Next Generation Bio-Molecular Imaging and Information Processing

Manjunath, B.S.*, Singh, A.K.*, Rose, K., Wilson, L., Fisher, S.
University of California at Santa Barbara, Santa Barbara, CA, USA

Continuing technological advances in fluorescence and atomic-force microscopy allow scientists to observe molecular function, distribution, and interrelationships in living cells. However, a full understanding of tens of thousands of proteins and the complex molecular processes they engage in requires a voluminous amount of image data that currently must be analyzed by visual inspection. To provide the new imaging and information processing technologies that ultimately will significantly advance scientific understanding of how cells respond to stress, injury, aging, and disease, this project marshals the expertise of leaders in biological science, computer science, and engineering from UC Santa Barbara, UC Berkeley, Carnegie Mellon University, and MIT.

Researchers from the four participating institutions will focus on three main research thrusts, which will ultimately revolutionize not only the information processing technologies but also progress in basic scientific research. First, **next-generation intelligent imaging** will focus on information processing at the sensor level to enable high speed and super-resolution imaging. The end result will transform nanoscale imaging using scanning probe microscopes, enabling biologists to study cellular processes at resolutions in time and space that are not possible with current technologies. Building on their expertise in microscopy and signal and image processing, the researchers will develop adaptive image acquisition methods for fluorescence microscopy that will provide higher spatial and temporal resolution by focusing on regions and time intervals necessary to obtain accurate statistical models of protein distributions in living cells.

The second research thrust is **pattern recognition and data mining**, which will focus on new tools for information processing. Salient features that characterize the underlying patterns in cells and tissues need to be computed for the vast volumes of images acquired through automated microscopy. The outcome of this research will lead to efficient representations of image and video features, researchers being able to search and browse through large collections of image and video data and look for similar patterns in such datasets, thus facilitating information discovery.

Third, this project will establish a **distributed interactive digital library for bio-molecular images.** Pattern recognition and data mining tools, together with new powerful methods for indexing, data modeling, and collaboration will come together in creating a unique infrastructure that will greatly facilitate image bioinformatics, thus complementing recent revolutionary advances in genomics. The impact of such an infrastructure will be through *large-scale biology* in which the results of a single experiment can be globally correlated with the results from other groups of scientists, thus accelerating discovery by narrowing the set of scientific hypothesis to be explored.

This project will foster the essential dialogue between biologists, computer scientists and engineers that is needed to discover the dynamic relationships between structure and function in complex biological systems. Unraveling the mysteries of such systems (e.g., the central nervous system) will benefit society through the development of treatments for diseases such as Alzheimer's disease, Parkinson's disease, age related blindness, stroke and other injuries to the central nervous system.

The project has been funded (beginning 2003) by a large ITR (Information Technology Research) award from the National Science Foundation.