University of California, Santa Barbara
Center for Bio-Image Informatics

2005 Summer Research Program

The UC Santa Barbara Center for Bio-Image Informatics provided a summer outreach research program for undergraduate students and high school students in Biology and Information Technology. The Center's mission is to establish a searchable digital library for bio-molecular images and to develop new information processing technologies for a better understanding of complex biological processes at the cellular and molecular level. Nine undergraduate students from CSU San Bernardino, CSU Fresno, University of the Virgin Islands, and UC Santa Barbara participated in the internship, as well as 2 students from local high schools.

The Summer Undergraduate Research fellows and high school interns interacted with faculty and graduate student mentors involved in this project. The undergraduates also attended various trainings designed to develop the specific skills necessary for success at the graduate level. The participants were involved in cutting edge research in biology, statistics, engineering, and computer science. Each student completed a research project under the supervision of one of the laboratories associated with the program. A student was paired with a mentor (a graduate student or postdoc) who was affiliated with his or her assigned laboratory. The students also interacted with other students and faculty who were part the project. They attended a weekly workshop on communication skills, presentations covering both the project and general scientific topics, and weekly project meetings. Additionally, the students received guidance and information regarding opportunities in science and how to pursue graduate education. The undergraduates learned to write scientific abstracts and summaries of their work, designed and created posters for their project, and produced and presented PowerPoint presentations. The final abstracts were incorporated into the UCSB Summer Undergraduate Research Colloquium and their posters were displayed at this campus-wide event. The high school students also produced a final presentation covering their project. The Bio-Image project held a separate Summer Intern Symposium where the students presented the work they accomplished to faculty and students.

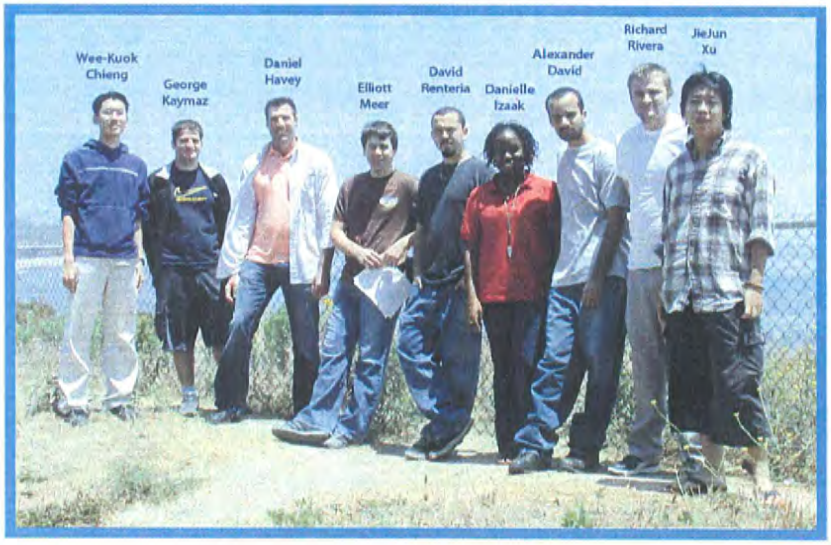
We would like to extend our appreciation to Dr. Fiona Goodchild, Liu-Yen Kramer, Wendy Ibsen, Maura Jess, and Trevor Hirst for their advice and support. We would also like to acknowledge and thank the graduate mentors and faculty for providing the research opportunities for the interns and their supervision. This summer program was supported by the National Science Foundation Information Technology research (ITR) program and Grant No. 0331697. The College of Engineering, the College of Letters and Science, and the Graduate Division provided additional funding.

B.S. Manjunath Director
Coordinator Center for Bio-Image Informatics

Mark Verardo
Education and Outreach NPNC03

University of California, Santa Barbara
Center for Bio-Image Informatics

2005 Summer Research Program



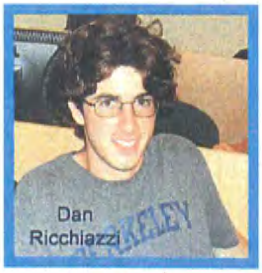


Table of Contents:

Enabling Rapid Data Ingest: The Scientists Digital Notebook 1
- Daniel Havey -

Enabling Microscopy in the Macro Scale: The Imaging Wall in the Biological Laboratory 3
- Wee-Kuok Chieng -

High Resolution Imaging of Microtubules using DIC and AF 6
- Elliot Meer -

Accurate Detection of Microtubules in Image Series 9
- JieJun Xu -

Time Series Analysis on Microtubule Behavior 11
- Richard Rivera -

Retinal Image Production & Analysis: Sample Preparation and Nuclei Counting Tool 15
- Danielle Izaak -

Visual Vocabulary Construction Using Principal Component Analysis 17
- David Renteria -

Edge Detection and Feature Extraction of Retinal Images 20
- Alexander David -

Invisible Embedding of Meta-Data into Biological Images 23
- George Kaymaz -

Enabling Rapid Data Ingest: The Scientist Digital Notebook

Daniel M. Havey, Computer Systems, California State University, San Bernardino
Mentor: Dmitry V. Fedorov, Electrical and Computer Engineering
Kristian Gustav Kvilekval, Computer Science
Faculty advisor: B.S. Manjunath, Electrical and Computer Engineering

Biological images stored in the bio-image database have various metadata associated with them. Currently, during image acquisition, associated metadata is stored in scientist's notebook and digitized during database ingestion procedure. Furthermore the metadata is frequently recorded in formats that although clear to the individual scientist may not be understandable to others. One of the goals of the bio-image project is to develop rapid and consistent data ingest procedures. To accomplish this goal we are developing the Digital Notebook application, which will assist a scientist to input pertinent metadata for an image being acquired and become their digital notebook. The application can be easily tailored to the individual needs of the scientists through the use of a configuration file. Furthermore, application will allow scientists to upload acquired images directly to the database. Since the application may not be connected to the database at the time of image acquisition the application is designed to work in both disconnected and connected modes. While acquiring the image the scientist can use the Digital Notebook in disconnected mode, recording the metadata in a database friendly XML file. Later when a network connection to the database is available, the scientist can use the application in the connected mode uploading the images and their metadata directly into the database. Thus, the time and effort required to ingest data is reduced. Another problem addressed with this application is information continuity when similar metadata for a series of images is acquired. The scientist can use the application in batch mode to enter the data only once and select a number of files to save. The development of applications such as the Digital Notebook should promote the free flow of information and facilitate dissemination of knowledge among scientists.

Enabling Microscopy in the Macro Scale: The Imaging Wall in the Biological Laboratory

Wee-Kuok Chieng, Electrical and Computer Engineering, California State University, Fresno
Mentor: Dmitry V. Fedorov, Electrical and Computer Engineering
Kristian Gustav Kvilekval, Computer Science
Faculty advisor: B.S. Manjunath, Electrical and Computer Engineering

The resolution of bio-images acquired from various kinds of microscopes is very large. The sizes of these images are far greater than the resolution of typical computer displays. Display panels that are available now can only show a fraction of an entire bio-image simultaneously or display them so small that resolution is lost. We will try to visualize the global pattern of biological structure while preserving the high-resolution features. This will lead to a better understanding of the underlying patterns and processes that determine function in biological structures for biologists. In order to solve the problem, we propose to display images on walls with banks of video screens. However, we have a few constraints. We want our system to be highly cost effective without scarify of performance. As a result, we have decided to use commodity display panels and tile them into a display wall. The high resolution bio-images will be shown on the display wall. Apart from the performance issue, we need to consider three things in terms of software that could unify the whole display into a single large display. First of all, scalability is a major problem we face. We can integrate two displays on a single machine, however, that will limit to the number of physical devices that can co-exist in a single machine. Scalability became a serious issue. Secondly, we want our video wall to be able to display images from our database and internet. Networking is the second issue that we want to work on it. Lastly, we need to deal with application flexibility issues. We don't want our video wall serve a single purpose. Apart from viewing the bio-images, we want our video wall to be able to run other software as well. In this project, we basically compile the DMX (Distributed Multihead X) in the Linux operation system. There are three reasons why we want to use the DMX as our software to unify commodity display panels. First of all, DMX (Distributed Multihead X) creates a "virtual" Linux workspace which is highly "application friendly". That means that we are free to compile and run almost all Linux compatible software on the video wall we develop. Secondly, DMX gives the compatibility to integrate Linux open source software like Chromium (3D viewing framework), Blockbuster (video screening software), etc, on top of it. Apart from that, DMX's framework allows "unlimited" connections of back-end X client (Display node). It basically solves our scalability issue. So, what is the structure of the DMX? And how it works? The overall structure of the distributed Multihead X (DMX) project is as follows: A single front-end X server will act as a proxy to a set of back-end X servers, which handle all of the visible rendering. X clients will connect to the front-end server just as they normally would to a regular X server. To the client, everything appears as if they have a single large display (DMX design document, 2003). At the end of this internship, we were able to build a low cost highly scalable display system using commodity hardware. We also included some very useful hardware like wireless gyra-mouse which will give the freedom to the user to interact with the array of monitors within 100 feet. The video wall will encourage the collaboration of scientists and researchers. They will be able to analyze the image in front of the video wall and share their opinions. However, there are still much that can be done to improve the prototype video wall that we developed this summer. We definitely want to develop imaging software that robustly allocates the memory of the image to different display

nodes so that we can speed up the image loading process. Moreover, we can always integrate the available 3D object viewing software like Chromium into the master node of the video wall.

Enabling Microscopy On A Macro Scale

Wee-Kuok Chieng, CSU Fresno,
Daniel Havey, CSU San Bernardino
Dmitry Federov, Kristian Kvilekval, B.S. Manjunath, ECE UCSB

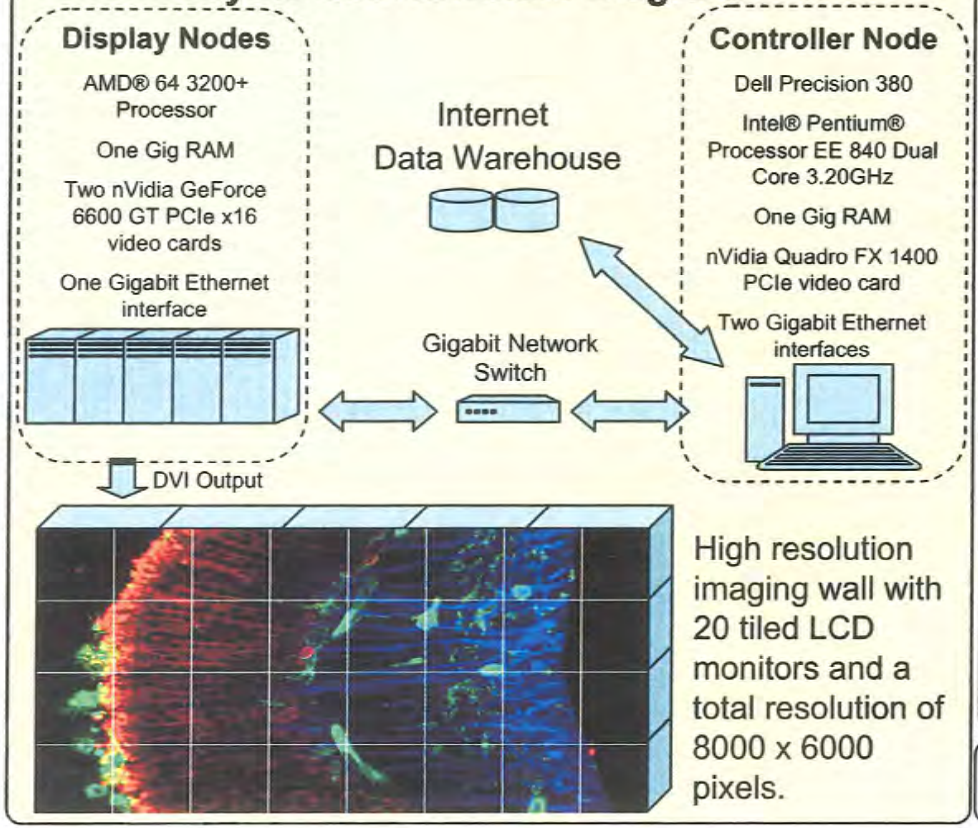
Motivation

- Assist scientists in gaining a better understanding of the underlying patterns and processes that determine function in biological structures.
- Visualization of the global structural pattern while preserving high resolution features.

Objectives

- Develop a high resolution Imaging Wall using commodity hardware and open source software.
- Develop software that allows comfortable visualization of high resolution images (100,000 x 100,000 pixels), provides necessary tools for visual comparison and meta-data integration

System Architecture Diagram



High resolution imaging wall with 20 tiled LCD monitors and a total resolution of 8000 x 6000 pixels.

Results

- Using commodity hardware we are able to build a low cost highly scalable system.
- We are developing cross-platform imaging software, that through the use of tiling allows the display of high resolution images. The software is written using C++ using Trolltech's Qt library.
- Open source software such as Distributed Multihead X on Linux.
- Dual DVI PCIe-16 video cards drive four monitors per display node lowering cost per display unit.

Conclusion

This system can be used to assist scientists in many areas of biological research such as retinal injury, cancer and Alzhiemer's disease.

This research was supported by NSF
Information Technology Research grant #0331697

High Resolution Imaging of Microtubules using DIC and AFM

Elliot Meer, Cell and Developmental Biology, University of California, Santa Barbara
Mentor: Austin Peck
Faculty advisor: Stuart Feinstein, Neuroscience Research Institute

Back in the early 1500's compound microscopes were just coming into use. The first of such microscopes had a magnification of about 3x-10x and all images were recorded by hand. We have come a long way since then nNow with light and fluorescence microscopy a single protein in a sea of thousands can be labeled and tracked. Electron microscopes allow us to view objects at over 1 million times their native size, and with resolution down to the nanometer scale. And with all of these new advancements in microscopes we have also found new ways of recording the images. Computers now allow one to simply click and save images of just about anything, including microtubules.

Microtubules (MTs) play many important roles in the cell including cell structure and transport. MTs are composed of 10-15 protofiliaments, which form a hollow tube. These protofiliaments are in turn made up of alpha and beta monomers of tubulin. Our biggest interest in MTs is how and why their regulation goes awry. Such malfunctions have been implicated as the causes for many neurodegenerative diseases. Mis-regulated MAPs are a defining characteristic of Alzheimer's; "tau tangles". The way in which our group imaged these MTs was by Differential Interference Contrast Microscopy (DIC) and by Atomic Force Microscopy (AFM).

DIC works by using an inverted light microscope, then taking the image that is generated and computer enhancing it. The principle used to generate this enhanced image is the differences in refractive indices inherent to the sample being imaged. As light passes through different parts of the sample the phase of that light is changed. These phase changes can then be converted into amplitude changes, which can be used to produce contrast.

AFM on the other hand works much light a record player. Topographic images are produced by responding to minute forces produced between the tip and the sample. This force can be measured by a laser beam bouncing off the back of a cantilever to which the tip is attached. For a comparison of these methods, see the Table below:

	DIC	AFM
Pros	Allows for live MT tracking Can understand MT dynamics Diagnostic before using AFM	High Resolution High Magnification Native sample environment Live
Cons	MT can go in/out of focus Need multiple people to confirm results	No Tracking Not simple to get good images

Several challenges using theses imaging methods are being overcome. Focusing on MTs and keeping them in focus was almost impossible. This meant when trying to track the Mts, in order to understand their dynamics, keeping the end of the MT in sight was virtually impossible. However this does not mean it cannot be done it just means that more time is needed in perfecting our technique for getting good images.

On the AFM better results were obtained. We were able to get very high resolution images, even being able to view individual protofiliaments. This will allow for future experiments in looking at microtubule associated proteins, and observing how they physically bind to MTs.

The results of our work are proof of the need for better image processing and pattern recognition. If we can use computers more effectively we can get more accurate and more detailed images. This will lead to decreased workload on the scientist part and increased ability to focus on interpretation of the image. For MT work this means with the DIC, increased understanding of MT dynamics. And with the AFM, increased understanding of MAP/MT interactions.

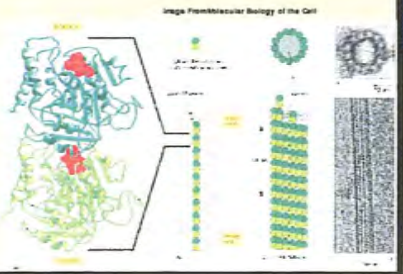
High Resolution Imaging of Microtubules using DIC and AFM

Elliott Meer, Cell and Developmental Biology, UCSB
Austin Peck, Peter Markiewicz, Les Wilson, Stuart Feinstein, NRI

This work supported by the National Science Foundation Information and Technology Research (ITR) program and Grant No. 0331697

Project:

- Prepare microtubule samples and acquire images to:
  - increase the utility of biological images acquired and...
  - facilitate their incorporation into a growing database



Interest:

- MT Dynamics are of central interest in the fields of neurodegenerative disease and many cancer chemotherapeutic strategies.
- Image enhancement facilitates easier, quicker analysis.
- Cataloguing by database allows rapid, quantitative comparisons of test conditions.

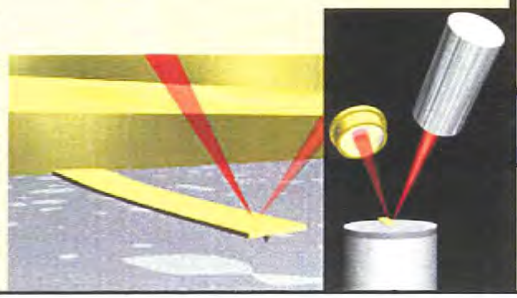
Methods:

DIC: A type of light microscopy, images are produced by differences in refractive index within the sample.

- phase changes are converted into amplitude changes, from which the contrast of the image is derived.
- technique is valuable because it allows background interference to be subtracted
- Microtubules viewed under DIC were prepared by either of two methods:
  - MT were assembled using taxol, in this case DIC was used only to verify MT polymerization
    - They were then placed on aptes coated glass cover slips for viewing under AFM
  - MT were assembled using axonemes, here DIC was used to view MT dynamics

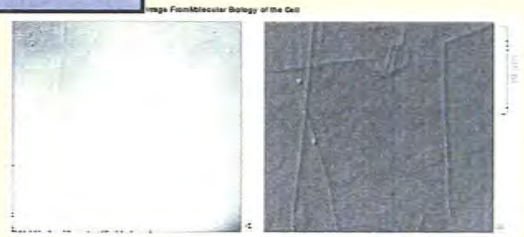
AFM: Atomic force microscopy generates topographic images of a sample by responding to minute forces between the tip and the sample.

- The movement of the tip is monitored by a light lever consisting of a laser beam bouncing off the backside of a cantilever to which the tip is attached.
- This generates resolution which far surpasses optical microscopy. Unlike electron microscopy, AFM can image samples in their native environment.

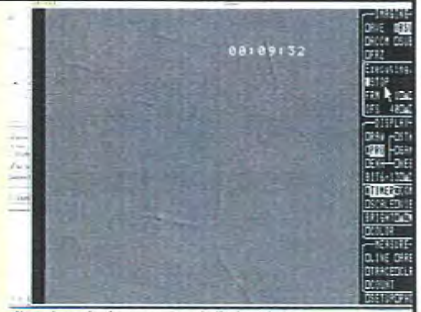


Results:

DIC:



Raw Image — Computer Enhanced/Background Subtracted



Above image is of axoneme "seeded" microtubules.

- As seen from the above DIC images , viewing MT with this sort of microscope has its limitations. DIC is fairly easy to use, although producing a decent image is more of an art than a science. If MT are grown using axonemes they have the ability to float in and out of the focal plane thereby making them almost impossible to track. But to its credit DIC can be used to image dynamicity, a major characteristic of MTs.
- On the other hand AFM (below) has the ability to take extremely high resolution images, even showing the structure and # of the individual protofilaments that makeup the MT. AFM too takes a trained technician to master but the controls are much more precise.

AFM:

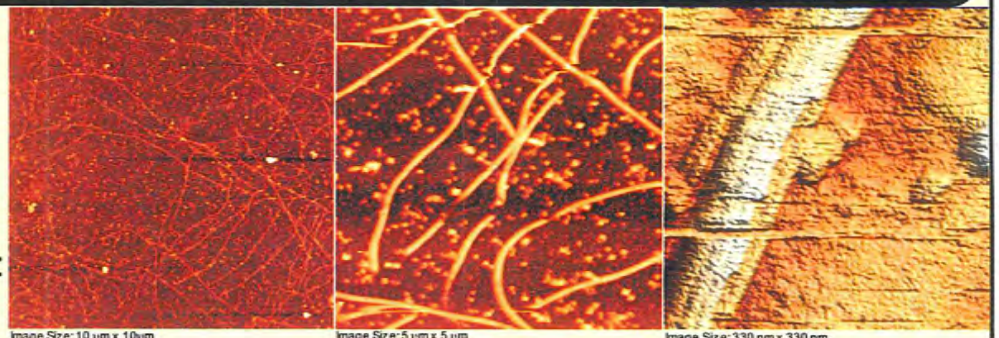


Image Size: 10 um x 10um — Image Size: 5 um x 5 um — Image Size: 330 nm x 330 nm

Future Goals:

- Use AFM to view how disease-implicated microtubule associated proteins, tau and stathmin physically bind MTs
  - Test the hypothesis that MAP binding induces conformational change in MTs

Accurate Detection of Microtubules in Image Series

JieJun Xu , Computer Science, California State University, Fresno
Mentor: Alphan Altinok
Faculty advisor: Kenneth Rose, Electrical and Computer Engineering

Introduction: Microtubules are proteins that form cylindrical hollow structures which are commonly distributed throughout cells. Traditionally microtubules are tracked manually. This is very tedious and time consuming, and it is impossible to manually track all of the microtubules in a dataset. Therefore data, such as the growth rate and shortening rate, extracted from one image can only be representative of a very limited portion of the image and not sufficient to show the global information in an image set. This project is an attempt to automate the microtubule detection process currently done by hand. It is also an attempt to develop an efficient method to detect all microtubules in one image globally instead of keeping track of only a few microtubules. Though there are many well known object detection techniques in image processing, none of them alone produces adequate results on microtubule images due to variations in background lighting and uneven fluorescence of microtubules in the image plane. Therefore, we tailor our own detection method for this specific type of image.

Methodology: The input is an image sequence which contains multiple images taken over a certain time period. What we want to deliver is the extracted statistical result of the microtubules activities such as the growth and shortening rates. Basically, our method is attempting to find the accurate sum of differences (change in microtubules lengths) between every two successive images by using a combination of image processing techniques such as adaptive thresholding and Reduction of later motion.

Adaptive thresholding: The idea of adaptive thresholding is to divide an image into subregions, then process (threshold) only those regions of an image that meet a certain criterion, and interpolate parameters of the processing function for the region which do not meet the criterion. The brief steps of adaptive thresholding are described as follows. Each image in the image sequences was divide into K regions by placing a m*n grid with 50

Reduction of Later Motion: By subtracting one binaried image from the other, we obtain an image which shows the changes in microtubules. In other word, this image contains all the statistical information about the growth and shortening events of the microtubules. However, this image also contains some information that affects the statistics, which is the effect of lateral motion. In order to reduce lateral motion, we applied an algorithm called Block Matching on the image, to clean up any misleading data and leave only the growth and shortening data in the image. The resulting images were then ready for further statistical analyses.

Conclusion: With the completion of this method, biologists no longer need to spend hours sitting in front of the microscope tracking microtubules manually. also a more accurate statistical result can be extracted from image sequences.

Accurate Detection of Microtubules in Image Sequences

JieJun Xu, CSU Fresno
Alphan Altinok CS, Kenneth Rose, ECE UCSB

Objective:

- Improving steps in detecting microtubule activity:
  - Segmenting microtubules from image sequence.
  - Enhancing images by eliminating lateral motion.

Background

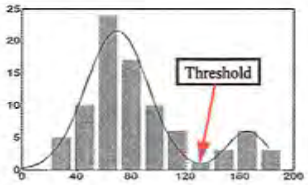
Known object detection techniques do not produce adequate results on microtubule images due to variations in background lighting and uneven fluorescence of microtubules in the image planes. In this work, we attempt to use adaptive segmentation techniques for developing an ad hoc solution.

Method

Adaptive thresholding

The idea is to process only those regions of an image that meet a criterion and interpolate parameters of the processing function for the regions which do not meet the criterion.

Each image was divided into K regions by placing a $m \bullet n$ grid with 50% overlap. After all histograms were computed, a test of bimodality was performed to select the regions for thresholding. Optimum threshold levels for bimodal regions were obtained by fitting two Gaussian density curves, and taking the interception of these two curves.



The criterion for selecting candidate regions for thresholding was

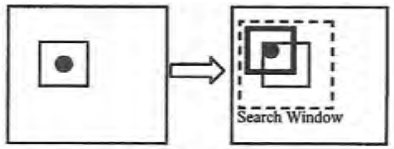
$$D(\mu_1,\mu_2) > \sigma_1 + \sigma_2$$

Once the regions were selected for processing and optimum levels were found, we interpolated threshold levels for the rest of the regions. A second interpolation was carried out point by point using neighboring threshold levels so that every point in the image had been assigned a threshold at the end of the procedure.

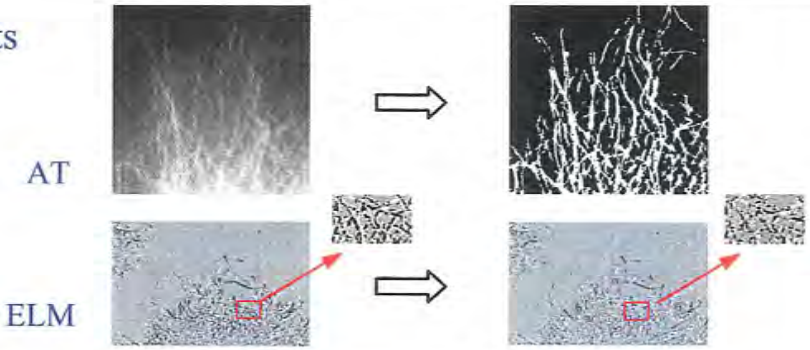
Finally, a binary decision was carried out for each pixel using the respective level.

Eliminating Lateral Motion (ELM)

Block Matching Algorithm is used in order to eliminate lateral motion in microtubule images. The displacement of an object in frame k is found by considering an $n \bullet n$ block centered about the object, and searching frame $k+1$ for the location of the best matching block of the same size.



Results



Future tasks

- Discontinuities in detected microtubules were found, therefore, we need to improve the processing part of adaptive thresholding.
- Content driven processing of microtubule images to be investigated.

This work was supported by the National Science Foundation Information and Technology Research (ITR) program Grant No. 0331697.

Time Series Analysis on Microtubule Behavior

Richard Rivera, Computer Science, California State University, San Bernardino
Mentor: Arnab Bhattacharya
Faculty advisor: Ambuj K. Singh, Computer Science

The study of microtubules, polymers of tubulin found inside of cells aiding in cell division, helps biologists gain a better understanding of degenerative diseases and their function in cells. Tau, a protein associated with microtubules which aids in neuronal cell polarity and the maintenance of axonal morphology is at the center of this research. Dysfunction of this tau protein is linked to Alzheimer's disease, frontol temporal dementia and other neurodegenerative diseases. Studying this protein is imperative in understanding cell dynamics and its dysfunction causing neuron death.

My research involves creating a program that accepts time series data from the length of a microtubule over time, clustering selected conditions, and building trees from this data. Various conditions including concentrations of tau to tubulin, 3 repeat tau (3 imperfect amino acid repeats separated by 13-14 amino acid inter-repeats) and 4 repeat tau (4 amino acid repeats separated by 13-14 amino acid inter-repeats) are included in this database. This data is parsed from text files that contain micron-pixel conversion ratios, time samples for the given microtubule, and its x-y coordinates at sampled time intervals. A necessary hurdle in obtaining reasonable results is data cleaning and compensation for unevenly sampled time series. This data skewness can be caused by time intervals from one microtubule not being similar to the rest of the microtubules in a given condition; which happens due to manual sampling.

Organizing the microtubules into a workable database was done in Perl by parsing Excel files provided by biologists. Scripts to maintain this database, which consists of 1,935 microtubules over 77 conditions, have also been written in a manner that if more microtubules were added from different condition sets it automatically create logs of missing, extra microtubules and file destinations for all data sets, including timestamps. These logs are imperative in bookkeeping and location of missing microtubule .RTM files. For instance, if a microtubule condition set is missing, a log containing the missing files can be sent to the supplier of the data and the data can be located, which happened to occur in the creation of the database. A script to rename mislabeled microtubules has also been created which accepts a file with replacement names as its argument and traverses the database performing requested name changes.

Two types of interpolations are performed on the microtubules. The first interpolation is implemented by means of comparing thresholds and interpolating gaps in the time series. Errors such as gaps in the time series can be caused by losing the microtubule while tracking. Also, tracking the wrong microtubule for an amount of intervals posed many problems in the pursuit of credible results. To fix these errors two mathematical techniques are implemented. First, if there is an extended time interval in the subjected data, linear interpolation is used to fit points to a line that agrees with the trend of existing points around it. Second, spikes in data that did not fall within the range of user-selected thresholds of growth and shortening are linearly interpolated by means of finding a trend in closely related time points that are within the threshold range and fitting the point in relation to these points. For the Lomb-Scargle Transformation, offending data was thrown out due to Lomb-Scargle's invariance of time intervals. This interpolation is the preferred method for post-steady state microtubule analysis due to its ability of being able to represent catastrophes and rescues without smaller, out of threshold level, errors.

The second interpolation implemented is the preferred method for pre-steady state (excess of tubulin in a microtubule causing primarily growth) microtubules. Since the microtubules are in this pre-steady state there are few attenuation, and rarely any shortening events, so a piecewise linear fit can be implemented. This is performed by breaking the time series into two components, each of which were two "lines" that represent the least squares evaluation of the data most fittingly. From these two components the data is evaluated with other segments in its respective component n-1 times and the average of these evaluations returns the best representation of the time series. The two components are then merged and best fit data is returned to the transformations. Two components are implemented in the case of attenuation occurring in the pre-steady state time series. Also, this best fit method removed data errors that may have been assumed to be within thresholds of the previous interpolation but do not fit the data series as a single unit.

To normalize the time series data frequency extraction is used. This includes Discrete Fourier Transformation, Discrete Cosine Transformation and Lomb-Scargle Transformation, all three of which are user selected. These transformations break the time series into components of frequencies and return the occurrence of frequencies in a series of coefficients. Performing these transformations aid in relating microtubules by features of growth and shortening and not by the length of time elapsed in the time series sample. The Lomb-Scargle transformation shows the most consistent results due to it's inclusion of a time offset that compensates for uneven sampling, which this database has a large amount of.

Clustering and tree structures are based on three algorithms-Kmeans, Neighbor Joining (NJ) and Unweighted Pair Group Method with Arithmetic Mean (UPGMA). NJ and UPGMA use a distance matrix to compute nearest nodes and group them until only two nodes remain, each connected by a single branch, while Kmeans uses centroids for its groupings. Centroids for Kmeans are chosen based upon the number of desired clusters and the microtubules chosen in clustering. Choosing values for the centroids is accomplished by means of sorting each dimension for the microtubule dataset and assigning an even interval between the microtubule dimensions related to the desired amount of clusters. This method creates an imaginary microtubule made up of dimensions from multiple microtubules, thus, eliminating randomness in assigning centroids and the possibility of choosing an outlying centroid. After the microtubules are assigned to groups the centroid dimensions are recomputed by taking the average distance to the centroid dimensions; this is repeated until clusters maintain the same microtubule groupings over repeated iterations.

Logging is performed for all microtubules evaluated. This data contains all of the operations performed on the raw microtubules. These logs provide the life history of all microtubules, which include time intervals that are out of range as well as catastrophes and rescues out of a preset threshold range for each microtubule. Matlab scripts are also generated that plot the time series and its extracted coefficients. These Matlab scripts give the user a visual look into the microtubules' behavior while it was being tracked and can also be used to compare a microtubule to other microtubules. These logs give the user the ability to manually look up microtubules in the database and view all operations and analysis performed on a microtubule of choice.

A Graphical User Interface (GUI) has been written that makes all of these features in a simple and user friendly manner. The GUI consists of study listings that contain condition listings of the database. These conditions are represented by a scrollable list. Three radio buttons represent which type of transformation is to be performed and two clickable buttons perform all of the operation of the program. The first button,"Compute", collects all microtubules in the user-highlighted list, performs the specific interpolation on the data depending on which transformation is selected, and

then sends the result to the user-selected transformation. This button also writes log files, including Matlab scripts to graph each microtubule. The second button is the "Draw" button. First, the user selects which type of clustering/tree algorithm is to be used and how many clusters are desired. Microtubule clusters or a tree representation for all microtubule conditions are then calculated and drawn on the screen. The interface is very easy to navigate and with only essential functionality it can be learned in a matter of minutes.

Results of the trees and clustering have been very good with most results being biologically correct for complete data sets. An example is the $3Rm\_wt$ and $4Rm\_wt$ tau isoforms conditions. For a two cluster selection, 3Rm occurred in cluster 1 96% while 4Rm occurred in this cluster 14

Since multiple sequences of conditions can be selected, similarities (or differences) that are not observed while the lab work is being performed can be determined. This project therefore can give researchers a better understanding of microtubule dynamics, and also better insight on cell division in neurodegenerative diseases.

TIME SERIES ANALYSIS ON MICROTUBULE BEHAVIOR

Richard Rivera, Computer Science, CSU San Bernardino
Arnab Bhattacharya, Ambuj K. Singh, Computer Science

Research Objectives:

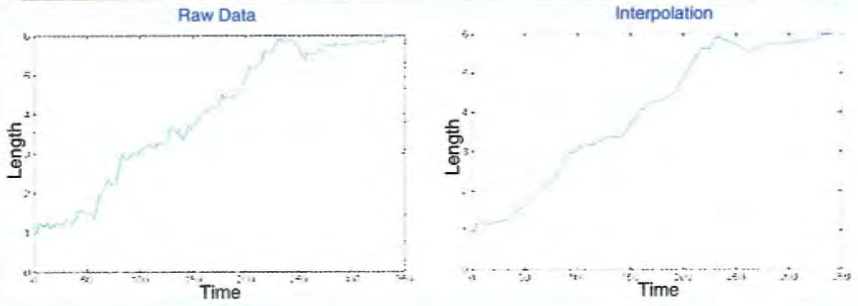
- Create a program that clusters and builds trees from unevenly sampled microtubule time series data.
- Remove spikes beyond given threshold to minimize variance.
- Create logs of each microtubules life history.

Goals:

- Obtain a better understanding of microtubule dynamicity.
- Compare different proteins applied to microtubule incubations and their effect on cell division.

Methodology:

- Data out of threshold range omitted in Lomb-Scargle Periodogram, interpolated in Discrete Cosine and Discrete Fourier transformations.
- Time variance interpolation in DCT and DFT.



- Frequency extraction using DCT, DFT, or Lomb-Scargle Transformation.
- Clustering and Tree Structures based on K-means, Neighbor joining, or UPGMA.
- Error logs and symbol files are created for each microtubule.

Lomb-Scargle:

- Results are most significant when Lomb-Scargle coefficient extraction is used.
- Time variance is handled in the equation and not by interpolation.

Equation:

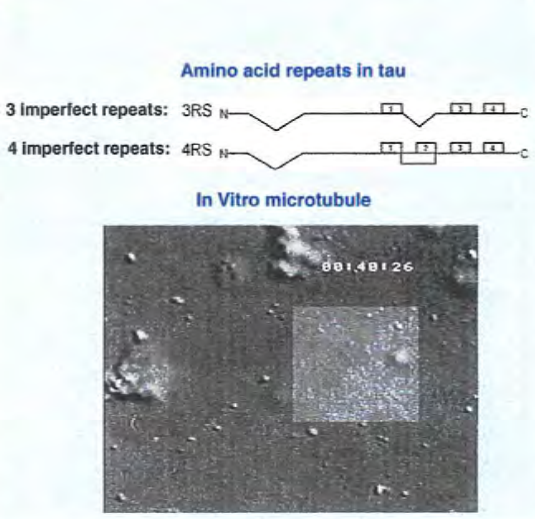
$$P(\omega)=\frac{1}{2\sigma^2}\left\{\frac{\left[\sum_{n=1}^{N}x(t_n)\cos\omega(t_n-\tau(\omega))\right]^2}{\sum_{n=1}^{N}\cos^2\omega(t_n-\tau(\omega))}+\frac{\left[\sum_{n=1}^{N}x(t_n)\sin\omega(t_n-\tau(\omega))\right]^2}{\sum_{n=1}^{N}\sin^2\omega(t_n-\tau(\omega))}\right\}$$

Invariance to time translation:

$$\tau(\omega)=\frac{1}{2\omega}\arctan\left\{\frac{\sum_{n=1}^{N}\sin 2\omega t_n}{\sum_{n=1}^{N}\cos 2\omega t_n}\right\}$$

Microtubules:

- Polymers of tubulin with diameter of ~25 nm found within cells.
- Dynamics are very important in cell division.
- Microtubule associated protein tau linked to Alzheimer's disease and front temporal dementia.



Results:

- Most conditions show strong divergence from tubulin only (no tau).
- 3R and 4R tau clustering is good for complete condition sets.
- Some groupings may be skewed because of missing microtubules in certain conditions.

Examples:

Wild-Type Isoform

Clusters	1	2
3RM	.96	.04
4RM	.14	.86

4RS 1:20 – 1:40 tau concentration

Clusters	1	2
1:20	.14	.86
1:40	.73	.27

Future Work:

- Port into an online applet that requires no files to be manually installed.
- Implementation of a user friendly system that differentiates dynamicity of a microtubule from human error more consistently.

This work supported by the National Science Foundation Information and Technology Research (ITR) program and Grant No. 0331697.

Retinal Image Production & Analysis: Sample Preparation and Nuclei Counting Tool

Danielle A. Izaak, Chemistry, University of the Virgin Islands
Mentors: Mark Verardo, Neuroscience Research Institute
Jiyun Byun, Electrical and Computer Engineering
Faculty advisors: B. S. Manjunath, Electrical and Computer Engineering
Steven K. Fisher, Neuroscience Research Institute

The retina is a multi-layered, heterogeneous tissue specialized for detecting light and responsible for the first steps in visual image formation. Retinal detachment is the separation of the inner sensory layer of the retina and the retinal pigment epithelium which can lead to significant retinal changes. These changes disrupt cell structure ultimately lead to loss of vision. This disruption may be caused by trauma or result from injury or disease,. Understanding the biological mechanisms behind the loss and recovery of vision following retinal detachment and reattachment has been the focus of the retinal research at the University of California, Santa Barbara, for many years. The Hypothesis of my project is "The number of cells decrease in detached retinas." To count the number of cells, I did sample preparation and visualized the nucleus within the retina. Images were taken with a confocal microscope. I then developed a cell counting tool to provide quantitative cell number information. The first part of this project is generating confocal microscopy images from normal and detached retinas. In order to prepare the tissue for imaging, established immunocytochemistry protocols are used. Before proceeding with these protocols, the tissue needs to be prepared correctly. The retina tissue utilized for the images is obtained from different species such cat or mouse within a regulation. For such a task there are different antibodies that can be used to portray these changes that take place in the cells of the retina. Seeing that we are interested in looking at the difference in cell numbers within the retina, the cells are stained with a nucleic binding cyanine dye. Images are then taken with a laser scanning confocal microscope. Therefore, the images that show these cells with the nucleic cyanine dye TOPRO, which stains the nuclei of the cells, allows us to visually see these cells. The resulting images are stored in a database where they can be accessed online. Currently there are thousands of images and they are currently painstakingly analyzed by researchers by looking at them. For the second part of my project, we developed an automated cell counting tool. In order to understand changes such as cell death, degeneration or damage to the retina over a period of time and obtain statistically valid observations, a lot of images should be collected from different samples and different animals. This is done in large scaled datasets. Presently, the number of cells per image is counted manually and such an analysis is not repeatable. This process is tedious and very time consuming. In order to make the process of counting cells easier, a program needs to be developed to perform this automatically. A current method is developed in Matlab to automatically count the cell with 3.67% accuracy for the outer nuclear layer (ONL). The method aids in quantifying the difference in number of the cells between the images of normal and detached retinas. My part is to develop a plug-in of ImageJ. ImageJ is an Image Processing and Analysis Tool developed by the NIH in Java. The tool which incorporates the counting method, will be applicable not only to retinal images but to different types of images such as a brain image.

Retinal Image Analysis: Sample Preparation & Nuclei Counting Tool

Danielle A. Izaak, Chemistry, University of the Virgin Islands
Mark Verardo[1], Jiyun Byun[2], Steven Fisher[1,3], B.S. Manjunath[2]
Neuroscience Research Institute[1], Molecular, Cellular and Developmental Biology[3], Electrical and Computer Engineering
University of California, Santa Barbara

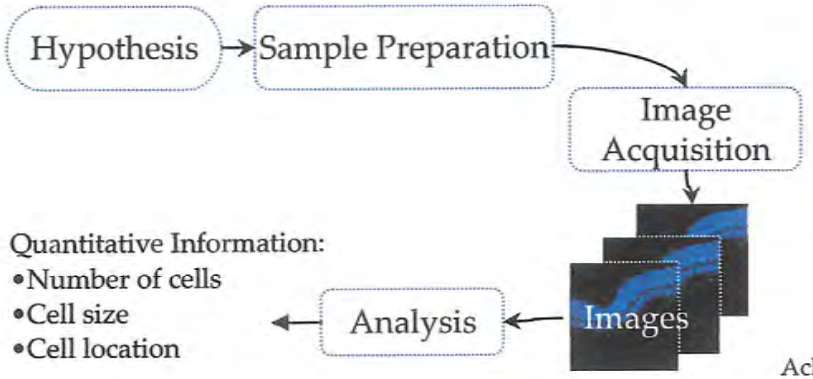
Objective:

- Understand the biological mechanisms behind the loss and recovery of vision following retinal detachment.

Introduction & Overview:

- What happens to cells before, during, and after detachment?
- Scientific Hypothesis Based Research
- **Hypothesis**: There are less cells in the retina after retinal detachment.
- Number of cells is one of the important features for supporting the hypothesis.
- Necessary to develop an automated tool for counting the cells.

Retinal Cell Research: Methodology



Sample Preparation:

❶Fixation

Agarose Embedding of tissue

❷Immuno. Run

Tissue stained with 1° & 2° Antibodies

❸Acquisition

Laser Scanning Confocal Microscope

❹Images

Example of section image labeled with TOPRO

Tool Development:

- Develop a tool that automatically counts the number of cells in an image.
- The tool is a plug-in of ImageJ (Image Processing and Analysis Tool developed by NIH).
- Plug-in available at http://rsb.info.nih.gov/ij/plugins/index.html

Snapshot of the cell counter plug-in in ImageJ

Snapshot of result image

Summary:

- Prepare samples for acquisition using the laser scanning confocal microscope.
- Develop a tool that automatically counts cells.
- Assist in understanding the differences between the cells of normal and detached retinas.

Future Work:

- Develop a tool that utilizes cell size and/or distribution to analyze images.

Acknowledgement: Research was supported by NSF Information Technology Research Grant# 0331697

Visual Vocabulary Construction Using Principal Component Analysis

David Renteria, Computer Science, California State University, San Bernardino
Mentor: Vebjorn Ljosa
Faculty advisor: Ambuj K. Singh, Computer Science

Visual vocabulary (Vivo) is a new idea to analyze biomedical images in a database. This research involves the implementation of a visual vocabulary to analyze laser scanning confocal microscope images of the retina. This program can help biologists to see the difference in patterns caused by different conditions.

Although biological differences are difficult to capture with a computer, the use of visual vocabulary allows a computer to capture differences between images without human assistance. First the images are partitioned into tiles that contain characteristic textures of each image. From each tile a feature vector is extracted based on color structure. Next we derive a set of symbols from the feature vectors using principal component analysis (PCA). In text this is similar to grouping documents by topic. This results in axes that correspond to meaningful biological differences. We refer to the axes as the visual vocabulary. Finally each image is expressed as a combination of the re-formed tiles that now makes up the image. Then each image is represented by a number of coefficients, one for each axis. The difference of the coefficients expresses the differences between the images.

In order to apply PCA to the set of images the mean is removed from the feature vectors, giving us our adjusted data. The adjusted data is just the feature vectors shifted around the origin. The result from applying PCA to the adjusted data gives us eigenvectors and eigenvalues. We choose the top 6 eigenvectors based on their eigenvalues and use them to express our visual vocabulary, which consist of 12 Vivos. We multiple the chosen eigenvectors transposed by the adjusted data transposed which gives us our Vivos transposed.

The Vivos can also be symbolically represented. Since none of the original tiles exactly equal any of the Vivos we show what they look like by display a tile. We choose a tile that has the greatest magnitude of the coefficient for that Vivo to express its characteristics. This is displayed in Figure 1.

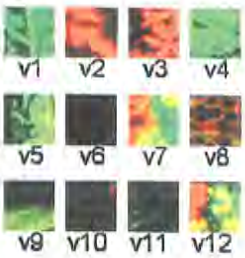


Figure 1: 12 primitive tiles for VIVO

We are also able to calculate the most discriminating Vivos between two conditions. The conditions for this dataset are normal (n), 1 hour detachment (1h), 1, 3, 7, and 28 day detachment (1d, 3d, 7d, and 28d respectively), 3 month detachment (3m), 1 day detachment followed by 6 day

reattached with increased oxygen (1d+6dO2) and 3 day detachment followed by 28 day reattached with increased oxygen (3d+28dr). To find the most discriminating Vivos between two conditions (C_1 and C_2), first the average of each Vivo for all the images with that condition must be determined. For example, we would take all the normal condition images and sum up each of their coefficient values for each Vivo and solve for the average of each Vivo. Next we would calculate the ratio of the difference squared of each Vivo by the square of their Euclidean distance (L_2), the difference between points. If m equals the total number of Vivos and $C_jV_i(avg)$ equals the average of that Vivo for all the images with a particular condition C_j then (L_2) equals $\sum_{i=1}^{m}\sqrt{(C_1V_i(avg)-C_2V_i(avg))^2}$. Next to find the discriminating Vivos we calculate the ratio of $(C_1V_i(avg)-C_2V_i(avg))^2/(L_2)^2$. The larger this value the stronger the discriminating power is for that Vivo.

Figure 2 shows the most discriminate Vivos by number in order of their discriminating power. The boxes represent the 9 different conditions. The edges represent the discriminating Vivos.

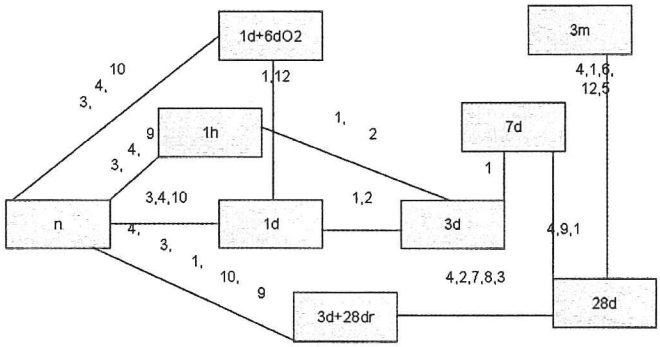


Figure 2: Most discriminate VIVOS

Future direction for this project could be to use independent component analysis (ICA) instead of PCA. Highlighting regions of interest could also be a very useful addition for biologist. This would highlight a region of an image and bring it to the biologist attention so that they can decide its usefulness. A graphical user interface (GUI) would also help to make it easier for people to use the program. A searching algorithm that would allow the user to choose one to several Vivos and would then display or list all the images that have a high coefficients for the Vivos selected would also be very useful.

Two useful programs that I wrote in matlab that could be used independent of this are dbgettable.m and dbputtable.m. Dbgettable.m is a matlab script that can retrieve values from a table of any size on a psql database. It retrieves all the fields and they can only be of numerical values. The values are put into a matrix that matlab can use and you can manipulate. Dbputtable does the reverse, it first empties the table and then put some numerical values back into the table.

One of the main benefits of using the visual vocabulary method is that computers can now identify meaningful differences of images. Visual vocabulary is not only limited to biological images but can be applied to any set of images.

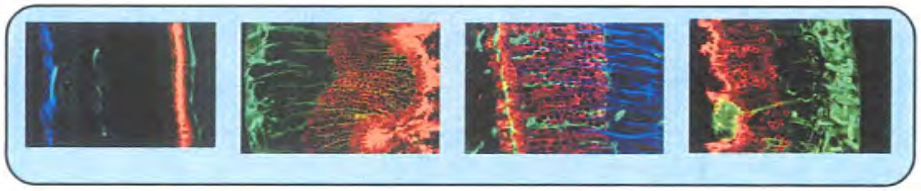
VISUAL VOCABULARY CONSTRUCTION USING PRINCIPAL COMPONENT ANALYSIS

David W. Renteria, Computer Science, California State University, San Bernardino
Vebjorn Ljosa, Ambuj K. Singh, Computer Science

Research Objective

- Produce a set of visual term to describe a set of images.
- Allow a computer to capture differences between images without human assistance.
- Analyze the images.
- Have the images in a web based database.
- Apply it toward any set of Images.

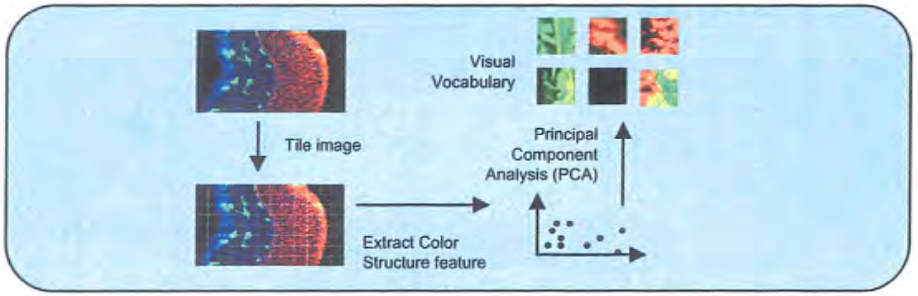
Retinal Images:



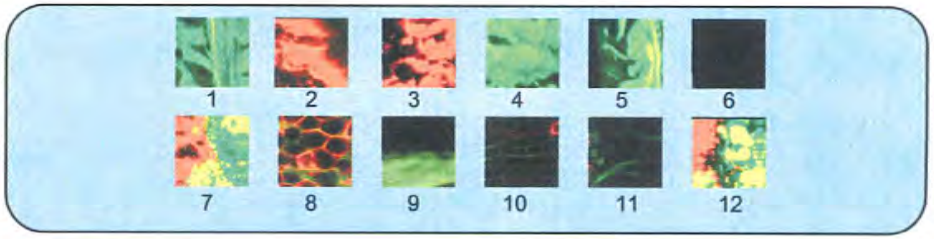
- Example images of retina showing normal (n), 3 day detachment (3d), 1 day detachment followed by 6 day reattached with increased oxygen (1d+6dO2), and 3 month detachment (3m), respectively.

Constructing the Vivo

- First the image is partitioned into non-overlapping tiles that contain characteristic textures of the image.
- Second the characteristics are extracted and a feature vector is constructed representing its image content.
- Next a set of symbols is derived from the feature vectors by applying principal component analysis (PCA). These symbols are called Vivos and a set of symbols is called Visual Vocabulary.
- Finally each image is expressed as a combination of the tiles that makes up the Image.

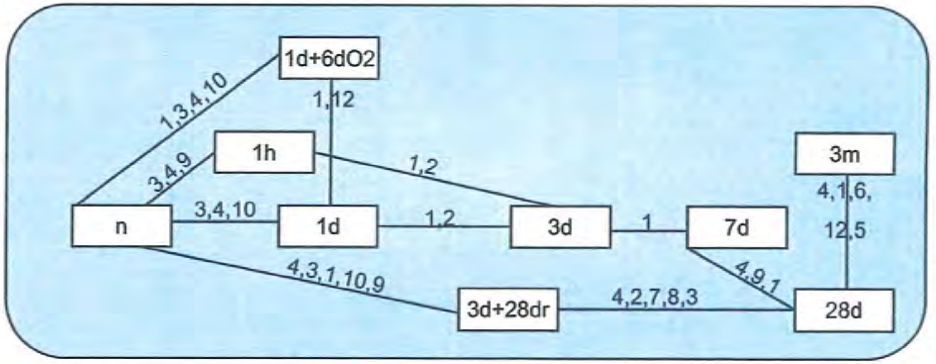


Symbolic Representation of Vivos



- The Vivos do not exactly equal any of the original tiles.
- We symbolically represent the Vivo by displaying the tile that strongly expresses its characteristics.

Most Discriminate Vivos



- To find the discriminating power of a vivo between two conditions we must find the average of each of the two conditions.
- Next we calculate the ratio of the difference squared of each Vivo by the square of their Euclidean distance.
- The larger the value the stronger the discriminating power is.
- The boxes above represent the nine different conditions.
- On the edges are the most discriminate Vivos by number in order of their discriminating power.

Future Direction

- Use independent component analysis (ICA).
- Highlighting interesting regions by Vivos.
- Different methods for analyzing the data.

This work is supported by the National Science Foundation and Technology Research (ITR) program Grant No. ITR-0331697.

Edge Detection and Feature Extraction of Retinal Images

Alexander David, Computer Science. University of the Virgin Islands
Mentor: Baris Sumengen, Electrical and Computer Engineering
Faculty advisor: B.S. Manjunath, Electrical and Computer Engineering

Analyzing retinal images plays an important role as scientists try to understand and treat retinal detachment. This is a condition that occurs either because of physical injury or disease, where the retina becomes detached from the inner wall of the eye. This condition almost always leads to vision impairment or loss. During detachment the retina under goes a great change in its physical structure that scientists have dubbed "remodeling". It is these changes that scientists are trying to better understand in order to develop new techniques for retinal reattachment and/or diagnosis. This is where image processing techniques come in handy. Using various image processing techniques, scientists have been able to study detached retinas in ways that were previously impossible. So with this in mind, the purpose of our research this summer was to enhance some of the ways that scientists can study these images. Our goals were to extract a set of features from these retinal images that can be used for a similarity search and to test a state of the art edge detection algorithm on our retinal images.

Feature extraction is the process of extracting numerical values from images that can later be used for image analysis. To extract our features first we set a global threshold for all of our images in an attempt to enhance the foreground of the image. This process sets all pixel values that are less than the threshold to zero. Once this is done we separate the image into its connected components in order to extract features from each connected component. We then computed areas for each component and removed all components from the image that are less than 20

For the other part of our project we implemented a state of the art edge detection algorithm on our retinal images. Edge detection is the process of finding the edges within an image in an attempt to pull out significant objects in that image. Generally this is accomplished as follows: values are assigned to the pixels in an image and the distances between them are computed using distance functions like the Euclidean distance. The maxima among these distances are found and recorded and then all other values besides the maxima are set to zero. The result is an image consisting of thin lines that represent the edges that the program detected from the image. The edge detection algorithm that we used this summer is called the Compass Edge Operator. How this program works is that it overlays a circle (of a user defined radius) onto the image and computes signatures for each half of the circle. There are different criterions for determining what values to assign for the signatures but for the program we used, the basis was how much of a certain pixel value or color was in the half circle. Once it has the signatures for each half, it computes the distances between these values using the Earth Movers distance. Once this is done, the program records the maxima, rotates the diameter and then does all of the previous calculations again. It continues to rotate the diameter until it reaches it original orientation. It records all the maxima for all the orientations and then it repositions the circle on the image and repeats the calculations. It continues to move the circle over the image until the entire thing is processed, and once this is done it returns and image that represents the edge strengths over every orientation. The pixel brightness in this new image represents how strong the edge is at that point; the brighter the pixel, the stronger the edge. Once we acquired this image we ran it through an algorithm that performs non maxima suppression, and so from this program we got the true edges detected by the Compass

Edge Operator.

The results of our research were promising indeed. The features we extracted served as a good measure of similarity over all. The center of gravity wasn't very robust against images with similar shape but different locations in the (x,y) axis. The edges extracted from the Compass Operator were very good although they were not perfect. In the future we will extend the GUI to handle data files of already extracted features, and we will make it accessible on the web. We also will connect the edges extracted from the Compass Operator and extract features from these new images and use them in the similarity search.

Edge Detection, Segmentation, and Feature extraction on Retinal Images

Alexander David, University of the Virgin Islands

Baris Sumengen, B.S. Manjunath, Computer Science, UCSB

Motivation

Automated analysis of retinal images plays a vital role as scientists try to understand retinal detachment and discover better means of reattaching retinas. The purpose of this project is to:

- advance techniques in image processing by which these images can be studied

Detached Retinal Sections

1 day after detachment | 3 days after detachment | 7 days after detachment

Research Objectives

- Extract numerical features that can be used for a similarity measure
- Test a state of the art edge detection algorithm on retinal images and try to use these edges for a similarity measure
- Develop a Graphical User Interface (GUI) to be used for the similarity search

Retinal Images

- Get images from microscope

Image Processing

- threshold images to enhance foreground
- edge detection

- extract features: center of gravity, spread and average intensity value

Red circles indicate intensity spreads

Spreads change as thresholds are changed

Edge Detection

- test a state of the art algorithm on retinal images.
- Extract edge features

Edges were extracted using two separate algorithms

Similarity Search

- Use distances between features as a similarity between images
- Implement a GUI for biologist to utilize for searching

Thresholding and Feature Extraction

- Separate the image into RGB channels for individual analysis
- Set a global threshold to enhance the foreground in channel and then extract the connected components from each channel

no threshold | thresh = 15 | thresh = 30 | thresh = 50

- Extract features for each connected component: center of gravity, spread, and average intensity value and store them in a feature vector for each connected component
  Ex. [45, 61.34, 169.3, 24]
- Compute the distances between these features using the Euclidean or the City Block distance functions

Euclidean distance

$$D(x,y) = \sqrt{\sum_{i=1}^{m}(x_i - y_i)^2}$$

City Block distance

$$D(x,y) = \sum_{i=1}^{m} |x_i - y_i|$$

Search Algorithm

- Computes features and number of components for the target image (T) and all of the other images (I(x)) in the search
- If the difference between the number of connected components for T and I is more than a certain value, then I(x) is disregarded
- Else let the number of components for T and I(x) be equal to the lesser number of two components
  ex. IF n < m THEN let m = n OR IF n > m THEN let n = m
- Normalize the feature values by dividing each feature by the standard deviation over the entire set of that feature
  ex. Spread / std(all Spread)
- Checks for user input to decide which distance to use
  ex. IF x == 1 THEN compute City Block distance between feature vectors
  Else compute Euclidean distance between feature vectors
- Sort I(x) images in descending order in terms of distance away from T, and return a list of image names

Feature Extraction Results

- Using different thresh holds, we were able to see different effects on the center of gravity and spread
- Center of gravity, spread and average intensity served as good measures of similarity between images
- Spread and average intensity were more robust against outliers in the data

- GUI was developed in Matlab
- It displays any image in a given file or directory and saves it for comparison
- This GUI will allow biologists to use the similarity search on large number of images
- Returns a list of image names that are similar to the chosen image
- Names are in descending order in terms of distance from T

Edge Detection

- Edge detection is the process by which the edges of an image are extracted in order to give information about the different objects
- Traditionally edge algorithms compute the distances between color intensities and then compute the depravities for these values.
- Then the algorithm looks for maximums among these values recognizes them as edges, then the edges are thinned by suppressing all non maximum values around them to 0.
- We used a state of the art algorithm (Compass Edge Operator) on our retinal images.

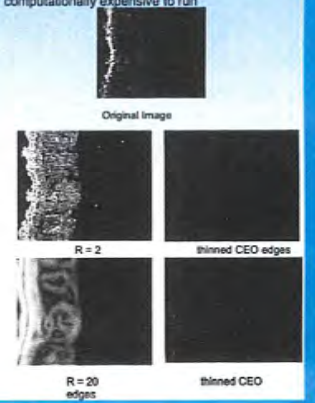
Compass Edge Operator (CEO) Algorithm

- Given a circle on the image (of a user determined radius) divided into two halves
- Assign pixel weights to each pixel in the circle
- Compute the Earth Mover's Distance between the two halves
  The numerical value of how much work it would take to change one value into the other corresponding value ex. EMD for 3, 5 would be 2.
- Record these values and then rotate the diameter of the circle a certain number of degrees so you have new halves

- Repeat steps 1 through 4 for every rotation until initial orientation is reached
- Find the maximums along every orientation and record several of the highest values along that orientation along with the orientation
- Move the circle and repeat all previous steps. Continue until the entire image has been processed.
- Suppress all other values besides the maxima to zero
- Return the edge image

Edge Detection Results

- The edges we extracted using the CEO algorithm were good but not perfect
- When compared to the Canny algorithm, the CEO algorithm performed better on our images
- Canny algorithm is however less computationally expensive to run



Original Image | R = 2 | thinned CEO edges | R = 20 edges | thinned CEO

Conclusions and Future Work

- The features we extracted are well suited for distinguishing between image patterns.
- Center of gravity wasn't robust in terms of being able to accurately distinguish between images that have similar shape but different locations in the axis, because it is more a measure of location than shape
- Edges extracted by the Compass Edge Operator were well defined but the algorithm is costly to run.
- Edges for the Compass Edge Operator still should be connected in order to possibly yield new information about the images.
- Features can be extracted from edge images and used in the similarity search algorithm.
- Develop GUI to support data files with metadata about images. Instead of computing features for each image, it will simply extract the data from files. Include option to choice different preset thresholds that correspond to different data files.

References and Acknowledgments

Funding provided by NSF ITR research grant 0331697

Mark Verardo, UCSB, and rest of the ITR staff and faculty for all their help with my work and making this a great summer and an amazing learning experience

Ruzon, Mark A., Tomasi, Carlo. Edge, Junction, and Corner Detection Using Color Distributions. IEEE PAMI, Vol 23 No. 11. Nov, 2001

Invisible Embedding of Meta-Data into Biolological Images

George Kaymaz, Computer Science, California State University, San Bernardino
Mentor: Kaushal Solanki, Electrical and Computer Engineering
Kenneth Sullivan, Electrical and Computer Engineering
Faculty advisor: B.S. Manjunath, Electrical and Computer Engineering

Introduction: The main idea of the data hiding scheme for bioimages is twofold. First, we want to reduce any visual distortions the image undergoes during the data hiding process. Secondly, we want to associate with every bioimage a tag, which will help in identifying how an image was acquired. The implementation of a tagging scheme will also help identify copyright violations. Thus, the two latter methods allow for easy transaction tracking for the bioimages. Work in implementing this state-of-the-art data hiding technique was previously done at the vision research laboratory, but the adaptation for bioimages is not fully completed. The goal for this project is to complete the data hiding scheme with an adaptation for bioimages. Currently the method for accomplishing this task is hiding data into the bioimages, and experimenting on the effect of varying some parameters such as, compression level, threshold, and the amount of embedded data. The main focus will be on minimizing the distortion to the image and maximizing the amount of data which is embedded. After accomplishing this task, we will have a base set of parameters we can use for a data hiding system integrated into the bioimage database. Work previously done contributed the higher thresholding scheme, which turned into the selective embedding scheme for certain transform coefficients. This enabled bioimages with a large amount of black low frequency space not to incur noise or blurring. As a result of lowering the number of transform coefficients chosen to embed data, a smaller amount of data is allowed for hiding. Using this scheme allowed a number of 200 characters (bytes) to be successfully embedded without any serious degradation to the bioimage. We want to achieve a higher amount data embedding and maintain high fidelity. As stated above, the selective embedding scheme hides less data. This problem is solved by selectively setting parameters which attribute to the image quality factor, code redundancy, and number of bits hidden per 8x8 pixel block. Keeping a higher threshold and adjusting the parameters will correctly allow us to achieve a higher volume of embedding.

Results: The following Table 1 shows my preliminary results with perturbing the encoder/decoder image parameters. This set of data represents particular factors where no decoding failure is achieved, and includes the maximum amount of hidden data.

From this experiment it is gathered that the set of data has three sets of parameters corresponding with images that have high, medium, low amount of black frequency space. We want to achieve one set of parameters for all these varying types of images. Since we are not considering compression attack for the bioimages, we can use higher quality factors, we also want to achieve a lower code-redundancy so we can embed more data. Another trail set of parameter factors were chosen, quality factor is now set to 95 and the code redundancy is adjusted so that no decode failure occurs (see Table 2).

This second set of data shows that we can embed more than 1KB of data within the tested bioimages. The varying set of images now have one set of parameters which embedded the maximum number of bytes without incurring major noise or blurring. The cost is incurred with a higher image

Table 1: High thresholding method with two different quality factors

Images names	TRAIL1 (B)	Code-Redundancy (q)	TRAIL2 (B)	Code-Redundancy (q)
1DMIP1	505	40	755	35
3CCNP1	329	54	694	38
3GVP23	1369	15	1369	15
41CABC01	155	123	296	87
41NFGF04	1369	15	1396	15
7CAP31	1369	15	1396	15
NCDP7	1369	15	1396	15
NRHGF53	505	40	755	35
NPNC03	505	40	755	35

*Threshold is kept at 1. B: Bytes. TRAIL1: quality = 75 and coef/block = 27. TRAIL2: quality = 85 and coef/block = 35.

Table 2: High thresholding method with quality factor 95

Images names	TRAIL (B)
1DMIP1	1173
3CCNP1	
3GVP23	
41CABC01	
41NFGF04	
7CAP31	
NCDP7	
NRHGF53	
NPNC03	

Threshold is kept at 1. B: Bytes.
Quality = 75 , coef/block = 35 and q=23.

quality factor, but as stated above compression attacks will not occur, so it's a factor which can be allowed. The following are two examples of images which have undergone histogram equalization, which pick up and intensify pixel values. Each represents a class of image represented by low-to-high black frequency space.

We found from this set of images that before and after the encoding process there is no significant noise or blurring of the images. We can now use one set of parameters to embed a considerable amount of data in a large variety of bioimages.

Conclusion and Future Work: We have achieved a higher volume of embedding in-conjunction with the selective coefficient embedding scheme. The higher threshold allowed the image to keep a high fidelity while scarifying a potential for higher payload of embedding. Future work in achieving higher volumes of embedding can focus on attempt in hiding data in all three color channels in RGB color space. In our scheme only one channel is utilized to embedded data, which is the gray-scale channel. One concern here is that more noise will be added with the watermark, but three

times the embedding volume can be achieved. Further work in achieving higher volumes of data embedding can also be achieved by compressing the message before encoding. Depending on the range and diversity of the data being embed higher compression of the message can be achieved, and later attribute a higher volume of data embedding.

Invisible Embedding of Meta-Data into Biological Images

George Kaymaz, CSU San Bernardino
Kaushal Solanki, Kenneth Sullivan, B. S. Manjunath , ECE UCSB

Research Objectives

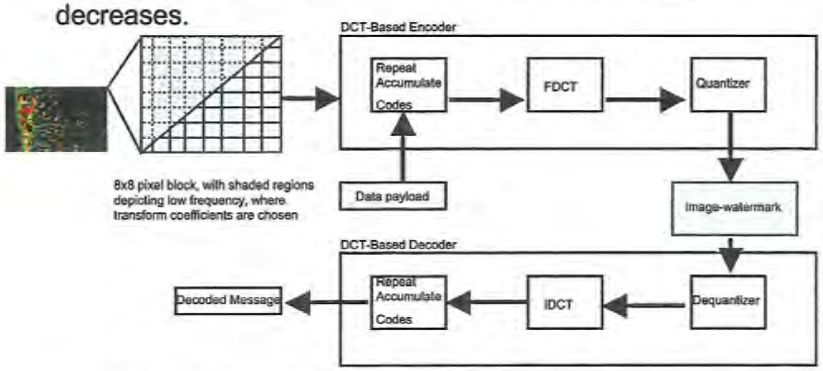
- Embed a high volume of data and at the same time maintain image fidelity.
- Robust method for data hiding for a diversified set of biological images.

Impacts

- Transaction Tracking method for biological images, by tagging a watermark.
- Reduction in database schema for Biological Images.

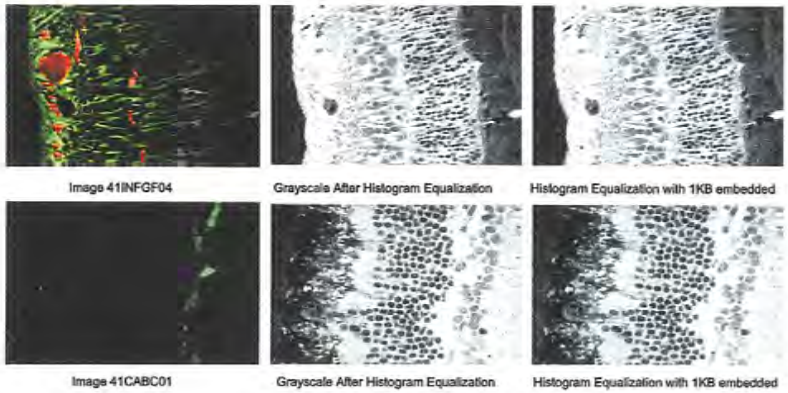
Methodology

- Higher thresholds improve quality of image, but less transform coefficients are used to embed data, so number of bits decreases.



- By testing a large grouping of biological images, and perturbing image quality, number-bits per 8x8 block, and redundancy codes. Higher embedding is achieved.

Result



- Notice after embedding and applying histogram equalization, image fidelity is achieved with a 1KB payload.

Conclusion

- In conjunction with the selective embedding scheme a proper set of parameters can now embed data for a large set of biological images.

Future Work

- Utilize all color channels in RGB color space for the data hiding scheme.

This research is supported by NSF
Information Technology Research grant #0331697