2014 Summer Undergraduate Research Experience
Project Showcase

13 August 2014
University of California, Santa Barbara
Harold Frank Hall, Room 1132

Center for Bio-Image Informatics

Funded by the National Science Foundation (IIS-0808772), with additional research support from the U.S. Department of the Navy (N00014-12-1-0503)

PI: Prof. B.S. Manjunath, Electrical and Computer Engineering

Network Science / Network Analysis and Synthesis

Funded by two grants from the National Science Foundation: an IGERT in Network Science (DGE-1258507), and a project on the querying and mining of dynamic graphs (IIS-1219254).

PI: Prof. Ambuj Singh, Computer Science

Above: Interns, faculty, and graduate student mentors from the 2014 Summer Research Experience. With thanks to Prof. Arturo Concepcion and the California State University, San Bernardino, for their continued support of this program.
Agenda:

Wednesday, August 13

9:30 am Welcome and Introductions
  Program overview and presentation procedures

9:45 am Student Presentations
  
  Open Source Hardware-Software Interface for a Pressure Mat
  Sebastian Rojas, Worcester Polytechnic Institute

  Segmentation of Light Microscopy Images
  Marsha Perez, CSU San Bernardino

  Object Recognition Online Demo
  Alex Krause, CSU San Bernardino

  Unconstrained Activity Recognition in an Office Environment
  Christopher Ray Ramirez and Parker Sankey, CSU San Bernardino

  A Comparison of Methods for Network Summarization
  Ali Hajimirza, University of Oklahoma, Norman
  Jason White, CSU San Bernardino

  Interpreting Ecological Network Data
  Sean Nguyen, UC Santa Barbara

  Multiscale Modeling of Biological Networks
  Kara Goodman, CSU San Bernardino
  Austen Piers, UC Santa Barbara

  Tweet Classification
  Yuanlin Xu, CSU Los Angeles

  Classification of Traumatic Brain Injury Using Diffusion Tractography
  Haraldur Tómas, UC Santa Barbara

12:30 pm Lunch and Internal Poster Presentations

Thursday, August 14 – Poster Sessions in the Bren Courtyard

  -- Presenters please arrive at least 30 minutes early to set up --

12:30 pm Poster Session 1 (Sebastian, Marsha, Alex, Parker and Chris)
2:00 pm Refreshments served
2:30 pm Poster Session (Ali and Jason, Kara and Austen, Sean, and Yuanlin)
4:00 pm Everyone helps clean up.
Project Abstracts:

Open Source Hardware-Software Interface for a Pressure Mat

Sebastian Rojas, ECE & Biomedical Engineering, Worcester Polytechnic Institute
Archith Bency, Prof. B.S. Manjunath, Department of ECE

Pressure mats are useful sensors in a variety of applications that range from screening for high pressure body sores and pressure wounds to monitoring for sleep therapy. The pressure mats themselves are affordable, but the hardware and software interfaces required to extract data are expensive, and their proprietary nature impedes improvements. The main objective of this project is to develop an open-source alternative that works with the Beaglebone Black, a widely available single-board computer that runs Linux. The developed interface visualizes the mat measurements as a grayscale image. This alternative will provide an important impetus for a wider adoption of pressure mats in research to solve healthcare and elderly care problems.

Segmentation of Light Microscopy Images

Marsha Perez, Bioinformatics, CSU San Bernardino
Renuka Shenoy, Prof. B.S. Manjunath, Prof. Kenneth Rose, Department of ECE

Due to the large volume of data present in many biological datasets, manual annotation is often impractical. This project focuses on accurate semi-automated segmentation of cells in light microscopy images. The data consists of six consecutive sections from the RC1 connectome, a volume of rabbit retinal data from a 0.25 mm diameter section imaged at 70 nm resolution using the Computational Molecular Phenotyping (CMP) paradigm. Each section is probed using a unique molecular marker, resulting in an intensity image indicating protein activity. The six images together give information on the different cell types and are all used in our formulation. The images are downsampled from their original size (65536-by-65536 pixels) to 2048-by-2048 pixels for ease of computation.

We preprocessed the images using median filtering to denoise the data. Standard k-means with simple manual initialization was used to obtain clusters corresponding to each cell type. The result from k-means suffered from both over- and under-segmentation in different regions. Different strategies were used to address these problems. We constructed a region adjacency graph to indicate neighboring cell segments. Similar adjacent cell segments were combined based on both size- and shape-based merging criteria. Further, we used marker-based watershed along with the distance transform to separate clumped cells.

We manually annotated ground truth for cell boundaries using labels obtained from the proprietary Viking software. When validating our results against the ground truth, we observed accurate segmentation of cells.
Object Recognition Online Demo

Alex Krause, Computer Science and Computer Engineering, CSU San Bernardino
Niloufar Pourian, Prof. B.S. Manjunath, Department of ECE

Object recognition has been a topic of intense research for many years. The majority of the existing work focuses on global image representation. New methods are used to capture spatial information into visual features to better represent an image/object. The purpose of this project is to provide an online tool for visualization of these new features using the BISQUE platform (an environment for handling and analyzing images). This application allows users to select an image from a database and provides corresponding segmented regions with an overlay outlined by vertices and localized features mapped to those segments. This program uses these new features to retrieve database images with the highest visual similarity scores. This enables one to better understand the new visual features and leads to feature enhancement. Improved localized visual features benefit many computer vision applications such as image classification, search and retrieval, and activity recognition.

Unconstrained Activity Recognition in an Office Environment

Christopher Ray Ramirez and Parker Sankey, Computer Engineering, CSU San Bernardino
Amir M. Rahimi, Prof. B.S. Manjunath, Department of ECE

In today’s world almost all surveillance cameras existing today are pan-tilt-zoom (PTZ) cameras. We present a systematic approach to model human activities in Vision Research Lab located in Harold Frank Hall at the University of California Santa Barbara. Our setup includes two PTZ network cameras that track faces using trained Haar features, and recognize people using linear discriminant analysis (LDA) with Fisherface; LDA maximizes the separation of features of similar objects, hence, it can distinguish differences between each person. The PTZ cameras follows the subject on-screen using CamShift and updates the pan, tilt, and zoom parameters in real-time with the objective of keeping the person in the center of the screen. Our office dataset includes 863 samples covering five actions. Our action bank includes interaction between two or more people, walking, sitting, writing on a whiteboard, and getting coffee from two different viewpoints. We then extract DenseTrack features that contain Histogram of Oriented Gradients (HOG), Histogram of Optical Flow (HOF), Motion Boundary Histogram (MBH), and Trajectory. In order to validate our recognition we use support vector machine (SVM) to train and test DenseTrack features. Current literature has not address activity recognition with mobile cameras due to challenging nature of camera movements. Despite the challenge, our work is a good estimate of unconstrained activity recognition with mobile cameras.

*WINNER: Best Student Presentation Award*
A Comparison of Methods for Network Summarization

Ali Hajimirza, Computer Science, University of Oklahoma, Norman
Jason White, Computer Science, CSU San Bernardino
Arlei Silva, Sourav Medya, Prof. Ambuj Singh, Department of Computer Science

Networks are ubiquitous – they model numerous complex structures and processes. A network is a set of nodes associated with dynamic attributes and a set of edges that represent relationships between nodes. Networks generated by real-world complex systems are extremely large. For example, Twitter consists of over 40 million users (or nodes) and over 1.4 billion connections (or edges) between users. In order to analyze such networks, the size of the network data must be reduced and, therefore, summarization becomes important. We experimented with three methods of compression: Slice Tree, Spectral Graph Fourier, and Spectral Graph Wavelets. Slice Tree partitions a network into smooth regions such that each region can be compactly represented by a single value. This single value summarizes the values of the nodes inside the region. Spectral Graph algorithms, in general, summarize a graph by performing a transformation from the graph spectral domain into the signal domain and filtering for important signal values. We examine each method for its scalability, accuracy, and sensitivity to input parameters on real and synthetic datasets. We discover that the Slice Tree algorithm is scalable and outperforms the Spectral Graph methods when the network values change smoothly across the topology.

Interpreting Ecological Network Data

Sean Nguyen, Biology, UC Santa Barbara
Kyoungmin Roh, Sourav Medya, Prof. Ambuj Singh, Department of Computer Science, Prof. Hillary Young, Department of Ecology, Evolution, and Marine Biology

In recent years, there has been an increasing number of studies on complex networks within ecological systems. Those focusing on food web networks have produced simple models, allowed assessments of network robustness, and other properties. However, it is difficult to analyze a food web without a complete set of necessary data. To accurately describe such networks, either additional observational data or an effective and efficient way to infer missing network data is needed. Our current network includes three species of parasites and six host species; however, field data on one parasite species is missing. To find the missing data, we first validated the Expectation Maximization (EM) algorithm across sample data. We then applied the EM algorithm to our current network in order to find missing parasite parameters. Our findings provide a simple and fast approach that has time, labor, and monetary saving implications when describing incomplete ecological food web networks.
Multiscale Modeling of Biological Networks

Kara Goodman, Computer Science, CSU San Bernardino
Austen Piers, Computer Science, UC Santa Barbara
Xuan Hong Dang, Sourav Medya, Hongyuan You, Kyoungmin Roh, Prof. Ambuj Singh, Department of Computer Science

A genetic network consists of gene expression levels and the genes' underlying PPI (protein-protein interaction) network. The project’s goal is to identify a small number of sub-network biomarkers within three genetic networks that predict a phenotype. Our data consists of microarray data from breast and liver cancer patients, as well as cell proliferation in Caenorhabditis elegans. The collected microarray data has features in the low thousands; allowing for a large number of possible sub-networks, which, in turn, makes the search for discriminative sub-networks NP-hard. Our lab’s machine learning algorithms MINDS (MINing Discriminative Subgraphs) and SNL (Sub-Network spectral Learning) are two methods that overcome this intractability. MINDS performs MH (Metropolis Hastings) sampling to discover discriminative sub-networks that are used to create NCDT (Network Constrained Decision Trees), which classify network snapshots. SNL uses regularized subspace learning under network topology constraints to discover discriminative sub-networks. Both SNL and MINDS reveal influential genetic biomarkers of the underlying phenotype with accuracies above 70 percent, respectively.

Tweet Classification

Yuanlin Xu, Computer Science, CSU Los Angeles
Michael Busch, Prof. Ambuj Singh, Department of Computer Science

An automatic and robust tweet topic classification technique plays an important role on many challenging research problems such as information retrieval, document categorization and social network analysis. The purpose of our project was to build an automatic tweet classification system with a domain-specific Wikipedia Knowledge Base. In this project, we used a 48.7GB Wikipedia Latest Articles data set, which was last modified on June 2014, to train our tweet classification system. We tested the system performance on 6000 tweets that were labeled from five topic labels in a Mechanical Turk survey. We constructed a Bag of Words model, a topic-specific vocabulary model, based on the Wikipedia Latest Articles data set and implemented a Naive Bayes’ classifier to classify arbitrary text in a given tweet. We also provided a least-squares regression approach for mapping the scores of Wikipedia categories to a topic-specific tweet as well.

Classification of Traumatic Brain Injury Using Diffusion Tractography
Traumatic brain injury (TBI) significantly changes mental health and it is most commonly caused by vehicle accidents, violence, and sports. Diagnosis of TBI with standard imaging techniques remains elusive as it has no clear biomarkers to test for, especially for mild versions of TBI. We propose a classification algorithm using features derived from diffusion tractography, which traces white matter fiber tracks in the human brain using MRI technologies. Such an objective measure of TBI can be used to assess whether athletes can safely return to their sport or military personnel are fit to return to battle. The features selected using machine learning methods will be useful in studying the pathologies and symptoms of TBI.