Andrew H. Fagg
Research Statement

My research centers around the symbiotic relationships between humans and machines. I study computing machines as models of representation and learning of motor and cognitive skills in biological systems; primates as inspiration for new robot control and learning techniques; and the interaction of humans with machines. Central to all of these problems are the issues of constructing rich representations of the state of the agent, the local environment, the task, and the skills; and of using various forms of available training information to refine these representations. In this study of *symbiotic computing*, I draw from the disciplines of robotics, artificial intelligence, machine learning, computational neuroscience, and interactive art.

Grasping and Manipulation in Robots and Primates

Manipulating one's world in very flexible ways is a skill that is shared only by a small number of species. Humans are particularly capable of applying their manipulation abilities in novel situations using a range of effectors from hands and other parts of the body to tools. The fact that grasping and manipulation comes so naturally might suggest a degree of triviality to the problem. However, a substantial amount of the primate neocortex is dedicated to creating the representations to support these activities and to subsequently turning these representations into robust action.

My current work is focused in part on the problem of how robots organize and learn knowledge representations for solving grasping and manipulation problems in unstructured environments. J. J. Gibson (1966, 1977) suggests that these representations should be partitioned into *what* can be done with a particular object and *why* the object should be manipulated in a certain way. The first of these, which Gibson terms *object affordances*, captures the details of what can be done with the object by the agent. The second captures information about how these individual manipulation skills are to be put together in order to solve a specific task. The task-neutral affordance representation is important in that it can provide an agent with a menu of actions/activities that are possible with a given object – whether the current task is well known to the agent or is a new one.

For a given object, how might an agent learn grasp affordance representations? Ultimately, one must establish a mapping from perceivable visual and haptic features to a set of parameterized grasping actions (specific positions and orientations for the hand, as well as configurations for the fingers) that are expected to be successful if executed. My group's current approach factors the grasp affordance problem into two components: one that first attempts to visually recognize the identity and pose of an object, and the second that identifies detailed descriptions of legal hand configurations for grasping that object.

Learning Visual Models for Object Pose Recognition. Inspired by the challenges that human infants must face in learning about objects, we take an interactive approach to the problem of constructing visual models of objects. A camera observes an object as it is manipulated through a range of viewing angles. The challenge is to discover visual operators that can be used to robustly identify the angle from which the object is being viewed. Significant complications arise from the fact that parts of the object may be occluded during the manipulation sequence and that some images may be dominated by extraneous lighting effects.

My student, Di Wang, and I approach this problem by first sampling a visual feature that recognizes the appearance of a component of the object, and then constructing a probabilistic model that describes the viewing angles over which the feature is observable (Wang, 2007). We sample a set of these features in order to represent the appearance of the object from all possible viewing angles. Given a novel image of an object and a set of features that are observable in the image, we then compute the likelihood of possible viewing angles. Having this estimate will ultimately allow a robot to plan reach-to-grasp actions for the object.

Learning Canonical Hand Configurations for Grasping Objects. Once an object has been visually identified and localized in space, how can an agent describe the set of possible grasping actions that may be taken? Because the set of possible actions will be ultimately used for planning and for exploratory learning, we are motivated to make this set as small as possible. My former student, Charles de Granville, and I developed an approach to constructing this representation from a large set of successful example grasps. This set of examples could be derived from manipulation sequences produced by the robot itself, or by a human acting directly on the manipulated object or acting through the agent via teleoperation. Each example grasp is described using the position and orientation of the hand relative to the object and the joint angles of the fingers of the hand. We cluster these examples using a *mixture of probability density functions* approach, in which each cluster corresponds to a canonical hand configuration that describes all three of these components (de Granville et al., 2006, 2008, and submitted; de Granville, 2008).

Key to this approach is our use of probability density functions (PDFs) defined over the unit quaternion space to capture orientations in three dimensions. The Dimroth-Watson distribution allows us to describe "mean" orientations and variations around this mean. We also employ a generalization of this distribution that allows us to compactly describe all possible orientations of the hand when grasping, for example, a cup from the side. Experimental results demonstrate that our clustering approach reliably identifies a small, intuitive set of canonical grasps. Coupled with knowledge of an object's pose, each of these contains sufficient detail to enable the robot to position the hand near a valid grasp configuration.

Learning Spatiotemporal Representations of Task. In addition to providing a menu of possible actions, the affordance representation can be used to recognize the actions of humans as they are made with respect to the objects in the environment. We have demonstrated that such an affordance approach can be used to "parse" a sequence of arm and hand movements as a teacher demonstrates a set of pick-and-place actions with a group of objects (Fagg et al., 2004; Brock et al., 2005). This approach provides a means of teaching the robot a rote sequence of actions that can be replayed under similar conditions.

We are working to develop a more general approach to representing tasks. Given a *set* of rote sequences corresponding to a single task, we would like to identify a compact description of state that captures 1) the important properties of objects that enable them to play particular roles in the task and 2) the specific spatial configurations of these objects. Because there exists a degree of variation of object and spatial properties across examples, a natural way to represent state is in terms of joint PDFs over a set of properties. In particular, the relative position and orientation PDFs that we developed for the representation of affordances are also applicable here.

Intelligent Prosthetics and Brain-Machine Interfaces

Prosthetic limbs have been used for many centuries. Despite the recent advances in this area, current devices fall short in significant ways. First, these devices often rely on the patient to use unnatural channels of control to command the limb (e.g., activation of unrelated muscles). Second, the patient largely relies on visual feedback to confirm that the movements are being executed as intended. Taken together, functional prosthetic arms/hands require a substantial amount of practice to achieve a high level of skill. In contrast, we want "intelligent" prosthetic limbs to make use of existing information to make inferences about the patient's intended movement and to translate this intent into action.

One source of information about the intent of action is the brain. In collaboration with Dr. Nicholas Hatsopoulos (University of Chicago) and Dr. Lee Miller (Northwestern University), my group has been working on the problem of predicting (or *decoding*) "intended" arm motion from the activity of the primary motor cortex, and then using this prediction to control a prosthetic arm. Two dominant approaches to modeling this transformation from neural activity to arm motion include Kalman filters (e.g., Wu et al., 2002) and tapped delay line (or Wiener) filters (e.g., Warland et al., 1997). There exist discrepancies in the literature as to which of these approaches leads to high quality and robust prediction. My former student, Shamim Nemati, and I explain these differences in terms of the amount of information that is available to construct the transformations (Nemati et al. 2007, and in revision). In short, the Wiener filter models tend to overfit small training data sets, but can outperform Kalman filter models when larger training sets are available. In addition, we have shown that overfitting of the Wiener filter can be addressed by including both prediction error and functional smoothness components to the cost function (i.e., ridge regression, Björck 1996). The result is a decoder that can perform at least as well as both the Kalman and traditional Wiener filters, no matter the training set size. When used for prosthetic arm control, this approach also yields decoders that produce much smoother movements than the traditional solution to the Wiener filter.

BMIs for Dynamic Control. The traditional approach to limb control via brain-machine interface is to infer the intended position or velocity of the hand (Wessberg et al., 2000; Taylor et al., 2002). We have shown that torques being produced at the limb joints can be predicted at an accuracy that is comparable to that of hand position and competitive with that of velocity (Miller et al., 2007; Fagg et al., submitted) . This result has important implications for enabling patients to not only position prosthetic limbs in appropriate configurations, but to also apply forces to the world around them. Hence, this approach could extend the types of tasks that patients will be able to perform.

Toward Complete BMIs. In continuing work, my group is planning several directions. First, we are working to address the problem of how a neural decoder can automatically calibrate itself as the patient wears and works with her prosthetic limb. Second, we are planning to examine the activity of neurons within higher-order motor areas for the purposes of inferring the intended movement. These areas can give us hints about the coming movements before they are executed and can give us details about how a prosthetic hand might be shaped in order to grasp an object. Third, we plan to develop techniques that leverage other contextual cues in the interpretation of neural activity. For example, gaze direction can give hints as to the possible targets for a reach, while acceleration of the torso and shoulders could indicate when a movement should be initiated. Furthermore, knowledge about the objects sitting in front of the patient and about the type of task in which the patient is currently engaged can further reduce the number of possible reach targets and even suggest what will be done with the object once it is grasped. This work will make use of the affordance and task models developed in our robotics work.

Interactive Art

In addition to my work with robot and primate grasping and manipulation, I am interested in computing machines that can make connections with humans on an intellectual or social level. This interest has played out in the area of interactive art. I am motivated to create interactive art pieces that 1) provide personal experiences for the visitors, and 2) motivate the visitor to have extended interactions with the piece. Although one aspect of approaching these challenges is to alter the behavior of the art piece as a function of time, it is not sufficient to simply rely on randomly-generated behavior. Instead, we strive to develop behavioral techniques that enable a visitor to easily grasp the surface structure of the behavior, but also make it challenging for the visitor to tease out the details.

Along with my collaborator, Adam Brown, I have been working on a number of interactive art pieces. Our first piece, *Bion*, is a 1000-node sensor network of sculpted forms. Each node is capable of sensing and reacting to the presence of a visitor, and then communicating the event to the rest of the network. The behavior of the network is subject to the dynamics of a chaotic attractor. The effect is that the network will always respond to a visitor in a way that exhibits a certain degree of structure. However, there can be a dramatic variation from one encounter to the next. In practice, this chaotic behavior exudes a degree of playfulness that almost invites visitors to interact for extended periods of time.

We are currently developing a new piece, *Orgonome*, that will consist of a small set of three-legged walking creatures. Their behavior is sculpted using *reinforcement learning* based on interaction with visitors. The explicit goal of the learning process is to develop behavior that will attract and maintain the attention of future visitors.

References

Björck, A. (1996). Numerical methods for least squares problems. *Society for Industrial and Applied Mathematics*, page 49.

Brock, O., Fagg, A. H., Grupen, R. A., Karuppiah, D., Platt, R., and Rosenstein, M. (2005). A framework for humanoid control and intelligence. *International Journal of Humanoid Robotics*, 2(3):301–336.

de Granville, C. (2008). Learning grasp affordances. Master's thesis, School of Computer Science, University of Oklahoma, Norman, OK.

de Granville, C. and Fagg, A. H. (submitted). Learning grasp affordances through human demonstration.

de Granville, C., Southerland, J., and Fagg, A. H. (2006). Learning grasp affordances through human demonstration. In *Proceedings of the International Conference on Development and Learning*. electronically published.

de Granville, C., Wang, D., Southerland, J., and Fagg, A. H. (2009). Grasping affordances: Learning to connect vision to hand action. In Sukhatme, G., editor, *to appear*.

Fagg, A. H., Hatsopoulos, N. G., de Lafuente, V., Moxon, K. A., Nemati, S., Rebesco, J. M., Romo, R., Solla, S. A., Reimer, J., Tkach, D., Pohlmeyer, E. A., and E., M. L. (2007). Biomimetic brain machine interfaces for the control of movement. *Journal of Neuroscience*, 27(44):11842–11846.

Fagg, A. H., Ojakangas, G., Miller, L. E., and Hatsopoulos, N. G. (submitted). Kinetic trajectory decoding using motor cortical ensembles.

Fagg, A. H., Rosenstein, M. T., Platt, Jr., R., and Grupen, R. A. (2004). Extracting user intent in mixed initiative teleoperator control. In *Proceedings of the American Institute of Aeronautics and Astronautics Intelligent Systems Technical Conference*. Electronically published.

Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. Allen and Unwin.

Gibson, J. J. (1977). The theory of affordances. In Shaw, R. E. and Bransford, J., editors, *Perceiving, Acting, and Knowing*. Lawrence Erlbaum, Hillsdale.

Nemati, S., Fagg, A. H., Hatsopoulos, N. G., and Miller, L. E. (2007). A comparison of linear and Kalman filter models for arm motion prediction. In *Proceedings of the Spring Meeting on the Neural Control of Movement*. electronically published.

Nemati, S., Hatsopoulos, N. G., Miller, L. E., and Fagg, A. H. (submitted). Constructing robust neural decoders using limited training data.

Taylor, D. M., Tillery, S. I., and B., S. A. (2002). Direct cortical control of 3D neuroprosthetic devices. *Science*, 296:1829–1832.

Wang, D. (2007). A 3D feature-based object recognition system for grasping. Master's thesis, School of Computer Science, University of Oklahoma, Norman, OK.

Warland, D., Reinagel, P., and Meister, M. (1997). Decoding visual information from a population of retinal ganglion cells. *Journal of Neurophysiology*, 78(5):2336–2350.

Wessberg, J., Stambaugh, C. R., Kralik, J. D., Beck, P. D., Laubach, M., Chapin, J. K., Kim, J., Biggs, S. J., Srinivasan, M. A., and Nicolelis, M. A. (2000). Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature*, 408:361–365.

Wu, W., Black, M., Gao, Y., Bienenstock, E., Serruya, M., and Donoghue, J. (2002). Inferring hand motion from multi-cell recordings in motor cortex using a Kalman filter. *SAB - Workshop on Motor Control in Humans and Robots: On the Interplay of Real Brains and Artificial Devices, Edinburgh, Scotland (UK)*, pages 66–73.