). Development of perception of affordances. *Advances in Infancy Research*, 8:51–98.

Aloimonos, J., Weiss, I., and Bandopadhay, A. (1987). Active vision. *International Journal on Computer Vision*, 2:333–356.

Arsenio, A., Fitzpatrick, P., Kemp, C., and Metta, G. (2003). The whole world in your hand: Active and interactive segmentation. Accepted for publication in the proceedings of the Third International Workshop on Epigenetic Robotics.

Aslin, R., Woodward, J., LaMendola, N., and Bever, T. (1996). Models of word segmentation in fluent maternal speech to infants. In Morgan, J. and Demuth, K., editors, *Signal to Syntax: Bootstrapping From Speech to Grammar in Early Acquisition*. Lawrence Erlbaum Associates: Mahwah, NJ.

Bajcsy, R. (1988). Active perception. *Proceedings of the IEEE*, 76(8):996–1005.

Ballard, D. (1989). Behavioral constraints on animate vision. *Image and Vision Computing*, 7:1:3–9.

Ballard, D. H. (1991). Animate vision. *Artificial Intelligence*, 48(1):57–86.

Bard, E. and Anderson, A. (1994). The unintelligibility of speech to children: effects of referent availability. *Journal of Child Language*, 21:623–648.

Basu, S., Essa, I., and Pentland, A. (1996). Motion regularization for model-based head tracking. In *Proceedings of the International Conference on Pattern Recognition*, Vienna, Austria.

Bazzi, I. and Glass, J. (2000). Modeling out-of-vocabulary words for robust speech recognition. In *Proceedings of the Sixth International Conference on Spoken Language Processing*, Beijing, China.

Beardsley, P. A. (1998). A qualitative approach to classifying head and eye pose. In *Proceedings of the IEEE Workshop on Applications of Computer Vision*, pages 208–213, Florence, Italy.

Billard, A. (2001). Imitation: a means to enhance learning of a synthetic proto-language in an autonomous robot. In Dautenhahn, K. and Nehaniv, C. L., editors, *Imitation in Animals and Artifacts*. MIT Press.

Billard, A. and Dautenhahn, K. (1997). Grounding communication in situated, social robots. Technical Report UMCS-97-9-1, University of Manchester.

Birchfield, S. (1998). Elliptical head tracking using intensity gradients and color histograms. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 232–237.

Black, M. and Yacoob, Y. (1995). Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion. In *Proceedings of the International Conference on Computer Vision*, pages 374–381.

Block, N. (1978). Troubles with functionalism. *Perception and Cognition: Issues in the Foundations of Psychology, Minnesota Studies in the Philosophy of Science*, 9:261–325.

Bloom, P. (2000). *How Children Learn the Meaning of Words*. Cambridge: MIT Press.

Blumberg, B. (1996). *Old Tricks, New Dogs: Ethology and Interactive Creatures*. PhD thesis, MIT.

Borchardt, G. C. (1993). Causal reconstruction. Technical Report AIM-1403, MIT Artificial Intelligence Laboratory.

Boykov, Y. and Kolmogorov, V. (2001). An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 359–374.

Breazeal, C. (2000). *Sociable Machines: Expressive Social Exchange Between Humans and Robots*. PhD thesis, MIT Department of Electrical Engineering and Computer Science.

Breazeal, C. and Aryananda, L. (2000). Recognition of affective communicative intent in robot-directed speech. In *Proceedings of the International IEEE/RSJ Conference on Humanoid Robotics*, Cambridge, MA.

Breazeal, C., Edsinger, A., Fitzpatrick, P., and Scassellati, B. (2001). Active vision for sociable robots. *IEEE Transactions on Systems, Man, and Cybernetics, A*, 31(5):443–453.

Breazeal, C., Edsinger, A., Fitzpatrick, P., Scassellati, B., and Varchavskaia, P. (2000). Social constraints on animate vision. *IEEE Intelligent Systems*, 15:32–37.

Breazeal, C. and Fitzpatrick, P. (2000). That certain look: Social amplification of animate vision. In *Proceedings AAAI Fall Symposium, Socially Intelligent Agents - The Human in the Loop*, North Falmouth, MA, USA.

Breazeal, C. and Scassellati, B. (1999). A context-dependent attention system for a social robot. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, pages 1146–1151, Stockholm, Sweden.

Breazeal, C. and Scassellati, B. (2000). Infant-like social interactions between a robot and a human caretaker. *Adaptive Behavior*, 8(1):49–74.

Breazeal, C. and Velasquez, J. (1998). Toward teaching a robot "infant" using emotive communication acts. In *Proceedings of the Simulated Adaptive Behavior Workshop.*

Brent, M. and Siskind, J. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, 81:B33–B44.

Brooks, R. A. (1986). A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, RA-2:14–23.

Brooks, R. A. (1991a). Intelligence without reason. In *Proceedings of the 1991 International Joint Conference on Artificial Intelligence*, pages 569–595.

Brooks, R. A. (1991b). Intelligence without representation. *Artificial Intelligence Journal*, 47:139–160. originally appeared as MIT AI Memo 899 in May 1986.

Brooks, R. A., Breazeal, C., Irie, R., Kemp, C. C., Marjanović, M., Scassellati, B., and Williamson, M. M. (1998). Alternative essences of intelligence. In *Proceedings of the American Association of Artificial Intelligence*, pages 961–968.

Brooks, R. A., Breazeal, C., Marjanovic, M., and Scassellati, B. (1999). The Cog project: Building a humanoid robot. *Lecture Notes in Computer Science*, 1562:52–87.

Brooks, R. A. and Stein, L. A. (1994). Building brains for bodies. *Autonomous Robots*, 1(1):7–25.

Bullowa, M. (1979). *Before Speech: The Beginning of Interpersonal Communicaion*. Cambridge University Press, Cambridge, London.

Burnham, D., Francis, E., Kitamura, C., Vollmer-Conna, U., Averkiou, V., Olley, A., and Paterson, C. (1998). Are you my little pussy-cat? acoustic, phonetic and affective qualities of infant- and pet-directed speech. In *Proceedings of the Fifth International Conference on Spoken Language Processing*, volume 2, pages 453–456.

Butterworth, G. (1991). The ontogeny and phylogeny of joint visual attention. In Whiten, A., editor, *Natural Theories of Mind*. Blackwell.

Canny, J. F. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):34–43.

Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., Douville, B., Prevost, S., and Stone, M. (1994). Animated conversation: Rule-based generation of facial expression, gesture, and spoken intonation for multiple conversational agents. *SIGGRAPH'94*.

Chabat, F., Yang, G. Z., and Hansell, D. M. (1999). A corner orientation detector. *Image and Vision Computing*, 17(10):761–769.

Chapman, D. (1990). Vision, instruction, and action. Technical Report AITR-1204, MIT AI Laboratory.

Chapman, D. and Agre, P. E. (1987). Pengi: An implementation of a theory of activity. In *Proceedings of the Sixth National Conference on Artificial Intelligence*, pages 268–272.

Chen, J., Sato, Y., and Tamura, S. (2000). Orientation space filtering for multiple orientation line segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(5):417–429.

Chen, Q., amd T. Shioyama, H. W., and Shimada, T. (1999). 3D head pose estimation using color information. In *Proceedings of the Sixth IEEE International Conference on Multimedia Computing and Systems*, Florence, Italy.

Cole, R. and Yap, C. (1987). Shape from probing. *Journal of Algorithms*, 8(1):19–38.

Colombetti, M. and Dorigo, M. (1994). Training agents to perform sequential behavior. *Adaptive Behavior*, 2(3):247–275.

Connell, J. (1989). A colony architecture for an artificial creature. Technical Report AITR-1151, MIT AI Laboratory.

Cordea, M., Petriu, E., Georganas, N., Petriu, D., and Whalen, T. (2000). Real-time 2.5D head pose recovery for model-based video-coding. In *Proceedings of the IEEE Instrumentation and Measurement Technology Conference*, Baltimore, MD, USA.

DeCarlo, D. and Metaxas, D. (1996). The integration of optical flow and deformable models with applications to human face shape and motion estimation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 231–238.

Deming, W. E. (1993). *The New Economics for Industry, Government, and Education.* MIT Center for Advanced Engineering, Cambridge, Massachusetts.

Dennett, D. C. (1987). *The Intentional Stance.* MIT Press.

Duda, R. and Hart, P. (1973). *Pattern Classification and Scene Analysis.* Wiley, New York.

Fasulo, D. (1999). An analysis of recent work on clustering algorithms. Technical report, University of Washington.

Ferrell, C. (1998). Learning by scaffolding. MIT Ph.D. Thesis Proposal.

Ferrell, C. and Kemp, C. (1996). An ontogenetic perspective to scaling sensorimotor intelligence. In *Embodied Cognition and Action: Papers from the 1996 AAAI Fall Symposium.* AAAI Press.

Firby, R. J. (1994). Task networks for controlling continuous processes. In *Proceedings of the Second International Conference on AI Planning Systems*, Chicago IL.

Fitzpatrick, P. and Metta, G. (2002). Towards manipulation-driven vision. In *IEEE/RSJ Conference on Intelligent Robots and Systems.*

Fogassi, L., Gallese, V., Fadiga, L., Luppino, G., Matelli, M., and Rizzolatti, G. (1996). Coding of peripersonal space in inferior premotor cortex (area F4). *Journal of Neurophysiology*, pages 141–157.

Folsom, T. C. and Pinter, R. B. (1998). Primitive features by steering, quadrature, and scale. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1161–1173.

Freeman, W. (1992). *Steerable filters and the local analysis of image structure.* PhD thesis, Media Arts and Sciences, MIT.

Freeman, W. T. and Adelson, E. H. (1991). The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(9):891–906.

Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 119:593–609.

Gansner, E. R. and North, S. C. (2000). An open graph visualization system and its applications to software engineering. *Software: Practice and Experience*, 30(11):1203–1233.

Garland, A. and Lesh, N. (2001). Learning hierarchical task models by demonstration. Technical report, Mitsubishi Electric Research Laboratories.

Gat, E. (1996). ESL: A language for supporting robust plan execution in embedded autonomous agents. In *Plan Execution: Problems and Issues: Papers from the 1996 AAAI Fall Symposium*, pages 59–64. AAAI Press, Menlo Park, California.

Gibson, J. J. (1977). The theory of affordances. In Shaw, R. and Bransford, J., editors, *Perceiving, acting and knowing: toward an ecological psychology*, pages 67–82. Hillsdale NJ: Lawrence Erlbaum Associates Publishers.

Glass, J., Chang, J., and McCandless, M. (1996). A probabilistic framework for feature-based speech recognition. In *Proceedings of the International Conference on Spoken Language Processing*, pages 2277–2280.

Glass, J. and Weinstein, E. (2001). SPEECHBUILDER: Facilitating spoken dialogue systems development. In *Proceedings of the Seventh European Conference on Speech Communication and Technology*, Aalborg, Denmark.

Goldberg, D. and Mataric, M. J. (1999). Augmented markov models. Technical report, USC Institute for Robotics and Intelligent Systems.

Goldberg, M. E. (2000). The control of gaze. In Kandel, E. R., Schwartz, J. H., and Jessell, T. M., editors, *Principles of Neural Science*. McGraw-Hill, 4rd edition.

Goodman, N. (1983). *Fact, Fiction, and Forecast*. Harvard University Press, Cambridge, Massachusetts.

Gorin, A., Petrovksa-Delacrtaz, D., Riccardi, G., and Wright, J. (1999). Learning spoken language without transcriptions. In *Proceedings of the IEEE Automatic Speech Recognition and Understanding Workshop*, Colorado.

Granlund, G. H. (1978). In search of a general picture processing operator. *Computer Graphics and Image Processing*, 8(2):155–173.

Graziano, M. S. A., Hu, X., and Gross, C. G. (1997). Visuo-spatial properties of ventral premotor cortex. *Journal of Neurophysiology*, 77:2268–2292.

Grice, H. (1974). Logic and conversation. In Cole, P. and Morgan, J., editors, *Syntax and semantics*, volume 3, pages 41–58. Academic Press, New York.

Grimson, W. E. L. and Huttenlocher, D. P. (1990). On the sensitivity of the Hough transform for object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(3):255–274.

Haigh, K. Z. and Veloso, M. M. (1998). Planning, execution and learning in a robotic agent. In *Proceedings of the Fourth International Conference on Artificial Intelligence Planning Systems*, pages 120 – 127.

Halliday, M. (1975). *Learning How to Mean: Explorations in the Development of Language*. Elsevier, New York, NY.

Harnad, S. (2002). Symbol grounding and the origin of language. In Scheutz, M., editor, *Computationalism: New Directions*, pages 143–158. MIT Press.

Harville, M., Rahimi, A., Darrell, T., Gordon, G., and Woodfill, J. (1999). 3D pose tracking with linear depth and brightness constraints. In *Proceedings of the International Conference on Computer Vision*, pages 206–213.

Hauser, M. D. (1996). *Evolution of Communication.* MIT Press.

Hazen, T. and Bazzi, I. (2001). A comparison and combination of methods for OOV word detection and word confidence scoring. In *Proceedings of the International Conference on Acoustics*, Salt Lake City, Utah.

Heinzmann, J. and Zelinsky, A. (1997). Robust real-time face tracking and gesture recognition. In *Proceedings of the International Joint Conference on Artificial Intelligence*, volume 2, pages 1525–1530.

Held, R. and Hein, A. (1963). Movement-produced stimulation in the development of visually-guided behavior. *Journal of Comparative and Physiological Psychology*, 56:872–876.

Hirschberg, J., Litman, D., and Swerts, M. (1999). Prosodic cues to recognition errors. In *Proceedings of the Automatic Speech Recognition and Understanding Workshop*.

Horn, B. K. P. (1986). *Robot Vision.* MIT Press.

Horprasert, T., Yacoob, Y., and Davis, L. S. (1997). An anthropometric shape model for estimating head orientation. In *Proceedings of the Third International Workshop on Visual Form*, Capri, Italy.

Horswill, I. (1993). *Specialization of Perceptual Processes.* PhD thesis, MIT.

Hu, M. K. (1962). Visual pattern recognition by moment invariants. In *IRE Transactions on Information Theory 8*, pages 179–187.

Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160:106–154.

Hueckel, M. H. (1971). An operator which locates edges in digitized pictures. *Journal of the Association for Computing Machinery*, 18(1):113–125.

Hurley, S. L. (2001). Perception and action: Alternative views. *Synthese*, 129:3–40.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., and Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science*, 286:2526–2528.

Jacobs, D. (1996). Robust and efficient detection of convex groups. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(1):23–37.

Jacobs, D., Rokers, B., , and Rudra, A. (2001). Fragment completion in humans and machines. In *Neural Information Processing Systems*.

Jeannerod, M. (1997). *The Cognitive Neuroscience of Action.* Blackwell Publishers Inc., Cambridge Massachusetts and Oxford UK.

Jeanrenaud, P., Ng, K., Siu, M., Rohlicek, J., and Gish, H. (1993). Phonetic-based word spotter: Various configurations and application to event spotting. In *Proceedings of the European Conference on Speech Communication and Technology (EUROSPEECH)*.

Jia, Y.-B. and Erdmann, M. (1998). Observing pose and motion through contact. In *Proceedings of the IEEE International Conference on Robotics and Automation.*

Johnson, S. P. (in press). Building knowledge from perception in infancy. In Gershkoff-Stowe, L. and Rakison, D., editors, *Building object categories in developmental time*. Hillsdale, Mahwah, NJ.

Jusczyk, P. (1997). *The Discovery of Spoken Language*. Cambridge: MIT Press.

Jusczyk, P. and Aslin, R. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, 29:1–23.

Kaelbling, L. P., Littman, M. L., and Moore, A. P. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, 4:237–285.

Kaelbling, L. P., Oates, T., Hernandez, N., and Finney, S. (2001). Learning in worlds with objects. In *AAAI Spring Symposium.*

Kamp, H. (1981). A theory of truth and semantic representation. In J, G., Janssen, T., and Stokhof, M., editors, *Formal Methods in the Study of Language*, pages 277–322. Mathematical Center Tract 135, Amsterdam.

Kass, M. and Witkin, A. (1987). Analyzing oriented patterns. *Computer Vision, Graphics, and Image Processing*, 37:362–385.

Kirkham, N. Z., Slemmer, J. A., and Johnson, S. P. (2002). Visual statistical learning in infancy: evidence of a domain general learning mechanism. *Cognition*, 83(2):B35–B42.

Kirsh, D. (1995a). Complementary strategies: Why we use our hands when we think. In *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society*, Hillsdale, NJ. Lawrence Erlbaum.

Kirsh, D. (1995b). The intelligent use of space. *Artificial Intelligence*, 73:31–68.

Kirsh, D. (1996). Adapting the environment instead of oneself. *Adaptive Behavior*, 4(3/4):415–452.

Klarquist, W. and Bovik, A. (1998). FOVEA: A foveated vergent active stereo system for dynamic three-dimensional scene recovery. *IEEE Transactions on Robotics and Automation*, 14(5):755–770.

Klingspor, V., Demiris, J., and Kaiser, M. (1997). Human-robot-communication and machine learning. *Applied Artificial Intelligence Journal*, 11:719–746.

Koga, Y., Kondo, K., Kuffner, J., and Latombe, J.-C. (1994). Planning motions with intentions. *Computer Graphics*, 28:395–408.

Konishi, S. M., Yuille, A. L., Coughlan, J. M., and Zhu, S. C. (2003). Statistical edge detection: Learning and evaluating edge cues. *Pattern Analysis and Machine Intelligence*, 25(1):57–74.

Kozima, H. (1998). Attention-sharing and behavior-sharing in human-robot communication. In *IEEE International Workshop on Robot and Human Communication (ROMAN-98)*, pages 9–14.

Kozima, H., Nakagawa, C., and Yano, H. (2002). Emergence of imitation mediated by objects. In *The Second International Workshop on Epigenetic Robotics*, Edinburgh, Scotland.

Kubota, T. (1995). *Orientational filters for real-time computer vision problems.* PhD thesis, School of Electrical and Computer Engineering, Georgia Institute of Technology.

Kubota, T. and Alford, C. O. (1995). Computation of orientational filters for real-time computer vision problems i: Implementation and methodology. *Real-Time Imaging*, 1(4):261–281.

La Cascia, M., Sclaroff, S., and Athitsos, V. (2000). Fast, reliable head tracking under varying illumination: An approach based on robust registration of texture-mapped 3D models. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 22(4), pages 322–336.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind.* University of Chicago Press, Chicago, Illinois.

Lamiroy, B. and Gros, P. (1996). Rapid object indexing and recognition using enhanced geometric hashing. In *Proceedings of the 4th European Conference on Computer Vision*, volume 1, pages 59–70, Cambridge, England.

Lenat, D. B. (1995). CYC: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11):33–38.

Marjanovic, M. (1995). Learning functional maps between sensorimotor systems on a humanoid robot. Master's thesis, MIT Department of Electrical Engineering and Computer Science.

Marjanović, M. J. (2003). *Teaching an Old Robot New Tricks: Learning Novel Tasks via Interaction with People and Things.* PhD thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, Cambridge, MA.

Marjanović, M. J., Scassellati, B., and Williamson, M. M. (1996). Self-taught visually-guided pointing for a humanoid robot. In *From Animals to Animats: Proceedings of 1996 Society of Adaptive Behavior*, pages 35–44, Cape Cod, Massachusetts.

Markman, E. M. (1989). *Categorization and naming in children: problems of induction.* MIT Press, Cambridge, Massachusetts.

Martin, D., Fowlkes, C., and Malik, J. (2002). Learning to detect natural image boundaries using brightness and texture. In *Sixteenth Annual Conference on Neural Information Processing Systems*.

Mataric, M. J. (1990). A distributed model for mobile robot environment-learning and navigation. Technical Report AITR-1228, Massachusetts Institute of Technology.

Mataric, M. J. (2000). Getting humanoids to move and imitate. *IEEE Intelligent Systems*, 15(4):18–24.

Matsusaka, Y. and Kobayashi, T. (1999). Human interface of humanoid robot realizing group communication in real space. In *Proceedings of the Second International Symposium on Humanoid Robots*, pages 188–193.

McCarthy, J. and Hayes, P. J. (1969). Some philosophical problems from the standpoint of artificial intelligence. In Meltzer, B. and Michie, D., editors, *Machine Intelligence 4*, pages 463–502. Edinburgh University Press.

McDermott, J. (2000). Experiments on junctions in real images. Master's thesis, University College London.

McGrenere, J. and Ho, W. (2000). Affordances: Clarifying and evolving a concept. In *Proceedings of Graphics Interface.*

McKenna, S. and Gong, S. (1998). Real-time face pose estimation. *International Journal on Real Time Imaging, Special Issue on Real-time Visual Monitoring and Inspection.*, 4:333–347.

Mester, R. (2000). Orientation estimation: conventional techniques and a new non-differential method. In *European Signal Processing Conference*, Tampere, Finland.

Metta, G. (2000). *Babybot: a study into sensorimotor development.* PhD thesis, LIRA-Lab, DIST.

Metta, G., Sandini, G., and Konczak, J. (1999). A developmental approach to visually-guided reaching in artificial systems. *Neural Networks*, 12:1413–1427.

Minsky, M. (1985). *The Society of Mind.* Simon and Schuster, New York.

Minsky, M. (1990). Logical vs. analogical or symbolic vs. connectionist or neat vs. scruffy. In Winston, P. H., editor, *Artificial Intelligence at MIT, Expanding Frontiers*, volume 1. MIT Press. Reprinted in AI Magazine, 1991.

Mitchell, T. (1980). The need for biases in learning generalizations. Technical report, Computer Science Department, Rutgers University.

Moll, M. and Erdmann, M. A. (2001). Reconstructing shape from motion using tactile sensors. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Maui, HI.

Mou, X. and Zue, V. (2001). Sub-lexical modelling using a finite state transducer framework. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, Utah.

Münch, S., Kreuziger, J., Kaiser, M., and Dillmann, R. (1994). Robot programming by demonstration (RPD) – using machine learning and user interaction methods for the development of easy and comfortable robot programming systems. In *Proceedings of the 24th International Symposium on Industrial Robots.*

Murata, A., Gallese, V., Luppino, G., Kaseda, M., and Sakata, H. (2000). Selectivity for the shape, size, and orientation of objects for grasping in neurons of monkey parietal area AIP. *Journal of Neurophysiology*, 83(5):2580–2601.

Murphy, K. P. (1996). Passively learning finite automata. Technical report, Santa Fe Institute.

Nayar, S. K., Nene, S. A., and Murase, H. (1996). Real-time 100 object recognition system. In *Proceedings of IEEE International Conference on Robotics and Automation.*

Needham, A. (2001). Object recognition and object segregation in 4.5-month-old infants. *Journal of Experimental Child Psychology*, 78(1):3–22.

Needham, A. and Baillargeon, R. (1997). Object segregation in 8-month-old infants. *Cognition*, 62(2):121–159.

Nelson, R. C. and Selinger, A. (1998). A cubist approach to object recognition. In *Proceedings of the International Conference on Computer Vision*, pages 614–621.

Nicolescu, M. and Mataric, M. J. (2001). Experience-based learning of task representations from human-robot interaction. In *IEEE International Symposium on Computational Intelligence in Robotics and Automation*, pages 463–468, Alberta, Canada.

Nothdurft, H. C. (1993). The role of features in preattentive vision: Comparison of orientation, motion and color cues. *Vision Research*, 33:1937–1958.

Oates, T. (1999). Identifying distinctive subsequences in multivariate time series by clustering. In *Knowledge Discovery and Data Mining*, pages 322–326.

Oates, T., Eyler-Walker, Z., and Cohen, P. (2000). Toward natural language interfaces for robotic agents: Grounding linguistic meaning in sensors. In *Proceedings of the 4th International Conference on Autonomous Agents*, pages 227–228.

Oates, T., Jensen, D., and Cohen, P. (1998). Discovering rules for clustering and predicting asynchronous events. In *Predicting the Future: AI Approaches to Time-Series Problems*, pages 73–79. AAAI Press.

Oztop, E. (2002). *Modeling the Mirror: Grasp Learning and Action Recognition.* PhD thesis, University of Southern California.

Oztop, E. and Arbib, M. A. (2002). Schema design and implementation of the grasp-related mirror neuron system. *Biological Cybernetics*, 87(2):116–140.

Papert, S. (1966). The summer vision project. Memo AIM-100, MIT AI Lab.

Paulos, E. (1999). Fast construction of near optimal probing strategies. Master's thesis, University of California, Berkeley.

Pelz, J. B. (1995). *Visual Representations in a Natural Visuo-motor Task.* PhD thesis, Brain and Cognitive Science, University of Rochester.

Pepperberg, I. (1990). Referential mapping: A technique for attaching functional significance to the innovative utterances of an african grey parrot. *Applied Psycholinguistics*, 11:23–44.

Perona, P. and Malik, J. (1990). Detecting and localizing edges composed of steps, peaks and roofs. In *Proceedings of the Third International Conference of Computer Vision*, pages 52–57, Osaka, Japan.

Pilu, M., Fitzgibbon, A., and Fisher, R. (1996). Ellipse-specific direct least-square fitting. In *IEEE International Conference on Image Processing*, Lausanne.

Pratchett, T. (1986). *The Light Fantastic.* St. Martin's.

Pratchett, T. (1989). *Diggers.* Transworld Publishers.

Pratchett, T. (1990). *Moving Pictures.* Gollancz.

Pratchett, T. (1991a). *Reaper Man.* Gollancz.

Pratchett, T. (1991b). *Witches Abroad.* Gollancz.

Pratchett, T. (1992a). *Only You Can Save Mankind.* Doubleday.

Pratchett, T. (1992b). *Small Gods.* Gollancz.

Pratchett, T. (1996). *Hogfather.* Gollancz.

Pratchett, T. (1999). *The Fifth Elephant.* Doubleday.

Quine, W. V. O. (1960). *Word and object.* Harvard University Press, Cambridge, Massachusetts.

Rahimi, A., Morency, L.-P., and Darrell, T. (2001). Reducing drift in parametric motion tracking. In *roceedings of the International Conference on Computer Vision.*

Rizzolatti, G. and Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, 21:188–194.

Ross, M. G. (2000). Exploiting texture-motion duality in optical flow and image segmentation. Master's thesis, Massachusetts Institute of Technology.

Roy, D. (1999). *Learning Words from Sights and Sounds: A Computational Model.* PhD thesis, MIT.

Roy, D., Gorniak, P., Mukherjee, N., and Juster, J. (2002). A trainable spoken language understanding system for visual object selection. In *Proceedings of the International Conference of Spoken Language Processing.*

Roy, D., Hsiao, K., and Mavridis, N. (2003). Conversational robots: Building blocks for grounding word meanings. submitted to HLT-NAACL03 Workshop on Learning Word Meaning from Non-Linguistic Data.

Roy, D. and Pentland, A. (2002). Learning words from sights and sounds: A computational model. *Cognitive Science*, 26(1):113–146.

Sandini, G., Gandolfo, F., Grosso, E., and Tistarelli, M. (1993). Vision during action. In Aloimonos, Y., editor, *Active Perception*, pages 151–190. Lawrence Erlbaum Associates, Hillsdale, NJ.

Scassellati, B. (1998). A binocular, foveated active vision system. Technical Report 1628, MIT Artificial Intelligence Lab Memo.

Scassellati, B. (1999). Imitation and mechanisms of joint attention: A developmental structure for building social skills on a humanoid robot. In Nehaniv, C. L., editor, *Computation for Metaphors, Analogy and Agents*, volume 1562 of *Springer Lecture Notes in Artificial Intelligence.* Springer-Verlag.

Scassellati, B. (2000). Theory of mind for a humanoid robot. In *Proceedings of the First International IEEE/RSJ Conference on Humanoid Robotics.*

Scassellati, B. (2001). *Foundations for a Theory of Mind for a Humanoid Robot.* PhD thesis, MIT Department of Electrical Engineering and Computer Science.

Schneiderman, H. and Kanade, T. (2002). Object detection using the statistics of parts. *International Journal of Computer Vision.*

Schwartz, E. L., Greve, D. N., and Bonmassar, G. (1995). Space-variant active vision: Definition, overview and examples. *Neural Networks*, 8(7-8):1297–1308.

Selinger, A. (2001). *Analysis and Applications of Feature-Based Object Recognition.* PhD thesis, University of Rochester, Rochester, New York.

Sherrah, J. and Gong, S. (2000). Fusion of perceptual cues for robust tracking of head pose and position. In *Pattern Recognition*, volume 34.

Shi, J. and Tomasi, C. (1994). Good features to track. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 593 – 600.

Sigal, L., Sclaroff, S., and Athitsos, V. (2000). Estimation and prediction of evolving color distributions for skin segmentation under varying illumination. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 152–159.

Sinha, P. (1994). Object recognition via image invariants: A case study. *Investigative Ophthalmology and Visual Science*, 35:1735–1740.

Smith-Mickelson, J. (2000). Design and application of a head detection and tracking system. Master's thesis, MIT.

Souvignier, B., Kellner, A., Rueber, B., Schramm, H., and Seide, F. (2000). The thoughtful elephant: strategies for spoken dialog systems. *IEEE Transactions on Speech and Audio Processing*, 8(1):51–62.

Sperry, R. W. (1952). Neurology and the mind-brain problem. *American Scientist*, 40:291–312.

Stauffer, C. (1999). Adaptive background mixture models for real-time tracking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 246–252.

Stauffer, C. and Grimson, W. E. L. (1999). Adaptive background mixture models for real-time tracking. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Fort Colins, CO.

Steels, L. (1996). Emergent adaptive lexicons. In *Proceedings of the fourth international conference on simulation of adaptive behavior*, pages 562–567, Cape Cod, MA.

Strom, J., Jebara, T., Basu, S., and Pentland, A. (1999). Real time tracking and modeling of faces: An EKF-based analysis by synthesis approach. In *Proceedings of the Modelling People Workshop, in the IEEE International Conference on Computer Vision*.

Swain, M. J. and Ballard, D. H. (1991). Color indexing. *International Journal of Computer Vision*, 7(1).

Swerts, M., Hirschberg, J., and Litman, D. (2000). Corrections in spoken dialogue systems. In *Proceedings of the International Conference on Spoken Language Processing*, Beijing, China.

Tarr, M. and Black, M. (1994). A computational and evolutionary perspective on the role of representation in vision. *CVGIP: Image Understanding*, 60(1):65–73.

Thrun, S. (1998). A framework for programming embedded systems: Initial design and results. Technical Report CMU-CS-98-142, Carnegie Mellon University.

Tomasello, M. (1997). The pragmatics of word learning. *Japanese Journal of Cognitive Science*, 4:59–74.

Trevarthen, C. (1979). Communication and cooperation in early infancy: a description of primary intersubjectivity. In Bullowa, M., editor, *Before Speech: The beginning of interpersonal communication*. Cambridge University Press.

Ullman, S. (1984). Visual routines. *Cognition*, 18:97–159. (Also in: Visual Cognition, S. Pinker ed., 1985).

Vaidya, P. (1988). Geometry helps in matching. In *Proceedings of the twentieth annual ACM symposium on theory of computing*, pages 422–425. ACM Press.

Varchavskaia, P., Fitzpatrick, P., and Breazeal, C. (2001). Characterizing and processing robot-directed speech. In *Proceedings of the International IEEE/RSJ Conference on Humanoid Robotics*, Tokyo, Japan.

Viola, P. and Jones, M. (2001). Robust real-time object detection. Technical report, COMPAQ Cambridge Research Laboratory, Cambridge, MA.

Vygotsky, L. (1962). *Thought and language*. MIT Press, Cambridge, MA.

Wang, C. and Brandstein, M. (1998). A hybrid real time face tracking system. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Seattle.

Wang, X. (1996). Planning while learning operators. In Drabble, B., editor, *Proceedings of the Third International Conference on Artificial Intelligence Planning Systems*, pages 229–236. AAAI Press.

Weng, J., Hwang, W. S., Zhang, Y., Yang, C., and Smith, R. (2000a). Developmental humanoids: Humanoids that develop skills automatically. In *Proceedings of the first IEEE-RAS International Conference on Humanoid Robots*, Cambridge, MA.

Weng, J., McClelland, J., Pentland, A., Sporns, O., Stockman, I., Sur, M., and Thelen, E. (2000b). Autonomous mental development by robots and animals. *Science*, 291(5504):599–600.

Werker, J., Lloyd, V., Pegg, J., and Polka, L. (1996). Putting the baby in the bootstraps: Toward a more complete understanding of the role of the input in infant speech processing. In Morgan, J. and Demuth, K., editors, *Signal to Syntax: Bootstrapping From Speech to Grammar in Early Acquisition*, pages 427–447. Lawrence Erlbaum Associates: Mahwah, NJ.

Williamson, M. (1995). Series elastic actuators. Master's thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

Williamson, M. (1998a). Neural control of rhythmic arm movements. *Neural Networks*, 11(7-8):1379–1394.

Williamson, M. (1999). *Robot Arm Control Exploiting Natural Dynamics*. PhD thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

Williamson, M. M. (1996). Postural primitives: Interactive behavior for a humanoid robot arm. In *Fourth International Conference on Simulation of Adaptive Behavior*, pages 124–131, Cape Cod, Massachusetts.

Williamson, M. M. (1998b). Exploiting natural dynamics in robot control. In *Fourteenth European Meeting on Cybernetics and Systems Research (EMCSR '98)*, Vienna, Austria.

Winston, P. (1970). Learning structure descriptions from examples. In *The Psychology of Computer Vision*, pages 157–209. McGraw-Hill, New York.

Wolfe, J. M., Oliva, A., Horowitz, T. S., Butcher, S. J., and Bompas, A. (2002). Segmentation of objects from backgrounds in visual search tasks. *Vision Research*, 42(28):2985–3004.

Wolfson, H. and Rigoutsos, I. (1997). Geometric hashing: an overview. *IEEE Computational Science and Engineering*, 4:10–21.

Wong, D. and Flynn, M. (1992). Fast division using accurate quotient approximations to reduce the number of iterations. *IEEE Transactions on Computers*, 41(8):981–995.

Wren, C., Azarbayejani, A., Darrell, T., and Pentland, P. (1997). Pfinder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Y. Wu, K. T. and Huang, T. S. (2000). Wide-range, person- and illumination-insensitive head orientation estimation. In *Proceedings of the International Conference on Face and Gesture Recognition*, Grenoble, France.

Yip, K. and Sussman, G. J. (1997). Sparse representations for fast, one-shot learning. In *Proceedings of the National Conference on Artificial Intelligence*.

Ziou, D. and Tabbone, S. (1997). Edge detection techniques- an overview. Technical report, Dept Math & Informatique. Universit de Sherbrooke, 1997.

Zue, V. and Glass, J. (2000). Conversational interfaces: Advances and challenges. *Proceedings of the IEEE, Special Issue on Spoken Language Processing*, Vol. 88.

Zue, V., Glass, J., Plifroni, J., Pao, C., and Hazen, T. (2000). Jupiter: A telephone-based conversation interface for weather information. *IEEE Transactions on Speech and Audio Processing*, 8:100–112.