Visual Computing Group
visualization principles applied to science challenges

Selected accomplishments
- developed novel algorithms for rendering volumetric data in real time
- developed MizBee (above), a tool for answering evolutionary questions by exploring genomic data
- helped Harvard Medical School systems biologists develop methods for understanding relations between gene expression and genomic data in *Drosophila*

Launched 2007
Status Continuing in SEAS
Participants
Moritz Bächer, ETH Zürich/IIC/SEAS
Christian Ledergerber, ETH Zürich/IIC
Miriah Meyer, IIC
**Hanspeter Pfister**, IIC/SEAS

A screenshot of the multiscale synteny browser MizBee, visualizing comparative genomics data of the stickleback and pufferfish.
From its start in 2007, the Visual Computing Group focused on providing visual tools and methods to help scientists better understand large, multidimensional data sets. The group has tackled problems from generating patient-specific models for medical simulations to exploring gene-expression data in the embryos of fruit flies. Through a range of projects and collaborations, the VCG has determined that visualization research targeted to biological data, specifically genomics data, has the potential to make a large impact in both the visualization and scientific communities.

This research lies at the intersection of visualization, computer graphics, and interaction design. In all of our projects, the group has found it important to balance theoretical basic research, the development of practical tools and methods, and the application of the tools to problems that solve a scientific need. An interesting finding is the challenge of cross-disciplinary translation from biological questions to abstract data-specific characterizations.

**Group projects**
The VCG, a small experimental group headed by Hanspeter Pfister, consisted of student interns Christian Ledergerber and Moritz Bächer and Fellow Miriah Meyer. Bächer is now a SEAS graduate student, and Meyer is a SEAS postdoctoral fellow.

Early on, the VCG focused on traditional scientific visualization research, building on the expertise of the group members. This field develops algorithms and methods for processing and understanding volumetric data, such as that coming from medical scanning devices. One of the first projects, done in collaboration with researchers at the University of Utah, looked at the problem of generating patient-specific models for use in downstream medical simulations.

Starting from volumetric data where each data point is labeled according to its appropriate material, such as bone, lung, or heart, our novel approach first creates a mathematical description of the surfaces of the different materials. Then, by simulating the interactions of virtual particles across these surfaces, we produce a very regular set of point samples that are used to generate a 3-dimensional model of the entire data set.

In conjunction with the virtual model work, we tackled the long-standing problem of how to mathematically reconstruct smooth representations of volumetric data in a general, scalable, and well-characterized framework. Data coming from scanning devices such as MRI or ultrasound machines, as well as results from simulations in fields like computational fluid dynamics, geophysics, and biomedical computing, are often represented volumetrically. These volumes, however, can be defined over regular or irregular grids, giving rise to several distinct classes of rendering algorithms and a plethora of implementations. By applying the method of moving least squares to volumes for the task of generating high-quality visualizations, we developed a unified framework for representing and rendering any type of volumetric data in a real-time system. Figure 2 shows several results from this system.

In the spring of 2008, members of the VCG developed new collaborations with biologists at Harvard Medical School and the Broad Institute. These collaborations stemmed from the need of the scientists to make sense of large, multidimensional genomics data sets, the scale of which is a result of recent advances in sequencing and imaging technologies.
Our initial conversations with these biologists pointed to a much larger need in the biology community for tools and methods that enable the exploration and understanding of large genomics data sets. Currently these tools are often designed by biologists without a rigorous understanding of, or attention to, the principles of visualization and user-interface design. The results are tools that, generally speaking: do not scale well, are inefficient, can easily mislead, and often are constrained by traditional methods for viewing and representing biological data. We saw the opportunity to make an impact by applying the group’s expertise in visualization to this scientific community.

Recently, the VCG completed a year-long project with comparative biologists at the Broad Institute who are looking for large sections of DNA that are maintained over evolutionary time. The size, location, and similarity of these sections provide answers to questions about the relatedness of species, the resulting function of groups of genes, and how species diverge through evolution. The project resulted in a visualization tool called MizBee that the VCG developed based on a detailed characterization of the questions the scientists were interested in asking, and how these questions related to the actual data. For one of our collaborators, this tool provided new insights that altered his data generation approach completely. A screen shot of MizBee is shown on the front cover of this chapter.

Ongoing work
The VCG is currently working with researchers at Harvard Medical School who are studying the regulation of genes during embryo development in multiple Drosophila species to understand how regulatory information is encoded in the genome, how it is deciphered as gene expression patterns in space and time, and how it changes during evolution to contribute to organismal diversity.

We are collaborating with these scientists to develop tools that enable more efficient exploration of the high-dimensional gene-expression space over the layout of an embryo, as well as methods for understanding and determining differences between embryos of different species. In tandem, we are also looking at new visual encodings for understanding interspecies differences in the regulation regions for genes. The ultimate goal of this work is to combine these visual systems into a multifaceted view of the regulation of genes in Drosophila.

Through our current and previous collaborative work, we recognize that an interesting and challenging aspect of our role as visualization experts is making the translation from specific scientific questions to an abstract data characterization. The correct translation requires an understanding of the science domain and language, the ability to generalize tasks, and good interviewing skills. Traditional training in neither computer science nor science provides education in all these facets. We believe that the visualization community is beginning to address this challenge, and the VCG is working to develop a principled analysis of the process of translation between biology and computer science.
Future directions
Our collaboration with biologists at the Harvard Medical School for understanding gene expression regulation in *Drosophila* embryos is an ongoing project with an expected lifetime of one to five years. The biology team is in the early stage of data collection and faces the daunting task of comparing and understanding gene expression data of 12 species, each with several thousand cells, for 20 genes at 6 time points. Understanding this data set, as well as what differences are important, will require very efficient and intuitive data exploration tools, with the long-term goal of linking expression differences to differences in the genomic sequences of the flies. This is a project rich with visualization challenges and the potential for a significant impact in the field of systems biology.

We are also at the early stage of working with several groups at the Broad Institute. These projects range from the development of tools for understanding the differences between pathogenic virus strains, such as the dengue virus and HIV, as they mutate geographically and in the context of drug therapies, to exploring the life cycles of thousands of “bar-coded” cancer cell lines. A small visualization group is developing at the Broad Institute that will work together to tackle these, and other, visualization problems at this cutting-edge sequencing center.

Publications and posters

Presentations
- Visualization in Biology, DePace/Megason Seminar 2009
- Particle-based Sampling and Meshing of Surfaces in Multimaterial Volumes, Visualization Conference 2008
- Volume MLS Ray Casting, Visualization Conference 2008
- Designing Visualizations with Biologists, University of Massachusetts, Lowell

Software

Key collaborations
- University of British Columbia Tamara Munzner, associate professor
- Harvard Medical School Angela DePace, assistant professor; Zeba Wunderlich, postdoc; Janet Iwasa, lecturer in molecular visualization
- Broad Institute Li-Jun Ma, scientist; Manfred Grabherr, scientist; Bang Wong, creative director; Daniel Kohn, artist in residence
- MIT Media Lab Peter Torpey, graduate student
- IBM Martin Wattenberg, director of visual communication lab
- University of Utah Mike Kirby, associate professor; Ross Whitaker, professor
- INRIA Gael Guennebaud, research associate
- ETH Zürich Markus Gross, professor