Research Statement

"The purpose of computing is insight, not numbers."
R. Hamming, founder and president of the Association for Computing Machinery, 1962

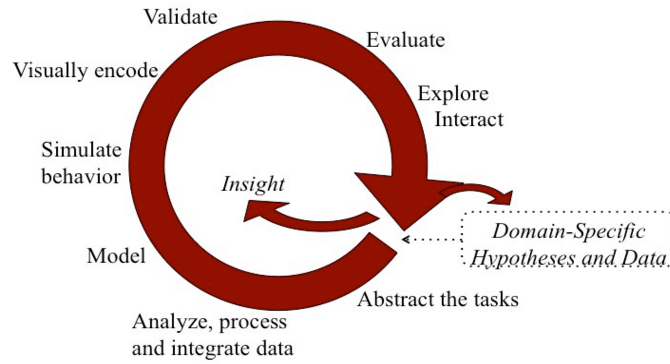

Figure 1. Modeling and Visualization Research Loop (adapted from a D.H. Laidlaw presentation, 2005)**.**

The incredible range of novel imaging and measurement technologies available to the scientific community is opening up new worlds of data, changing fundamentally our understanding of physical and natural processes. For example, it is now possible to measure complex systems and phenomena — from combustion engines igniting to human spines bending. However, scientific data acquisition marks only the first step. To turn *numbers* into insight, computational models can help us replicate complex systems and make predictions about their behavior. Along this very line, computer graphics and visualization techniques and algorithms — from image processing to geometric modeling to physically-based simulation and visual analysis — harness the immense power of the human visual perception system to make *insights* into complex processes possible.

Computer graphics and visualization generally denotes the creation, storage, manipulation, and simulation of geometric models and images. Such models come from a diverse and expanding set of fields including physical, mathematical, biological domains, and even the humanities. To lead to insight (Figure 1), geometric modeling and visualization requires analyzing and abstracting the domain tasks and the captured data; proposing the appropriate level of modeling-detail necessary to study the domain hypotheses; constructing, simulating, visualizing and validating the geometric models and their behavior, often at very large scales; providing exploratory visualization techniques to facilitate the formulation of new hypotheses; and then iterating again through the loop with the new hypotheses, until insight emerges. To build, verify, and understand geometric models requires thus *fundamentally interdisciplinary and exploratory skills* [1], in addition to meeting the *analysis, scalability, modeling, interaction and visual perception challenges* [2] typical of this process.

My research explores novel geometric representations, computational modeling, and visual analysis tools that are needed for the simulation and analysis of complex scientific phenomena. Following select topics in the exploratory visualization-for-insight loop, my first research component focuses on intelligent algorithms for analyzing spatial biological data. To address limitations in the input data, the second component pursues data-driven geometric biological models. Closing the loop, the third component focuses on effective large-scale visual paradigms and interactive techniques for spatial and non-spatial science and engineering data.

My work has lead so far to five grants (2 PI, 3 co-PI), including an **NSF CAREER award**; to 8 journal publications; more than 25 conference publications and system demonstrations; 2 book chapters; and one provisional patent filing. Four of the publications have received conference awards, most recently a **Best Paper Award at IEEE VisWeek BioVis 2011**, the premier visualization in biology venue. In addition to reviewing articles, book chapters, and grant proposals for several federal and international venues, to service on three program committees, three organizing committees (outstanding public commendation), and co-organizing a workshop, I have guest-edited the *IEEE Computer Graphics & Applications* journal's special issue on biomedical applications, and helped formulate the top challenges in the visualization of engineering tensor-fields (Dagstuhl).

This statement describes the main ideas behind my research and my approach to building a strong interdisciplinary research group.

Intelligent Algorithms to Accurately Reconstruct 3D Articulated Motion

Computational models depend in large measure on the input they use, such as images of biological structures. However, imaging techniques have limited resolution and accuracy due to imaging artifacts such as partial volume effects related to the Nyquist sampling theorem; as a result, extracting accurate measurements often relies on expert-human intervention. Overcoming such limitations requires solid physical and numerical treatment.

In [3][4] we introduced *unsupervised tissue classification techniques* to reduce biological imaging partial volume effects. Using a Bayesian framework, we generate from each volume image a super-resolution distance field — a scalar field that specifies, at each point, the signed distance from the point to a material boundary. The distance fields and a geometric model are then used to register an object through the sequence of medical images and thus recover motion information. Evaluation showed expanded capabilities and average accuracy improvements of 74% over the previous state-of-the-art technique; among other indirect results, this advancement enabled the creation of a database documenting wrist motion over a large number of subjects, advancing knowledge in orthopedic surgery. Our current work focuses on *intelligent algorithms for extracting motion* information from *dynamic* images. The state of the art techniques rely on extensive manual labor (up to 9 hours of processing required for each hour spent collecting data), making the time and cost for data analysis prohibitive for clinical use. Using image filtering and an intelligent hierarchical, temporal-coherent model of bone overlap, we have been able to reduce the amount of manual labor by a factor of 9 (!) while maintaining expert-level accuracy [5]. The software (partially funded through my **NSF CAREER award**) has been deployed at the University of Pittsburgh Medical Center (UPMC) Biodynamics Lab and is used to generate the first database of 3D spine motion in the world; with tremendous potential impact on the health of millions of US citizens.

Our future work seeks to improve the accuracy of the tracking procedures through a classification image-processing approach. The approach builds on the temporal coherence in the dynamic data; if successful, it may enable future non-invasive studies of soft-tissue deformation with motion.

Latent, Data-Driven Geometric Models in Biology

The first challenge in computational modeling for biology is deciding what level of modeling detail is necessary, and developing appropriate representations and approximations so that we obtain biologically significant measurements, while keeping the resulting models efficient to simulate. The second challenge is that many of the inputs we need to build realistic models are not measurable directly in live individuals or systems; for example, elastic tissue properties. Input entities that are not measurable directly become, in fact, latent variables in our models. These variables need to be inferred from directly measurable data.

In [6][7][8][9] we have shown that *a latent, data-driven approach* can overcome long-standing limitations in humanoid articulation modeling. To address the second challenge, the key idea behind this approach is to use sampled dynamic motion data to infer unknown data such as soft-tissue geometry and behavior. To address the first challenge, the resulting hybrid geometric models blend distance fields, manifolds, and meshless representations to enable the predictive simulations of articulations, with far reaching insights in robotics and rehabilitative therapy. In [10][11][12] we have further shown that *scalable visual abstractions* of molecular system behavior can be *algorithmically* built from rule-based models of lab observations. These representations (partially funded through the Pitt Clinical Translation and Science Institute fellows program) can then be used to reason and make predictions about the system behavior in a variety of settings — for example, in drug design.

Future work (partially funded through my **NSF CAREER award**) researches computational tools for estimating and modeling soft-tissue geometry and behavior from skeletal dynamics; computational models of humanoid articulations; and integrating visual abstractions with the exploratory simulation of rule-based models.

Effective Visual Abstractions for Large-Scale, Multivariate Data

Big Data come with unique scalability challenges to visualization, from the sheer size of datasets (petascale) to the multitude of different measurements (multivariate data) available for the same system behavior. Many of these data feature both spatial (i.e., characterized by Cartesian coordinates) and non-spatial traits, for example in bioinformatics, geospatial analysis, neuroimaging, and even engineering (e.g., annotations detailing simulation parameters). Yet, to enable insight, most applications require the *effective* and interactive visual analysis of large-scale, multiple fields simultaneously; where *effective* denotes visualizations which capture the important aspects of the data, while being intuitive to the domain experts.

In [10][11][12] we describe the user requirements, scalable visual paradigms (e.g., interactive contact maps) and design decisions behind RuleBender, a novel system for the rule-based integrated visualization of *large scale, multivariate biochemistry data*. The system emphasizes visual global/local model exploration and integrated execution of simulations. RuleBender has been adopted as both an educational and a biology research tool at numerous research labs and universities, including CMU and Yale, and has lead to a **Best Paper Award** at IEEE BioVis 2011. In [13] we introduce a web-based, client-server approach to assist the interactive navigation of large-scale astronomy observations (**NSF CDI Award**). A spatial index-structure enabling prefix matching of spatial objects, in conjunction with pixel-based overlays, allows fetching, displaying, panning and zooming of gigabit images of the sky in real time. Current work supports the *spatial and non-spatial visual integration and mining of large-scale, multidimensional astronomy data*, including catalog spectra and annotations. A prototype implementation into a system named Astroshelf has drawn enthusiastic feedback from the astronomy community. Steps further, in [14] we pursue the *visual integration, comparison, and exploration of spatial and non-spatial high-dimensional geriatric research data* (**Pitt Multidisciplinary Grant**). These neuroimaging data span both the spatial, volumetric domain – through medical imaging volumes – and the non-spatial domain, through variables such as age or walking speed. A linked-view design geared specifically at interactive visual comparison integrates spatial and abstract visual representations to enable the users to effectively generate and refine hypotheses in a large, multidimensional, and fragmented space. Collaborations with colleagues across departments have lead to several other interactive screening tools and publications. For example, in [15] we introduced the Chinese Room, an interactive visual analysis system to mine *multivariate, ambiguous data in machine translations*, while in [16] we visualize *large-scale and dense symmetric-tensor fields* used in turbulent combustion. Collectively, these works advance the state of the art in visualization; not only through novel visual encodings, but also through the domain analyses, discussions, and design lessons they contribute.

Future work pursues the design of scalable visual abstractions and exploratory techniques for large-scale, multivariate scientific data; and principles for the visual integration of spatial and non-spatial data. For the first direction, we look to use machine learning clustering techniques to extract potential features from large datasets (**NSF CBET Award**); and to develop novel visual techniques for browsing large collections of observations. An important lesson emerging for the second direction is that users with different backgrounds may employ a spatial and non-spatial visualization application differently. This may be particularly important for the design of collaborative meeting environments, in which users contribute complementary expertise.

Advising and Research Group

The interdisciplinary research group that I have built [17] helps me explore these research topics, as well as provide a continuing path to research for the graduate and undergraduate students enrolled in my courses (**Pitt Innovation Award**). The group currently includes 3 doctoral students, 1 masters student, and several undergraduate students, as well as 12 collaborators among faculty and postdocs across the campus, spanning almost as many disciplines. In addition, 15 undergraduate and masters students have worked with the group and successfully graduated, 9 of them with a thesis or project; I purposely raise funding for undergraduate research through the **NSF REU** program. Many group members work in interdisciplinary teams. This is the first graphics research group at Pitt in 15 years; as such, building it has required creative recruiting and training strategies.

The group meets weekly for talks in which one member presents some aspect of their research and receives peer-feedback not only on the research, but also on the presentation. Since communication skills are essential in interdisciplinary collaborations, each year I recruit, train and hire Public Speech coaches (through the Pitt Writing Center) to work with my students on improving their oral and written communications skills. The group also relies on interest-based weekly subgroup meetings; sometimes on daily round-robin updates; on outings to celebrate achievements; and on tri-annual short elevator-speech public presentations — an experiment in quick dissemination of information. This latter approach supports the group as a whole; and facilitates recruiting and involving undergraduate students into research. We also regularly organize outreach activities together, such as K-12 presentations or workshops. I often receive enthusiastic feedback, commenting on how articulated, excited and motivated the young researchers in my group are about their projects.

To help my advisees grow into independent researchers, each term I require them to assemble semester-long research plans, complete with goals, a timeline, and bi-weekly deliverables. The plan typically requires several guided iterations; we monitor the progress weekly. At the end of each semester, we evaluate accomplishments against the original plan. This process encourages the students to identify their priorities, think through their approach and allocate resources adequately; it helps identify recurring problems early; and also documents the students' progress towards their degree. Deliverables often include conference and group presentations. One of my advisees (doctoral student Tim Luciani, who had joined the group as a sophomore) was recently awarded an **NSF Graduate Research Fellowship**, partly in recognition of his undergraduate research with the group. Other group alumni have placed in top software companies such as Google and Microsoft.

Conclusion

I research automated techniques and computational representations for scientific modeling and data visualization. My work draws on and extends ideas from computer graphics and visualization, medical imaging, differential geometry, and applied mathematics. Collaborative work with colleagues across domains leads to richer computer science research and provides a mechanism for evaluating the usefulness and robustness of results.

My research impacts both the computational domain — through novel abstractions and automations to make tractable geometric and visual computations that would otherwise be impossible — and the application areas, through an improved understanding of the underlying phenomena in a wide range of systems. The tools and collaborations I build help solve practical scientific problems in disciplines such as biology, orthopedics, epidemiology, astronomy and mechanical engineering. Beyond the scientific impact of this work, we also achieve a better understanding of the process of interdisciplinary collaboration and what makes such collaborations succeed.

References:

[1] Frederick P. Brooks. The Computer Scientist as a Toolsmith II. Communications of the ACM, Mar 1996

[2] Chris Johnson. Top Scientific Visualization Research Problems. IEEE Computer Graphics and Applications, Aug. 2004.

[3] G.E. Marai, D.H. Laidlaw, J.J. Crisco. Super-Resolution Registration Using Tissue-Classified Distance Fields. IEEE Transactions on Medical Imaging, Feb 2006.

[4] D.H. Laidlaw, G.E. Marai, K.E. Fleischer, A. Barr. Partial Volume Segmentation and Boundary Distance Estimation with Voxel Histograms, in Handbook of Medical Imaging: Processing and Analysis, 2nd edition, Academic Press 2010.

[5] Md. A. Haque, W. Anderst, S. Tashman, G.E. Marai. Hierarchical Dynamic Model-based Tracking of Cervical Vertebrae from Fluoroscopy Images. Medical Physics Engr, 2012. In Review

[6] G.E. Marai, D.H. Laidlaw, C. Demiralp, S. Andrews, C.M. Grimm, J.J. Crisco. Estimating Joint Contact Areas and Ligament Lengths from Bone Kinematics and Surfaces. IEEE Trans. Biomed. Engr, May 2003.

[7] G.E. Marai, C.M. Grimm, D.H. Laidlaw. Arthrodial Joint Markerless Cross-Parameterization and Biomechanical Visualization. IEEE Transactions on Visualization and Computer Graphics, Sep 2007.

[8] G.E. Marai, J.J. Crisco, D.H. Laidlaw. A Kinematics-Based Method for Generating Cartilage Maps and Deformations in the Multi-Articulating Wrist Joint From CT Images. IEEE 2006 Engineering in Medicine and Biology (EMBC'06), Sep 2006.

[9] G.E. Marai, J.J. Crisco, D.H. Laidlaw. Development of a Kinematic 3D Carpal Model to Analyze In Vivo Soft-Tissue Interaction Across Multiple Static Postures. IEEE 2009 Engineering in Medicine and Biology (EMBC'09), Sep 2009.

[10] W. Xu, A. Smith, J.R. Faeder, G.E. Marai. RuleBender: A Visual Interface for Rule-Based Modeling. Bioinformatics Journal, Apr 2011.

[11] A.M. Smith, W. Xu, Y. Sun, J.R. Faeder, G.E. Marai. RuleBender: Integrated Visualization for Biochemical Rule-Based Modeling. IEEE Visualization 2011, IEEE BioVIs: Symposium on Biological Data Visualization, Oct 2011.

[12] A.M. Smith, W. Xu, Y. Sun, J.R. Faeder, and G.E. Marai. RuleBender: Integrated Modeling, Simulation and Visualization for Rule-Based Intracellular Biochemistry. BMC Bioinformatics, Jun 2012.

[13] T. Luciani, S. Myers, B. Sun, B. Cherinka, W.M. Wood-Vassey, A. Labrinidis, G.E. Marai. Panning and Zooming the Observable Universe with Prefix-Matching Indices and Pixel-Based Overlays. IEEE VisWeek 2012, LDAV'12, Oct 2012.

[14] A.Maries, M.Olsen, K.Wong, W.Layton, C.Rosano, H. Aizenstein, R. Boudreau, S.L. Yilmaz, G.E. Marai. GRACE: A Visual Comparison Framework for Integrated Spatial and Non-Spatial Geriatric Data. IEEE TVCG, 2012, In Review

[15] J.S. Albrecht, R. Hwa, G.E. Marai. The Chinese Room: Visualization and Interaction to Understand and Correct Ambiguous Machine Translation. Computer Graphics Forum, Jun 2009.

[16] A. Maries, Md. A. Haque, S. L. Yilmaz, M. B. Nik, G. E. Marai. Interactive Exploration of Stress Tensors Used in Computational Turbulent Combustion. Springer-Verlag, Dagstuhl - Visualization and Processing of Tensors, 2011.

[17] The Pitt Interdisciplinary Visualization Research Lab, http://vis.cs.pitt.edu