From Individuals to Populations: Biologically-Informed Multi-Modal Situation Understanding with Sensor Networks.

Proposal submitted in response to BAA 06-028, Topic 3: Disparate Sensor Network Based Situation Understanding

Principal Investigator: **Andrew Y. Ng**. Stanford University, Computer Science Department, Stanford, CA 94305-9010. (650)387-9561. ang@cs.stanford.edu

Other Universities: Carnegie Mellon University

Agency to which proposal is submitted: Office of Naval Research (ONR)

Funds requested: 3-year base: $2,997,972. 2-year option: $1,998,939. 5-year total: $4,996,911.

Technical Personnel: **Carlos Guestrin, Takeo Kanade, Michael Lewicki** and **José Moura** (CMU); **Leonidas Guibas, Daphne Koller, Andrew Y. Ng** and **Sebastian Thrun** (Stanford University).

Project summary: The demands of the GWOT require that we develop a new level of situation understanding, ***combining a large amount of information from disparate sources into global, high-level conclusions in dynamic environments***. These multi-modal sources include recently developed low-cost sensing modalities, such as images, audio, FLIR, LADAR, and also text logs, intelligence reports, and real-time, on-site warfighter observations. The integration of these sources to obtain high-level conclusions challenges us to develop novel methods and algorithms for representing and analyzing this information.

To accomplish this, we propose to develop a ***biologically-informed unified mathematical foundation for multi-modal scene understanding***. Biological organisms effortlessly use multiple sensors such as vision, audio and touch to perceive the world, and are able to discover associations between these very different input modalities as well as identify the most salient features in each. Our integrated research program comprises four tasks: (i) ***Neo-cortex informed algorithms for automatically learning succinct representations*** for any sensor or set of sensors. (ii) ***Probabilistic information fusion***, where we pervasively identify hierarchies/grouping of elements to automatically extract symbolic descriptions of a scene, such as "paths," "actions," "behaviors," and "intentions" of elements. We will also use this to generate ***human-understandable multimedia scene summary documents***. (iii) ***Robust, distributed probabilistic inference algorithms***, with ***Query Anywhere*** capability, to allow a distributed network of warfighters to efficiently access both local and global information about a scene. (iv) Testbed evaluation of all of these integrated ideas on ***two real, not laboratory, sensornet deployments*** in CMU (wide-area deployment) and Stanford (surveillance of a large, crowded indoor space) that mimic future DoD applications.

Our high-level scene understanding methods will provide the foundation for ***more effective surveillance for the GWOT***. In the medium term, it will be directly applicable to problems such as airport surveillance and border security, where we must fuse multiple sensing modalities, including cameras, microphones, seismic sensors and RADAR, to not only detect and individuals, but also reason about groups of individuals and their intentions, and detect anomalous behaviors. We also expect our approach will be useful for current deployments, e.g., the *Command Post of the Future*, now in use in Iraq for combining multiple sources of information, but requiring significant user intervention. Our proposed methods will significantly decrease the amount of user intervention, identify hazards more quickly, and improve the efficacy of decision-making at all levels.