



Lister Hill National Center for

Biomedical Communications

An Intramural Research Division of the U.S. National Library of Medicine

A Report to the Board of Scientific Counselors April 2014

3D Informatics

Terry S. Yoo, PhD
Michael J Ackerman, PhD
David T. Chen, PhD
Bradley C. Lowekamp
Oleg Kuybeda, PhD
Joseph Santoroski

Office of High Performance Computing and Communications

U.S. National Library of Medicine, LHNCBC
8600 Rockville Pike, Building 38A
Bethesda, MD 20894



U.S. National Library of Medicine



Table of Contents

1	Introduction: Open Science through 3D Informatics	2
2	Overview: 3D Informatics	3
3	Background – Embracing Data Science.....	4
4	Resources – Facilities and Personnel	4
5	Status Report.....	5
5.1	The Visible Human Project.....	5
5.2	The Insight Toolkit	6
5.3	ISP – Interactive Scientific Publication	6
5.4	The Insight Journal	7
6	3D Informatics Projects	7
6.1	3D High Resolution Electron Microscopy.....	7
	Transmission Electron Tomography for Molecular Biology	8
	Ion-Abrasion Scanning Electron Microscopy for Whole Cell Imaging.....	9
6.2	SimpleITK – SITK.....	10
6.3	Computational Photography Project for Pill Identification (C3PI).....	10
6.4	3D Printing at NLM	11
7	Mentoring and Training in 3D Informatics.....	12
8	Program Significance and Project Plan.....	12
9	Summary	13
10	Publications.....	14

1 Introduction: Open Science through 3D Informatics

This report describes ongoing efforts at the Lister Hill National Center for Biomedical Communications and the National Library of Medicine to promote, advance, and to provide leadership in non-text, data-intensive informatics. This program concentrates heavily on information science surrounding image data of 2, 3, and higher dimensions. Our mission reaches beyond high dimensional image processing or medical computer graphics; it attempts to synthesize semantic information and dynamic user navigation with visualizations created from real medical data. In the cases where we use models of biological structures, all of these models are generated from sources of real data, firmly grounding this work in the practical aspects of medicine. To adequately describe this research area, we have coined the new term, *3D Informatics*, to encompass the blended notions of information science and medical visualization.

The program's underlying principle is, "Open Source + Open Data = Open Science." We hope to promote scientific accountability, accelerate discovery, and encourage collaborative team research through data sharing, public software, and multidisciplinary research initiatives. We have conducted a series of initiatives in acquiring additional shareable, public data as well as extensive projects in open source software development. These efforts have resulted in unqualified successes across the medical imaging research community worldwide. We briefly review these developments including:

- an update on the Visible Human Project Data (over 3,400 licenses worldwide);
- the current status and an assessment of the VHP Insight Tool Kit (ITK) (an international community with over 1,500 users, including the NIH Roadmap National Centers for Biomedical Computing);
- a review of the Interactive Scientific Publication project conducted with the Optical Society of America (OSA) to develop data-full documents; and
- a review of the Insight Journal, a sponsored effort to promote reproducible results.

Beyond a review of our established successful external projects, we describe our current internal/intramural directions in 3D Informatics. We cover:

- a progress reports on data analysis for 3D nanometer-scale imaging with the High Resolution Electron Microscopy group at the NCI Center for Cancer Research;
- HPCC software development of SITK (SimpleITK), a project to promote ITK programming across multiple disciplines as well as connect with undergraduate programming education;
- an introduction to the Computational Photography Project for Pill Identification (C3PI); and
- an overview of 3D printing technologies at NLM including the *HHS Ignites* award-winning collaborative project, the NIH 3D Print Exchange with the Computational Biology Section of the NIAID Bioinformatics and Computational Biosciences Branch.

We conclude with projections for future projects, including extending our explorations in modern high performance computing by synthesizing multi-GPU (graphics processing unit) systems with many-core multi-threaded processing for single-particle microscopy and the sponsoring of a public open source computer vision challenge in pill identification.

2 Overview: 3D Informatics

Since 1999, the NLM Lister Hill Center's Office of High Performance Computing and Communications has been developing a high-performance computing program to manage and analyze complex, high-dimensional medical data. More than a single project, this program is built on "pillars of excellence," supporting a broad research umbrella. The general areas of (A) multidimensional image processing, (B) digital modeling and shape analysis, and (C) scientific and medical visualization are the three columns supporting the comprehensive effort in understanding and communicating visual information. Underlying all of this is a traditional informatics foundation of data collection, distribution, and content analysis of volume image datasets (See Figure 1). The individual research areas support the larger program, and like many endeavors, the whole is greater than the sum of its parts. We describe each of the pillars as a support or leg of a structure, and each is required to uphold the mission. Many projects comprise the separate research areas, and each area advances at its own pace depending on the availability of resources and personnel to carry the work forward. Together incrementally, the research areas lift and strengthen our overall effort at the Lister Hill Center, elevating us to provide clearer vision and leadership in the informatics community.

This program is not limited to just three dimensions, and we routinely invoke time varying data and multimodal radiological information in our investigations. Much of our work has been performed in collaboration with faculty and research scientists from across the country as well as across NIH. We have begun the exploration of the infrastructure and organizational issues surrounding volume data archives. We are continuing our research in volume image segmentation, registration and analysis. Our ultimate goal is to broaden our focus to incorporate information and models at a variety of scales, permitting the

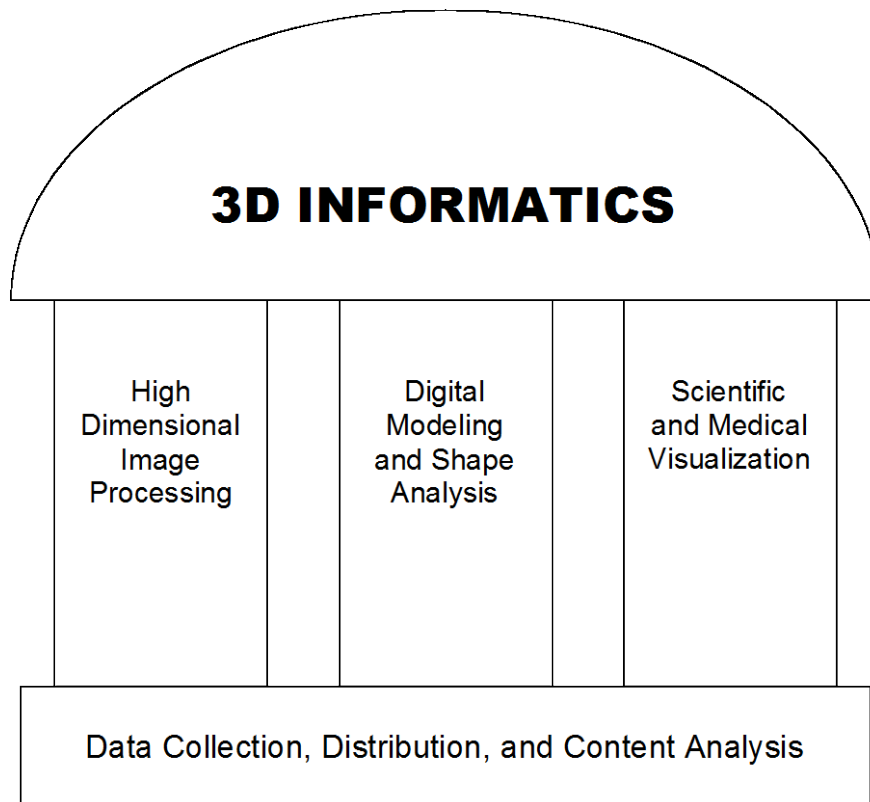


Figure 1. The position of 3D Informatics and its relation to its research subfields. 3D Informatics is defined as the gathering, manipulation, classification, storage, retrieval, representation, navigation, and display of complex, high-dimensional medical data.

analysis of complex information across disciplines linking non-visual information with images and admitting the use of information that is studied at the microscopic or nanometer scale to be incorporated in models and simulations that are performed at the system or organism level.

3 Background – Embracing Data Science

In recent months, the scientific community has been placing a great deal of emphasis on the concepts of *big data* and endorsing an area of investigation called *data science*. The challenges of big data are not the sheer size of the information being studied, but rather the heterogeneity of the sources of information. Our position is that *images*, 2D, 3D, 4D and higher dimensional data, form some of the most complex pervasive heterogeneous data that must be managed (classified, indexed, made searchable). Linking incommensurate data, including images, will be a factor in data science as the field unfolds. From high-dimensional input, our team is able to explore complex analyses and generate visualizations to better understand images as data.

4 Resources – Facilities and Personnel

Since the year 2000, the Office of High Performance Computing and Communications has worked to create a comprehensive internal laboratory in 3D Informatics, we have grown from a single researcher to a small lab with six investigators (two principal and four contract personnel). Our laboratory has ongoing research associations with laboratories and centers at Harvard Medical School, the Ohio State Univ., Brigham Young Univ., UNC-Charlotte, and the University of Maryland Baltimore County. In the past fourteen years we have hosted dozens of summer graduate fellows and NIH Summer Interns in addition to Summer Faculty Fellows. Notable alumni from our programs have gone on to serve in the radiology residency program at the Mayo Clinic as well as to receive a doctorate from Oxford University in England studying the magnetic resonance properties for imaging Carbon-13. We maintain a variety of associations with neurosurgeons, radiologists, forensic pathologists, and computer scientists nationwide. Dr. Terry Yoo has directed all the work performed by the Program on 3D Informatics.

Our current staff in the 3D Informatics Group includes:

- Michael J. Ackerman, PhD, Chief, OHPCC
- Terry Yoo, PhD Principal
- David T. Chen, PhD, Senior Scientist
- Bradley C. Lowekamp, Research Scientist
- Joseph Santoroski, Systems Architect
- Oleg Kuybeda, PhD, Scientific Programmer

An important step in creating an imaging lab for the Lister Hill Center has been the establishment of an applications development center. The former Learning Center for Interactive Technology was redesigned as the OHPCC Collaboratory where a variety of computing and telemedicine resources are now available. OHPCC has employed a succession of high performance computing facilities over the years. Currently, the 3D Informatics Group is exploring modern high performance computing, performing advanced image analysis on small-footprint systems employing a combination of multi-core technologies and Graphics Processing Units (GPUs) for both rendering and for computation. Notably, our computer and visualization services include:

- a rack-mounted multi-core compute server with 4 CPUs with 10 cores each, 10 virtual cores for a total of 80 cores supported by 256 GB RAM;
- a desk-side dual-CPU visualization server with 8 cores along with 64GB RAM an internal GPU and 4 external GPUs connected via 2 expansion chassis, each with 2 NVidia GeForce Quadro Plex GPUs each with 512 shader cores and 6GB video memory;
- a desk-side compute server with 24 virtual cores, 96 GB RAM, and 4 NVidia GeForce Titan

- GPUs each with 6GB of video memory and 2,688 shader cores;
- a fiber channel Storage Area Network (X-SAN) supporting 8 machines and 50 TB of online storage allowing data-attached storage I/O speeds to any of our compute servers or workstations;
- three immersive display systems with stereo viewing capabilities (large rear-projection passive stereo, an active stereo 4K display, and a desktop WUXGA display with active stereo); and
- two 3D printers: a 3DS Cube and a ZCorp Z650.

The reconfigured and equipped Collab has been essential to the prosecution of the initiatives described in this report. This resource makes possible the participation of summer faculty, associates, fellows and interns by providing space and unique computing and visualization equipment for visiting researchers and students. With the creation of successful initiatives and the establishment of effective development spaces, our program has met its research and development goals.

5 Status Report

The Office of High Performance Computing and Communications has supported a series of internal and external programs in 3D Informatics. These efforts are interrelated, supporting and complementing one another in a general mission of directing the field through data collection and open-source software development. Though ongoing, OHPCC has had considerable success in this domain, creating and sponsoring tools that have begun to influence not only our target domain, but which have also been adopted in other areas of computer science.

Several projects are either ongoing or have been the subject of previous reports to the Lister Hill Center Board of Scientific Counselors. We briefly provide a status report for these projects.

5.1 The Visible Human Project

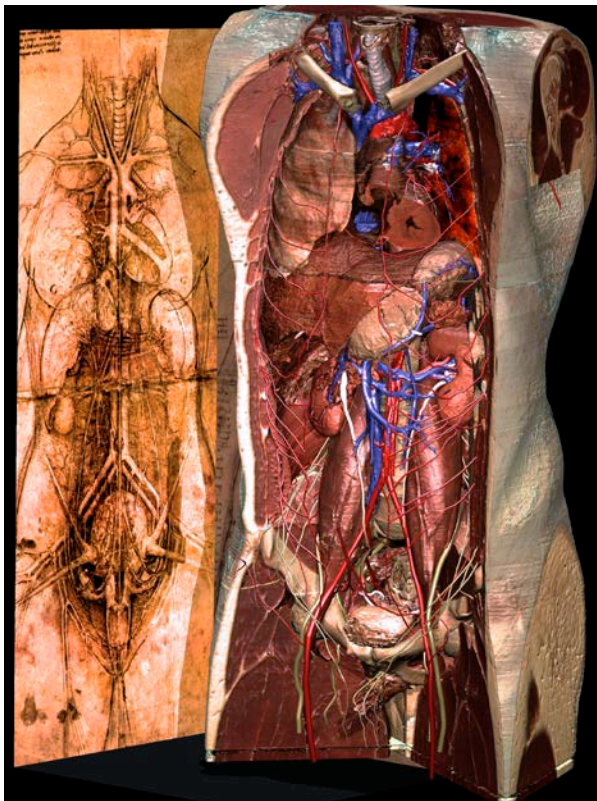


Figure 2. Rendering of the segmented Visible Human Project male dataset using Voxelman, a volume rendering project from the Technical University at Hamburg, led by Dr. Karl Heinz Hoehne, principal investigator.

The Visible Human Project® (VHP) enables education, diagnosis, treatment planning, virtual reality, and virtual surgeries, through the use of its publicly-available Visible Man and Visible Woman reference data sets obtained from human cadavers. While VHP data acquisition was completed in 1996, the VHP data sets are still being licensed to researchers and research articles are still acknowledging their use of VHP data.

The Visible Human Project® was an outgrowth of NLM's 1986 Long-Range Plan. Its long-term goal was to produce a system of knowledge structures that transparently link visual knowledge forms to symbolic knowledge formats such as the names of body parts.

The Project involved the creation of complete, anatomically detailed, three-dimensional representations of the normal male and female human bodies. It included the acquisition of transverse CT, MRI, and cryosection images of representative male and female cadavers. The male was sectioned at one millimeter intervals, the female at one-third millimeter intervals.

The University of Colorado Health Sciences Center served as the contractor for this project.

5.2 The Insight Toolkit

The Insight Toolkit, ITK, is NLM's long-running experiment in open source software for medical image analysis. ITK began in 1999 with six research and development contracts to design and develop a library of computer algorithms for the digital analysis of the Visible Human Project data. That effort is entering its fifteenth year. It has expanded to apply to surgical planning, cancer detection and diagnosis, remote sensing of satellite imagery, segmentation and analysis of cosmological data, 3D microscopy and microanalysis, and even handwriting recognition. Nearly 1.5 million of lines of C++ code comprise the ITK software library with over 5,000 downloads every month to over 25 countries (1,274 downloads during the week of February 10, 2014).

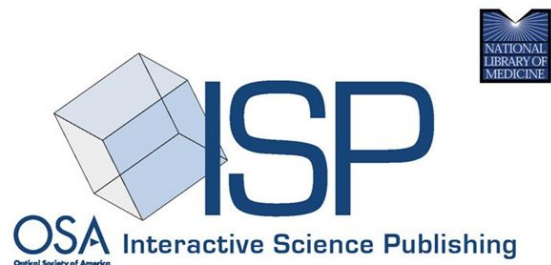


In 2010, the Office of High Performance Computing and Communications sought support from the American Reinvestment and Recovery Act (ARRA) to revise and update ITK. With the help of stimulus funding, six new research and development contracts and ten new software applications awards were funded and managed by the OHPCC 3D Informatics Group. The resulting major release which utilizes multi-core and GPU technology, ITK version 4.0, was announced in 2012. Currently in version 4.5.1, ITK remains a cornerstone to the foundations of medical image processing research worldwide.

5.3 ISP – Interactive Scientific Publication

Interactive Scientific Publishing (ISP) was developed and demonstrated as a joint project between the Optical Society of America and the National Library of Medicine at NIH. It allows authors to electronically publish papers which are linked to the referenced 2D and 3D original image datasets. These image datasets can then be viewed and analyzed interactively by the reader. ISP provides the software for authors to assemble and link their source data to their

publication. But more important is that it provides readers with image viewing and analysis tools based on ITK. The goal of ISP is to improve learning and understanding of the information being presented. ISP was shown to have enough educational value that readers were willing to invest in the needed set-up and learning phases. The social aspects of data sharing and the enlarged review process may be the hardest obstacles to overcome.



5.4 The Insight Journal

In 2004, NLM endorsed an experiment in interactive publication where open peer-reviewed articles on image analysis algorithms are linked with working source code. This online publication, called the Insight Journal, is an exploration that combines open access journalism with open-source software, promoting reproducible computer science. Computer programmers communicate largely through programs, not through text. By publishing both the article and the programming that underlies the results, the Insight Journal seeks to accelerate discovery in computer science.



Entering its ninth year, the Insight Journal continues to thrive. Hundreds of articles have been published since December 2005 when the Insight Journal began accepting combined articles with linked source code. New ideas can be submitted to the journal as well as deliverables for NIH-funded research contracts and grants, preserving the contribution not only as a peer-review manuscript, but also as the original programming that accompanies it. The Insight Software Consortium has adopted the Insight Journal as the gatekeeper and arbiter of new contributions to the toolkit. All proposed additions to ITK must first be submitted to the Insight Journal, and reviewed through a public forum. The Insight Journal is proving to be a valuable experiment in open access publishing as well as a vital clearinghouse for medical image processing algorithms.

6 3D Informatics Projects

While many of our projects have been continuing for a number of years and previously reported to the BoSC, the 3D Informatics Group has begun new investigations as old ones have been retired or as resources and staff have become available. One characteristic shared by these projects is that they have been structured to provide driving problems for the existing programs within our group. Just as the Insight Toolkit was originally intended to exercise and promote the data generated by the Visible Human Project, we have designed a series of initiatives that will promote, enable, extend, and explore uses of our software programs as well as improve our capacity in high performance computing, computer vision, and biomedical visualization.

6.1 3D High Resolution Electron Microscopy

We have chosen the processing of 3D high-resolution electron microscopy data as a driving problem. When we were developing the Insight Toolkit, one of our advisors, Bruce McCormick of Texas A&M University, offered that we would know when we succeeded when people used our toolkit for purposes other than segmenting human anatomy. We have adopted that premise in our own research, exploring new domains where our tools and knowledge can be applied. As a research domain for exercising our image processing software, 3D microscopy has the advantage of very large, often noisy data posing extraordinary problems of registration, segmentation, modeling, and visualization. We have engaged the Laboratory of Cell Biology of the Center for Cancer Research in the National Cancer Institute as partners who pose the driving biological problems and generate microscopy data. Working at the scale of nanometers, the problems posed come from two types of microscopy, *transmission electron tomography* and *ion-abrasion scanning electron microscopy*.

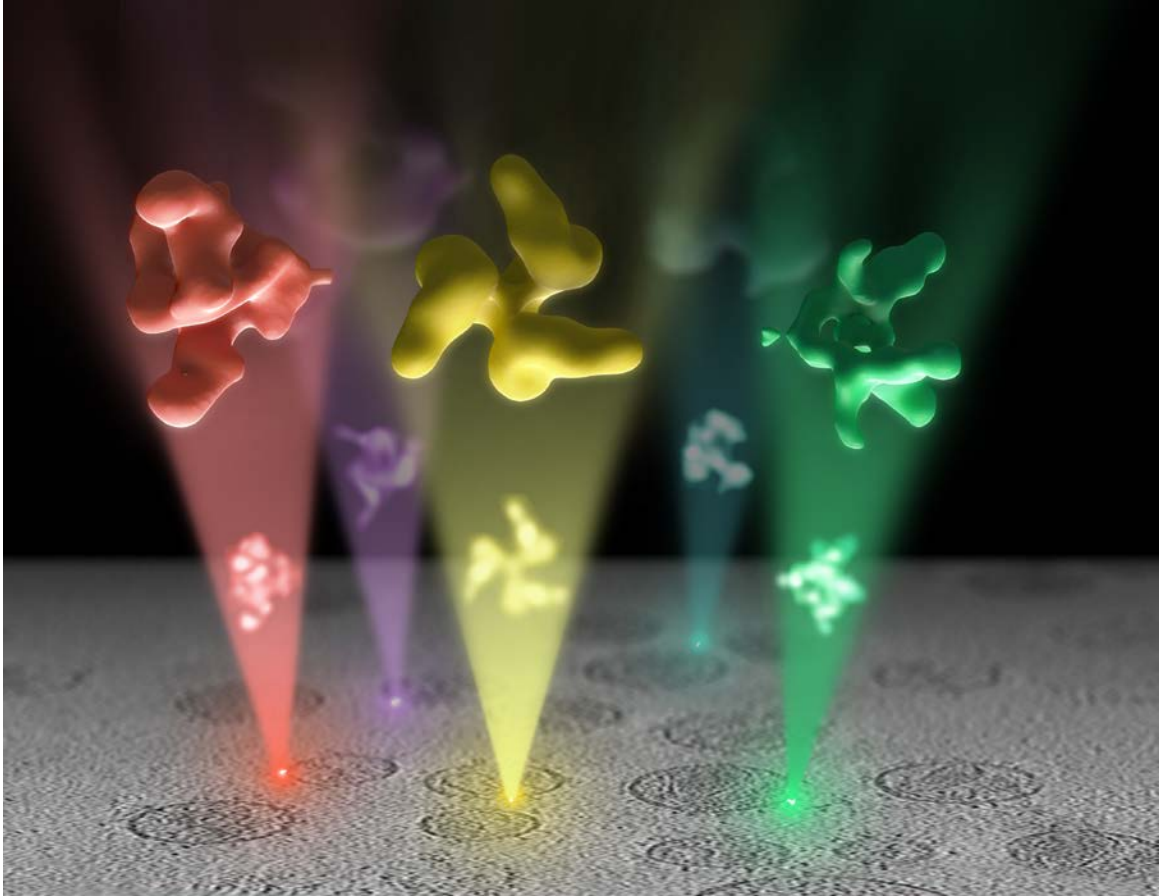


Figure 3. Resolving Structure from Chaos: This image depicts super-resolution HIV glycoprotein trimeric spike reconstruction from transmission electron tomography projections through general purpose computing on GPUs. The horizontal plane depicts a single tomographic projection showing several HIV virions. In this process, multiple projections are used to create 3D volumes, and multiple sub-volumes are accurately aligned and averaged to decode the physical structure of the trimeric spike. The colored rays represent the process by which combinations of volume datasets are combined to resolve structures with resolutions of 10-20 angstroms. This process has decoded the structure of several spikes that are embedded on the surface of HIV. (Cover image: Journal of Structural Biology, 181(2), Feb 2013)

Transmission Electron Tomography for Molecular Biology

By tilting the sample stage of a transmission electron microscope as much as ± 70 degrees, scientists are able to perform computed tomography (Transmission Electron Tomography or TET) on the resulting data and create 3D reconstructions of small volumes at a scale of 10 nanometers. By leveraging a property of the Shannon-Hartley sampling theorem, our team is able to improve the resolution of these images by an order of magnitude through averaging repeating sub-volumes in the data, extracting molecular shapes on the order of 10 Angstroms. To accomplish this, the lead scientific programmer for this project, Dr. Kuybeda, has created a hybrid HPC architecture using GPUs to solve the complex problems of linear algebra and controlling the flow of the computation using the MATLAB Parallel Toolbox. The GPU work is executed using nVIDIA CUDA kernels programmed in C++. The combination of GPU computing and scripting in MATLAB allows the rapid development and adjustment of algorithms while preserving the rapidity of computation. We have used these techniques to explore the structure of the trimeric glycoprotein complexes that are the infectious element found on the surface of HIV virions (depicted above in an artist's rendering of the process).

Ion-Abrasion Scanning Electron Microscopy for Whole Cell Imaging

An alternate approach to generating 3D microscopy data is through block-face imaging where a focused gallium ion beam is used to ablate or abrade surface at intervals as small as 3 nanometers, and the exposed surface imaged with a scanning electron microscope with an X-Y resolution of 3x3 nanometers. The entire process is analogous to the milling process used for the Visible Human Project data for gross anatomy, but performed at nanometer scales. The microscopic process is known as FIB-SEM (focused ion beam-scanning electron microscopy) or IA-SEM (ion-abrasion-scanning electron microscopy). IA-SEM is capable of scanning a much wider field of view than transmission electron tomography. While tomography is used to resolve individual protein structures on viruses, IA-SEM is used to examine entire cells, the spatial relationships among pathogens and organelles within a whole cell at resolutions of 5 nanometers.

We use the Insight Toolkit to examine these data. As a collection of algorithms for segmentation and registration of 3D data, ITK is an ideal resource for the quantitative analysis of IA-SEM data. The single dataset sizes are often on the order of 20 gigabytes, which far exceeds the data sizes posed by 3D radiological scans. These data place demands on ITK, forcing us to examine the software, and strengthen our capacity for out-of-core computing, multi-threaded algorithms, and other improvements to the algorithm library. The image below shows a macrophage from a cell line that has been chronically infected with HIV. The membrane surface of the macrophage was segmented using a watershed method from ITK. HIV virions were segmented using a Hough transform filter from ITK, and illuminated in red. The inset shows the 2D view of a serial section of the data, demonstrating the difference in perception; the 2D view implies the structure could be a filopodia, but the 3D rendering shows a membrane ruffle.



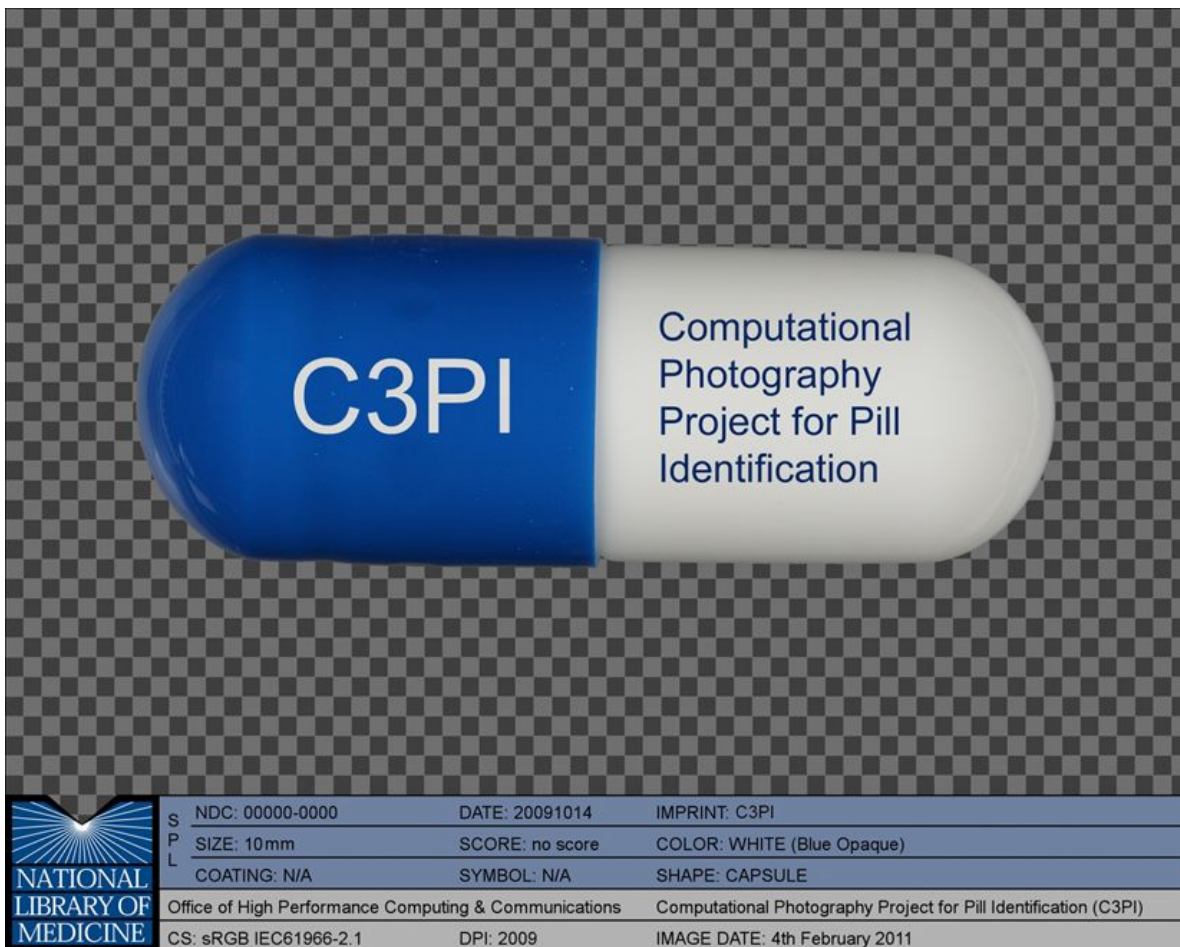
Figure 4. Representation of the surface and interior of an HIV-infected macrophage obtained with newly developed tools for 3D imaging using ion-abrasion scanning electron microscopy. Sections that would appear to contain “filopodia” when imaged by transmission electron microscopy of individual sections can actually correspond to large wavelike membrane processes as illustrated by the cut-away view of a central slice. The membrane protrusions may potentially fold back to the surface of the cell, creating viral compartments (viruses shown in red) by trapping the contents of the aqueous environment within the invaginated folds of the membrane. (Cover image: www.iavireport.org | IAVIRreport 13(6) November-December 2009)

6.2 SimpleITK – SITK

As part of the ARRA-funded ITK-version4 effort from 2010, HPCC initiated a project to make ITK easier for non-expert programmers to use the available powerful collection of algorithms. Known as SimpleITK or SITK, the 3D Informatics Group has now taken the leadership of the programming, design, and support of this project. Research Scientist, Bradley Lowekamp, is the principal architect for SimpleITK. SITK supports multiple language bindings, permitting access to ITK algorithms using accessible scripting languages such as Python. Scripting languages enable the rapid development and adjustment of algorithms or combinations of imaging filters. SimpleITK is now appearing as part of 3D Slicer, the imaging console for the National Alliance for Medical Image Computing (NA-MIC), one of the NIH Roadmap National Centers for Biomedical Computing. SimpleITK targets domain scientists such as microbiologists with modest programming skills, and helps them accelerate their science.



6.3 Computational Photography Project for Pill Identification (C3PI)



The Computational Photography Project for Pill Identification (C3PI) is developing an information infrastructure for identifying unknown oral solid dosage pharmaceuticals (pills) from digital photographs. A byproduct of this effort is the generation of a comprehensive portfolio of digital images of those pills along with supplementary metadata for computer vision research. The verse and obverse pill images in

this collection are already supporting other NLM databases such as RxNav as well as the FDA's effort to include a pill image in a pharmaceutical company's Structured Product Labeling (SPL) submission.

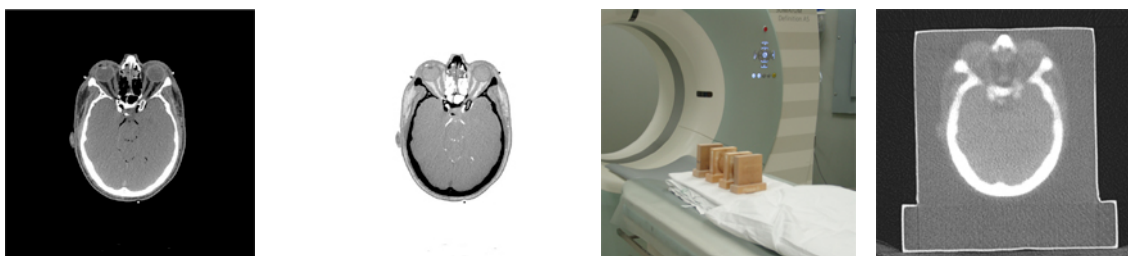
In this effort, OHPCC computer scientists are seeking object identification metrics and methods that are invariant with respect to camera angle, lighting, and the color transfer functions often found in digital cameras. We will be using the Challenge mechanism to enlist community input to solving this problem.

Toward developing new tools for pill identification, the C3PI project is generating a comprehensive portfolio of digital images of oral solid dosage medications from the Nation's prescription formulary. Our goal is for this portfolio to become a public foundation for computer vision research in content-based information retrieval and image-based search.

In addition to these image data, C3PI is generating information on color classification, image segmentation, pill description and dimensions, as well as other metadata. The data collection and the derivative digital science are an investment in consumer safety.

6.4 3D Printing at NLM

Since 2007, HPCC has been pursuing a research and development program using 3D printing for visualization as well as for the generation of data models. Aside from the visualization of complex data, we have engaged in two prominent projects using these technologies.



a. VHP head CT Scan b. Inverted CMY colors c. Imaging the models d. Resulting CT scan

Figure 5. Results of the first demonstration case of 3D CT phantom printing. A slice from a head CT scan (a) is inverted (b) and extruded into a 3D model, printed and scanned (c). The resulting image (d) clearly shows hard and soft tissue matching the original data.

Rapid prototyping of X-ray CT Phantoms – Over recent years, we have been investigating the use of rapid prototyping technologies for the creation of witness objects or phantoms that can be imaged with X-ray CT scanners and that mimic human tissue. Our printers use inkjet mechanisms, so we are able to achieve subtle gradations of X-ray contrast. We are able to model metastatic invasive disease as well as solid masses in our models. The purpose of this work is to provide research models for the accurate development and assessment of automated methods for cancer detection.

The NIH 3D Print Exchange – In 2013 following our commitment to open source and open data, the 3D Informatics Group joined with groups from NIAID and NICHD to successfully compete in the inaugural HHS Ignites program to promote shared biomedical modeling for 3D printing. We were able to secure modest funding from HHS and with the help of our partners deploy an online portal dedicated to the dissemination of free printable models featuring biomedical content. NLM contributed content, but also programming pipelines using ITK and VTK that converts CT data to printable models. This project was successfully presented to the HHS Idea Lab before the HHS Chief Technology Officer and his board in February 2014. (<http://3dprint.nih.gov/>)

7 Mentoring and Training in 3D Informatics

The 3D Informatics group has hosted and provided mentorship for a number of fellows, interns, and summer faculty over the years. Our alumni include:

Corey Arnold, PhD*
Jesus Caban, PhD
Alejandra Costello
Catagay Demiralp, PhD*
Haixia Du, PhD
Russell Ellis
Alex Fletcher
Trevor Hamilton
Okan Irfanoglu, PhD*
Neda Jahanshad, PhD*
David Leiman, MD*
David Liao
Jianfei Liu, PhD*

Raghu Machiraju, PhD
Matthew Madden
Ankush Manchanda
Jonathan Morris, MD*
Bryan Morse, PhD
Kwansik Nho, PhD
Marc Olano, PhD
Marie Schroeder, PhD*
K.R. Subramanian, PhD
Kayla Sukri
Leah Sukri
Emily Watt, PhD*
Zhan Zhang, PhD

*Graduate or undergraduate students who worked with 3D Informatics and later acquired their MD or PhD

Some noteworthy alumni from our program deserve specific distinction:

Dr. Jianfei Liu began his doctoral studies on a project at NLM, and continued with Dr. Yoo as his co-advisor until he completed his PhD. He later went on to hold a postdoctoral fellowship at NIH in the Department of Diagnostic Radiology. Presently he is a Research Assistant Professor of Radiation Oncology at Duke University.

Dr. Jonathan Morris was a medical informatics rotation student who conducted and published research with the 3D Informatics group. He received his MD from Howard University and matched for his residency in radiology at the Mayo Clinic where he remains today as staff.

Dr. Jesus Caban was a postdoctoral fellow with the 3D Informatics group for two years. Upon leaving NLM, he assumed the head of medical imaging at the National Intrepid Center of Excellence on Traumatic Brain Imaging at the National Naval Medical Center. He is expected to receive a faculty appointment at UMBC in the fall of 2014.

Dr. Neda Jahanshad was a summer graduate student with the 3D Informatics program. She conducted research one summer in conjunction with NICHD. She is now Assistant Professor of Neurology at UCLA.

Dr. Marie Schroeder was an undergraduate in biomedical engineering at Duke University. She went on to receive a Fulbright scholarship, matriculated in the doctoral program at Oxford University where she remains today as a lecturer.

8 Program Significance and Project Plan

In our 2007 report to the Board of Scientific Counselors, we suggested that the field of bioinformatics may be at a crossroads. The traditional research areas of gene sequence storage, searching, and retrieval may be slowing as a burgeoning research area. The shift from infrastructure building to applications development is becoming profound as commercial ventures and academic institutions alike begin repositioning their research in drug discovery, patient specific medical treatment, and basic research in

gene expression. To meet these new challenges, it may not be enough to simply pursue a strategy of data mining. Rather, the modeling and simulation of biological, medical, and physiological processes as well as the visual representation of these mechanisms at all scales (from nano to macro) may emerge as some of the key research areas of the next ten years. We are planning for the following projects to continue the mission of the 3D Informatics group:

A GPU-Multicore HPC Architecture for Single Particle Microscopy – We plan a departure from sub-volume averaging for TET, and instead are choosing to begin applying our techniques to single particle microscopy. Single particle imaging is more flexible, but requires an additional dimensional analysis, exploding the computational problem. We are designing a new hybrid multi-core/GPU HPC architecture including as many as 12 GPUs for general purpose computing.

C3PI Grand Challenge – Under the HHS Competes program, enabled by the America Creating Opportunities to Meaningfully Promote Excellence in Technology, Education, and Science Act of 2007, we intend to solicit and compare programs for identifying prescription oral solid dose medications. Using the portfolio of digital images from the C3PI effort, we will distribute training images and data for the development and refining of pill-identification programs. We will collect working programs along with source code as open source software. The results of this challenge will be disseminated to support the development of new mobile applications for the rapid identification and verification of prescription medications.

High-throughput microscopy – the work of microbiologists is often the qualitative observation of natural phenomena, and the quantitative analysis of events or measurement of structures can be tedious and prone to error. Using SimpleITK and open source imaging consoles such as 3D Slicer, our group intends to make a committed effort to automating quantitative analysis, developing modules specifically for IA-SEM data.

9 Summary

The research area we call 3D Informatics is a synthesis of medical image processing, computer graphics, modeling, and data storage and retrieval. It is an underserved and underrepresented research area that deserves investment. We have proposed and delivered a comprehensive research program for the Lister Hill Center. This concept is built up of multiple strengths or research pillars, and like a table with three legs, will stand stably only with the strength and support of all three. We believe that a program that is *not* comprehensive, that does not integrate the breadth of volume data analysis, modeling, and representation will focus too tightly on the narrow confines of their mathematics and techniques, and so be incapable of adequately supporting the practice and research endeavors of medicine. We believe that we have grown a fledgling program with this vision, supplying the Lister Hill Center with a new opportunity to lead in the medical community.

Our goal is to establish one of the premier laboratories in medical image analysis and medical computer graphics in the country here at the Lister Hill Center. Our opportunities to attract visiting researchers and sponsor collaborative research have been demonstrated in the past few years as our program has been building from its early successes. In addition, since the research domain of connecting medical volume data, visualization, user interaction, and data modeling is relatively young, our program can make salient changes to the field. Through our emphasis on open, reproducible science we have shown that we can influence this field using our external programs, directing the community toward open source and open access.

10 Publications

- Terry S. Yoo, Bradley C. Lowekamp, Oleg Kuybeda, Kedar Narayan, Gabriel A. Frank, Alberto Bartesaghi, Mario Borgnia, Sriram Subramaniam, Guillermo Sapiro, and Michael J. Ackerman. 2014. Accelerating discovery in 3D microanalysis: Leveraging open source software and deskside high performance computing. In *Proceedings of Microscopy and Microanalysis 2014*.
- Jianfei Liu, Kalpathi R. Subramanian, and Terry S. Yoo. 2013. A Robust Method to Track Colonoscopy Videos with Non-informative Images. *International Journal of Computer Assisted Radiology and Surgery*, **8**(4): 575–592.
- Jianfei Liu, Kalpathi R. Subramanian, and Terry S. Yoo. 2013. An Optical Flow Based Approach to Tracking Colonoscopy Video. *Computerized Medical Imaging and Graphics*, **37**(3): 207–223.
- Bradley C. Lowekamp, David T. Chen, Luis Ibáñez, and Dan Blezek. 2013. The Design of SimpleITK. *Frontiers in Neuroinformatics* **7**:45.
- Bradley C. Lowekamp B and David T. Chen. 2013. BinShrink: A multi-resolution filter with cache efficient averaging. *Insight Journal*. Web. 18 November 2013 <<http://hdl.handle.net/10380/3450>>.
- Keydar Narayan, Danielson CM, Lagarec K, Bradley C. Lowekamp, Coffman P, Laquerre A, Phaneuf MW, Hope TJ, and Sriram Subramaniam. 2013. Multi-resolution correlative focused ion beam scanning electron microscopy: Applications to cell biology. *Journal of Structural Biology*. 2013 Dec 1. pii: S1047-8477(13)00314-6.
- David M Schauder, Oleg Kuybeda, Jinjin Zhang, Katherine Klymko, Alberto Bartesaghi, Mario J Borgnia, Mark L Mayer, Sriram Subramaniam. 2013. Glutamate receptor desensitization is mediated by changes in quaternary structure of the ligand binding domain. *Proceedings of the National Academy of Sciences*. **110** (15), 5921-5926
- Audray K Harris, Joel R Meyerson, Yumiko Matsuoka, Oleg Kuybeda, Amy Moran, Donald Bliss, Suman R Das, Jonathan W Yewdell, Guillermo Sapiro, Kanta Subbarao, Sriram Subramaniam. 2013. Structure and accessibility of HA trimers on intact 2009 H1N1 pandemic influenza virus to stem region-specific neutralizing antibodies. *Proceedings of the National Academy of Sciences*. **110**(12), 4592-4597.
- Oleg Kuybeda, Gabriel A Frank, Alberto Bartesaghi, Mario Borgnia, Sriram Subramaniam, and Guillermo Sapiro. 2013. A collaborative framework for 3D alignment and classification of heterogeneous subvolumes in cryo-electron tomography. *Journal of Structural Biology*. **181** (2), 116-127.
- Gabriel A Frank, Oleg Kuybeda, Alberto Bartesaghi, Mario J Borgnia, Guillermo Sapiro, and Sriram Subramaniam. 2013. Computational separation of conformational heterogeneity using cryo-electron tomography and 3D sub-volume averaging. *Biophysical Journal* **104**, 350.
- Joel R. Meyerson, Erin EH Tran, Oleg Kuybeda, Weizao Chen, Dimiter S Dimitrov, Andrea Gorlani, Theo Verrips, Jeffrey D Lifson, and Sriram Subramaniam. 2013. Molecular structures of trimeric HIV-1 Env in complex with small antibody derivatives. *Proceedings of the National Academy of Sciences* **110** (2), 513-518.
- Audray K Harris, Joel R Meyerson, Yumiko Matsuoka, Oleg Kuybeda, Amy Moran, Donald Bliss, Suman R Das, Jonathan Yewdell, Guillermo Sapiro, Kanta Subbarao, and Sriram Subramaniam. 2013. Molecular structures of native HA trimers on 2009 H1N1 pandemic influenza virus complexed with neutralizing antibodies. *Biophysical Journal* **104**, 414.
- Erin EH Tran, Oleg Kuybeda, Alberto Bartesaghi, Jacqueline LS Milne, and Sriram Subramaniam. 2013. Binding of neutralizing antibodies results in distinct quaternary conformations of trimeric HIV-1 envelope glycoproteins. *Biophysical Journal* **104**, 352.
- Erin EH Tran, Mario J Borgnia, Oleg Kuybeda, David M Schauder, Alberto Bartesaghi, Gabriel A Frank, Guillermo Sapiro, Jacqueline LS Milne, and Sriram Subramaniam. 2012. Structural mechanism of trimeric HIV-1 envelope glycoprotein activation. *PLoS Pathogens* **8**(7), e1002797.
- Gabriel A Frank, Alberto Bartesaghi, Oleg Kuybeda, Mario J Borgnia, Tommi A White, Guillermo Sapiro, and Sriram Subramaniam. 2012. Computational separation of conformational heterogeneity using cryo-electron tomography and 3D sub-volume averaging. *Journal of structural biology* **178**(2), 165-176.
- Terry S. Yoo, Donald Bliss, Bradley C. Lowekamp, David T. Chen, Gavin E. Murphy, Kedar Narayan, Lisa M. Hartnell, Thao Do, and Sriram Subramaniam. 2012. Visualizing cells and humans in 3D: biomedical image analysis at nanometer and meter scales. *IEEE Computer Graphics and Applications*. Sep-Oct 2012, **32**(5), 39-49.
- Jesus J. Caban, Adrian Rosebrock, and Terry S. Yoo. 2012. Automatic identification of prescription drugs using

- shape distribution models. In *Proceedings of ICIP 2012: IEEE International Conference on Image Processing*, October 1-3 2012.
- Jianfei Liu, Kalpathi Subramanian, and Terry S. Yoo. 2012. A Phantom Design for Validating Colonoscopy Tracking. In *Proceedings of SPIE Medical Imaging*, Feb. 4-9, 2012, San Diego, CA.
- Jesus J. Caban, Penny Rheingans, and Terry S. Yoo. 2011. An Evaluation of Visualization Techniques to Illustrate Statistical Deformation Models”, *Computer Graphics Forum*, **30**(3), May/June, 2011.
- Jesus J. Caban, Jianhua Yao, David Liao, Daniel J. Mollura, and Terry S. Yoo. 2011. Enhancing image analytic tools by fusing quantitative physiological values with image features. In *Proceedings of Society for Imaging Informatics in Medicine (SIIM)*, June 2011.
- Terry S. Yoo, Trevor Hamilton, Darrell Hurt, Jesus Caban, David Liao, and David T. Chen. 2011. Toward quantitative x-ray CT phantoms of metastatic tumors using rapid prototyping technology. In *Proceedings of ISBI 2011: IEEE Computer Society International Symposium on Biomedical Imaging*, Xiaochuan Pan and Michael Liebling, eds. (30 March- 2 April 2011, Chicago, IL).
- Jesus J. Caban, David Liao, Jianhua Yao, Daniel J. Mollura, Bernadette Gochuico, Terry S. Yoo. 2011. Enhancing image classification models with multimodal biomarkers. In *Proceedings of SPIE Medical Imaging*, **7963**, 1-9, Feb 2011.
- Jianfei Liu, Kalpathi Subramanian, and Terry S. Yoo. 2011. Temporal Volume Flow: An Approach to Tracking Failure Recovery. In *Proceedings of SPIE Medical Imaging 2011*, Feb. 12-17, 2011, Orlando, FL.
- David Mann, Jesus J. Caban, Philipp J. Stolka, Emad Boctor, and Terry S. Yoo. 2011. Multidimensional transfer functions for effective visualization of streaming ultrasound and elasticity images. In *Proceedings of SPIE Medical Imaging*, **7964**, 1-8, Feb 2011.
- Jianfei Liu, Kalpathi Subramanian, and Terry S. Yoo. 2010. Region Flow: A Multi-Stage Method for Colonoscopy Tracking. In *Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2010*, Sept. 20-24, 2010, Beijing, China. Lecture Notes in Computer Science(LNCS) **6362**, Springer.
- Michael J. Ackerman, Elliot Siegel, Fred Wood. 2010. Interactive Science Publishing: A joint OSA-NLM project. *Information Services & Use*, 30:39-50.
- Jesus J. Caban, Penny Rheingans and Terry Yoo. 2010. An MRF-based Statistical Deformation Model for Morphological Image Analysis, *Mathematical Models in Biomedical Image Analysis (MMBIA)*, CVPR, 31-38, June 2010
- Jianfei Liu, Kalpathi Subramanian, Terry Yoo. 2010. Accurate motion parameter estimation for colonoscopy tracking using a regression method. In *Proceedings of SPIE Medical Imaging*, **7624**, 13-18 February, 2010, San Diego, CA
- Adam E. Bennett, Kedar Narayan, Dan Shi, Lisa M. Hartnell, Karine Gousset, Haifeng He, Bradley C. Lowekamp, Terry S. Yoo, Donald Bliss, Eric O. Freed, and Sriram Subramaniam. 2009. Ion-abrasion scanning electron microscopy reveals surface-connected tubular conduits in HIV-infected macrophages. *PLoS Pathogens* **5**(9):Sep 2009. e1000591. doi:10.1371 / journal.ppat.1000591