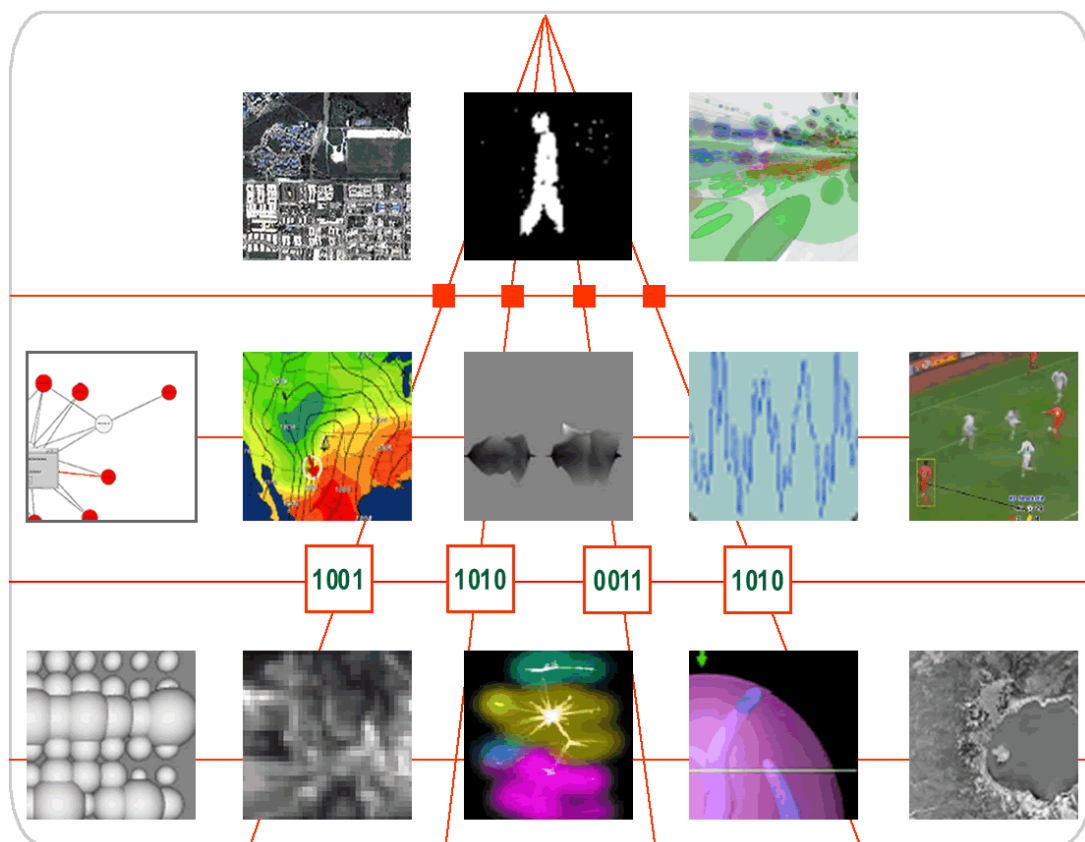


Interactive Digital Multimedia IGERT Annual Research Review

Year 1: 2003-2004



UNIVERSITY OF CALIFORNIA, SANTA BARBARA

Proceedings

**Interactive Digital Multimedia IGERT
Annual Research Review**

Year 1: 2003-2004

October 8, 2004
University of California, Santa Barbara

Supported by:

National Science Foundation Grant #0221713
“Digital Multimedia: Graduate Training in Interactive Digital Multimedia”

Interactive Digital Multimedia IGERT Annual Research Review

Year 1: 2003-2004

Table of Contents

Program Organization	7
Year 1 Report	9
Data Organization and Visualization in a Web Community Environment	15
E. Kaplan & S. DiVerdi	
Aesthetic Techniques for the Visual Display of Information	18
A. Black	
The Effect of Context Cues in Saccadic Decisions	23
B. Drescher & L. Boucheron	
Realism and Perceptions of Data Quality in Computer-Displayed Maps.....	27
A. Boughman	
Map Generalization for Mobile Display: Evaluation and Automation	29
J. Dilleuth & N. Vu	
Clustering Web Images Using Linked Text	33
A. Villacorta	
An Integrated System of 3D Motion Tracker and Spatialized Sound Synthesizer	34
M. Quinn, J. Thompson, & M. Li	
Visually Optimized Quantization Parameters in AVC/H.264	40
J. Hu	
Placement of Annotations in Video Feeds	44
V. Thanedar	
Map Design and Perceptual Salience	47
K. Goldsberry	
The Influence of Spatial Ability and the Use of Dynamic, Interactive Animation	49
in a Spatial Problem-solving Task	
C. Cohen	
Is Student Engagement Dependent on Lecture Relevance? A Study of Student	52
Engagement in Multimedia Classrooms	
M. Bulger	
Experiments in Sound Granulation and Spatialization for Immersive	55
Environments	
D. Thall	

OnKai: Sculpting Three-dimensional Objects for Control in Computer Music	58
Composition	
S. Morita	
Warehousing and Integration of Biological Databases	60
K. Hawkins	
MATConcat: An Application for Exploring Concatenative Sound Synthesis	64
Using MATLAB	
B. Sturm	
A Real-Time Application Demonstrating the Interaction of Atmosphere and.....	66
Ocean Using Sound and Image	
B. Sturm & A. Black	

2003-2004 Interactive Digital Multimedia IGERT Participants

Fellows/Trainees:

August Black, Media Arts and Technology
Laura Boucheron, Electrical and Computer Engineering
Tony Boughman, Geography
Stephen DiVerdi, Computer Science
Julie Dillemoth, Geography
Barbara Drescher, Psychology
Ethan Kaplan, Media Arts and Technology
Michael Quinn, Electrical and Computer Engineering
John Thompson, Music
Alex Villacorta, Statistics
Nhat Vu, Electrical and Computer Engineering

Research Associates:

Monica Bulger, Education
Cheryl Cohen, Psychology
Kirk Goldsberry, Geography
Jing Hu, Electrical and Computer Engineering
Mary Li, Electrical and Computer Engineering
Satoshi Morita, Media Arts and Technology
David Thall, Media Arts and Technology
Vineet Thanedar, Computer Science

Undergraduates and Mentors:

Kevin Hawkins, Computer Science
Vebjorn Ljosa, Computer Science (Undergraduate Mentor)

Faculty Advisors:

Kevin Almeroth, Computer Science
Keith Clarke, Geography
Miguel Eckstein, Psychology
Sara Fabrikant, Geography
Jerry Gibson, Electrical and Computer Engineering
Mary Hegarty, Psychology
Tobias Hollerer, Computer Science
S. R. Jammalamadaka, Statistics
JoAnn Kuchera-Morin, Media Arts and Technology/Music
George Legrady, Media Arts and Technology
B.S. Manjunath, Electrical and Computer Engineering
Rich Mayer, Psychology
Stephen Pope, Media Arts and Technology
Curtis Roads, Media Arts and Technology
Kenneth Rose, Electrical and Computer Engineering
Ambuj Singh, Computer Science
Matthew Turk, Computer Science

Postdoctoral Researchers:

Matthias Kölsch, Computer Science
Xinding Sun, Electrical and Computer Engineering

New IGERT Trainees in 2004-2005 (and degree institutions):

Maria del Mar Alvarez, Computer Science
B.S., University of Puerto Rico Mayaguez
Carlos Castellanos, Media Arts and Technology
B.A., San Francisco State University
Peter Khooshabeh, Psychology
B.A., University of California, Berkeley
Justin Muncaster, Computer Science
B.S., University of California, Santa Barbara
Dan Overholt, Music - Interdisciplinary PhD
M.S., Massachusetts Institute of Technology
B.A/B.S., California State University, Chico
Wesley Smith, Media Arts and Technology
B.A., Johns Hopkins University
Bob Sturm, Electrical and Computer Engineering
M.S., University of California, Santa Barbara
M.A., Stanford University
B.S., University of Colorado, Boulder

External Advisory Board:

Dr. Michael Century, Chair of Arts Department, Rensselaer Polytechnic Institute
Dr. Tony Chan, Dean of Physical Science, College of Letters and Science, and
Professor of Mathematics, University of California, Los Angeles
Dr. Fiona Goodchild, Education Director, California Nanosystems Institute, University of
California, Santa Barbara
Dr. Ramesh Jain, Rhesa S. Farmer Distinguished Chair in Embedded Experimental
Systems, School of Electrical Engineering and College of Computing, Georgia
Institute of Technology
Dr. Alvy Ray Smith, former Graphics Fellow at Microsoft and co-founder of Pixar



Faculty and students from the IGERT Program in Interactive Digital Multimedia
June 4, 2004

Interactive Digital Multimedia IGERT Annual Research Review

Year 1: 2003-2004

Bringing the Program Together

Interactive Digital Multimedia involves a range of technologies centered on the creation, encoding, transmission, storage, presentation, and analysis of multimedia data, as well as the study of human interaction with multimedia systems. Given the diversity of applications that include learning, communication, music, arts and entertainment, biology, medicine, and many other areas, a multidisciplinary approach to the subject is essential.

The National Science Foundation's Integrative Graduate Education and Research Traineeship Program is also a vehicle for interdisciplinary activity. At its core IGERT seeks to train students to address questions more global than those tackled by conventional PhDs. IGERT's provide collaborative research, a range of courses and seminars, internships in industry, experience in management and entrepreneurship, and training in ethics and career management, all in an environment that transcends traditional departmental boundaries.

Hence the setting for an IGERT in Interactive Digital Multimedia at UCSB.

Program Overview

The digital media IGERT began in October 2003 with 11 students spanning eight academic disciplines: Electrical & Computer Engineering, Computer Science, Media Arts & Technology, Statistics, Geography, Psychology, Education, and Music. One can easily imagine how ambitious an endeavor it was to gather so many together into one program. Each discipline presented a unique definition of research; simply getting everyone to recognize similarities in their academic agendas was a challenge. We found cohesion by organizing around three main areas:

- (1) *Multimedia systems*: storage, transmission, and programming of digital multimedia.
- (2) *Multimedia content*: representation and analysis of multimedia information, and advanced tools for creating, analyzing and manipulating multimedia content.
- (3) *Interactivity*: presentation of, and human interaction with, multimedia information in the context of interactive multimedia applications.

Seminars and Program Requirements

We also created a seminar series for the program to help unify the program. The series provided a platform for IGERT faculty to discuss their own interdisciplinary research. It also drew a number of speakers from other well-known institutions. Here is a list of the invited speakers who visited in 2003-2004 to speak of the past, present and future of artist/engineer collaborations:

11.14.2003 Dr. Wei-Ying Ma, Microsoft Asia
01.23.2004 Dr. Ramesh Jain, Georgia Tech

02.13.2004 Dr. Mark Sylvester, Mixed Grill
02.20.2004 Dr. Osmar Zaiane, University of Alberta
02.27.2004 Dr. Lawrence Rabiner, Rutgers
03.05.2004 Dr. Mark Hansen, UC Los Angeles
03.12.2004 Dr. Victoria Vesna, UC Los Angeles
03.12.2004 Dr. James Gimzewski, UC Los Angeles
04.09.2004 Dr. Thomas Huang, University of Illinois, Urbana-Champaign
04.23.2004 Dr. V. A. Uspenskiy, Moscow University
05.07.2004 Dr. Dominic Massaro, UC Santa Cruz
05.14.2004 Dr. Michael Century, Rensselaer Polytechnic
05.28.2004 Mrs. Dana Plautz, Intel

In addition to the seminar series, the digital media IGERT maintains a list of requirements to help students realize cross-disciplinary potential. These requirements, intended for completion while receiving IGERT support, include taking two elective courses from outside their area of study, training in ethics, and the realization of an interdisciplinary dissertation topic. Finally, students are required to work with their colleagues and with faculty on interdisciplinary group projects. The goal with many of the projects is to produce something, an installation perhaps, that advances contemporary research while at the same time maintaining aesthetic merit.

Program Involvement

The program is comprised primarily of IGERT Trainees and their faculty advisors. ‘Trainee’ is an NSF classification for a graduate student who has received an IGERT fellowship. In an attempt to scale the program to a larger community, we included eight research associates in the summer of 2004. Associates are students who, for one reason or another, did not receive an IGERT fellowship, but who are still interested in interactive digital multimedia research. They contributed to the program just as the other IGERT fellows did – that is, by collaborating with IGERT faculty and students on various projects over the summer and fall terms. Research associates are especially important because they provide a larger skill set with which to collaborate. This helps to expand the resources available to all participants (students and faculty) and improve the overall program.

Finally, with respect to program improvement, on June 4th, 2004, IGERT students and faculty met at a retreat to assess their first academic year. The students were able to present project ideas to a faculty audience with varied research interests and outlooks. The faculty helped the students shape their ideas into something more interdisciplinary. In turn, the students offered constructive feedback regarding the program to their advisors in a supportive, group environment. Because IGERT’s intent is to improve upon the graduate student experience, to be entirely successful, the opinions of the students must be considered. They deserve an opportunity to shape the program to their benefit. To this extent, everyone considered the retreat to be a success.

The Future

Although we are still setting some of the groundwork for the program, 2004 has been both exciting and rewarding. We have seen successful collaboration on research projects, some of

which will be presented at national and international conferences and workshops. More importantly, we have established a centralized IGERT lab with plans to expand that space into a fully equipped interactive media lab in early 2005. There is now a defined base for interactive digital multimedia research at the University of California, Santa Barbara.

The digital media IGERT is also helping the Media Arts and Technology Graduate Program develop a serious research presence. With a number of IGERT faculty heavily involved in the development of MAT, the two programs are very closely related. Indeed, IGERT is something of a forerunner of the Media Arts and Technology doctoral track, and IGERT students are seen by many as models for potential MAT Ph.D. candidates.

We end our first year with 20 participating students (11 Trainees, eight Research Associates, and one undergraduate scholar) and two Postdoctoral Researchers. We begin our sophomore term with seven new Fellows and the guarantee of more Associates later on. This is a significant increase in the diversity and potential for interdisciplinary research. Moreover, our students bring diversity of a different sort. Considering that eight of our 18 IGERT Trainees come from traditionally underrepresented backgrounds, the second year of the Interactive Digital Multimedia program promises to be even more exciting and productive. Following is a summary of the current research projects being conducted by the IGERT students, the initial ideas of which were discussed during the one-day project retreat on June 4, 2004.

Acknowledgements

We would like to thank Dr. Matthew Tirrell, Dean of College of Engineering, Dr. Charles Li, Dean of Graduate Division, Dr. David Marshall, Dean of Humanities and Fine Arts, Dr. Martin Moskovits, Dean of Mathematical, Life, and Physical Sciences in the College of Letters and Science, and Chancellor Henry Yang, for their encouragement and support of the interdisciplinary Interactive Digital Multimedia IGERT. We would also like to thank Professor JoAnn Kuchera-Morin, Chair, Media Arts and Technology Program, and Professor Umesh Mishra, Chair, Electrical and Computer Engineering Department, for their help in building this program. Finally, we wish to give special thanks to our external board of advisors for their guidance. This IGERT project is supported by the NSF Award #0221713.

B.S. Manjunath, Director
Interactive Digital Multimedia IGERT

Interactive Digital Multimedia IGERT Student Research Projects - 2004

Data Organization and Visualization in a Web Community Environment

Ethan Kaplan
IGERT Fellow
Visual Arts

Stephen DiVerdi
IGERT Fellow
Computer Science

Abstract—We present pStruct, a radical new content organization architecture constructed around philosophical theories and informed by the chaotic, social interactions of multitudes of users. pStruct is a tool for allowing disparate data to form relationships between each other autonomously in a web community environment. The server architecture is built on massively multi-threaded programming concepts in a java runtime environment. To gain insight to the internal structure that emerges within pStruct, we also present a visualization component which ascribes a 3D shape to the graph. Using real-time 3D rendering techniques, we generate an artistic and informative interactive means to navigate the clouds of data contained within pStruct. The result of this project is a new concept in data organization and interaction that is applicable to a wide variety of content storage and retrieval application areas.

Index Terms—web forum community, data organization, multi-threaded programming, data visualization, real-time rendering

I. INTRODUCTION

TRADITIONAL web forums enforce a strict taxonomy of conversation organization, which users must conform to in order to participate in the discussion. Messages are relegated to conforming to this strict taxonomy, and forever are presented in accordance with its organization. This behavior does not allow for new connections to be made among threads. Instead, the system is immutable, with context enforced by a set structure outside of the users' control. pStruct, as a responsive system allows content to self-organize into logical informational clusters of data, linked by the users' interactions with the data. Thus, new users visiting the forum will see a new topic emerge as the data is reorganized to better suit users' needs and in response to other users' interaction with said data.

pStruct is a social software framework that is based around the concepts of connectionism and post-structuralism. Connectionism and post-structuralism are concepts which stipulate that pieces of information or texts have no meaning in and of themselves, and instead meaning is only derived from a text's relationship with other texts in any given system [1]. pStruct uses these philosophies to address the problems inherent to online communities, in particular, the fact that information must conform to strict taxonomies in traditional computer mediated communication systems. At the core level, the purpose of the pStruct project is to remove the usual threaded-conversation format imposed on web community interaction, and instead allow data the opportunity to organize itself more intelligently in response to user interaction.

The result of a pStruct web forum is a community in

which data interaction is guided by the relationships of the users to the data. Individual users subtly influence the data's organization, so the actions of the entire community shape the forum into a previously unforeseen structure reflecting its interests.

II. ARCHITECTURE

pStruct is architected as a massively multi-threaded system with influence from distributed multi-agent system architectures [2]. pStruct works by allowing data and users to exist in a specific ontology as micro-programs, each operating independently and competing for a fixed number of resources. Internally, the forum state is represented as an undirected graph. Nodes, representing things such as users or posts, have edges that connect nodes to each other with a strength representing the aggregate amount of activity across that connection.

At the core level of pStruct is the World, which houses Nodes, Factories, Protocols and Services (see Figure 1). Nodes within pStruct are micro-programs which run in a manner similar to agent systems, in that each node has beliefs, desires and goals independent of every other node. Nodes can represent people, responses to data, stories, images or any other type of content. Behaviors drive nodes to find nodes with related information, thus forming a self-organized system. Factories are sub-programs which manage the creation and destruction of objects. A NodeFactory manages node creation and maintenance while the Connection Factory manages the maintenance of connections and their strengths.

Protocols are small sub-programs, which engender communication between nodes in the pStruct system. Included are a messaging protocol, which enables node-to-node messaging on a global level, as well as an affordance protocol, which enables messaging on a local connection level. Messages are used for Nodes to communicate with each other in a deterministic fashion, such as a Node announcing its presence to every other node, or a user instant-messaging another user. Affordances allow probabilistic communication, such as a Node asking for specific data, but not caring whether or not another node answers the request. Affordance requests are relayed locally through node-to-node connections, while messages are sent directly to recipients.

Affordance requests and messages carry the data necessary for the pStruct system to self-organize. For example, in order for textual content to self-organize, each node containing text sends out an affordance requests to its connected nodes containing in it the concordance data for

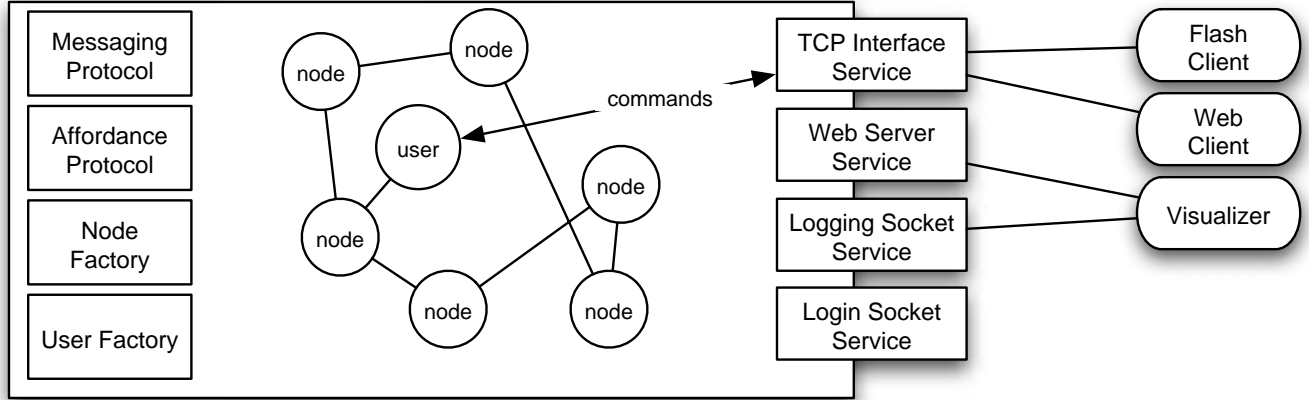


Fig. 1. Overview of the pStruct system architecture. Protocols and Factories provide support for a large, undirected graph of Nodes. The structure and content are communicated via a set of Services to external clients for visualization and interaction.

that node. Nodes that are n links away from the sending node receive the request, compare the concordance data to their own, and connect to the source node if the concordances are sufficiently similar. Affordance requests propagate through the undirected graph's structure, in a node-oriented peer-to-peer fashion. Messages are used for communication of data from a given node to all nodes in the system, nodes of a particular type, or a specific node. Messages enable nodes to communicate with other nodes regardless of the system organization by directly delivering the information.

Services are sub-programs which perform different activities within pStruct, such as enabling communication with the outside world, handling user registration and handling database communication. Services operate as independent applications within the system, and perform activities independently of the activities of nodes within the system. The systems within pStruct include a server for getting snapshots of the system state as well as the content of individual nodes, a logging server for tracing system activity, and the client-interface server to send commands to nodes.

III. VISUALIZATION

As pStruct is a self-organized graph, responding to user and data interaction within a web forum, the resulting structure is unknown, unlike in the traditional web forum case, where the structure is externally imposed. While users can interact with pStruct through a forum style interface, only a limited view of the graph's structure is made apparent, and the overall shape and characteristics are lost. Therefore, an additional component of the pStruct system is necessary to visualize the internal organization of the forum, to intuitively present the effects of the new design philosophy of pStruct.

The visualizer is a separate entity from the pStruct system. It passively renders the graph structure of the forum, by reading the xml log feed from the server. Log events such as *new-node* or *update-connection* are used to maintain a duplicate of the forum graph in a rendering-friendly

representation.

The first challenge in rendering an unstructured graph is to choose reasonable 3D positions for each graph node, which is called an embedding [3]. The distance between two adjacent nodes should represent the strength of the edge connecting them, while unconnected nodes should spread out to maximize visibility. To embed the graph in accordance with these constraints, the visualizer calculates a physical simulation of dynamic forces on the graph nodes until a stable configuration is reached. Edges are modeled as springs, with spring constant and rest length proportional to the strength of the edge. Nodes exert a repulsive force on each other, to maximize spread. To keep the embedding from growing unbounded, an overall attractive force pulls nodes towards the center of mass of the graph. The simulation then computes an embedding that satisfies these constraints (see Figure 2a).

Once the embedding is complete, the graph is ready to be rendered. The visualizer uses OpenGL to create an image of the graph's representation in 3D. The nodes and edges are rendered as billboarded polygons textured with a 2D gaussian distribution function. Alpha blending is used so that nearby nodes will blend together to form an apparent single, larger component. Two different blending techniques can be used for the node sprites - additive or transparency. Additive blending simply adds the new pixels into the old frame buffer pixels, rather than replacing them. This creates an effect in which the nodes look like light sources, and a set of closely adjacent nodes looks like a brighter light source (see Figure 2c). The result is attractive, but due to the depth-independent nature of the final image, occlusion cues are absent. Occlusion is important for determining the foreground from the background. Alternately, transparency blending simulates physical transparent objects, which partially occlude objects behind them. The result makes nodes look like small drifts of smoke, which coalesce into larger clouds, obscuring other clouds behind them (see Figure 2b). Edges are then always rendered on top with additive blending, so the

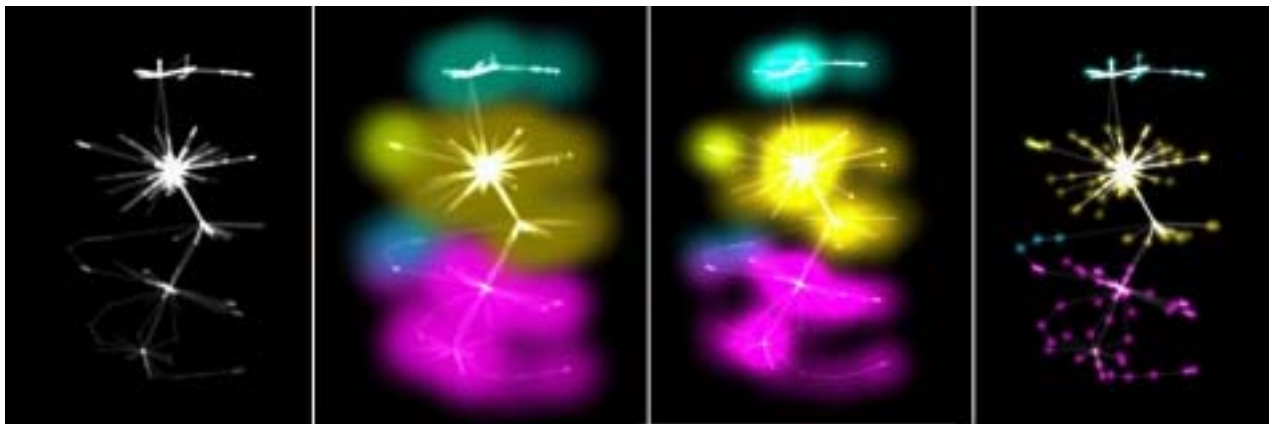


Fig. 2. Example visualizations of a simulated dataset. *from left to right*: (a) The 3D structure of the spring embedding. (b) Clusters visualized with transparency blending. (c) Clusters visualized with additive blending. (d) Small sprites for easy individual node identification.

complete structure of the graph is always visible, regardless of occluding nodes.

The last step is to reconstruct semantic information about the forum. Specifically, in a traditional forum, threads and topics of discussion are clearly marked, but in pStruct, the self-organization allows new topics to converge out of common themes and through user interaction. An overview of the forum should then be able to identify the components that make up cohesive topics. This can be posed as a graph analysis task of determining clusters. The visualizer use a technique called normalized cut [4], which partitions a graph into two components of similar size, with the minimum cost cut (sum of the weights of the cut edges). Recursive application of normalized cuts segments the graph into k clusters of related content. One problem with this technique is that it does not gracefully handle incremental updates to the graph (new node, remove node, etc.), so it must be rerun each update. For incremental graph updates, pStruct can alternately use a different clustering technique based on the spatial positions of the nodes, called k -means clustering [5]. k -means iteratively clusters data into k clusters based on a spatial distance metric, and can handle incremental updates to the data with little additional cost.

IV. CONCLUSION

pStruct is a radical departure for computer mediated communication systems, and the hope is that through its use, the traditional regimented taxonomies of discourse that have become endemic to web communities will be rendered obsolete. Through self-organization and a responsive environmental ontology, online communities will no longer be subject to a macro-based structure, and instead develop organically according to usage. The challenge with an open-ended system like pStruct is in the area of representation, specifically how to enable interaction with such a complex system. Through our research, we have presented one method of interacting with the system using a new method of visualizing undirected graphs. Future development will provide further methods of interacting

with the system, including web-based interfaces and tangible computing devices. pStruct is designed in such a way as to enable it to function as a system for dealing with any complex set of data, and future projects will involve using pStruct as a method of analysis for complex systems beyond web discussion forums.

REFERENCES

- [1] Paul Cilliers, *Complexity & Postmodernism*, Routledge, London and New York, 1998.
- [2] Doug Lea, *Concurrent Programming in Java Second Edition, Design Principles and Patterns*, Addison-Wesley, Boston, 2000.
- [3] T. Kamada and S. Kawai, "An algorithm for drawing general undirected graphs," *Information Processing Letters*, vol. 31, no. 1, pp. 7–15, 1989.
- [4] Jianbo Shi and Jitendra Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.
- [5] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264–323, 1999.



Ethan Kaplan is a media artist and researcher working specifically in the fields of virtual communities, fanaticism and online identity issues. He received his B.A. in Visual Media, Film and Computer Studies from University of California, San Diego in 2001 and is currently pursuing both a Master of Fine Arts and Master of Arts in the Art and Media Art and Technology departments respectively. His IGERT advisors are Tobias Höllerer and George Legrady.



Stephen DiVerdi actively researches in the fields of real-time graphics and augmented reality. He received his B.S. in Computer Science from Harvey Mudd College in 2002, and is currently pursuing a Ph.D. in Computer Science at University of California in Santa Barbara. His IGERT advisors are Tobias Höllerer and George Legrady.

Aesthetic Techniques for the Visual Display of Information

August Black

IGERT Fellow

Media Arts and Technology

Abstract—IGERT summer research, conducted by August Black and George Legrady is described and summarized, emphasizing the methods of the research and showing visual samples of the results.

Index Terms—IGERT, research, visualization, data analysis, aesthetics.

I. INTRODUCTION

The main objectives of this summer's research were to develop visualizations of static and time-based data sets for the purpose of mining out un-perceivable information and knowledge. For this research, there are two separate projects involved. One is the Seattle Public Library project, where call numbers of books are being recorded and the amount of checked-out books per half-hour, per Dewey decimal category are to be tracked and visually displayed in quasi real-time. The second project is concerned with Pockets Full of Memories, where quantifiable descriptions of objects have been collected from exhibitions in Paris, Rotterdam, Linz, Budapest, and Helsinki. The task, here, would be to find criteria and mechanisms for searching the database and displaying the results in the form of multiple graphic visualizations.

The guiding principles are inherently artistic, however, and so being, the research should provoke aesthetic as well as informative results. To do this, a large space of possible visualizations must be "scanned", articulated, and evaluated. "Scanning" in this sense, means looking at the various sorts of informative aesthetics that make up our contemporary visual landscape. Articulation is then the production and materialization in the form of software sketches that display the "gist" of a visual representation. Evaluation is the process of assessing the visual to see what, if anything at all, it says about the data it represents and to define in what respect it shows something new or enlightening. Through this, an aesthetic statement should be formed that paints a picture of the data in such a way that it can hold the attention of an audience.

Another thing that can be said about the research is that it was done according to methods more associated with studio type work (i.e. as a work of art), so there is no hypothesis to test or objective outcome to establish. The research can be seen as a preliminary stage for a larger production, one that will be actualized and made public.

II. AESTHETICS, TRUTH, AND INFORMATION

Data and information, as well as the algorithms that are chosen to format, categorize, or measure that data, have no visual identity per se. The crux of the research is to interpret the data in a way that can be seen on a 2 dimensional flat screen. This is not always easy given that most data is naturally multi-variate. [11] Also, given any set of data, it is impossible to show all things

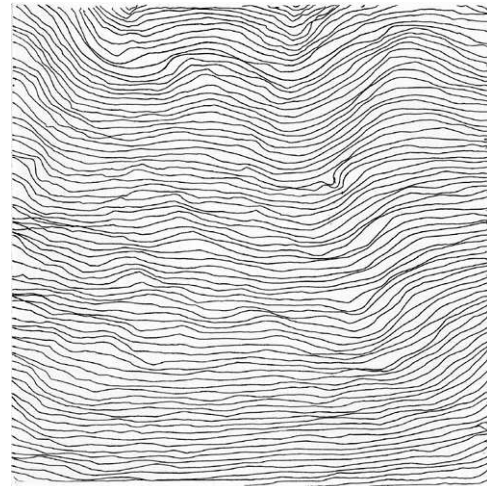


Fig. 1. ©Sol Lewitt - No Straight Lines. [9]

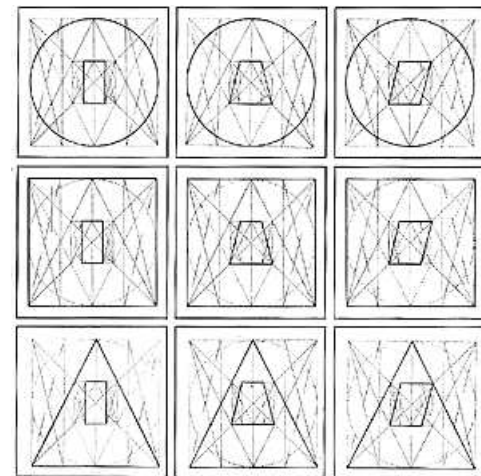


Fig. 2. ©Sol Lewitt - Geometric figures within geometric figures. [8]

about that data in one visual representation. Mapping characteristics from the data to observable characteristics in an image requires some sort of bendable distortion. Small things like scale and perspective can easily twist how one views the information attained from the data. [10] Additionally, choosing what aspects of the data to show places heavy editorial constraints on the outcome and this certainly frames how information is to be exposed from the data source. The visuals can only project truth to a certain degree, and it is within this degree of freedom that aesthetics for the piece should be derived.

On the one hand, an attempt is made to present the data in a way that illustrates important features. By basing much of

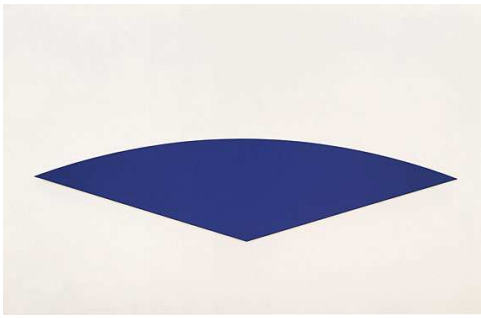


Fig. 3. ©Ellsworth Kelly - Dark Blue Curve. [3]

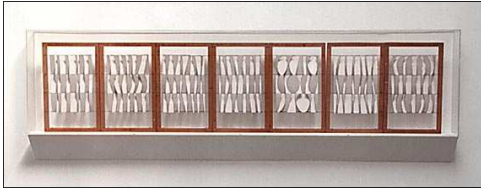


Fig. 4. ©Ellsworth Kelly - Sculpture for a large wall. [4]

the visual outcome on standard plotting mechanisms such as histograms, bar charts, and scatter plots much of the informative aspects of the data remain in tact. But, on the other hand, a larger visual context is considered. Figures 1 through 4 show a small sampling of the vast amount of visual references that go into framing this research as a work of art with elements of design and a general look and feel.

III. POCKETS FULL OF MEMORIES

"Pockets Full of Memories" is an interactive installation that consists of a data collection station where the public takes a digital image of an object, adds descriptive keywords, and rates its properties using a touchscreen. The data accumulates through-out the length of the exhibition. [6]

This project involves the quantitative description of everyday objects. Gallery-goers are asked to put personal objects on a scanning station where an image of the object is scanned and placed in a database. He or she then must describe the object using an entry form on a computer. The person can give the object



Fig. 5. Pockets Full Of Memories - installation at Pompidou Center, Paris. ©George Legrady

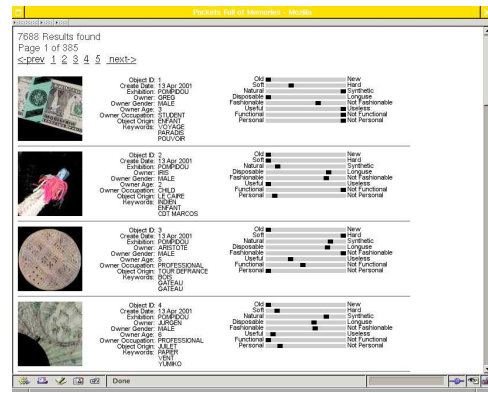


Fig. 6. Pockets Full Of Memories - data browser.

keywords and assign it scalar values between two descriptive captions: Old - New, Soft-Hard, Natural-Synthetic, Disposable-Long Use, Fashionable -Not Fashionable, Useful -Not Useful, Functional- Not Functional, Personal - Not Personal.

Since it's initial premiere, Pockets Full of Memories has accumulated over 7000 pieces of data. The data analysis will focus on showcasing relationships and differences in the collective perspectives of all of the communities represented through the contributions from the exhibitions in their respective cities of Paris, Linz, Budapest, Rotterdam and now in Helsinki. However, for the time being, most of the work has been to merge the databases from the 4 exhibits into a single data warehouse. Additionally, a web browser was written that can search and display the entries. [7]

Initially, the research was to span both projects evenly - Pockets Full of Memories and the Seattle Public Library. But, as time went on and our tasks unfolded in front of us, the urgency of the library project took precedence, leaving little time for Pockets Full of Memories. Now that the data warehouse has been constructed, visualizations of the data can be created, time permitting.

IV. SEATTLE PUBLIC LIBRARY

This project is a proposal for an installation in the mixing chamber of the Seattle Public Library that will visually map on a daily basis and over time the circulation of non-fiction books, revealing the collective reading interests of the library's patrons. The project's intent is to create a work that reflects the dynamic nature of contemporary society, a work that is informative and provides a stimulating, aesthetic experience. The visualization will be displayed on 6 plasma screens positioned horizontally across the span of the 24' glass structure behind the librarians' reference desk.

The numerically labeled topics (call numbers), as well as the semantic relationships (titles), and associative relationships (group of books checked out by a single person) will be recorded every half hour, mined for potential patterns, and graphically displayed on the 6 plasma screens.

Figures 9 through 22 show images of the various software sketches produced during the summer. Figures 9-11 are web applications that can be used to browse the data on demand.



Fig. 7. photo ©2004 The Seattle Public Library.



Fig. 8. The mixing chamber inside the SPL where the plasma screens will be installed. photo ©George Legrady.

The rest are animated demonstrations and can only be viewed as moving images. For this report, screen shots are provided.

V. CONCLUSION

What was conducted this summer is the first phase of a multi-phase project. At this point, the goal was to explore the characteristics and shape of the data being analyzed and produce multiple visualizations. But, the project is continuing its development until it will be installed in February of 2005. The next stage is to evaluate the positive and negative aspect of what has been done so far and compose a final draft.

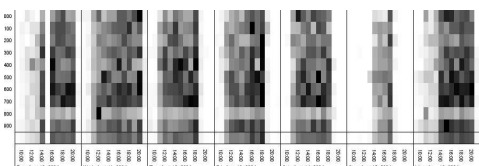


Fig. 9. SPL - histogram of amount of books checked out per hour over a week. Time is plotted in respect to Dewey categories. The lightness or darkness of a given unit represents the amount of books taken out in that category for that hour. With the default settings, black will represent a maximum value and white will represent a minimum value.

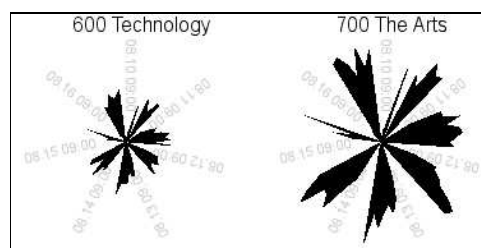


Fig. 10. SPL - Radar spatter plot showing 7 days worth of data in the 600's and 700's. This plot allows one to view any number of days together in one leaf-like image. The relationship between the days is emphasized.

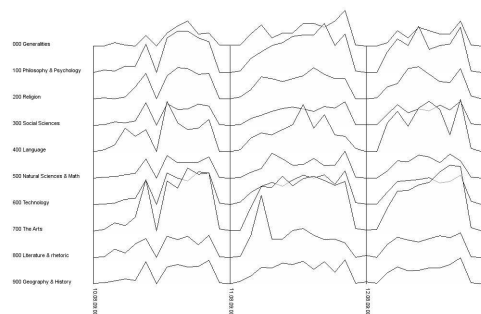


Fig. 11. SPL - waveform of amount of books per hour over 3 days. The waveform plot can show values that are not limited to the 8-bit grey scale values of the histogram.

ACKNOWLEDGMENT

This research was conducted with Professor George Legrady and exists as it is through dialogue with him. Also, thanks are due to Professor Tobias Hollerer for his suggestions and critique.

REFERENCES

- [1] S. Fortune. C program to compute 2d voronoi diagrams and delaunay triangulations. <http://cm.bell-labs.com/who/sjf/>.
- [2] T. Jones. *AI Application Programming*. Charles River Media, 2003.
- [3] E. Kelly. Dark blue curve. http://www.guggenheimcollection.org/site/artist_works_72_0.h
- [4] E. Kelly. Sculpture for a large wall. <http://www.artseensoho.com/Art/MARKS/kelly98/kelly1.html>.
- [5] S. Kloder. python source code to compute 2d voronoi diagrams and delaunay triangulations. <http://www-cvr.ai.uiuc.edu/~kloder/ece450/index.htm>.
- [6] G. Legrady. Pockets full of memories. <http://www.georgelegrady.com>.
- [7] G. Legrady. Pockets full of memories. <http://pocketsfullofmemories.com/>.
- [8] S. Lewitt. Geometric figures within geometric figures. http://www.parasolpress.com/lewitt_2.html.
- [9] S. Lewitt. No straight lines. <http://www.crownpoint.com/artists/lewitt>.
- [10] E. Tufte. *The Visual display of quantitative information*. Graphics Press, 1983.
- [11] E. Tufte. *Envisioning Information*. Graphics Press, 1990.

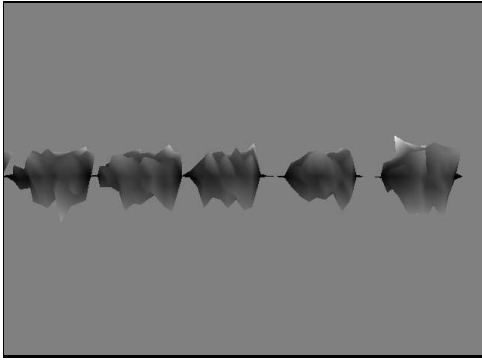


Fig. 12. SPL - OpenGL blob plot - Time is plotted along the center axis. The peaks around the radius are the amount of books checked out in a Dewey category. There are 10 points around the centered axis, each for the 000's, 100's, 200's, etc. The more books checked out for a given hour in a given category, the more white it's peak will be.

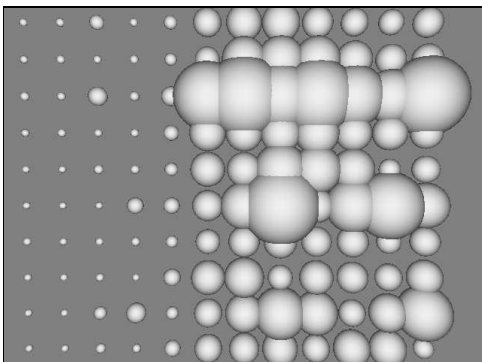


Fig. 13. SPL - OpenGL simulated voronoi histogram. Using overlapping spherical shapes where size of the sphere is dictated by value for that unit, this "clotting" histogram builds a general abstract form of a day's worth of values for outgoing books. From left to right is time in hours from 9:00- 20:00. From bottom to top are the Dewey categories segmented by the 100's. At most times, the diagram will "bulge" around the 700's category.

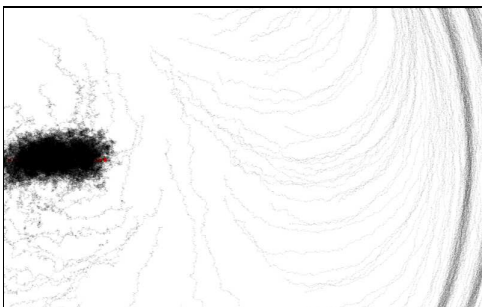


Fig. 14. SPL - stochastic walking fuzziness. Lines are drawn on the screen that are attracted to certain target areas. The target areas are the red points. They represent values for outgoing books. The image that is formed can be layered to build up a blackened smoky picture of daily, weekly or monthly data.

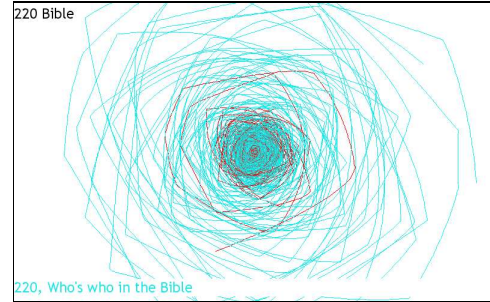


Fig. 15. SPL - stochastic spiral display. The daily data (log of all circulating books, circa 37,000) is interpreted. From this spiral shaped lines are created based on the characteristics of that Dewey category. Each line represents a single entry (book, CD, audio tape, etc.). The longer the line, the longer the book has been out in circulation. The incremental angle of the line segments is determined by the type of entry - books will have the largest angle increments. A bright blue line means that the call number of that book is closer to the minimum call number within that step size of Dewey call numbers. For example, if we take a Dewey category based on a step size of 10 (ex 50-59), a call number of 50 would be bright blue. A call number of 59 would be bright red. And, a call number of 55 would be grey. All the lines are plotted one on top of the other.

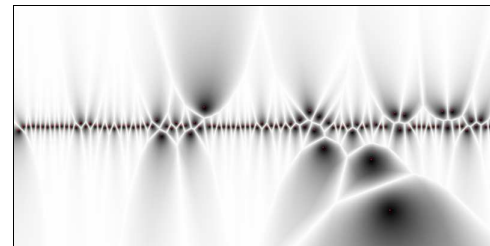


Fig. 16. SPL - voronoi image, 100's over the 10's. This shows a voronoi images (not a graph or diagram) based on a distribution of points within the image bounds. The images are constructed using a simple function to compute the closest and next closest point out of our initial setup of points. it does this for each pixel in the image. The resulting pixel value is the distance of the (closest divided by the next closest multiplied by 255. The red points represent Dewey categories segmented by the 100's and 100's. They are displaced from 1/2 (height of the image), based on the amount of books taken out in that Dewey category. The image of the 100's is overlapped with the 10's.

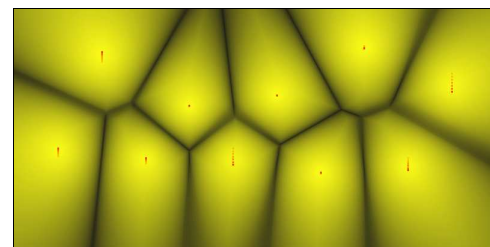


Fig. 17. SPL - voronoi diagram. Similar to the previous voronoi example, now with variables for color, interpolation of one set of data to the next (good for making animations), fade type, and whether or not to plot the actual voronoi tessellation. Tessellation diagram is drawn using a modified version of Steve Fortune's line-sweep algorithm. [1, 5]

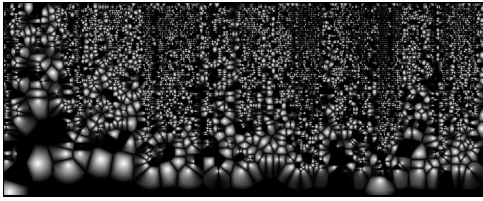


Fig. 18. SPL - voronoi diagram of circa 37,000 books in circulation on 08-18-2004. The voronoi tessellation which we see on screen is rendered around imaginary Delaunay points representing each and every book that is currently in circulation. This image shows approximately 37,000 entries. Each Delaunay point represents a single book entry. From left to right, the Delaunay points are plotted according to the respective call number of the individual book, 000 to 999. The Delaunay point is then displaced from the top of the screen according to how long it has been out of the Library at that date. The Delaunay point for a book that has just been checked out will be at the very top of the screen, where points representing books that have been out for over three weeks will be plotted nearer to the bottom of the screen.

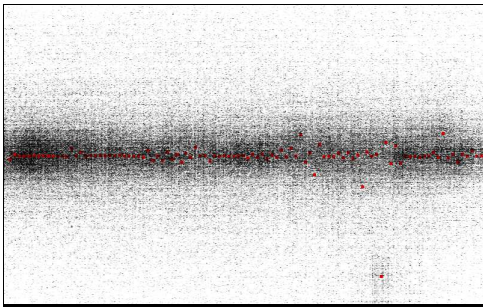


Fig. 19. SPL - genetic algorithm. [2] A genetic algorithm was constructed to breed and mutate dots on screen that would eventually converge on a target point. The resulting images show a Xerox-like blur around dots representing amount of books. This image shows the Dewey categories split by the 10's (red dots) for one hour's worth of data. Hourly images can be layered or animated to show variation.

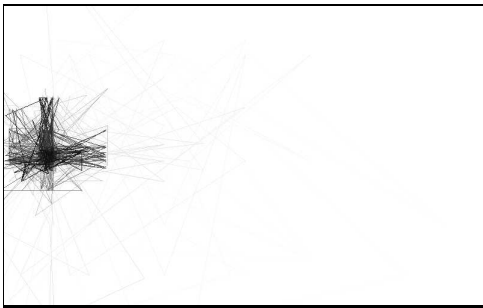


Fig. 20. SPL - genetic algorithm with lines. A similar algorithm is used to converge on a target area within the image, only now line segments are drawn through each successive population. The line segments are painted darker as the generations go from one to the next.



Fig. 21. SPL - genetic algorithm with bezier lines. The chromosomes are characterized by a set of x coordinates, a slope and a intercept. the y coordinates will be calculated using the line function $y = m \cdot x + b$ (where m = slope and b = intercept) Each line then acts as Bezier handles, for shaping the curve. Each "generation" is a single curvy contour with multiple Bezier handles (the lines) in between.

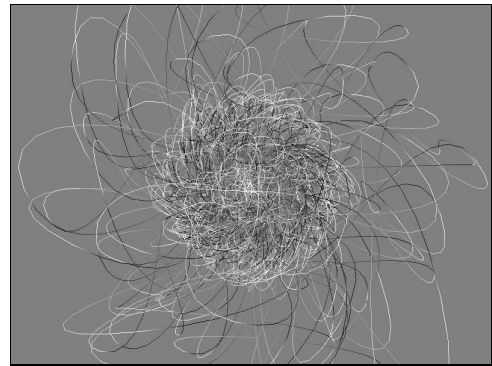
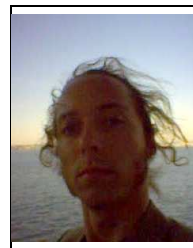


Fig. 22. SPL - OpenGL bezier blobs. SPL daily data for a Dewey category segmented by the 10's is used to draw this 3D image. The blob is formed by a single line split up into bezier line segments. Each line segment represents a single entry and can be seen as a black to white color transition on the line. The line starts at the center and rotates and grows outward. The radius is incremented based on the length of the books title (multiplied by some random variation). The angles in the y and z direction are based on whether the item is a book or CD or DVD.



August Black has an awful habit of calling himself an artist. Previously, this has meant making marks on paper and later on canvas. Now, this means almost anything concerning material, concept, and form. His research is based in the overlap of media, focusing mostly on the kinds of audiences that are created and induced by emerging conventions of observation and involvement.

He works in radio, television, software, networks, comics, text, and projected sound/light. Collaborating with others on various free radio stations in Austria, he's devised a technique for performing live radio on a shoestring budget from networked locations outside of the studio, transforming the location at hand into material and subject for conceptual play.

He is currently an IGERT research fellow at the University of California Santa Barbara.

<http://aug.ment.org>

The Effect of Context Cues in Saccadic Decisions

Barbara A. Drescher

IGERT Fellow

Psychology

Laura E. Boucheron

IGERT Fellow

Electrical and Computer Engineering

Abstract—We investigate the forces driving decisions about human eye movement during visual search tasks. Quick eye movements, called saccades, may be driven by bottom-up or top-down processes. Bottom-up processes include evaluations made by the visual system such as the color and orientation of objects. In contrast, top-down processes are cognitive in nature. They consider the context of a scene in order to determine the likelihood of an object’s presence or appearance at a given location.

The purpose of this research is to integrate information from bottom-up and top-down processes to create a more accurate computational and automated model of human saccadic behavior. Current models are utilized for the analysis of low-level features. We evaluate the contribution of these features to the saccadic decisions and weight them accordingly. Additional factors, weighted through experimentation, are added. In the future we would like to incorporate the ability to adjust the model according to the task. Our goal is to use the information gathered from this research to improve computerized object recognition and search.

Index Terms—top-down processing, bottom-up processing, saccade, low-level image features, computational vision model, object recognition, contextual cues

I. OVERVIEW

WHILE we perceive virtually all areas of the visual field, the details come more directly from our brains than our eyes. Receptors in our eyes are most concentrated in a small region at the center of retina called the fovea. As a result, the space at which this area is fixed gives us the greatest resolution and, therefore, the most information. In order to make sense of the entire visual field, we move this high resolution fovea to relevant regions of the scene [1]. These quick eye movements are called saccades. To decide where to move our eyes, however, some processing of information on the periphery must occur.

There is a large body of evidence that demonstrates that peripheral processing guides visual saccades [2–5]. The controversy lies in what type of information gathered in this manner most influences saccadic decisions. Eye movements may be driven by bottom-up or top-down processes. Bottom-up processes include evaluations made by the visual system such as the color and orientation of objects. In contrast, top-down processes are cognitive in nature. They consider the context of a scene in order to determine the likelihood of an object’s presence or appearance at a given location.

In recent years, several models have been constructed to evaluate images for low-level features in an attempt to predict human saccadic behavior [1–4, 6, 7]. These have enjoyed limited success for several reasons. First, these models fail to consider top-down processes or severely limit the contribution these processes make to saccadic decisions. In addition, the balance of the many factors involved in these decisions most likely varies greatly according to task.

The purpose this research is to integrate information from bottom-up and top-down processes to create a more accurate computational model of human saccadic behavior. Current models are utilized for many of the low-level factors to be analyzed. We evaluate the contribution these features make to the saccadic decision and weight them accordingly. Additional factors, weighted through experimentation, are added. We begin with a model that predicts behavior for visual search tasks. In the future we would like to incorporate the ability to adjust the model according to task.

II. THE INFLUENCE OF EXPECTATION

Many researchers have documented effects of low-level features on saccadic decisions in visual search tasks. Properties such as eccentricity and salience have been successfully modeled and tested [3, 4, 8]. However, objects in the real world tend to occur in somewhat predictable situations, or “expected” contexts [9].

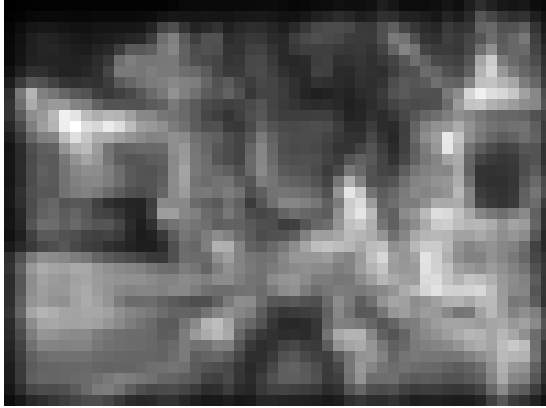
If peripheral processing (prior to eye movements) provides contextual information, then eye movements should be directed toward locations where the object is likely to appear. To test this hypothesis, we tracked the eye movements of 20 undergraduate students while searching for objects in natural scenes. Scenes consisted of photographs of background such as a hotel room, a beach, or a kitchen table. Objects such as a lamp, a sandcastle, and a salt shaker were placed within the scenes. Each background was viewed in three conditions, one in which the object appeared in an expected location, one in which the object appeared in an unexpected location, and the last without the object present.

Random assignment of the participants and stimuli into 3 groups ensured that each participant viewed each background once and only once. One-third of the images for each group contained objects in the expected location, one-third an unexpected location. The remaining third contained no object. Trials consisted of a fixation of 1 second followed by a word or phrase that described the target object for 2 seconds. Following another fixation, the image appeared for 2 seconds. Participants were instructed to indicate if the object appeared in the image.

Accuracy in this task for when the target was present (Expected and Unexpected) was measured as the distance to the endpoint of the first saccade from the target. Distance to target was significantly greater in the unexpected condition than in the expected condition. When no object was present, the first saccades were closer to the expected location of the object than the unexpected location.



(a) Canoe Picture (missing case)



(b) Saliency Map

Fig. 1. (a) Original missing image for “canoe” and (b) the corresponding saliency map

III. TOP-DOWN AND BOTTOM-UP MODELS OF VISUAL ATTENTION

There are two main directions that visual attention research has taken: bottom-up and top-down. Computer vision research in bottom-up visual attention focuses on the implementation systems to mimic biological responses to low-level features of the visual scene [10]. Top-down visual attention, on the other hand, seeks to develop a hierarchical model of the image, whereby visual search tasks may be better quantified at the image processing level [11, 12]. It should be noted, however, that some researchers (e.g. [10]) seek a biological basis for top-down visual attention.

The bottom-up visual attention model described in [10, 13] has been implemented in an intricate neuromorphic vision toolkit available by request from the Ilab at University of Southern California [14], and is used by many vision researchers to assist or begin their research. This algorithm relies on the use of a saliency map to determine the evolution of visual attention. In this case the saliency map is a weighted average of three features evaluated at multiple scales: color, intensity, and orientation [10]. For

example, Figure 1 shows one of our test images with the corresponding saliency map. While human vision may rely to some extent on both bottom-up and top-down visual features [11], we have found in our research that the influence of these low-level visual features, such as saliency, have a relatively small effect on visual attention in search tasks. Our experimental results are presented in Section IV.

As an illustration of the importance of contextual cues in visual processing, consider Figure 2. Object recognition is still a difficult and largely unsolved problem in computer vision, making the development of contextual descriptions



(a) “Car”



(b) “Person”

Fig. 2. Illustration of the importance of context in object recognition and image understanding. The two pictures “car” and “person” shown above actually include the same blob rotated by 90 degrees. It is our understanding of context (a car cannot stand on its nose and people don’t lie in the gutter) that leads to the individual interpretations of the two scenes. Note that this occurs even in the case of severely degraded images such as shown above. Images from [12].

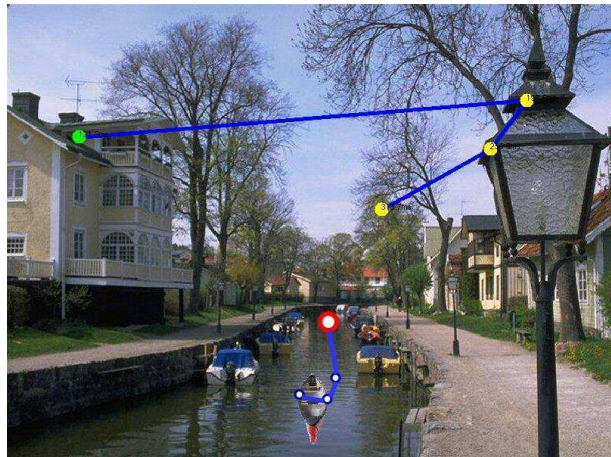
of natural images difficult to accomplish with current state of the art. As a result, there is concerted effort to achieve semantic descriptions of scenes using low level image features, such as color [15], the statistics of low level features [12], or a limited ontology [11]. There has been some limited success with such strategies, but such algorithms deal with a very constrained knowledge base [11, 12]. It is thus important to also consider methods of visual learning, several examples of which may be found in [12, 16, 17]. From the aspect of biologically based computer vision, human visual attention appears to exhibit learning and memory of visual cues to guide attention [18–21]. It is the implementation of these contextual descriptions that is sorely needed in computer vision. Using the behavioral data collected as described in Section II, we hope to improve both the contextual description and learning capabilities of our computer vision model. As an example of the significance the behavioral data lends to contextual cues, consider Figure 3.

IV. RESULTS

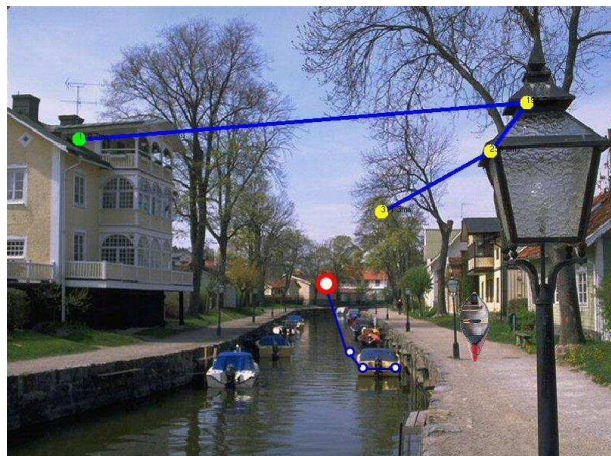
Principle Components Analysis was applied to the dataset to determine the relative contributions of several factors to the first saccades made by participants. This information was integrated into a computational model that weights factors of saliency, eccentricity, and context in predicting first saccades. The computational model allows a limited automated computer analysis of images for this purpose. This model will be modified through experimentation and computational refinement as additional data are collected. First, we plan to adjust distance algorithms to weight tiles in a nonlinear manner. For example, the current formula reduces probabilities as distance from the image center increases. This relationship, however, reverses at a mean distance. In addition, human saccades are not accurate, falling short more often than “overshooting” the target. The current model decreases probabilities at a stable rate as distance from the target increases. It is our ultimate goal to produce a program with the ability to automatically determine more variable values, in particular the expected location of objects, in an automated fashion.

REFERENCES

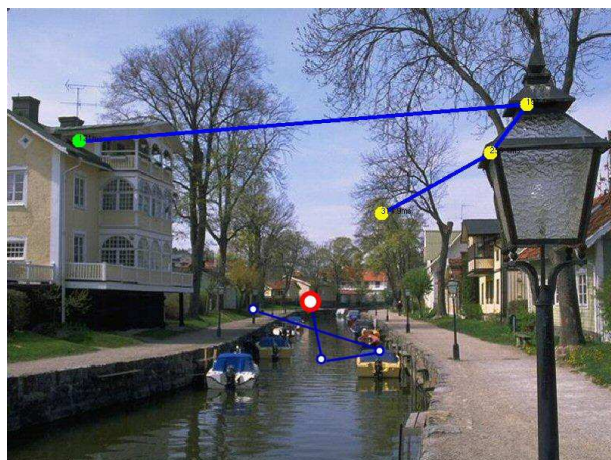
- [1] I. A. Ryback, V. I. Gusakova, A. V. Golovan, L. N. Podladchikova, and N. A. Shevtsova, “A model of attention-guided visual perception and recognition,” *Vision Research*, vol. 38, pp. 2387–2400, 1998.
- [2] D. Noton and L. Stark, “Scanpaths in eye movements during pattern perception,” *Science*, vol. 171, no. 3968, pp. 308–311, Jan. 1971.
- [3] L. Itti and C. Koch, “A saliency-based search mechanism for overt and covert shifts of visual attention,” *Vision Research*, vol. 40, pp. 1489–1506, 2000.
- [4] D. Parkhurst, K. Law, and E. Neibur, “Modeling the role of salience in the allocation of overt visual attention,” *Vision Research*, vol. 42, pp. 107–123, 2002.
- [5] K. Schill, E. Umkehrer, S. Beinlich, G. Kreiger, and C. Zetsche, “Scene analysis with saccadic eye movements: Top-down and bottom-up modeling,” *Journal of Electronic Imaging*, vol. 10, no. 1, pp. 152–160, Jan. 2001.
- [6] V. Navalpakkam and L. Itti, “A goal oriented attention



(a) Expected

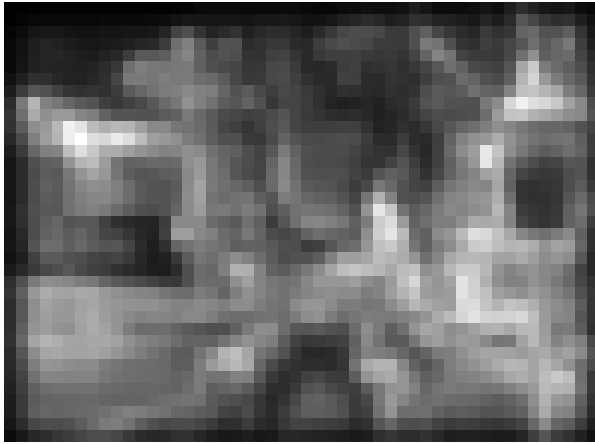


(b) Unexpected

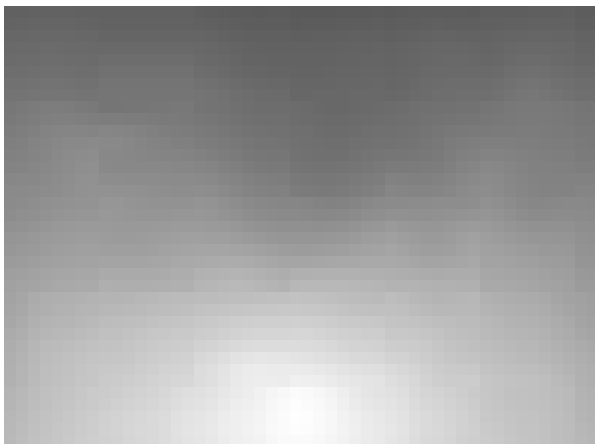


(c) Missing

Fig. 3. Comparison of the saliency model for visual attention with behavioral data. For the saliency model, the green dot indicates the first saccade, and the yellow dots the following three saccades. For the behavioral data, the red and white dot indicates the first saccade, and the blue and white dots the following three saccades.



(a) Original saliency map



(b) Resulting combination map

Fig. 4. The resulting weighted combination of the saliency map shown in (a) and the additional factors of eccentricity and context as previously discussed.

guidance model,” *Lecture Notes in Computer Science*, vol. 2525, pp. 453–461, Nov. 2002.

- [7] R. P. N. Rao, G. J. Zelinsky, M. M. Hayhoe, and D. H. Ballard, “Eye movements in iconic visual search,” *Vision Research*, vol. 42, no. 11, pp. 1447–1463, 2002.
- [8] C. Araujo, E. Kowler, and M. Pavel, “Eye movements during visual search: The costs of choosing the optimal path,” *Vision Research*, vol. 41, pp. 3613–3625, 2001.
- [9] I. Beiderman, “Perceiving real-world scenes,” *Science*, vol. 177, no. 4043, pp. 77–80, July 1972.
- [10] L. Itti, “Models of bottom-up and top-down visual attention,” Ph.D. dissertation, California Institute of Technology, Pasadena, January 2000.
- [11] V. Navalpakkam and L. Itti, “A biologically inspired scene based question answering agent,” in *Proc. 9th Joint Symposium on Neural Computation*, Pasadena, CA, May 2002.
- [12] A. Torralba, “Contextual priming for object detection,” *International Journal of Computer Vision*, vol. 53, no. 2, pp. 153–167, July 2003.
- [13] L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.
- [14] Ilab homepage at USC. [Online]. Available: <http://ilab.usc.edu>

- [15] A. Oliva and P. G. Schyns, “Diagnostic colors mediate scene recognition,” *Cognitive Psychology*, vol. 41, pp. 176–210, 2000.
- [16] B. Moghaddam and A. Pentland, “Probabilistic visual learning for object recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 696–710, July 1997.
- [17] T. O. Binford and T. S. Levitt, “Evidential reasoning for object recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 7, pp. 837–851, July 2003.
- [18] H. Intraub, “The representation of visual scenes,” *Trends in Cognitive Science*, vol. 1, no. 6, pp. 217–222, Sept. 1997.
- [19] M. M. Chun and Y. Jiang, “Contextual cueing: Implicit learning and memory of visual context guides spatial attention,” *Cognitive Psychology*, vol. 36, pp. 28–71, 1998.
- [20] J. M. Henderson, “Human gaze control during real-world scene perception,” *Trends in Cognitive Science*, vol. 7, no. 11, pp. 498–504, Nov. 2003.
- [21] A. Hollingworth and J. M. Henderson, “Accurate visual memory for previously attended objects in natural scenes,” *Journal of Experimental Psychology*, vol. 28, no. 1, pp. 113–136, 2002.
- [22] L. Itti, C. Gold, and C. Koch, “Visual attention and target detection in cluttered natural scenes,” *Optical Engineering*, vol. 40, no. 9, pp. 1784–1793, Sept. 2001.
- [23] D. Walther, L. Itti, M. Riesenhuber, T. Poggio, and C. Koch, “Attentional selection for object recognition—a gentle way,” *Lecture Notes in Computer Science*, vol. 2525, pp. 472–479, Nov. 2002.



Barbara A. Drescher received her B.A. in Psychology from CSU Northridge and her M.A. in Experimental Psychology from CSU Northridge. Her research interests include visual search, space perception in virtual environments, effects of color on human physiology and emotion, statistical methods, pseudoscience and belief, and interactive learning. Her academic advisor is Dr. Miguel Eckstein.



Laura E. Boucheron received her B.S. in Electrical Engineering in 2001 and her M.S. in Electrical Engineering in 2003 from New Mexico State University, Las Cruces, New Mexico. She is currently pursuing her Ph.D. in Electrical and Computer Engineering at UCSB. Her research interests include image processing in general, specifically image perception, image understanding, and the role of context in image processing and object recognition. Her academic advisor is Dr. B.S. Manjunath.

Realism and Perceptions of Data Quality in Computer-Displayed Maps

Tony Boughman
IGERT Trainee
Geography Department

Abstract— Issues in data quality are of interest as we continue to accumulate huge quantities of digital information. Creators of information are familiar with the inherent shortcomings in translating real-world observations into digital form, but do users understand the limitations of data? Cartographers are aware of the necessity of presenting an abstract view of the infinitely complex world. Improvements in computer technology allow increased realism in digital images and there seems to be a general movement toward more realistic display. More realistic visual rendering does not mean that the underlying data are more accurate and precise. Researchers are currently looking into the idea of communicating the degree of uncertainty in digital data to consumers by visualizing the uncertainty. Is the trend toward more realism doing the opposite and conveying the impression of more certainty in the data?

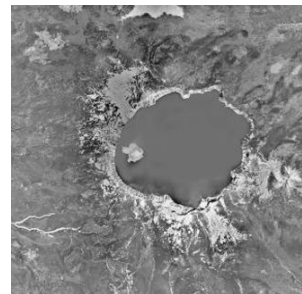
Do people infer greater accuracy and precision from maps that are more detailed and realistic? An experimental study will compare a more generalized, abstract map to a more detailed, realistic-looking map. The maps will be displayed on computer monitors and participants will be asked to make judgments on the relative accuracy and precision of the maps by rating them and by performing estimation tasks.

Index Terms—cartography, data quality, maps, realism, uncertainty.

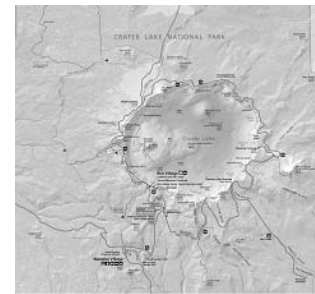
INTRODUCTION

ISSUES in data quality are of particular interest as we continue to accumulate huge quantities of digital information. Creators of geographic information are familiar with the inherent shortcomings in translating real-world observations into digital representations, but do users understand the limitations of data? Cartographers have long been aware of the necessity of presenting an abstract view of the infinitely complex world. Improvements in computer technology allow increased realism in digital images and there seems to be a general movement toward more realistic display [1]. This movement is expressed in plan-view, orthographic maps as well as 3-D visualizations, video games, and virtual environments. More realistic visual rendering does not necessarily mean that the underlying data are closer to reality, or more accurate and precise. Researchers are currently looking into the idea of communicating the degree of uncertainty in digital data to consumers by somehow presenting or visualizing the uncertainty [2]–[4]. Is the trend toward more realism doing the opposite and conveying, perhaps inadvertently, the impression of more certainty in the data?

This paper presents an experimental design to examine if people infer a higher level of data quality in terms of locational accuracy and precision from maps that are more realistic-looking. A comparison will be conducted among three types of images: a true color aerial photograph, a shaded-relief map with naturalistic coloring, and a simpler, more abstract map. The maps are typical of those used for locational reference or navigation.



a) image map¹



b) realistic-looking map²



c) abstract map³

Figure 1. Examples of maps at three points along the realistic—abstract spectrum.

The aerial photograph is an image captured by a camera rather than drawn by a cartographer yet because it is orthogonal and planimetrically correct (orthophoto) it may be considered to be a form of map [5]. This image map most closely represents realism—certainly photo-realism (Figure 1a). The realistic-looking map combines both raster and vector graphics and has contextual information such as water

¹ <http://terraserver.microsoft.com/>

² <http://www.shadedrelief.com/realism/>

³ <http://www.mapquest.com>

features, trails, railroads, buildings, naturalistic coloring, hill shading, land cover texture, etc. This map appears similar to how we would imagine the place to look if viewed from above (Figure 1b). The more abstract map is a vector graphic with simple colors and little added contextual information. It is modeled on the typical route-finding maps available on the World Wide Web (Figure 1c).

With the ascendance of the World Wide Web the Internet has surpassed printing as the leading medium of map delivery, therefore this study uses as its stimuli digital maps displayed on computer monitors [6]. Detail is limited by computer monitor resolution, typically 72 pixels per inch, while normal human visual acuity is much greater than this [7]. When viewing maps on computer displays or websites with interactive capabilities more detail is often enabled by zooming in to a larger scale; however, this study only looks at statically displayed maps at a fixed scale.

The maps in this study are designed with varying levels of realism and the concomitant levels of abstraction. Realism is a very complex concept with many interacting components. The stimuli used here vary through manipulations of context, detail, coloring, and topographical representation. A series of three maps (image, realistic, abstract) showing the same location at the same scale is used for each specific task or question. For each task or question, the level of generalization is held constant for the map feature involved. For example, a question asking to compare distances on a road network will use maps that show the same level of generalization for the roads [8].

This study is aimed at non-expert map users who most likely have no training in digital map creation. Participants drawn from the research pool of undergraduate students at the University of California, Santa Barbara are asked to make accuracy and precision judgments while performing common map use tasks. These tasks include making route decisions, estimating relative proximities, estimating relative sizes of map features, and judging how precisely distances are represented. Participants will rate each map in terms of the accuracy and precision of the information used to complete a given task. One question type, for example, asks participants to compare distances between pairs of objects on the map. They are asked “Is the distance greater from point A to point B or from point A to point C?” and given the option of “cannot be determined”. The differences in distance among pairs of objects will vary, with some pairs having an obvious difference (A to B is much farther than A to C). Other pairs will have less difference in distance, all the way down to A to B being equal in distance as A to C. In this way a threshold of where the “cannot be determined” responses begin can be found for each type of map. This method is used to establish how precisely distances are estimated.

The results of this study provide insight into the effects of cartographic realism on people’s inferences of the underlying

quality of the data. A relationship found here could then be employed in visualizing geographic data with regard to the conveyance of uncertainty. It may be feasible to use realism, or some component of realism, as a visual variable in situations that involve communicating the degree of data quality.

REFERENCES

- [1] Patterson, T. Getting Real: Reflecting on the National Park Service Maps. Presented at: International Cartographic Association (ICA), Mountain Cartography Workshop, Timberline Lodge, Mt. Hood, Oregon. <http://www.shadedrelief.com/realism/>. 2002.
- [2] Battenfield, B.P. Representing Data Quality. *Cartographica*, Vol. 30, No. 2-3. 1993.
- [3] McGranaghan, M. A Cartographic View of Spatial Data Quality. *Cartographica*, Vol. 30, No. 2-3. 1993.
- [4] Goodchild, M.F., and Clarke, K.C. Data Quality in Massive Data Sets. In Abello, J. et al. (eds.), *Handbook of Massive Data Sets*, 643—659. Netherlands, Kluwer Academic Publishers. 2002.
- [5] Muehrcke, P.C., and Muehrcke, J.O. *Map Use*. Madison, Wisconsin, JP Publications. 1998.
- [6] Petersen, M.P. “Trends in Internet Map Use.” *Proceedings of the 18th International Cartographic Conference 1997*, pp. 1635-1642. 1997.
- [7] Clark, R.N. “Experiments with Pixels Per Inch (PPI) on Printed Image Sharpness.” Clarkvision.com, <http://www.clarkvision.com/imagdetail/prnter-ppi>. 2003
- [8] McMaster, R.B. and Shea, S. *Generalization in Digital Cartography*. Washington, D.C., Association of American Geographers, 1992.



Tony Boughman has a background in farming, construction, and landscaping. He earned a B.S. in Geography in May, 2003, from Salisbury University, Maryland. He is currently in the Geography Department at UCSB, where his research interests are in GIS and Cartography. His research advisor is Dr. Sara Fabrikant.

Map Generalization for Mobile Display: Evaluation and Automation

Julie Dillemath

IGERT Fellow

Department of Geography

Nhat Vu

IGERT Fellow

Department of Electrical and Computer Engineering

Abstract—An understanding of effective maps for small digital displays and how users interact with maps while mobile is critical to map and mobile computer interface design. This project evaluates maps at different levels of generalization for an on-foot navigation task, and presents an approach for automatically creating a generalized map from an aerial photograph.

I. INTRODUCTION

Maps are key applications for mobile devices, from providing location-based services (LBS) to consumers and tourists via smart phones, to assisting field data collection and emergency services personnel with tablet PCs or wearable systems. While advances in systems development increase the capabilities of mobile devices in terms of graphical display, processing power, location-awareness and wireless communications, software and infrastructure development continues to bring more spatial data to these devices [1, 2].

However, very little research has been conducted on user interaction with maps on mobile devices in a field setting, especially with regard to map representation. The advent of digital media has introduced new research questions for digital maps, ranging from differences stemming from the medium itself, to technical factors including display size and capabilities of zooming and panning [4]. Mobility takes these issues still further, introducing a dynamic, outdoor work environment for which research studies conducted in static, indoor settings do not apply.

Our project begins to address some of these issues from two directions. First, a human-subjects experiment evaluates maps at different levels of generalization for the small displays of handheld computers through a field-based task. The second part of the project investigates methods of automatic generalization of features from an aerial photo, specifically vegetation. If the generalized map is shown to be more effective than the photorealistic one, a process to create a generalized version of an aerial photo automatically rather than manually can assist users with no cartographic training in creating a more effective map.

The experiment is a pilot study to investigate the effectiveness of map representation type, specifically with regard to level of generalization, by analyzing subject performance in an on-foot navigation task with a handheld computer used as a navigation aid. In examining subject time and performance accuracy as well as interaction with

the mobile device during the task, the results carry implications for map design for small, mobile displays and identify factors that affect the use of maps while moving.

II. PART 1: PILOT STUDY

A. Methods

The study evaluates maps for a small display at two different levels of generalization. At one extreme, an aerial photograph represents the least generalized condition, with photorealistic color and texture (Figure 1). As raw data, no cartographic design is applied, and shadows and radial distortion of vertical features contribute to the high level of detail. The second condition is a generalized map, a simplified and classified version based on the aerial photo, created by manually tracing polygons in a GIS and coloring each polygon according to the feature type it represents: building, sidewalk, paved road, grass/other ground vegetation, tree, sand or water. Colors are set to resemble those of the aerial photograph.

Sixteen subjects, graduate students at UCSB, completed the navigation task of following a route displayed on the map. All subjects did the task with both map conditions, and two different routes. Both the route order and map order were systematically varied among subjects. Routes were the same length, 0.74 km, covered similar-sized, non-overlapping areas of the UCSB campus, and contained an equivalent number of turns. Figure 1 shows both routes. The primary purpose of using a route marked on the map for the task was to require participants to continually interact with the digital map in determining which direction to walk and where to turn next, in order that data could be gathered on subjects' adjustments in level of zoom and panning for each map condition. In addition to a computer log file of all user interaction with the mobile device, data was collected by the researcher's observations of the subjects during the task, and through written questionnaires before and after the task. The hardware platform was a ruggedized tablet PC by Xplore Technologies, approximately the same dimensions as a sheet of paper, with a depth of 4 cm, and weight of 2 kg. The user interface was a Java application that displayed the map in a 6 x 6 cm window, surrounded by a pan frame, with incremental zoom-in and zoom-out buttons. The window size represented the limited display of a typical PDA-sized handheld computer. The advantage of using a tablet PC over a PDA for this research was that the Windows XP operating system allows the Java

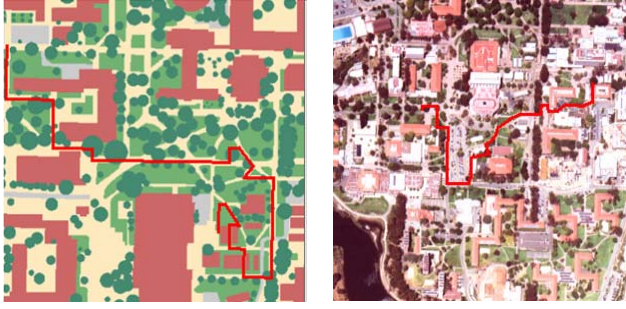


Figure 1. Route 1 with generalized map (left) and route 2 with aerial photograph (right)

application to be run without additional configuration for a PDA device.

The primary goal of the study was to evaluate the maps with regard to three dependent variables: time efficiency, amount of device interaction in the form of map browsing, and accuracy. Time efficiency considered the total amount of time it took each subject to complete the route for each map, walking at his or her normal pace. The amount of map browsing was defined as the amount of zooming and panning done by the subject in completing the task. Finally, the ability to complete the task accurately was seen as an important factor to consider. For the task, subjects were instructed to follow the route marked on the map as accurately as possible. Accuracy was measured by the number of errors subjects made along the route with respect to features. Walking around the wrong side of a tree, for example, or walking on sidewalk where the route indicates to walk on grass were counted as errors.

It was hypothesized that 1) the generalized condition would require less zooming and panning than the aerial photograph condition, given the comparatively lesser amount of detail and higher contrast among feature types, and 2) would result in faster time to route completion. 3) The aerial photograph condition was expected to result in better accuracy performance, since it contained more details and visual cues than the generalized map.

In addition, subjects' evaluative comments from the questionnaires provide information about the usability of the maps and interaction with the device.

B. Results

Overall, the generalized map performed better than the aerial photograph with regard to the variables that were tested, with a significantly shorter time to route completion ($t(15)=2.66$, $p=0.02$) and less map browsing ($t(15)=2.99$, $p=0.009$ for amount of zoom level changes; $t(15)=2.89$, $p=0.01$ for number of zoom levels used; and $t(10)=2.75$, $p=0.01$ for number of pans) with the generalized condition. These results support the hypotheses made regarding time and zooming/panning. Accuracy, however, was not shown to be substantially different between map conditions, though subjects made fewer errors with the generalized condition.

Subjects preferred the generalized map to the aerial photograph 10 to 4 (with 2 subjects undecided), reporting that the generalized map was easier to read, less complex and with less distracting or unhelpful details. Those who preferred the aerial photograph, however, liked that it was more realistic, had greater detail, and was easier to relate to the real world landscape.

From the data collected on the questionnaires, subjects reported sidewalks to be the most useful feature for orientation in both conditions, followed by vegetation and then buildings (Figure 2). In terms of the characteristic subjects used most to identify features, out of shape, color, size and texture, shape was most helpful. Texture and color were cited as least useful for identifying features.

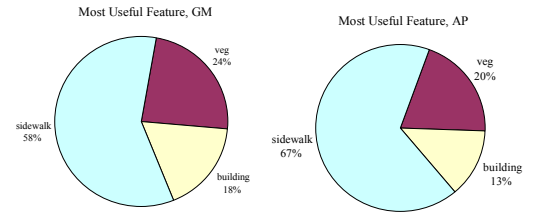


Figure 2. Most useful features for orientation as reported by subjects.

The results reported by subjects inform the second part of the project, automatic feature extraction and generalization. For this particular task in this landscape, shape of features is important to preserve, but texture and, to a lesser extent, color may not provide any additional useful information to a map user. Automatically creating a more generalized version of the aerial photo by removing unnecessary detail would create a more effective map from a raw dataset such as an aerial photograph. Sidewalks, the feature used most for the task by subjects, already look generalized on the aerial photo given their rectilinear shape and relatively uniform color. Vegetation, however, varies widely, from ground cover such as grass, to shrubs and trees of various species and heights. In the aerial photograph,

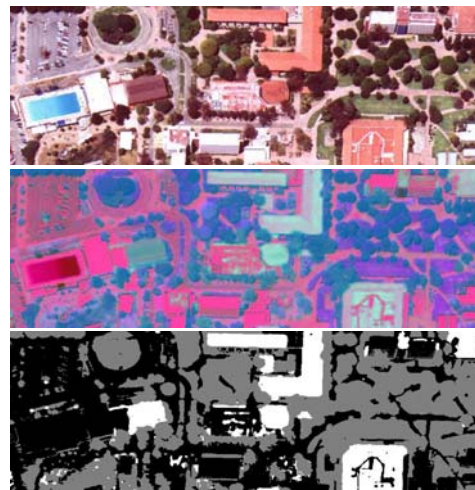


Figure 3. Training Image and Clustering Result

tree canopy occludes the ground below; thus, it is important to differentiate trees, which can overlap other features, from grass.

III. PART 2: AUTOMATIC SEGMENTATION

Manual segmentation and generalization of vegetation cover is time consuming. Thus having a method to extract vegetation automatically is ideal. Our first objective was to classify the land cover into two categories. One consists of all vegetation areas, and the other includes all remaining land cover types. Our second objective was to further separate vegetation into “high” vegetation, including trees and shrubs, and “low” vegetation composed mainly of grass. We speculate that for the purposes of this study, no further classification of the vegetation types is needed.

Most popular vegetation classification techniques use the Normalized Difference Vegetation Index (NDVI) as the defining feature in the classification process [3]. However, the NDVI requires near infrared data in addition to the traditional RGB color data. Although the NDVI provides an efficient means to accomplish our first objective, we decided to investigate the extent to which vegetation extraction can be successfully performed using images containing only RGB data, such as scanned aerial photographs. Furthermore, these images are readily available from the UCSB Map and Image Library.

A. Vegetation Extraction

An image covering approximately 15% of the UCSB campus was used as the training set for classification. Our initial task was to find features, on the pixel level, that could clearly distinguish vegetation from all other land cover types. Since the vegetation color in the RGB image varies over a wide range from dark brown to light green, we transformed the image into HSV, XYZ, CIE*Luv, and CIE*Lab colorspace to verify whether vegetation can be more homogeneously characterized. Next we used the

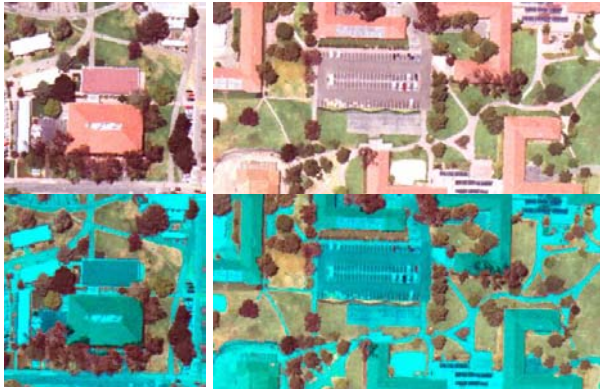


Figure 4. Nearest Neighbor Classification Result.

k-means clustering algorithm as a rough indicator to determine whether the features in the different colorspace would form natural groupings. Because k-means requires that we provide, as a guess, the number of clusters k , we

used $k = 2, 3, 4$, and 5 for each colorspace and visually assessed the results to determine the cluster number and colorspace that best separate vegetation from other cover types.

We found that the CIE*Luv and CIE*lab colorspace along with $k = 3$ resulted in the most accurate classification. Figure 3 shows the training image, its representation in CIE*Luv space and the k-means clustering result. It is apparent that the gray cluster represents vegetation. Subsequently using the final k-means centers, nearest neighbor classification was performed for the rest of the UCSB campus. Results for some areas are shown in Figure 4.

B. Tree Grass Classification

We found that although the CIE*Luv colorspace sufficiently provides enough feature separation for vegetation extraction, it does not clearly separate high and low vegetation. Thus, we must find new features that can meet our classification requirement. From visual inspection, the most distinguishing feature that separates high and low vegetation is the intensity, i.e. trees tend to be darker than grass. However, there are certain areas where this is not always true. A closer inspection of the saturation channel in HSV space

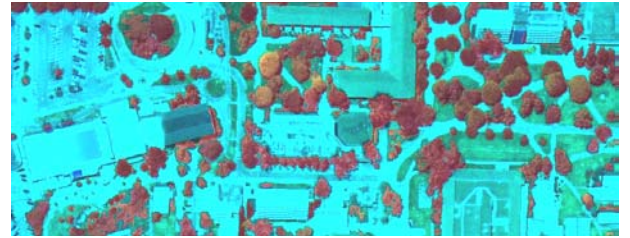


Figure 5. High Vegetation Classification Result.

reveals that most high vegetation areas have higher saturation values compared to those of lower vegetation areas. Accordingly, we used both the intensity and saturation values to classify high and low vegetation. The results of isolating just the trees are shown in Figure 5.

C. Further Improvements

There are some drawbacks to our algorithm. First, the vegetation extraction is not accurate for regions containing water, such as the UCSB lagoon. Thus, finding additional features that can more accurately characterize vegetation will improve our result. Secondly, results from the high/low vegetation classification show variations from region to region. Better preprocessing of the image or making the algorithm more adaptive may improve the high/low classification performance. Finally we note that all the features defined are on the pixel level. Obtaining larger regional feature characteristics such as texture or shape could yield more accurate results.

IV. CONCLUSION

The results of the pilot study indicate that generalization improved subject performance with a map on a small display used for a mobile task. Shape was shown to be an important characteristic for identifying features. Automating the generalization process through extraction and classification of features can assist in converting raw image data into a more useful map. Accuracy in the classification process, however, remains a challenge. The approach described here extracts particular feature types, such as vegetation, in order to then generalize that data to reduce unnecessary detail while preserving important characteristics. Further research can investigate mobile map use for different tasks and different landscape types, while additional work on the automation process can allow flexibility for generalizing maps for particular purposes.

REFERENCES

- [1] J. Casademont, E. Lopez-Aguilera, J. Paradells, A. Rojas, A. Calveras, F. Barcelo, and J. Cotrina. "Wireless Technology Applied to GIS." *Computers & Geosciences* vol. 30, 2004, pp 671-682.
- [2] Z.-R. Peng, M.-H. Tsou, *Internet GIS: Distributed Geographic Information Services for the Internet and Wireless Networks*. Hoboken, New Jersey, John Wiley & Sons, Inc., 2003.
- [3] M. M. Verstraete, B. Pinty, "Designing Optimal Spectral Indexes for Remote Sensing Applications.", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, no. 5, Sept. 1996, pp 1254-1265.
- [4] M. Wood, "Interacting With Maps", in *Human Factors in Geographical Information Systems*, D. Medykjy-Scott and H.M. Hearnshaw, Editors. Belhaven Press, London, 1993, 111-123.



Julie Dillemath is in the Geography Department, working with Keith Clarke and, for this research, Tobias Höllerer in the Computer Science Department. She earned her B.A. in Archaeological Studies from Yale University in 1998.



Nhat Vu received his B.S. in Electrical Engineering from Washington University in St. Louis in 2003. He is currently under the supervision of Dr. B. S. Manjunath in the department of Electrical and Computer Engineering.

Clustering Web Images Using Linked Text

Alexander Villacorta*

IGERT Fellow

Statistics and Applied Probability

Abstract—The focus of this work is to investigate the leverage gained by associating an image with its linked textual support. In this work, world wide web images are analyzed along with textual information found on pages linking to and from the pages containing these images. For any two images a total of ten different submeasures are calculated which reflect the different levels to which images share similarities. By utilizing the link structure of the web it is hoped that a more efficient and robust similarity measure may be established. In this framework, final clustering may be done by arbitrary schemes once the similarity measures are obtained.

Index Terms—Image Clustering, World Wide Web, Text Mining

I. INTRODUCTION

THE world wide web is an example of a vastly diverse collection of images which could greatly benefit from accurate image clustering. One of the most significant challenges to this task is defining similarity measures for images contained in multi-theme web pages in which several topics and images may be associated with a particular page. In this work we aim to overcome these challenges using a linked textual support approach. By considering multiple sources of information about a web image, it is hoped that a more accurate similarity measure can be established between any two images.

II. IMAGE REPRESENTATION AND TEXTUAL SUPPORT

The intuition behind this approach is that textual information found on web pages linking to and from a particular page offer insight on the semantic context of that page and thus the images contained therein. Also, by using multiple sources of textual content, this approach allows for different semantic meanings to be associated with a particular image while also correcting misleading keywords. Finally, since many images may not be associated with any type of textual information, we will also consider low level visual features of every image.

For each image, x_i , we extract four attributes which identify it in this framework. The first attribute, $x_{i,1}$, represents the nearby keywords which are associated with the page containing the image. This attribute contains the key words found in the surrounding text and also the keywords found in the meta tags of html pages such as the *alt* and *header* tag. The attribute $x_{i,2}$ is formed by considering all pages which link to the page containing the current image and combining the text in these pages to be considered as one document. The third attribute, $x_{i,3}$ is created in an

analogous manner. The last identifying attribute is the visual features of the image and in this study is taken to be the 144 dimension correlogram. In this setup, we do not require that attributes $x_{i,1}, x_{i,2}, x_{i,3}$ exist for every image.

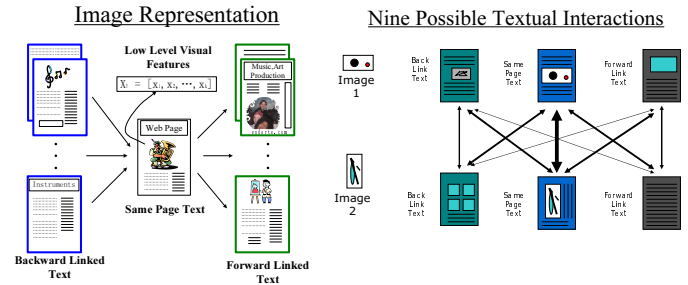


Fig. 1. Image Representation and Interactions

Next we investigate the 10 possible interactions between any two images. Figure 1 shows the representation of an image and the 9 possible textual interactions, the last interaction is between the visual features of the two images. For each of these interactions we define a subsimilarity measure which reflects the degree of similarity between each pair of attributes. For the textual features we use Latent Semantic Indexing to measure similarity and for the visual features we use Euclidean distance. As part of the general framework, we introduce a weight vector, $\beta = [\beta_1, \dots, \beta_{10}]$, which assigns an importance value to each subsimilarity measure. This weight vector allows the method to give more importance to nearby keywords and tags, while also incorporating linked text accordingly. Lastly, using these ten different subsimilarities we calculate the overall similarity measure.

III. CONCLUSION

In this work we have proposed a linked text framework for obtaining similarity measures where images are compared using 10 different subsimilarities. Current experiments include testing the sensitivity of the weight vector β and the effectiveness of the textual information of the linked pages. Also, various clustering methods will be used.



Alexander Villacorta is currently a graduate student in the department of Statistics and Applied Probability at UC Santa Barbara. He received his B.S. in Mathematics at the University of Michigan in 2000, and a M.S. degree in Applied Mathematics from the University of Colorado in 2002. His advisor is Dr. S.R. Jammalamadaka and research interests are multimedia datamining, classification and clustering.

*This work was done while at Microsoft Research Asia as a visiting student

An Integrated System of 3D Motion Tracker and Spatialized Sound Synthesizer

Mary Manli Li

IGERT Associate

Electrical and Computer Engineering

John Thompson

IGERT Fellow

Music

Michael J. Quinn

IGERT Fellow

Electrical and Computer Engineering

Abstract—In this paper, we present our work on building a system which tracks an object in real time using multiple cameras and outputs 3D motion data to a sound synthesis system. The synthesis system uses the data to produce audio which is played back in the space, creating an interaction between the subject and the system. Tracking is performed by first segmenting out moving objects in 2D images resulting in an object centroid from each camera. We then use the epipolar geometry method to determine the object's 3D real-world centroid location. We also compared the results from the epipolar method to the least squares estimation method. Results of each method were similar, with the epipolar results having a slightly lower error rate. 3D centroid data is then sent to the data management and visualization system (MAX/MSP/Jitter) that delivers the data to the audio synthesis server (SuperCollider 3). Tracked objects activate nodes that are distributed in the virtual space. Two categories of nodes, generative and transformative, are used to trigger and manipulate the synthesis and transformation algorithms. The parameters of these algorithms vary through time based on the history of tracked objects in the sensor space.

Index Terms—Sensor Networks, Surveillance, Motion Tracking, Musical Synthesis

I. INTRODUCTION

RECENT advances in sensor and communications technologies have made large networks of low-power dedicated sensors possible. The area of visual surveillance has benefited greatly from these advances, and real-time automatic surveillance systems are becoming more and more a reality. The goal of such a network is to track, as efficiently as possible and in real-time, objects of interest and report any useful information extracted by the sensors. This project aims to implement an interactive surveillance system which uses the 3D tracking data to control a music synthesis system. In the follow sections, we will discuss the overall system architecture, motion segmentation, camera calibration, 3D tracking and its performance and the integration of music synthesis.

II. SYSTEM ARCHITECTURE

The motion detection system consists of three Unibrain Fire-i firewire cameras connected to a single PC running Microsoft Windows XP Professional. The cameras deliver uncompressed 320 x 240 pixel RGB images. The motion detection software was written in C++ using Intel's OpenCV libraries. Camera calibration was also written using the OpenCV libraries. The 3D Tracker Filter library

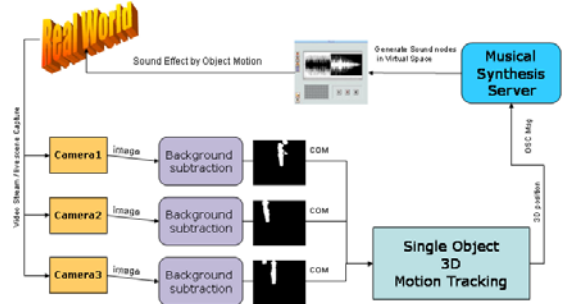


Fig. 1. Motion Tracking Sound Synthesizer System Diagram

included with OpenCV was modified for use in our application. Open Sound Control (OSC) was used for communication with the Sound Synthesizer. The sound synthesis/visualization system consists of an Apple Powerbook and an Apple G5. Fig. 1 shows the system architecture.

III. MOTION SEGMENTATION

In order to detect motion in a video stream, moving objects must be separated from any non-moving regions of the video. This is performed at the camera level. Two common methods of motion detection are Temporal Differencing and Background Subtraction.

A. Temporal Differencing

By subtracting consecutive video frames, we can determine what has changed in the time between them. This process is described by:

$$D(n) = I(n) - I(n - 1) \quad (1)$$

Fig. 2 shows an example of this process. The difference image is then thresholded to segment out areas of motion. One major drawback to this method is that if the object is a uniform color, often only the edges are detected, as seen in Fig. 2 (d). Thus this method is only used for a short time while a background model is being constructed. The output of this process (Fig. 2d) is dilated and then used as a mask for building the background model. Any areas of detected motion are not used to update the background model.



Fig. 2. Temporal Differencing Example. a) Previous Frame b) Current Frame c) Difference (normalized for display purposes) d) Thresholded Difference

B. Background Subtraction

A better and more robust method of motion detection utilizes a background model which is subtracted from the current frame. For a stationary camera, the background is defined by what the camera sees when no objects are present. When the background is subtracted from the current frame, any objects not present in the background model will be emphasized. In [1], several models are discussed and compared. For simplicity, we model each background pixel as a Gaussian random variable. The threshold value was chosen as 2 standard deviations from the mean.

The background model is initialized over the first 100 frames of video. The areas of motion are found via temporal differencing and are roughly masked in the averaging process. Once the 100th frame is reached, it is assumed that a sufficient background model has been created for background subtraction. For the next 100 frames, the background model is updated using background subtraction motion data to mask areas of motion. Once the system hits frame 200, the frequency of background update is reduced.

An example of background subtraction is shown in Fig. 3.

IV. 2D TRACKING

The system currently assumes that a single person is being tracked and so the person's location is determined by calculating the centroid of the entire image. By tracking a person we allow him/her to control the system output. Assuming little or no noise, the frame centroid is a good estimate of the person centroid. The person centroids from all cameras are used to calculate the position in 3D space.

In the future, for the purpose of tracking multiple moving objects, more complex methods will be required such as those that use shape or histogram information.

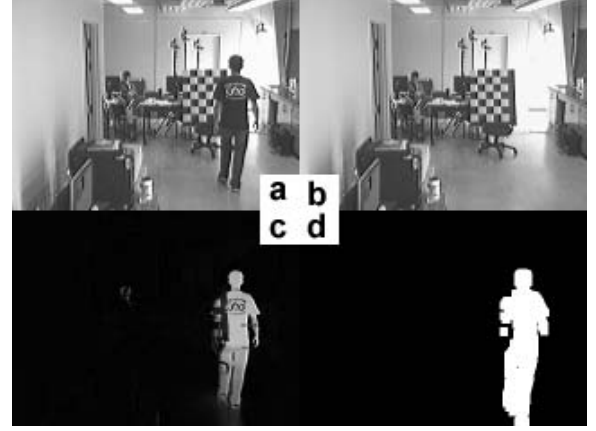


Fig. 3. Background Subtraction Example. a) Current Frame b) Current Background Model c) Difference (normalized for display purposes) d) Thresholded Difference(post morphology)

V. CAMERA CALIBRATION

Camera Calibration is the preparation for any 2D to 3D or 3D to 2D transformation. Calibration obtains the intrinsic and extrinsic parameters. The intrinsic parameters describe a camera's properties such as focal length, principal point and distortions. The intrinsic parameters are independent of the position of the cameras in the physical environment. The extrinsic parameters we use here describe the transformation between the camera coordinate system and the world coordinate system. For a group of cameras, the relative extrinsic parameters can also be obtained between the cameras. The combination of the intrinsic and the extrinsic parameters of a camera describe a relationship/transformation between an image point and its correspondence in real world. The following two equations describe this relationship:

$$X_{camera} = M_{ext}X_{world} \quad (2)$$

$$X_{image} = M_{int}X_{camera} \quad (3)$$

where X_{world} is a point in the world coordinate system, X_{camera} is a point in the camera coordinate system, X_{image} is a point in the image reference frame. The M_{ext} is a linear mapping between X_{world} and X_{camera} , M_{int} is a linear mapping between X_{camera} and X_{image} . In addition to individual calibration, we also calibrated between the cameras using Matlab Calibration Toolbox, these results will be used in the future to obtain a more accurate camera matrix for further research purposes. For the time being, only individual calibration results are used. In the following section, we will discuss the methods used in 3D tracking, and the experiment results.

TABLE I
ESTIMATION ERROR (CM)

	EPP-avg	EPP-std	LS-avg	LS-std
X	3.6	7.0	5.3	8.7
Z	5.9	33.5	5.7	30.0

VI. 3D TRACKING

As mentioned in the 2D Tracking section, we used the Center of Mass (COM) as the position of the single moving person. In our experiment, camera calibration sets the world coordinate system with +Z facing the center camera, and the X-Z plane parallel to the ground, and +Y direction pointing up. Given the results of COM in images from each of the cameras, we used two methods to locate real world moving object position. The first method is based on epipolar geometry (EPP) as described in [3] and the second method is based on the least square estimation (LS) as described in [4]. To collect the ground truth data, we had a person walked onto marked points and used their real world positions as ground truth. The marked points are located in a rectangular area of 1 by 2 meters. Figure 4 shows corresponding result. Note that the marked points correspond to frames (130-150,180-200,215-235,265-285). In the tracking process, the person changes his/her posture regularly. Since the height of the person will be stable within a short period, currently we only verify the result of X-Z coordinate. As shown in Table I, EPP's estimation error has an average of 3.6cm with standard deviation of 7.0cm at X direction; an average of 5.9cm with standard deviation of 33.5cm at Z direction. LS's error has an average of 5.3cm with a standard deviation of 8.7cm at X direction; an average of 5.7cm with standard deviation of 30cm at Z direction. In general, these errors are significantly small compared to the size of a person and it turns out to meet the requirement for sound synthesis as discussed in later sections. More results can be found at <http://www.engineering.ucsb.edu/~mmlee/research/MotionTracking.html>.

VII. SYSTEM PERFORMANCE

We ran several real-time experiments for varying lengths of time. The longest was about 70 minutes and ran a total of 8780 frames (from each camera). With OSC communication with the Music Synthesizer enabled, the system achieves a stable processing rate 2.09 frames per second(fps). The system also achieves about 7fps offline (reading from AVI instead of cameras) processing rate. In real time experiments, with all three cameras running on a single PC, the image capture process alone results on the average 3.7fps. Therefore much improvement on frame rate can be done by using distributed and synchronized camera network system. Table II shows a comparison between

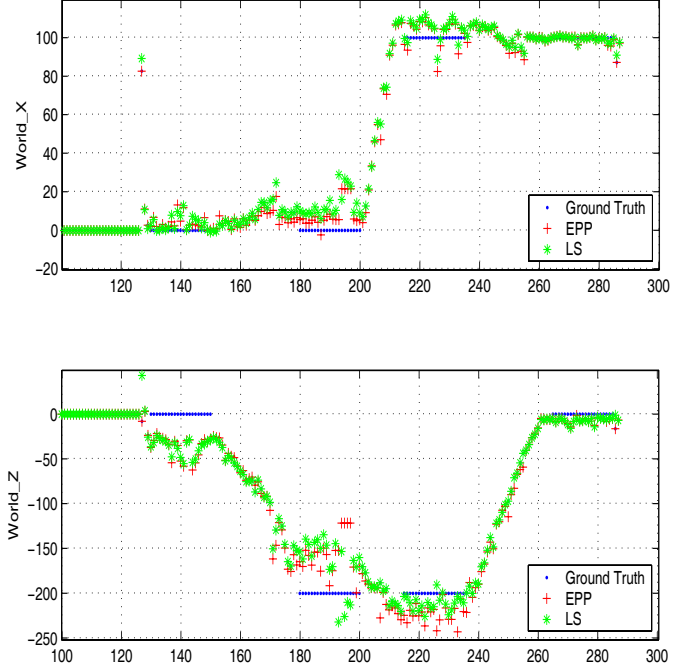


Fig. 4. Marked-point experiment results

TABLE II
SYSTEM PERFORMANCE (FPS)

Frames	Real Time	AVI (1)	AVI (2)
100	2.17	7.14	6.25
200	2.00	5.71	5.41
300	2.00	6.00	5.77
500	2.08	N/A	N/A
900	2.14	N/A	N/A
1400	2.12	N/A	N/A
1800	2.28	N/A	N/A

real time processing and offline processing rate based on experiments using directly captured real time images and video (AVI) sequence images.

VIII. MUSICAL MAPPING STRATEGIES: SPATIALLY DESIGNED SONIC POTENTIALS

A. Sonic Nodes in the Virtual Space

Nodes with various sonic functions are distributed in the virtual space. This forms a system of nodes comprised of Generative Nodes and Transformation Nodes. Each node has a surrounding spherical field known as the activation space. The activation space adjusts its activation level based on the distance of the moving object(user) from the center of the activation space's spherical field.

The use of nodes with various sonic functions allows for

the creation of a spatial layout of sonic potentials. The nodes adjust over time based on the activity of the space. This creates variation within a static spatial layout. The nodal system creates an open ended non-goal directed formal design that changes its structures based on the behavior of users within the system.

B. Generative Nodes

B.1 Chord Nodes

As the user moves in the space, she discovers and activates chord nodes. The chord nodes are arranged in the space such that the location of the user is likely to activate more than one chord node at a time.

The arrangement of localized chord nodes creates a spatial method of pitch organization. Because the user will activate these chord nodes in an unpredictable manner, the pitch sets are constructed and transformed to function within an ateleological framework.

Each chord nodes contains a four-note pitch set. Seventeen pitch sets are initialized at the onset. Fourteen of the seventeen sets are unique in their normal order. Although the pitch sets are diverse, they are closely knitted in their makeup. This lends a unified quality to the pitched verticalities of the sonic space. The seventeen pitch sets are distributed into thirty-four nodes so that each pitch set occurs twice. Each chord node, when activated, creates simple sine wave synthesizer for each pitch of the chord node. The instrument utilizes small random deviation of the parameters to create subtle variations.

As the user moves throughout the space, the sonic material subtly shifts, melding the space into a cohesive flow. As the tracked object leaves a history of its path in the space, the pitch sets transform in response. The sine wave instrument represents one extreme of the timbre continuum between simplicity and complexity. The simple timbre of the sine wave instrument creates a highly permeable sonic layer that can be easily incorporated with more complex timbres.

B.2 Chain of Synths Activated with Impulse

A second generative node utilizes enveloped impulses with frequencies straddling the border between audio and infra-audio rates (15 - 20Hz). These impulses are sent through a network of unit generators (see Fig. 6).

The sharp tuned impulses create a texture that easily penetrates the malleable timbres of the chord nodes. The frequency and duration of the impulses allow for effective spatialization results.

B.3 Sample Playback

When activated, the sample playback nodes play back an enveloped sampled sound. Each activated node selects a sound sample from a list, sets the pointer to a start position in the sample, sets the amplitude of playback, and

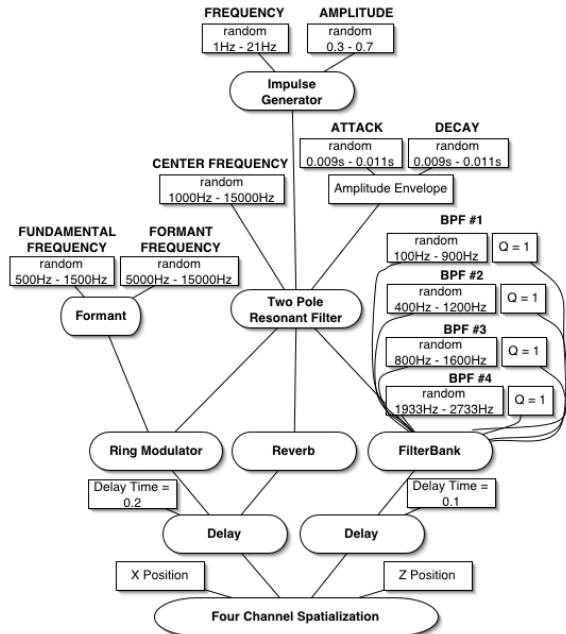


Fig. 5. Flow chart of the impulse based synthesis instrument

determines a play back rate. The sample playback material adds rich material to the space creating surprising textures in combination with other nodes.

C. Transformative Nodes

Certain nodes take audio input from generative nodes and transform the sonic material. These transformative nodes have no effect if they are activated in isolation, but if activated in combination with generative nodes, the sonic texture of the generative node's audio output is altered.

C.1 Convolution

In the audio world, the convolution of signals is a powerful tool for sonic transformation. Convolution marries two audio signals, creating a cross-synthesis of the sounds. In this implementation, previously recorded audio segments are convolved with the output of the chord nodes. The audio files, many of which contain human voices, create the effect of someone speaking the sonic space.

C.2 Pitch/Time Warping

The pitch-time warping node uses a granular method of shifting the pitch and time characteristics of a sound. The sound is broken into windowed grains. These grains can be played back at varying speeds while simultaneously duplicating or reproducing the grains to independently control the sound's pitch and time qualities. Pitch and time parameters are modified according to the position of the tracked object that triggered the node. In this implementation, the audio output of the chord nodes is fed to an eight second buffer when the pitch-time warping node is

activated. The result is a timbre transformation of the sonic space.

D. Spatialization

D.1 Localization and Distance Simulation

The system outputs quadraphonic audio distributed to speakers surrounding the sensor space. The position of the sounds within the sensor space is determined by the position of the tracked object. Distance is simulated through direct sound to reverberant sound mixture. This ratio is dictated by the following formulas:

$$DirectSoundAmplitude = 1/zPosition.abs \quad (4)$$

$$ReverbSoundAmplitude = 1/zPosition.abs.sqrt \quad (5)$$

The spatialization algorithms implemented lend a sense of additional movement and depth to the sonic material.

E. Distribution of the Nodes

E.1 Number Of Nodes

The motion tracking system measures the center of mass of the tracked object. This reduces the probability of very high or very low values along the Y-axis. A Gaussian distribution of nodes along the Y-axis compensates for the center of mass limitations on this axis. The X-axis and Z-axis each have a uniform distribution of nodes along their axes.

The nodes are distributed as described in Table III.

TABLE III
NODE DISTRIBUTION

Type of Node	Number of Nodes
Chord Nodes	34
Convolve Nodes	16
Pitch/Time Warp Nodes	64
Impulse Nodes	360
Sample Playback Nodes	50

F. Variations through Time

The system adapts its state and musical structures based on the activity of the space. The activity state of the system is evaluated every two seconds. Based on the evaluation, the system is determined to have a high activity state, a moderate activity state, a low activity state, or a sparse activity state. Each state determines what synthesis algorithms will be implemented in that state as well as whether the nodes should adjust their position in relation to the tracked object. Additionally, a tally is kept of activated nodes types. When a threshold value is reached

by any of the node types, the impulse frequency, the resonant frequency, and many other parameters of various algorithms are varied.

F.1 Pitch set transformation

When a threshold value is reached, each pitch set is multiplied by a value near to 1.0. Since the multiplier is a floating-point number, the pitches tend toward microtonal values. The exact value of the multiplier is determined by the time taken to reach the threshold value of any of the node types.

G. Visualization

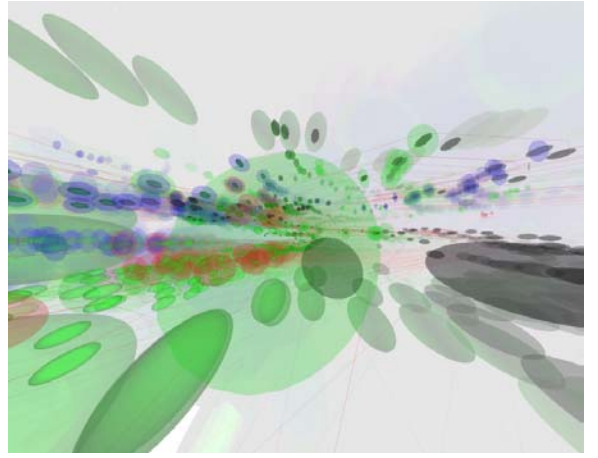


Fig. 6. A Captured Frame from the Visualization

G.1 Visualizing the Nodes

Initially, the nodes are displayed in the virtual space. As the tracked object activates nodes in the space, the positions where the nodes were activated are visualized as spheres and circles of various colors. The camera view of the virtual space flies just behind the tracked object. The path of the tracked object is shown as a red line that remains in the space along with the activated nodes.

IX. CONCLUSION

We implemented a basic motion capture and sound synthesis system. This system has paved the way for surveillance and camera sensor network research at UCSB. It has also provided a starting point for using physical position and features as input to multimedia systems.

X. FUTURE WORK

In the near future, we hope to increase the computation speed (frame rate) by implementing the system in a distributed network. Kalman filter can be implemented to improve the 3D estimation result. We are also moving toward implementing a more complex background maintenance method to account for environment changes. We

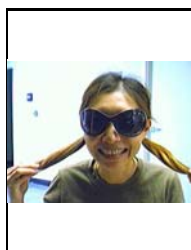
plan to modify the system to track multiple objects based on size, shape, and histogram information. This will require integration of the 2D and 3D tracking systems. Finally, we hope to eventually implement this system wirelessly. We will be looking into possible solutions in the near future.

ACKNOWLEDGMENT

The authors would like to thank Professors B.S. Manjunath, George Legrady, and JoAnn Kuchera-Morin for guidance and support throughout this project. They would also like to thank Dr. Xinding Sun for technical guidance. Thanks go to Lance Putnam and Satoshi Morita for providing sample C++ code for the OSC implementation. Finally, they would like to thank the NSF and UCSB's Interactive Digital Multimedia group for the opportunity to work on this project.

REFERENCES

- [1] K. Toyama, J. Krumm, B. Brumitt and B. Meyers. Wallflower: Principles and Practice of Background Maintenance. *Seventh International Conference on Computer Vision*, pages 255-261, September 1999.
- [2] W. Hu, T. Tan, L. Wang, and S. Maybank. A Survey on Visual Surveillance of Object Motion and Behaviors. *Systems, Man and Cybernetics, Part C, IEEE Transactions on*, Volume: 34, Issue: 3, Aug. 2004. Pages:334352.
- [3] Hartley and Zisserman. Multiple View Geometry, second ed. Cambridge University Press. 2003.
- [4] Olivier Faugeras, Three-Dimensional Computer Vision, MIT Press, 1993



Mary Manli Li received her B.S. degree with high honor in Computer Science with minor in Applied Mathematics from San Francisco State University in 2003. She participated summer researches at University of Michigan and UC Berkeley during her undergraduate studies. Her primary research interests are computer vision and image analysis. It is her second year of the MS/PHD program at the Department of ECE at UC Santa Barbara.



John Thomspson (b. 1974) is a composer and media artist. He studied music composition at the Hochschule für Musik, Carl Maria Von Weber in Dresden, Germany with Jörg Herchet, and obtained a Master of Arts degree in music composition at MTSU in Nashville, TN. He is keenly interested in the synaesthetic properties of various media and works to algorithmically explore the engines of cross-disciplinary creativity. Presently he is a graduate student at the University of California at Santa Barbara where he has studied music and media arts with JoAnn Kuchera-Morin, Curtis Roads, Marcos Novak, and Stephen Travis Pope.



Michael J. Quinn is currently a Ph.D. student in the ECE department of UC Santa Barbara. He received the B.S. and M.S. degrees in Electrical Engineering from the University of Missouri - Rolla in 1997 and 1999, respectively. His research interests include motion tracking, camera sensor networks, and super resolution. His academic advisor is Professor B.S. Manjunath.

Visually Optimized Quantization Parameters in AVC/H.264

Jing Hu

IGERT Associate

Electrical and Computer Engineering

Abstract—In this project we generated quantization parameter matrices which optimize perceptual visual quality in the Advanced Video Coding(AVC) standard/H.264. The luminance, texture and temporal masking models of human vision systems(HVS) are explored to conduct subjective experiments. The intra-coded and inter-coded macroblocks(MB) are studied separately to best correlate with the error propagation pattern. The optimal quantization parameter matrices are embedded and indexed in both the encoder and decoder and thus do not require extra bits to transmit.

Index Terms—perceptual quality, masking, quantization parameters, video coding, AVC/H.264

I. INTRODUCTION

OVER the last one and a half decades, digital video compression technologies have fundamentally changed the way we create, communicate and consume visual information. They have not only transformed existing applications and services like the distribution of entertainment video to the home but also spawned brand new industries and services like video-conferencing, direct-to-home satellite distribution, digital video recording, video-on-demand services, high-definition TV, video on mobile devices, streaming video, etc. Since the early 1990s, international video coding standards such as H.261, MPEG-1, MPEG-2/H.262, H.263, and MPEG-4(Part 2) have been powerful engines behind the commercial success of digital video compression [3]. The recently finalized Advanced Video Coding(AVC) standard, also named as ITU-T Recommendation H.264 and MPEG-4 Part 10, offers a coding efficiency improvement by a factor of two over previous standards (Figure 1) and its network abstraction layer(NAL) transports the coded video data over networks in a more “network-friendly” way [7]. Because of these two features, the AVC/H.264 standard is very likely to emerge as the method of choice for the next generation video networks.

With the fast evolution of video compression technologies, perceptual video quality metrics are becoming very important in evaluating, optimizing and comparing video compression schemes. They are also critical in monitoring the quality of video sequences sent over either the traditional television or the more recent IP networks to maintain a certain video quality level. Subjective video perceptual quality measurements have been conducted under standardized International Telecommunication Union(ITU) Recommendations: ITU-T P.910 [1],

ITU-R BT.500 [2], etc.. Subjective video perceptual quality measurements involve a huge number of experiments on human subjects so they are expensive, time-consuming and more critically, theoretically impossible to be implemented in real-time applications. The most commonly used objective video quality metric is the mean-squared error(MSE) or equivalently the peak signal-to-noise ratio(PSNR) of the distorted videos. This metric does not explore the perceptual effects of the distortion so is usually criticized for not correlating well with human perception. It fails when one tries to compare different types of artifacts and it does not take into account the difference caused by viewing conditions. To achieve better performance the sophisticated objective video quality metrics have been set up in the past few years based on the lower order processing of human vision systems(HVS) [4], [6]. Both general and coder-specific objective video quality metrics can be divided into three different categories — full reference, reduced reference and no reference, in terms of how much information about the reference video is available to the metric. However the subjective tests conducted by Video Quality Experts Group(VQEG) in the year 2000 concluded that none of these objective metrics is close to be able to replace subjective testing [5].

To improve the performance of the objective video quality metrics, a well-recognized approach is to design the metrics tightly based on the video compression systems. In this project we focus on AVC/H.264 because of its novel performance and its lack of perceptual quality measurements. We start from studying the intra-/inter-prediction and integer transform in AVC/H.264 codecs and error propagation through the video transmission system. We then conduct two different subjective experiments to explore the effects of luminance, texture and temporal masking models of HVS. We will conclude with quantization parameter matrices which optimize the perceptual quality of the compressed videos. Our work improves the performance of AVC/H.264. It serves as the first step of deriving full reference objective video quality metrics for AVC/H.264 and the understanding of error propagation patterns will find great application in designing reduced and no reference video quality metrics.

II. ALGORITHMS

IN the encoder of AVC/H.264, the incoming video I-frames are divided into arrays of 16×16 pixel mac-



(a) Compressed by AVC/H.264 at 199kbps. SNR(dB): Y-38.17, U-40.58 V-41.89



(b) Compressed by AVC/H.264 at 155kbps. SNR(dB): Y-37.10, U-39.72 V-41.02



(c) Compressed by MPEG-4 at 155kbps.

Fig. 1. news.cif compressed by AVC/H.264 and MPEG-4 at different bit rates.

roblocks(MB). Each MB (or a subset of it) then undergoes either intra- or inter-prediction. The mode index for intra-prediction, motion vector and quantized residual data are entropy coded and transmitted to the decoder through the channel. Distortion in the encoder is mainly caused in the uniform quantizer while intra-/inter-prediction significantly changes the effect of the distortion. We will treat intra- and inter-prediction separately in the following two subsections. In both cases 4×4 block sizes are used, however the algorithms developed easily generalize to other processing block sizes.

A. Quantization parameter matrices for the intra-coded MBs

Intra-prediction is a new feature embedded in AVC/H.264 which contributes to 6–9% in bit saving by removing the spatial redundancy in neighboring MBs. Each 4×4 luminance block is first predicted from surrounding reconstructed pixel values using one of the 9 modes illustrated in Figure 2 which yields a residue block \mathbf{X} with the smallest summation of absolute errors(SAE).

The exact integer transform (1) is used to avoid the mismatch between encoder and decoder in the discrete cosine transform(DCT)-based coding standards.

$$\mathbf{Y} = \left\{ \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & 1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix} \mathbf{X} \begin{bmatrix} 1 & 2 & 1 & 1 \\ 1 & 1 & -1 & -2 \\ 1 & -1 & -1 & 2 \\ 1 & -2 & 1 & -1 \end{bmatrix} \right\} \otimes \mathbf{E}_f. \quad (1)$$

The post-scaling factors contained in \mathbf{E}_f are combined with the uniform quantization following the forward transform. A total of 52 values of quantization step sizes(Qstep) are supported by AVC/H.264 and they are indexed by quantization parameters(QPs) as shown in Table I. Note that these values are arranged so that an increase of 1 in QP means an increase of Qstep by approximately 12% and roughly a reduction of bit rate by approximately 12%.

Texture masking of HVS refers to the reduction in visibility of one image component caused by the presence of

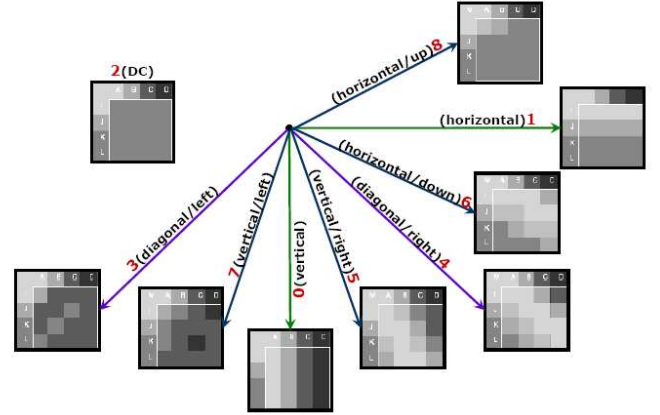


Fig. 2. The intra prediction modes in AVC/H.264.

another image component with similar spatial location and frequency content [4]. We generate first the 16 quantization distortion patterns in Figure 3 by extracting the 16×16 reconstructed matrix caused by quantizing only one frequency coefficient in \mathbf{Y} . By comparing them to the 9 intra-prediction blocks in Figure 2, we predict that the quantization distortion is texture-masked by the image content to different levels depending on the similarities between the distortion pattern and intra-prediction mode co-existing in one 4×4 block.

We therefore designed a subjective experiment to measure the QPs at threshold of each frequency coefficient for each intra-prediction mode. This experiment involves three subjects at the age of 20, 25 and 31 respectively and three images representing applications in news, sports and sceneries. Only the 4×4 blocks which are intra-coded using one of the 9 intra-prediction modes in the image are chosen at one time to add only one out of the 16 distortion patterns. The magnitude of the distortion is decided by QPs which keep increasing until the distortion becomes

QP	0	1	2	3	4	5	6	7	8	9	10	11	12	...
QStep	0.625	0.6875	0.8125	0.875	1	1.125	1.25	1.375	1.625	1.75	2	2.25	2.5	...
QP	...	18	...	24	...	30	...	36	...	42	...	48	...	51
QStep	...	5	...	10	...	20	...	40	...	80	...	160	...	224

TABLE I
QUATIZATION STEP SIZES IN AVC/H.264 CODEC.

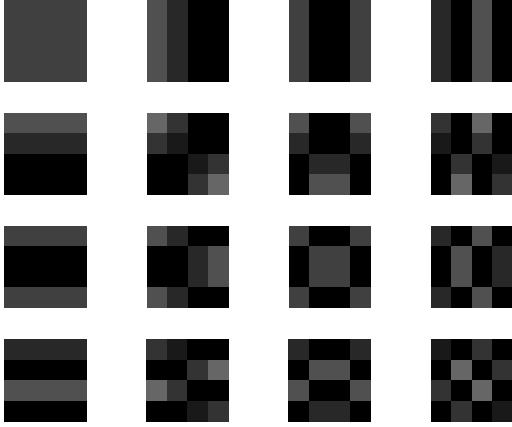


Fig. 3. Normalized quantization error patterns in AVC/H.264

perceivable to the subject and the QPs at threshold are recorded. One of the many distorted images presented in this experiment is shown in Figure 4. The three images are all of Common Intermediate Format(CIF) format shown on a 17 inch monitor with resolution 1152×864 pixels and average background luminance $170cd/mm^2$. The viewing distance is about 6 times the height of the images.

B. Quantization parameter matrices for inter-coded MBs

Inter-frame motion estimation and compensation has been the main scheme in almost all of the video coding standards. It removes the temporal redundancy in the video sequences. In the last subsection we described how we employed texture masking of HVS to generate the quantization parameter matrices for intra-coded MBs. In this subsection we will also make use of the three other masking types of HVS — luminance masking, baseline contrast masking and temporal masking. Texture masking plays a similar role for inter-coded MBs because the intra-prediction mode in the referenced intra-coded MB sets up the basic pattern in the following inter-predicted MBs. Luminance masking and baseline contrast masking, both being spatial characteristics of HVS and both assuming uniform background, deals with the perception of a uniform square and sinusoid waves respectively. They together indicate that the quantization parameter matrices

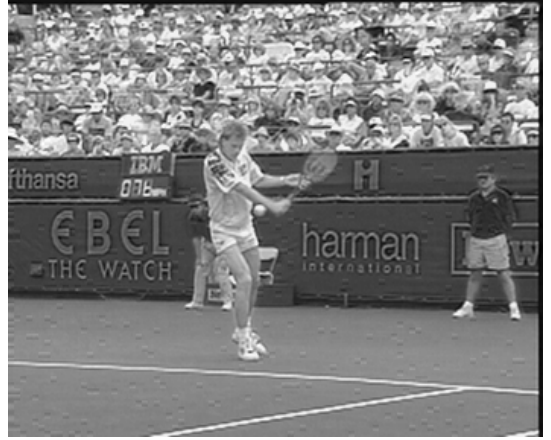


Fig. 4. stephan.cif in subjective experiment 1— intra-prediction mode tested=3, error pattern added I=2 J=1, QP=43

will depend on the local luminance of each 4×4 block. And also the distortion will have different perceptual effects when presented in a video sequence than presented in images. In the second experiment, we combine the luminance/baseline contrast and temporal masking concepts to generate quantization parameter matrices for inter-coded MBs and also to derive a practical rule to adjust the quantization parameter matrices by the local luminance.

We start from studying the error propagation caused by inter coding. The formulation is demonstrated in the following for two pixels, $x(m_1, n_1, t_1)$ in the reference frame and the corresponding $x(m_2, n_2, t_2)$ in the current predicted frame, where m_i , n_i and t_i are the row, column and frame indices. x , \hat{x} and \tilde{x} denote the pixel value in the original video, the reconstructed video in the encoder and the reconstructed video in the decoder respectively. In the encoder, motion estimation and compensation of $x(m_2, n_2, t_2)$ both refer to $\hat{x}(m_1, n_1, t_1)$ so the error does not propagate and new error generates because of quantization of the residual data. In the decoder, supposing the motion vector and quantized residual data are received without any loss through the channel, then

$$\tilde{x}(m_2, n_2, t_2) = \tilde{x}(m_1, n_1, t_1) + (\hat{x}(m_2, n_2, t_2) - \hat{x}(m_1, n_1, t_1)). \quad (2)$$

We define $\tilde{e} = \tilde{x} - \hat{x}$ as the error caused by transmission, then equation (2) becomes

$$\tilde{e}(m_2, n_2, t_2) = \tilde{e}(m_1, n_1, t_1), \quad (3)$$

which means that no new error generates but the error in the reference block propagates. The error propagation pattern can be described in the frequency domain as a three dimensional filter

$$H(u, v, f) = S(u, v) \cdot e^{-j2\pi(u(m_2 - m_1) + v(n_2 - n_1) + \frac{f}{f_s}(t_2 - t_1))}, \quad (4)$$

where f_s denotes the frame rate in *frames/second*. In AVC/H.264, the motion estimation resolution ranges over full-, half- and quarter-pixel, so the spatial filter $S(u, v)$ has different forms as shown in Figure 5.

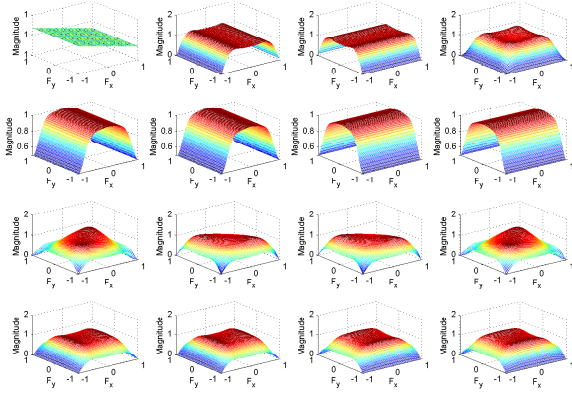


Fig. 5. Spatial filters of motion estimation in AVC/H.264

In the second subjective experiment, we simulate the error propagation using the model derived above. The reference frame index and spatial offsets are randomly chosen. Figure 6 shows a frame in one of the many video sequences presented in this experiment. The 9 intra-prediction modes and 6 luminance levels for each of them are included in one video sequence to reduce the number of video sequences. Four different frame rates — 5, 15, 20, 30 fps were employed to explore the temporal masking effect. Subjects and the viewing conditions are kept the same as in the subjective experiment 1.

III. CONCLUSION

THE QPs collected in both subjective experiments showed good correlation among different subjects and image. The averages of the QP samples were taken to form the recommended QP matrices for different frame rates and different luminance levels. All the QP matrices are embedded in both the encoder and decoder. The intra-prediction modes and frame rates are already transmitted so the only parameter required to be sent from the encoder to the decoder so that they both use the same QP matrix is the luminance level — an integer ranging from 1 to 6. This

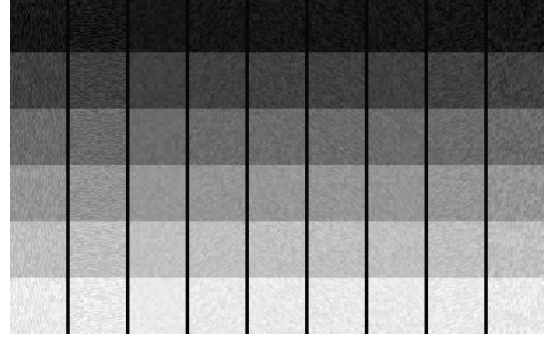


Fig. 6. One frame in subjective experiment 2— error pattern added: I=1 J=1, QP=30

employs fewer bits than sending the QP — an integer ranging from 0 to 52 in the current AVC/H.264 baseline/main applications which use same QP for all 16 frequency coefficients. These perceptually optimized QP matrices improve the performance of AVC/H.264. More subjective experiments are expected to be done in the future to polish the QP matrices and to establish their invariance for different viewing conditions.

REFERENCES

- [1] ITU-T Recommendation P.910, “Subjective video quality assessment methods for multimedia applications,” 1999.
- [2] ITU-R Recommendation BT.500, “Methodology for the subjective assessment of the quality of television pictures,” 2002.
- [3] A. K. Luthra, G. J. Sullivan and T. Wiegand, “Introduction to the Special Issue on the H.264/AVC Video Coding Standard,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, 2003.
- [4] T. N. Pappas and R. J. Safranek, “Perceptual Criteria for Image Quality Evaluation,” in *Handbook of Image & Video Processing* (A. Bivok eds.), Academic Press, 2000.
- [5] Video Quality Experts Group, “The Quest for Objective Methods: Phase 1, Final Report,” <http://www.its.bldrdoc.gov/vqeg/>, 2000.
- [6] Z. Wang, H. R. Sheikh and A. C. Bovik, “Objective video quality assessment,” *Chapter 41 in The Handbook of Video Databases: Design and Applications* (B. Furht and O. Marqure, eds.), CRC Press, pp. 1041-1078, 2003.
- [7] T. Wiegand, G. J. Sullivan, G. Bjøntegaard and Ajay Luthra, “Overview of the H.264/AVC Video Coding Standard,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 560-576, 2003.



Jing Hu graduated from Rice University with an M.S. in 2003 and has been a PhD candidate at UCSB since then, both in Electrical and Computer Engineering. Her research interests are video transmission over wireless networks and perceptual video quality measurement and optimization. Her advisor is Professor Jerry D. Gibson.

Placement of Annotations in Video Feeds

Vineet Thanedar

IGERT Associate

Department of Computer Science

Abstract—The augmentation of video footage with graphical information is rapidly gaining popularity and acceptance in a wide variety of applications. Visually overlaid annotations are very common in telecasts (e.g., statistics in sports broadcasts, banner advertisements), where they are strategically placed by hand, or incorporated into the video stream using elaborate camera tracking and scene modelling techniques. Such information is not commonly available for raw or unprepared videos. We look at the problem of automatically placing annotations in unprepared videos using computer vision and image analysis techniques. Specifically, we explore the placement of annotations in the most appropriate locations in arbitrary videos.

Index Terms—Annotation placement, perceptual descriptors, video pre-processing

I. INTRODUCTION

THERE has been a phenomenal growth in the amount of digital video material available over the past few years. Much of this video information is generally in its natural or *unprepared* form i.e., there is no augmentation of the visual component of video data and no metadata indicating the semantics of the involved scenes. Increasingly, in real-time video broadcasts, we are witnessing the live video feed being augmented with graphical overlays. Advertisements, game scores, and logos are a few of the several graphical items. Interactive TV has been gaining momentum over the past few years. This involves customizing the content delivered based on viewer profiles. The video data delivered to the consumer can be augmented with multiple pieces of information that are retrieved upon request.

We address the problem of automating the process of placing annotations in videos. Optimal annotation placement should avoid occlusion of moving objects and informative scene elements, and should aim to not distract the viewer. There is a newly emerging industry of commercial video processing services [1] [2] that enables insertion of 2D and 3D annotations into videos, such as the virtual first-down line in football broadcasts, or virtual advertising. Such material is placed using intricate tracking equipment, scene modelling, and/or considerable human intervention. We are interested in finding good solutions for the placement problem in videos, for which no background or supplementary information about the camera and 3D scene geometry is available. In this scenario, placement needs to be guided by low-level information available from the video stream alone.

II. MAIN IDEA

We preprocess the video to perform a low-level analysis of the stream. We identify elementary properties pertinent to our understanding of visual information and apply quantitative measures to analyze the video stream for these properties. We identify uniformity, motion, and clutter as basic properties of video regions. We refer to these as *Elementary Perceptual Descriptors* since each of these characteristic properties is distinctly perceiv-



Fig. 1. A player annotation in a soccer game. The annotation is placed automatically in an optimal position based on analysis of regions around the player.

able by the human visual system. Uniformity refers to homogeneity in intensity or color and motion refers to the movement of scene elements. We identify clutter as a property of regions that depict a high degree of detail such that the overall appearance of the region is ‘chaotic’ and hence it is difficult to make sense of information in a cluttered region when it is observed in isolation, i.e., without reference to what is around. For example, in Figure 1, the background crowd is a cluttered region. We derive a measure of the degree of importance of a region as a combination of these properties. In our system we identify uninteresting or relatively less important regions in the video, which are potential locations for placing annotations. We utilize this knowledge to direct the placement of scene annotations.

Our final annotation placement formulation is a combination of the values computed for each of the elementary perceptual descriptors. Let M_u , M_m , and M_c [3] be the values for the uniformity, motion, and clutter measures respectively. Our formulation for M_u computes the maximum mean intensity difference between adjacent pixels in a region as a measure of the region’s uniformity. Motion, M_m , is based on the percentage of pixels in a region exhibiting ‘apparent’ motion based on frame-to-frame intensity differences. We apply discrete wavelet transforms [4] to compute the value for clutter, M_c , in a region. We identify clutter as an important element driving annotation placement. Cluttered regions carry very little interesting or comprehensible information. The value for the clutter measure is computed for each candidate region in the scene. The different regions that are evaluated, are selected based on the annotation size and placement preferences. If the clutter value is greater than a user-defined threshold, τ (video specific), we assign that value as the final measure. If clutter is below that threshold, then we compute a measure for the region that takes into account the uniformity and motion in the region. Perceptually too, regions that are more uniform and exhibit lesser motion are more appropriate for annotation placement. Thus, the final measure value

for a candidate region is computed as follows:

$$\text{If } M_c > \tau \text{ then } M = M_c \text{ else } M = M_{um}$$

where,

$$M_{um} = \frac{k}{\epsilon + M_u + M_m + M_u \times M_m^2}$$

The regions are ranked in increasing order of the value of M , higher values indicating regions with a greater potential for placement. M_{um} is the combined metric that takes into account the uniformity and motion values for a region. Regions with higher values for the expression, M_{um} are preferred over others for annotation placement. Our uniformity and motion formulations [3] are such that lower values of M_u and M_m denote more uniformity and lesser motion respectively. A region with greater degree of uniformity will have larger values for the above expression and correspondingly a higher rank in the placement order. A higher value of the motion measure M_m results in a lesser value for the overall measure. Note that motion has a greater effect on the region selected (M_m^2), since non-occlusion of other moving objects was deemed a more important factor. Hence in the case where a region overlaps with moving objects, the effect of uniformity of the region decreases. The weight k is a positional preference weight that can be used to prioritize specific region choices (such as the top of a video frame) or reduce or magnify the influence of the combined homogeneity/motion value when combining this metric with others (such as with clutter). Figures 1 and 2 show examples of our automated annotation placement.

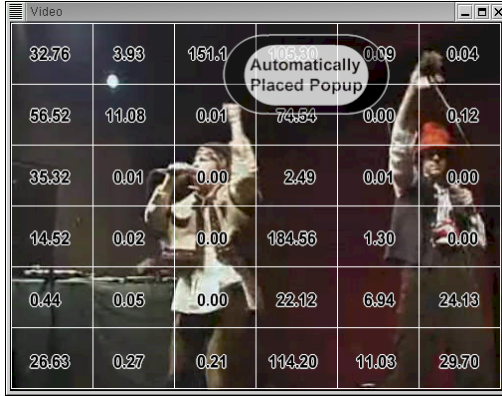


Fig. 2. A “Pop Up Video” style annotation, automatically placed in a music video. The numbers show our final placement criterion value for this specific frame for a grid of 36 discrete regions.

III. SYSTEM DESIGN

The system architecture is depicted in Figure 3. An offline preprocessing step is applied to the raw video feed to generate information required during the online placement of annotations. The preprocessing stage includes analysis of the video for the elementary perceptual descriptors. We divide the video frame space into 16x16 *micro-blocks* and apply descriptor formulations to each individual block. The result for each frame over the entire length of the video is then stored on disk. The

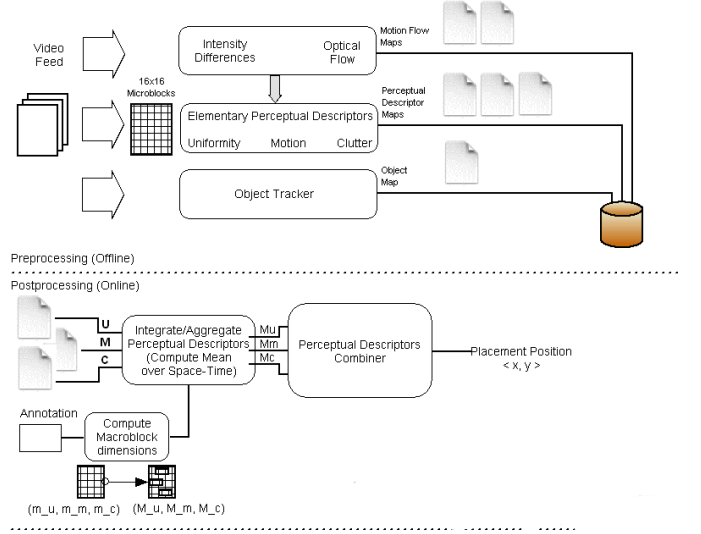


Fig. 3. System Architecture

preprocessing stage also includes allied modules such as object tracking and shot boundary detection and can be extended to incorporate additional ones.

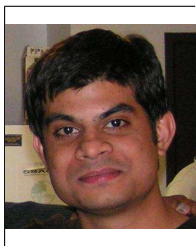
Online annotation placement is then determined at run-time by combining values for each descriptor from the preprocessed micro blocks into a *macro block* (annotation-sized) value. The system design supports online augmentation of the video feed by placement of any arbitrary-sized annotation. This approach is applicable to a variety of domains where the video feed is available for analysis purposes.

IV. CONCLUSION

We have presented a framework for the automated placement of annotations in arbitrary videos. The augmentation of general videos with annotations presents useful applications in sports and entertainment. Other useful application domains for video augmentation include instructional and training videos (a virtual classroom or distance learning scenario), video brochures (e.g., for tourism), or personal home videos - all of these can be annotated with relevant information. There exist interesting research aspects yet to be explored. For example, additional elementary descriptors, such as uniform texture can be factored into the placement decision.

REFERENCES

- [1] Princeton Video Image, “Virtual Product Integration,” <http://www.pvi-inc.com/>, 2004.
- [2] Sportvision, “Race F/X,” <http://www.sportvision.com/>, 2004.
- [3] V. Thanedar and T. Hollerer, “Semi-automated placement of annotations in videos,” Tech. Rep. 2004-11, Department of Computer Science, University of California, Santa Barbara, 2004.
- [4] S. G. Mallat, “A theory for multiresolution signal decomposition: The wavelet representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.



Vineet A. Thanedar Vineet is working towards his M.S degree in Computer Science at UC, Santa Barbara. Previously, he completed his Bachelor of Engineering (Computer Engineering) degree from the University of Pune, India. At UCSB, his IGERT as well as Masters thesis advisor is Prof. Tobias Höllerer.

Map Design and Perceptual Salience

Kirk P. Goldsberry
IGERT Associate
Department of Geography

Abstract— The eye-movement measurement methods developed in psychology have received renewed interest in human computer interaction research to study how people interact with graphical user interfaces. This research studies map users' eye movements while they are inspecting static meteorological map displays. Cartographers control visual variables in order to make thematically relevant information in a map display perceptually more salient. The goal of this research is to enhance our understanding of how adjusting certain visual variables will influence people's viewing behavior dependent on perceptual salience in the display. Specifically, we create research stimuli (maps) that adhere to cartographic conventions as well as enable empirical research into the effects of visual variables on perceptual salience. Several map displays are produced using different cartographic techniques. In order to isolate the effects of the design variables, multiple maps representing the same geographic data set are generated. The eye-movement methods will explore subjects' variable responses to these maps. The broader impacts of this research include progress toward an improved framework linking map-purpose and cartographic design decisions that would increase the communicative efficiency of geographic data displays.

Index Terms—Perceptual salience, Cartography, Eye-movement, Visual variables

I. INTRODUCTION

As digital maps become increasingly prevalent, map designers have a growing responsibility to optimize the graphic communication of geographic information. Previous cartographic research has attempted to suggest design principles applicable to the graphic representation of geographic information. Until recently however, a lack of empirical measurements has prevented researchers from confirming the validity of many common cartographic conventions. Fortunately, technological advancements have benefited the eye-movement measurement methods developed in psychology. State-of-the-art hardware and software can interact to collect large amounts of data relating to how people interact with computer map displays.

Map designers strive to optimally present geographic information to map-readers. This research attempts to measure the influence of cartographic visual variables on perceptual salience. Specifically, this research uses eye-movement measurement and saliency models to describe the effects of three visual variables: color hue, color value, and size. The goal is enhance our understanding of the relationship between thematic relevance and perceptual salience (Lowe, 1999, Lowe, 2003).

This research uses stimuli taken from previous psychological research (Hegarty, 2003). Hegarty (2003) measured interactions between subjects and static meteorological displays (see Figure 1). Her research involved measuring the differences in interaction and performance times between novices and experts. The

research here aims to improve overall subject performance by making thematically relevant information more perceptually salient.

II. METHODS

There are three “themes” present in Figure 1: Temperature (represented using color-filled isotherms), Air Pressure (represented using hollow isobars), and the base map data (state boundaries, water bodies etc.). Hegarty and Canham asked subjects questions involving air pressure although the map stimulus' representation of air pressure is clearly less salient than that of temperature. The ITTI saliency map shown in Figure 2 demonstrates the salience of the original stimulus. This research creates alternate representations of the same data represented in Figure 1. These alternate representations are created by adjusting three visual variables (color hue, color value, and size). The goal of the alternate representations is to make the thematically relevant information (Air pressure) more perceptually salient.

The ITTI model is used to initially evaluate salience differences between the Hegarty stimulus and the adjusted representations. The results are demonstrated in Figures 3-5. The ITTI output images present the most salient areas of the corresponding input image in white. The least salient areas are represented as black.

Unfortunately there are no eye-movement results at the time of writing. The expected results include both increased accuracy and improved performance times with the adjusted weather maps.

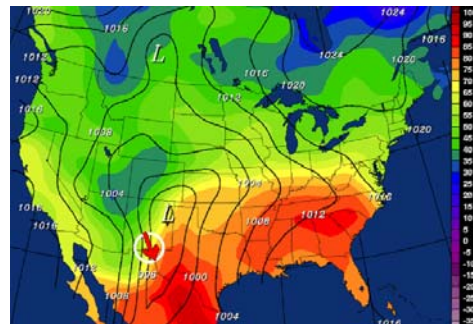


Fig. 1. An original stimulus from the Hegarty study. Air pressure is represented using black isobars. Temperature is represented using a “rainbow” color scheme.

Salience depends on cartographic design decisions. Manipulating the visual variables will influence what stands out most on maps. The ITTI model predicts which parts of an image will attract subjects' attention. Eye-tracking

experiments can empirically evaluate the interactions between subjects and map displays. Future cartographic experiments should evaluate implications of conventional design decisions. Ideally the results of eye-movement and interaction research will lead to practical benefits for the cartographic community.



Figure 2: An ITTI saliency output for Figure 1. The white areas on this output indicate that the yellow/orange temperature bands are the most salient elements of the original stimulus.

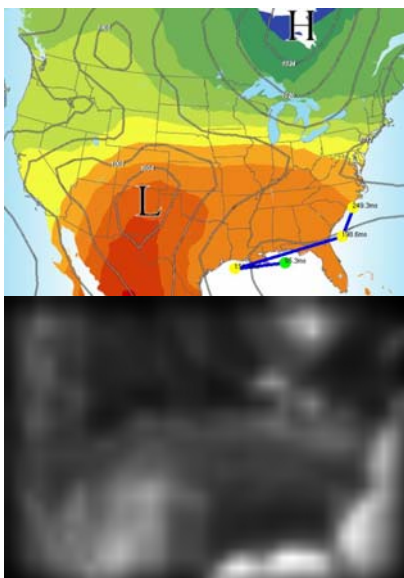


Figure 3: Adjusted the size of the isobars in an attempt to make air pressure more salient. This ITTI output indicates that the Gulf Coast is the most salient area of the map.

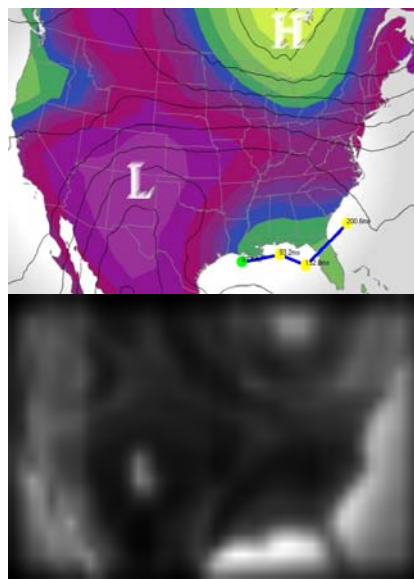


Figure 4: This map represents air pressure using filled “contours” and temperature using black lines. Still, the coastline appears to be the graphic’s most salient area.

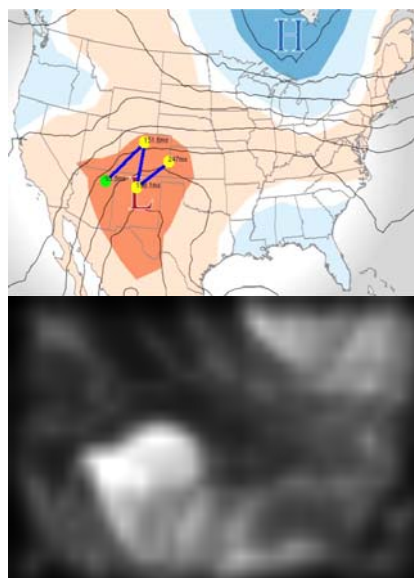


Figure 5: Adjusting the air pressure color scheme results in some major saliency changes. In this case, a 5-Class diverging red-blue color scheme successfully highlights areas of high and low pressure. Meteorological convention suggests red hues are associated with areas of low pressure, and blues are often associated with areas of high pressure.

REFERENCES

- [1] Bertin, J. (1983). *Semiology of Graphics: Diagrams, networks, maps*. Madison, WI, University of Madison Press.
- [2] Lowe, R. K. 1999. Extracting information from an animation during complex visual learning. *European Journal of Psychology of Education* 14:225-244



Kirk P. Goldsberry is a first-year PhD student in the UCSB Department of Geography. His primary research focuses on modern cartographic design. His faculty advisor is Dr. Sara Fabrikant. Dr. Fabrikant researches Cartography, GeoVisualization, Cognition, and Spatialization.

The Influence of Spatial Ability on the Use of Dynamic, Interactive Animation in a Spatial Problem-solving Task

Cheryl A. Cohen
IGERT Associate
Psychology

Abstract— This protocol study investigated the differences in problem-solving strategies used by participants with low- and high-spatial ability on a spatial visualization task. The task required participants to draw a cross-section of an unfamiliar 3D digital object. Three sources of data were used to analyze differences in task performance among low- and high-spatial participants: 1) frequency of animation use; 2) coded verbal reports, extracted from the participants' 'think-aloud' protocols; and 3) drawing accuracy. Participants with high spatial ability interacted with the animations more frequently than did low-spatial participants. Furthermore, high spatial subjects mentioned a greater variety and number of physical and spatial features of the stimulus object than did low spatial participants. Finally, high spatial participants drew more accurate representations of cross-sections than did low-spatial participants.

Index terms: Spatial ability, individual differences, interactive animation, protocol studies

I. INTRODUCTION

EDUCATORS in a wide range of disciplines see great potential in the use of dynamic, interactive computer visualizations for improving education and training. Yet cognitive psychology research suggests that individual differences among learners in the use of interactive visualizations influence the amount of learning that occurs. Spatial ability is one individual difference that can influence a learner's ability to extract information from dynamic, interactive animations. In three previous experiments, the author demonstrated that spatial ability and the frequency with which participants interacted with an animated computer model both made significant contributions to performance [1] on a spatial visualization task. High spatial participants interacted with the animated model more often and systematically than did low spatial participants. Many low spatial participants stated that they did not understand how to use the animated model to help them solve the task.

The goal of this protocol study was to investigate the strategies used by high- and low- spatial participants to perform a spatial task involving interactive computer visualizations. The performance task used in this study was identical to the one used in the author's previous experiments. A possible application of this research is to develop training programs to help individuals with low

spatial skills learn effective strategies for interacting with dynamic digital animations.

II. METHOD

Six participants (3 high- and 3 low-spatial ability) were recruited from graduate programs at the University of California, Santa Barbara. Potential participants were screened for spatial ability based on two psychometric tests: the Guay-Lippa Visualization of Viewpoints [4] and the Vandenberg Mental Rotation Test [5].

In preparation for the performance trials, participants read a definition of the term "cross-section," and viewed an instructional animation that demonstrated a cross-section being cut from an apple. During the performance phase of the study, participants completed 12 paper-and-pencil trials in which they drew the cross-section of an imaginary object. The stimulus figure used in the performance trials was egg-shaped, with a transparent exterior that revealed an internal network of duct-like structures. (Figure 1). The figure was modeled to resemble in shape and complexity the biliary ducts of the human liver. Pictorial depth cues, such as highlights, shadows, and visual occlusion suggested spatial depth in the figure. A fictitious figure was used in order to avoid any confound introduced by participants' recognition of a familiar object.

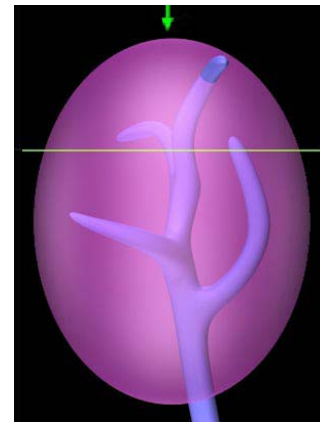


Fig. 1. The stimulus object used in all performance trials

Written instructions directed participants to imagine that the egg-shaped figure was sliced at a green line, and that they were viewing the figure from the perspective of an

arrow. Participants were further instructed to draw the cross-section of the stimulus figure that would result from the imagined slice, at the perspective, or view orientation, indicated by the arrow. During all performance trials, participants had unrestricted access to two interactive animations of the egg-shaped stimulus object. One animation showed the stimulus object rotating 360° around its vertical axis, while the second showed a 360° rotation around the horizontal axis. Each animation had a slider bar that allowed the participant to pause the rotation at a particular view of the stimulus figure. The slider bar also allowed participants to advance or reverse the rotation at a self-controlled speed.

III. PROCEDURE

The experimenter explained the meaning of the term “cross section” and showed the participants the instructional animation. Participants were asked to ‘think aloud’ as they completed the cross-section drawings, and were videotaped as they performed the performance measures. Participants were tested individually. There was no time limit for the performance trials.

Videotapes of each participant’s performance on the trials were reviewed and coded for frequency of animation use. A single instance of animation use was defined as the participant’s manipulation of either the horizontal or vertical animation, per trial. In addition, the verbal report of each participant’s ‘think-aloud’ protocol was transcribed and coded for the frequency with which they mentioned physical and spatial features of the stimulus figure. Table 1 lists the features that were coded from participants’ verbal protocols.

Object features
Shape (e.g., shapes of branches)
Angle (e.g., angles of intersection of branches)
Local position of features
Global position of features
Cutting plane features
Orientation (horizontal or vertical plane)
Location (top, bottom, right or left of figure)
View orientations
2D view
“Correct answer” view
Animation view

Table 1. Coded features of the stimulus object

IV. RESULTS

Three sources of data were used to analyze the differences in performance among high- and low-spatial participants: 1) frequency of animation use; 2) verbal report data (mention of physical and spatial features of the stimulus object) extracted from the participants’ ‘think-aloud’ protocols; and 3) cross-sectional drawing accuracy.

A. Frequency of animation use

Consistent with the author’s previous research, the high spatial participants used the animation more frequently than the low spatial participants (Fig. 2).

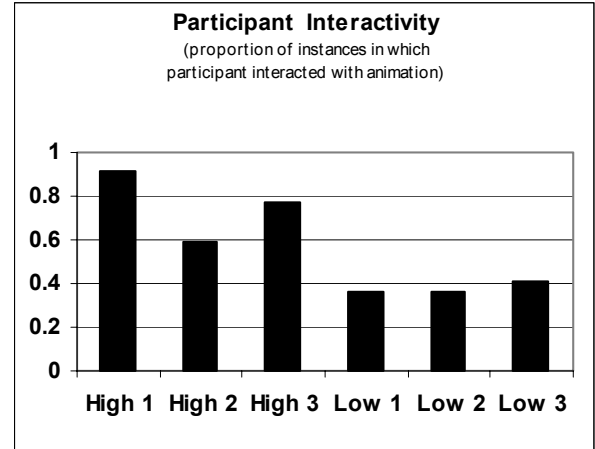


Fig. 2. Frequency of animation use by high and low spatial participants

B. Verbal Report

High spatial participants mentioned a greater variety and number of physical and spatial features of the stimulus figure than did low spatial subjects (Fig. 3). The greatest difference between high- and low-spatial participants on verbal coding was seen in the mention of the cutting plane; high spatial participants mentioned the cutting plane almost twice as often as low spatial participants. Of further interest is the trend for both high- and low- spatial participants to mention view orientations less frequently than they did features of the stimulus object and the cutting plane.

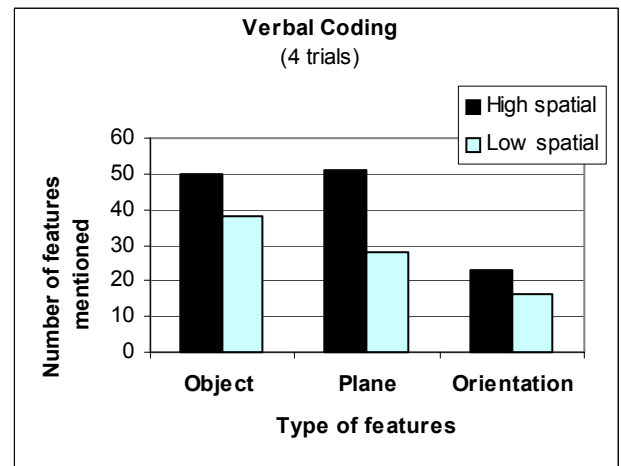


Fig. 3. Physical and spatial features mentioned by high- and low-spatial participants.

C. Cross-sectional drawing accuracy

Figure 4 compares the relative performance of high- and low- spatial participants on a composite variable, the mean of the outer shape and angular relationship variables. Fig. 4. Performance of high and low participants on an aggregate variable (mean of performance on outer shape and angular relationship variables).

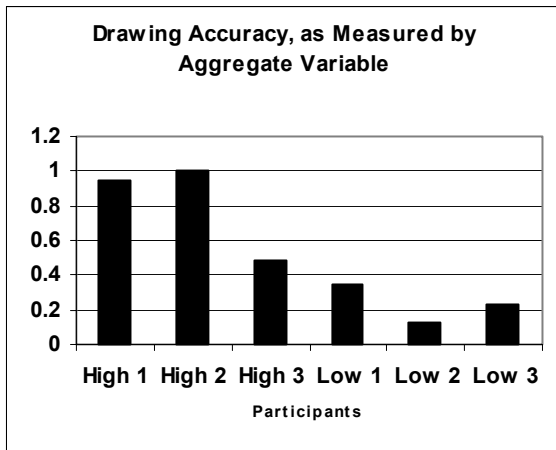


Fig. 4. Performance of high- and low-spatial participants, as measured by an aggregate variable (mean of outer shape and angular relationship variables.)

V. DISCUSSION

With regard to frequency of animation use, low spatial participants often expressed difficulty in understanding how to use the animations.

The fact that high spatial participants mentioned a greater variety and number of physical and spatial features may indicate that they are more aware of the spatial factors required to understand and visualize any three-dimensional object. There was also a qualitative difference in the type of object features mentioned by high- and low- spatial participants. Low spatial participants tended to mention surface features of the object, such as color and shading, while high spatial subjects more often mentioned shapes and angular intersections within the object. Although shading is an important two-dimensional depth cue, more revealing three-dimensional depth cues were available to participants who rotated either or both of the animations.

Of interest is the fact that both highs and lows mentioned view orientations less frequently than the other two classes of feature variables (object features and cutting plane features.) Mental representations of at least three different view orientations were required in order to imagine, and draw, the correct representation of a given cross-section. Only two of these three view orientations were visible at any given time: the 2D view of the object as seen in the paper-and-pencil problem, and the view of each animation, which would change only upon the participant's interaction. The third view orientation, the representation of the "correct

answer," must be imagined by the participant, or inferred from spatial information provided in the interactive animations. Perhaps this is the hardest element of the task for participants across abilities since it requires the participant to engage working memory and problem-solving resources.

Visualizing the unseen cross-section of an unfamiliar object involves mental transformations of given views, or, in the case of this study, interactivity with external visualizations. The initial coding of the verbal protocols suggests that low spatial participants were less aware of key spatial features and relationships than were high spatial participants. If the low spatial participants were unaware of certain spatial features and relationships, they presumably would not be motivated to interact with the animation in order to get more spatial information about these features. Their inability to use the animation effectively could be a result of difficulty mapping the two-dimensional features and relationships seen in the stimulus drawing onto the rotating animations. Coding of the behavior of both high and low participants will offer further data to synthesize with existing data.

REFERENCES

- [1] Cohen, C., Hegarty, M., Montello, D. & Keehner, M: (2004) *The Effects of Spatial Abilities and Animated Instruction on Representing Cross Sections*. Paper presented at the meeting of the American Educational Research Association, San Diego, CA, 2004.
- [2] Ericsson, K. and Simon, H. (1984). *Protocol Analysis: Verbal Reports as Data*. Cambridge, MA, MIT Press.
- [3] Lowe, R. (in press) Interrogation of a dynamic visualization during learning. *Learning and Instruction*.
- [4] Guay, R. & Mc Daniels, E. (1976), *The Visualization of Viewpoints*. The Purdue Research Foundation. West Lafayette, IN. (as modified by Lippa, Hegarty, & Montello, 2002).
- [5] Vandenberg & Kuse (1978). Mental rotations: Group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, 47, 599-604.



Cheryl A. Cohen is completing a PhD in Cognitive Psychology at UCSB. Her research interests focus on spatial cognition and problem-solving. Her IGERT advisor is Dr. Mary Hegarty and her expected date of graduation is June 2006.

Is Student Engagement Dependent on Lecture Relevance?

A Study of Student Engagement in the Multimedia Classroom

Monica Bulger
IGERT Associate
Gevirtz Graduate School of Education

Abstract— To determine whether student engagement was dependent on lecture relevance, a pilot study was conducted in a composition classroom at UCSB. Using monitoring software to record student activities, a comparison was made between patterns of student computer use and lecture activities. Phases of engagement reflect student sensitivity to the relevance of lecture content.

Index Terms—Engagement, computers and writing

I. INTRODUCTION

A review of the research literature concerning the topic of classroom engagement reveals that, although commonly thought to simply mean "involvement" or "participation" in classroom activities, engagement is a multi-faceted term that can be studied from behavioral, cognitive, and emotional perspectives. The way in which a researcher categorizes the term "engagement" influences the approach and design of his/her study.

Defined as "to engross wholly" and "to participate, to be involved in an activity," the term *engage* is a verb that can describe the degree to which an individual is involved in an activity [1]. Kearsley and Shneiderman (1998) define *engagement theory* as "students must be meaningfully engaged in learning activities through interaction with others and worthwhile tasks"[2]. Engagement theory reflects the ideals of Mayer et al's (1999) *constructivist learning*, which "occurs when learners actively construct meaningful mental representations from presented information"[3].

From a behavioral perspective, *engagement* includes positive conduct, involvement in learning and academic tasks, and participation in school-related activities [4]. As Lentz (1988) points out, categories for *on-task* behaviors are very specific, and include "orientation to task, teacher, board, or reciting student," whereas *off-task* behaviors generally include "everything else" [5]. Behavioral engagement is something that can be visually observed and evaluated. Students who are behaviorally engaged may ask questions in class, face the locus of classroom activity, or avoid disruptive activities. The behavioral perspective provides a starting point for this study.

Previous studies [6] of student engagement involve observation, interviews, and retention testing. In this study, engagement was measured instantaneously using monitoring software to record student computer activities. These activities were synched with a time-stamped video recording of the lecture and compared with observations of student behavior in the classroom.

II. METHOD

A. Participants

The participants were 19 undergraduate college students enrolled in a Writing 2 course at the University of California, Santa Barbara. All students enrolled in specific sections of Writing 2 taught by the same instructor during the summer of 2004 were invited to participate.

B. Materials

The computer classroom in which the study occurred held twenty-five computers. Computers were arranged in five vertical rows; most students would need to swivel their chairs sideways, away from the computer screen, to view the lecturer at the podium in the front of the classroom. The classroom was equipped with Dell Pentium III computers, Boss Everywhere monitoring software, Internet access, Microsoft Office, and graphic development software. Each computer was identically configured and permitted any student attending the class to login to an identical desktop. A video camera was operated by a technician and positioned in the back left corner of the classroom, near two unused computers.

C. Procedure

Participants were observed within the context of a classroom session. One week before the observation, students signed a voluntary informed consent form following APA guidelines for informed consent. On the day of the observation, as soon as each student entered the computer classroom, they could login to a computer of their choice and the monitoring software recorded their activities. The teacher's lecture was recorded using a time stamped videotape. In addition, minute-by minute student activities

were recorded by a single observer using handwritten observation. Handwritten observation consisted of noting where students were looking (at screen, at teacher, at notebook, around room) and whether the student left the computer (meet with teacher at front of classroom, leave classroom).

III. RESULTS

The monitoring software recorded user activities by minute, duration of time spent on each application, and periods of inactivity. To determine whether patterns existed in student computer use, I broke down the data into minutes and counted the number of users and actions per minute for each application. I created a table of this data and was able to see how many students were actively using the Internet or Microsoft Word at any given minute. I created a chart of actions per minute for Internet use and another chart for Word use. The charts allowed me to see patterns of computer activity.

I then transcribed the lecture and also broke it down by minute. After comparing the student activities with the lecture activities, I decided to categorize the parts of the lecture that were relevant and irrelevant to the central topic. For example, if the instructor began to talk about a recent vacation, or described difficulties she was having with the lab computers, I considered this irrelevant, whereas if she talked about supplemental materials for the lesson or described an upcoming assignment, I considered this relevant.

I compared the computer activities with the contents of the lecture. If students were engaged in activities that were not related to the course, such as searching ESPN, shopping on Ebay or participating in Instant Messaging, I labeled these actions as “off-task.” If the students were using Microsoft Word to take notes, or visiting a website at the request of the instructor, I labeled these actions as “on-task.”

Once the categories were in place, I started to observe specific phases of engagement. Patterns of student engagement reflected the flow of the class: prior to class beginning, an initial wind-up phase reflected high non-course relevant Internet use. Two minutes into the lecture, students began to reduce their Internet use. During the core ten minutes of lecture, a focused attention phase was evidenced by an increase in Microsoft Word use, a decrease in Internet use, and an additional decrease in use of open applications that were not relevant to the course (such as music, instant messaging, and e-mail). Five minutes before the lecture ended, students entered a restless phase in which they start completing housekeeping activities, such as checking their e-mail, finishing Internet searches, and closing applications.

Patterns of student engagement reflected the flow of the class: prior to class beginning, an initial wind-up phase reflected high non-course relevant Internet use.

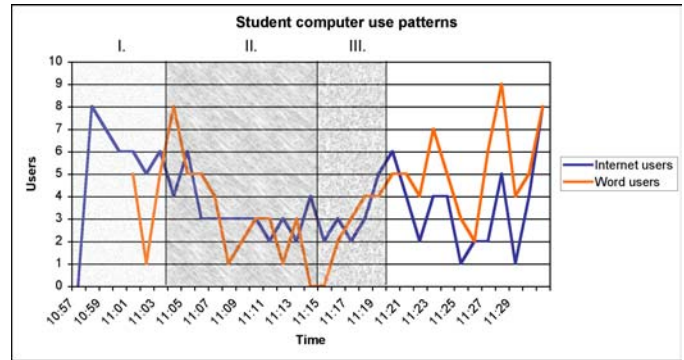


Fig. 1 Phases of Student Engagement. The number of Internet users and Word users during the three phases of classroom engagement.

Three minutes into the lecture, students begin to reduce their Internet use. During the core ten minutes of lecture, a focused attention phase is reflected by an increase in Microsoft Word use, a decrease in Internet use, and an additional decrease in use of open applications that were not relevant to the course (such as music, instant messaging, and e-mail).

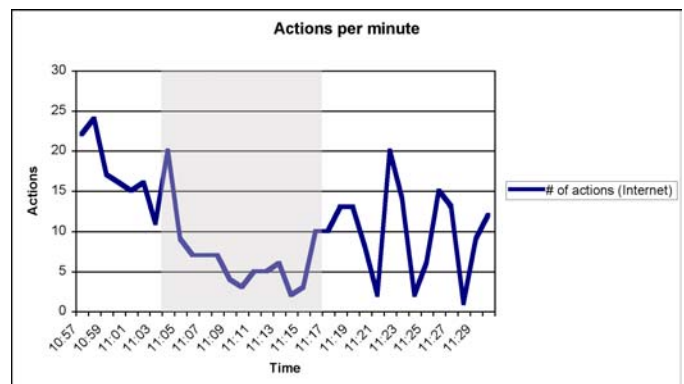


Fig. 2 Internet Actions per Minute. This graph shows a significant decrease in the amount of Internet actions during the focused attention phase.

Five minutes before the lecture ended, students entered a restless phase in which they appear to sense the end of the lecture and start completing housekeeping activities, such as checking their e-mail, finishing Internet searches, and closing applications.

DISCUSSION

The phases of engagement reflect student sensitivity to the relevance of the instructor’s lecture. Students did not significantly reduce off-task behavior such as Internet use until two minutes into the lecture. This could be a result of students finishing whatever off-task activities they had begun before the lecture. Another explanation could be that for the first two minutes, the instructor described a kayaking trip in Hawaii. Students may not have seen the relevance of this description until she connected it with the lecture topic, at which point Internet use significantly declined, going in a single minute from 20 actions per minute to less than 9.

Additionally, for a sustained ten minute period, the focused attention phase, actions per minute remain below 7. During this time, the instructor is explaining an upcoming assignment and introducing the concept of revision.

Recording student computer activities provides an instantaneous measure of engagement. Limitations of this study include a small sample size (19 students) and short lecture duration (17 minutes). This method of measurement is promising, and would be better applied to a larger sample. Phases of engagement could be tested with longer lecture duration and larger sample size to determine whether more phases would emerge or whether the phases would remain similar in proportion if occurring during a longer lecture.



Monica Bulger is a doctoral student in the Gevirtz Graduate School of Education. Her research interests include student engagement, composition instruction, and multimedia learning.

ACKNOWLEDGMENTS

Thank you to IGERT for providing this research opportunity. Thank you to my advisor, Charles Bazerman, and my IGERT mentors Richard Mayer and Kevin Almeroth for their time and direction.

REFERENCES

- [1] Meece, J.L., Hoyle, R.H., & Blumenfeld, P.C. (1988). Students' goal orientations and cognitive engagement in classroom activities. *Journal of Educational Psychology*, 80 (4), 514-523.
- [2] Kearsley, G. & Shneiderman, B. (1998). Engagement theory: a framework for technology-based teaching and learning. *Educational Technology*, 38, 20-23.
- [3] Mayer, R.E., Moreno, R., Boire, M., & Vagge, S. (1999). Maximizing constructivist learning from multimedia communications by minimizing cognitive load. *Journal of Educational Psychology*, 91 (4), 638-643.
- [4] Fredricks, J.A., Blumenfeld, P.C., & Paris, A.H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74 (1), 59-109.
- [5] Lentz, F. E. (1988). On-task behavior, academic performance, and classroom disruptions: Untangling the target selection problem in classroom interventions. *School Psychology Review*, 17 (2), 243-257.
- [6] Fredricks, J.A., Blumenfeld, P.C., & Paris, A.H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74 (1), 59-109.

Experiments in Sound Granulation and Spatialization for Immersive Environments

David Thall
IGERT Associate
Media Arts and Technology

Abstract— A new granular sound signal generator, processor, and control system is presented, developed, and discussed. The unifying principles that exist among past implementations are first examined. The result of this investigation is a systematic design that, through reclassification and generalization, combines and extends these early models. Sets of signal processing sub-routines are then identified, collected and combined into a general-purpose grain generator. A novel user interface is also presented along with various high-level control regimes. This assists the user in navigating the multi-dimensional parameter space inherent in granular transformations. The combination of fast and efficient processing with scalable and customized controllers results in a system designed exclusively for advanced granular synthesis and processing. This system could be used in real, virtual, and mixed immersive environments, providing a real-time synchronized stream of complex spatial ambient and environmental tones and noises alongside synthesized visual and/or tactile streams.

Index Terms—Granular sound synthesis and transformation, immersive environments, 3-D sound spatialization, ambient and environmental sound emulation.

I. INTRODUCTION

GRANULAR sound synthesis and sampling is a family of powerful, time domain techniques for sound design and composition [1]. They allow the composer to stream or scatter acoustic particles in multiple dimensions, either in real-time or by script. Using various granular synthesis and processing models, a set of parametric control data is generated and mapped to an underlying grain-scheduling algorithm. In this context, a ‘sound grain’ can be considered a finite time segment of an arbitrary waveform modulated (shaped) by an amplitude envelope.

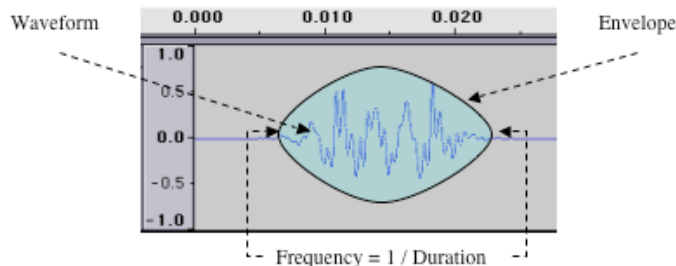


Fig. 1. Anatomy of a Sound Grain.

Historically, granular synthesis has been used to perform

such acoustic transformations as pitch shifting and time-scaling, and has been used for spectrum analysis [2], [3]. In its recent incarnations, digital audio programmers have utilized the technique to widen the range of possible sound enhancements and transformations. The ability to work with sound at the level of individual sound grains opens up the possibility to deconstruct and reassemble sounds anew, into innovative shapes and textures.

II. UNIFICATION

Various modern implementations of grain-based transformation systems have been proposed. Many of these software prototypes have expanded the parameter space available for sonic transformations; however, they fail to give the user a comprehensive control regime. Table 1 displays a breakdown of a select group of modern grain-based systems.

The commonalities and differences that exist among these systems are combined into a real-time, generalized granular processing and control model. *Cloud* is a general-purpose, multi-tiered granular synthesis and transformation plug-in written for the Supercollider 3 sound synthesis server developed by J. McCartney. *Emission Control* is a graphical user interface (GUI) and high-level control scheme designed exclusively for advanced granular transformations in time and space.

TABLE I
CATEGORICAL BREAKDOWN OF MODERN GRANULATION IMPLEMENTATIONS

	PODX	ISPW	CloudGen	constQ	Creatovox	PulsarGen	Reaktor	Reason	Emission Ctl
Stream Overlap Control	✓	✓		✓	✓	✓	✓		✓
Particle Density Control			✓	✓					✓
Implicit Parameter-Linking				✓	✓	✓		✓	
Explicit Parameter-Linking									✓
Synchronous Scheduling	✓	✓	✓	✓	✓	✓	✓	✓	✓
Asynchronous Scheduling			✓	✓	✓			✓	✓
Pattern Scheduling					✓			✓	✓
Arbitrary Scheduling									✓
Deterministic Processing			✓			✓		✓	✓
Statistical Processing	✓	✓	✓	✓	✓		✓	✓	✓
Arbitrary Processing									✓
Per-grain Filter				✓					✓
Per-grain Dynamics									✓
Multi-file		✓							✓
Multi-stream	✓	✓			✓			✓	✓
Multi-channel				✓	✓				✓

III. ALGORITHM

A software algorithm to schedule, generate and process acoustic grains must be designed to be fast, efficient, and scalable [4]. Fig. 2 displays the algorithm used. In order to be fast, many repetitive calculations within the sample-block loops have been replaced with memory accesses to stored function tables. A data structure has been created that

dynamically allocates and de-allocates grains from memory, allowing for the most efficient use of time and resources. Furthermore, the sections of the algorithm are scalable to allow for a user-specifiable ceiling limit of grain overlap and number of output channels.

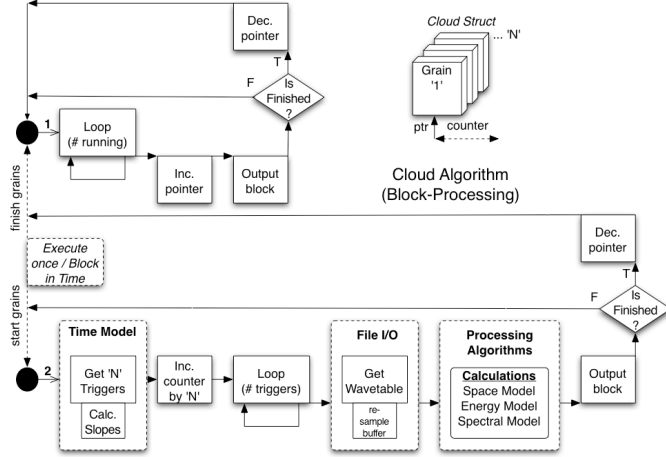


Fig. 2. The Cloud Algorithm. Two loops run in parallel. The first loop outputs currently running grains (i.e., grains that haven't finished processing since the previous sample block). The second loop schedules, calculates, and generates new grains.

In order to abstract the granulation algorithm into a generalized form that could describe a majority of the possible model variations and implementations, the architecture has been designed to combine and extend earlier systems (such as those described in Table 1). Multiple modular, decoupled processing models can be easily combined, reconfigured, and expanded as needed. Three such models (see Fig. 3) include calculations for arbitrarily-complex time-based grain scheduling, multi-channel per-grain spatialization (including per-stream velocity and distance cues [5], [6]), and different types of perceptual gain compensation applied to varying densities of overlapping grains.

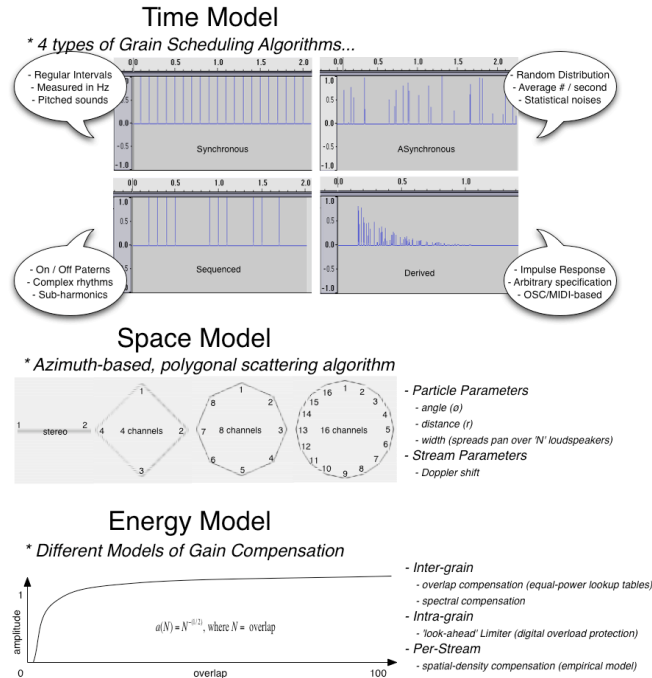


Fig. 3. Granular processing models of time, space, and energy. Other models (as well as extensions to those shown) are possible.

Due to the dynamic nature of the granulation process, CPU-load control is needed to handle the extreme ranges of stream density and grain durations. At any point in time, 10s or even 100s of grains may be created. Previous implementations have attempted to explicitly set limits on density and/or duration, or set a threshold above which grains will be dropped from the scheduler. An alternative to this is the adaptive *flow control* algorithm developed by the author (see Fig. 4). The integration of this process into the *Cloud* algorithm allows the system to maintain consistent real-time output in a variety of CPU-intensive situations.

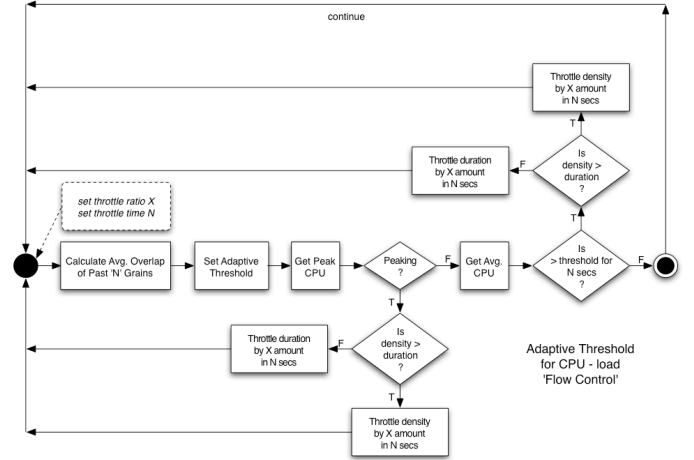


Fig. 4. An adaptive threshold algorithm for granular 'Flow Control'. The system can be thought of as a type of dynamic compressor that throttles the scheduler and the output grains according to the measured Peak and Average CPU.

IV. INTERFACE

A generalized control scheme to perform high-level, multi-dimensional granular transformations is needed. The system should be powerful, providing numerous options and specifications. Furthermore, the various control devices should allow for morphological composition planning on multiple time scales, with an emphasis in directly manipulating the advanced granular processing parameters through sets of advanced graphical views.

A set of novel user interface elements have been constructed that work together to control all aspects of the underlying grain processor, from making simple parameter updates to automating complex multi-dimensional parameter morphologies. A tree diagram of the class hierarchy can be viewed in Fig. 5.

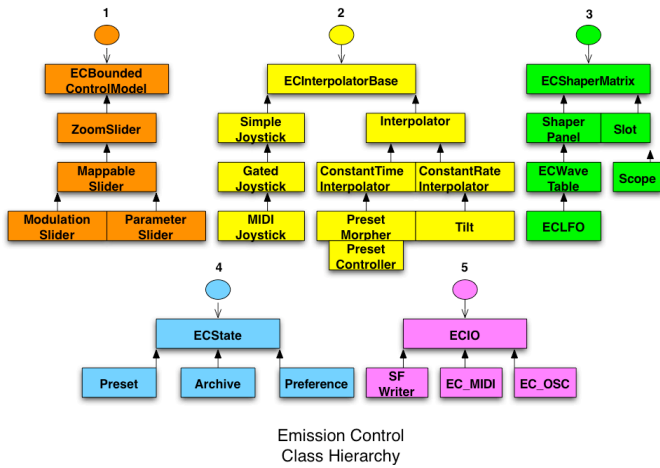


Fig. 5. Emission Control Class Hierarchy. The library can be broken into a five-part tree topology: Range-bounded controllers, multi-parameter interpolators, a modulation matrix, a system of presets and preferences, and input/output specifications.

The current user interface combines these basic elements into a unifying model. A select group of parameters has been chosen for real-time control. Each parameter can be modulated from an externally-routed control source, allowing for arbitrarily-complex granular transformations over multiple time-scales. Furthermore, multiple parameters can be linked together according to user-defined mathematical relationships, and controlled using high-level controls such as 2-D graphical joysticks and ‘N’-dimensional parameter interpolators.



Fig. 6. The current state of the Emission Control GUI.

V. FUTURE WORK

In the future, a spherical spatialization algorithm will be developed that could eventually be tested in the CNSI sphere under construction at the University of California, Santa Barbara. This algorithm extends the polygonal spatialization algorithm by subdividing the polygon a number of times and rotating the divisions up to 90° on the Z-axis. Coding a binaural spatial model will offset the difficulty of finding a setup for auditioning multi-channel granular sound transformations. A model of sound source occlusion may also be adapted into the spatialization routine, calculating the reflective or absorbent quality of objects within, for example, a visually-synchronized scene.

ACKNOWLEDGMENT

The author would like to thank Dr. Curtis Roads for his continued support and guidance.

REFERENCES

- [1] Roads, C. 2002. *Microsound*. Cambridge, Massachusetts: MIT Press.
- [2] Gabor, D. 1946. "Theory of communication." *Journal of the Institute of Electrical Engineers Part III*, 93: 429-457.
- [3] Otis, A., Grossman, G., Cuomo, J. 1968. "Four sound-processing programs for the Illiac II computer and D/A converter." *Experimental Music Studios Technical Report Number 14*. Urbana: University of Illinois. Paper by Otis, A. "Program 3. Time Rate Changing."
- [4] Bencina, R. 2001 "Implementing Real-Time Granular Synthesis." Unpublished manuscript. Revised and updated version "Implementing Real-Time Granular Synthesis" in Greenbaum (Ed), *Audio Anecdotes*, A.K. Peters, Natick. Awaiting publication.
- [5] Chowning, J. 1971. "The Simulation of Moving Sound Sources." *Journal of the Audio Engineering Society* 19(2-6).
- [6] Loomis, J., Klatzky, L., Golledge, G. 1999. "Auditory Distance Perception in Real, Virtual, and Mixed Environments." In Y. Ohta & H. Tamura eds., *Mixed Reality: Merging real and virtual worlds*. pp. 201-214. Tokyo: Ohmsha, 1999.



David Thall is currently a Lecturer in the Media Arts and Technology Program at the University of California, Santa Barbara. In 2004 he earned an M.S. in Multimedia Engineering from UCSB. His research interests include particle-based sound processing and user interface design. His IGERT advisor is Dr. Curtis Roads.

OnKai: Sculpting Three-dimensional Objects for Control in Computer Music Composition

Satoshi Morita
IGERT Associate
Media Arts and Technology

Abstract— The author has designed and developed a virtual reality system in C++ and SuperCollider, which enables composer to sculpt three-dimensional objects for control in computer music composition. System incorporates hand gestures and head movements as an input from the user by magnetic tracker and data glove, and enables users to interact with three-dimensional objects displayed in the head mounted display. Synthesis/Processing parameters, triggering and spatialization parameters can be controlled by manipulating orientation, location and shape of three-dimensional object. First implementation of the design has been realized and tested. It has demonstrated the benefit of representing related parameters compactly. Further improvement for object surface manipulation was suggested.

Index Terms— computer music, graphical user interface, human computer interaction, virtual reality.

I. INTRODUCTION

In computer music composition, composers need to manipulate various aspects of the composition, such as audio synthesis and processing parameters for manipulating the timbral characteristics of their instruments, triggering timing and volume changes of the instruments in time, and panning and reverberation parameters for mixing and spatialization. Organizing and manipulating these parameters over time has been a challenge of composing with computers, where many of the responsibilities such as instrument design, mixing and spatialization are not delegated to others but also in strict control of the composers. Currently, composers use either computer programming languages specially designed for audio synthesis and music composition or DAW (digital audio workstation) which has graphical user interface drawing analogy from musical scores, time-sheet and physical audio hardware. Composing by writing program brings flexibility but lacks intuitiveness, and composing by using DAW lacks flexibility of control and is difficult to understand the relationships among different parameters.

OnKai is a virtual reality system designed by author for manipulating, organizing and recording compositional parameters for computer music composition. Composers can manipulate their musical forms and sound synthesis/processing parameters by interacting with three-dimensional objects (spheres, cubes), which represents one or

multiple musical parameters by changing object's attributes (shape, location, orientation). Its aim is to research an alternate system for composers to realize their works in intuitive yet flexible manner by mapping parameters to three-dimensional objects for organization and manipulation.

II. SYSTEM DESIGN

User interacts with the system by typical virtual reality system components, such as magnetic trackers, data glove and Head Mounted Display (HMD). Kaiser Electro-Optics ProView HMD is used to display the virtual world which user is immersed in. Two Ascension Flock of Birds [1] magnetic trackers are used to capture the user's movement. One tracker is mounted on the HMD to capture head movement, and another tracker is mounted on the right hand for capturing hand movement. Head orientation is used to rotate the view of the user in the VR world, and right hand movement is used for navigating the pointer in the world. 5DT Data glove [2] is worn on the right hand for capturing the hand gestures, which are mapped to different modes of operation for manipulating three-dimensional objects. Open Sound Control (OSC) [3] is used for internal communication between different processes within a computer as well as for communication between processes in different machines.

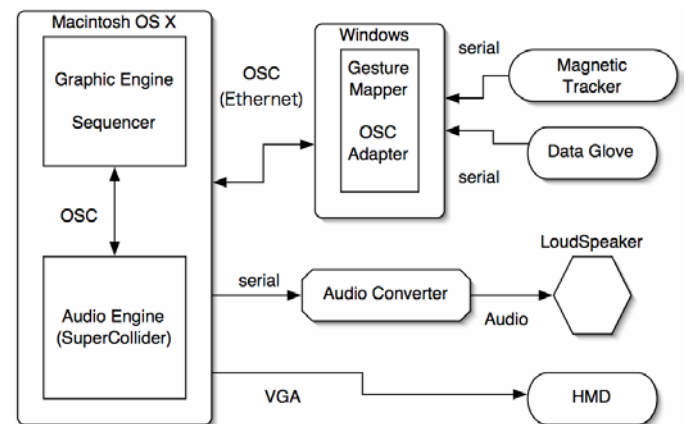


Fig. 1. System diagram of OnKai. Gesture input is handled by Windows machine, and graphics and audio outputs are handled by Macintosh OS X machine. OSC is used for internal communication between different processes within a computer as well as for communication between processes in different machines.

System designed consists of five main software components written in C++ and SuperCollider; gesture mapper, graphic engine, sequencer, and audio engine. Gesture mapper runs in Windows machine and graphic engine, sequencer and audio engine runs on Macintosh OS X machine. The choice of separating the system in two machines is due to lack of driver support for magnetic tracker and data glove for Macintosh which author has experience in.

Gesture mapper takes inputs from magnetic trackers and data glove and converts the hand gesture, head and hand positions to OSC messages and sends them to the sequencer in Macintosh over Ethernet connection. It also converts hand gestures from raw finger flexions to one message, such as “only index finger being bent”.

Graphic engine is implemented using OpenSceneGraph 3D graphical toolkit [4]. It takes OSC message containing user gestures and updates the display of the HMD. Three-dimensional objects are created and controlled by user commands and displayed accordingly.

Sequencer is storage of parameter change in time. It stores parameters determined by the user for playback. When user requests playback, the data stored are sent to graphics engine to update the display as well as to audio engine via OSC to control the audio parameters.

Audio engine is implemented using SuperCollider audio programming environment [5]. It takes OSC messages from sequencer and synthesizes audio controlled by it.

III. RESULTS AND OBSERVATION

In the current implementation, user can manipulate the orientation and location of a sphere and a cube as well as the position of its surface vertexes. Additive audio instrument with twenty sine oscillators connected to panner and reverberator are controlled by two objects. Location of the sphere in space controls the pan of the instrument, and radius controls the reverberation time. One side of the cube consists of 20 x 20 vertex points which user can manipulate its position in x-y-z coordinates. Out of six faces of the cube, side one is used to control the frequency ratio of sine oscillators by its vertex position. Side two is used to control the amplitude of sine oscillators. Side three is used to manipulate the triggering of the instrument. Side four is used to manipulate the velocity of the instrument upon triggering. Observation by experimenting with the system gave interesting result, that it has demonstrated the compactness of the parameter representation but lacks the intuitiveness for manipulating the object. This is due to its lacking feature of the system for manipulating object surface more intuitively, other than vertex position manipulation.

IV. CONCLUSION

System for sculpting three-dimensional object for control in computer music composition has been developed. It has

demonstrated the benefit of representing instrument parameters compactly. Current system needs improvement for object surface manipulation function, other than per vertex manipulation. Due to its initial development stage, further refinements and improvements are necessary. However, it showed a path for alternative method for organizing and manipulating parameters for computer music composition.

REFERENCES

- [1] “Ascension Technology Flock of Birds Manual”, ftp://ftp.ascension-tech.com/PRODUCTS/FLOCK_OF_BIRD/Flock_of_Birds_Manual-RevB.pdf, 2002
- [2] “5DT Data glove 5 Manual”, <http://www.5dt.com/downloads/5DTDataGlove5Manual.pdf>, 2000
- [3] M. Wright, A. Freed, A. Momeni, “Open Sound Control: State of the Art 2003,” *Proceedings of the 2003 Conference on New Interfaces for Musical Expression (NIME-03)*, Montreal, Canada, pp. 154-159, 2003.
- [4] “OpenSceneGraph,” <http://openscenegraph.sourceforge.net/>, 2004
- [5] J. McCartney, “Rethinking the Computer Music Language: SuperCollider,” *Computer Music Journal*, vol. 26, no. 4, pp. 61-68, 2002.



Satoshi Morita is currently in the Media Arts and Technology program at UCSB working towards a M.S. in Multimedia Engineering (expected graduation in 2005). His IGERT advisor is Stephen Pope and faculty advisor is Curtis Roads. Satoshi Morita's research interests are human-sound interface design and audio signal processing.

Warehousing and Integration of Biological Databases

Kevin Hawkins

Undergraduate Research Associate

Computer Science

Abstract—The amount of data available to biologists is growing exponentially. This data resides in many databases with differing content and semantics. Scientists need a way to use all relevant information from these databases to help determine what proteins or genes might be interesting to use in an experiment. This emphasis on data-driven research creates the need to integrate the available databases to allow querying across multiple heterogeneous datasets. I have created a data warehouse composed of local copies of the protein-protein interaction databases MINT, DIP, and BIND. The warehouse has a global schema that allows users to formulate a query without having knowledge about the specific databases involved. A web interface provides an easy way to query the data warehouse. Results returned to the user may be a combination of information from different sources. The interface identifies to the user the source of each result, and when applicable a confidence value for the result is also displayed.

Index Terms—data warehouse, data integration, protein-protein interaction, biological database

I. INTRODUCTION

TO further understand molecular biology biologists are studying how proteins interact within a cell. The process to create a protein begins with a piece of DNA, called a gene, being transcribed into mRNA within the nucleus of the cell. The mRNA, which contains the instructions for how to build a particular protein, is then translated into a protein outside the nucleus. Proteins are responsible for the majority of the work that takes place in the cell. For this reason biologists are interested in not just a single protein but in the network of interactions that take place in order for certain processes to occur (i.e., programmed cell death).

To study protein interactions biologists use a wide variety of experimental methods, including: affinity chromatography, coimmunoprecipitation and two-hybrid system. [1] Using these experimental methods biologists have created databases containing protein-protein interactions. These datasets are heterogeneous, with differing content and semantics, because research groups may conduct slightly different experiments or refer to proteins by different names. This means that there is overlap between the datasets and some databases are missing information that others have already collected. To integrate these protein-protein interaction databases would be very useful to biologists.

Although data integration has been studied for many years, biological data still provides many challenges. Some techniques that have been used to integrate biological data include: link-driven federation, view integration and warehousing. My research aims to learn about the two integra-

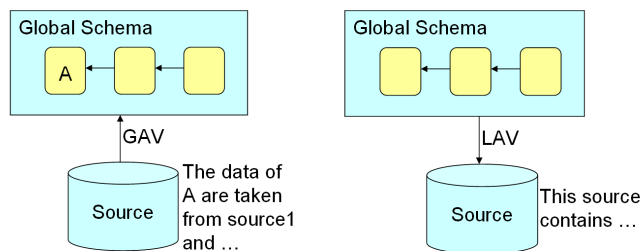


Fig. 1. Global-as-view (GAV), defining the global view in terms of the source vs. local-as-view (LAV), defining the sources in terms of the global view

tion techniques global-as-view and local-as-view and how these techniques can be used to create a data warehouse of biological data.

II. APPROACH

The three protein-protein interaction databases that I used are MINT, DIP, and BIND. MINT (a Molecular Interactions database) [2] is from Rome University and contains 15,570 experimentally verified protein-protein interactions involving 7,430 proteins. DIP (Database of Interacting Proteins) [3] is from the University of California, Los Angeles and contains 44,150 interactions with 17,047 proteins. BIND (Biomolecular Interaction Network Database) [4], from the Samuel Lumenfeld Research Institute at Mount Sinai Hospital affiliated with the University of Toronto, contains 44,275 interactions and 27,988 proteins.

The first decision I had to make is which integration technique to use, global-as-view or local-as-view. Global-as-view (GAV) [5] defines the global view in terms of the sources, describing exactly where to find each field from the global view within the sources. Local-as-view (LAV) [6] is the opposite. The sources are defined in terms of the global view. LAV tells exactly what is in the source but not necessarily how the global field is derived from the source. In order to get the global field some query planning is required. I chose GAV as a starting point to discover how it would work with biological data.

Another decision is whether to materialize the global view. A view is materialized if the data is actually put into tables which can in turn be queried, and non-materialized if the global view definitions are used to find the data at query time. The difference between the two is that materialized takes more space and non-materialized is not as

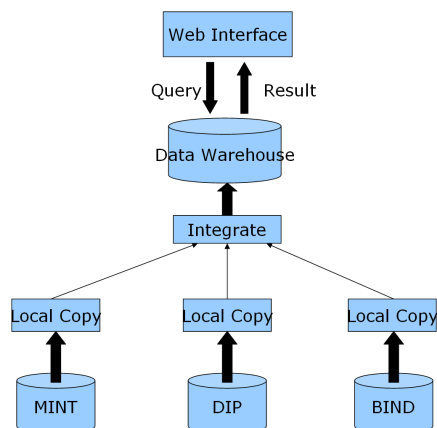


Fig. 2. General plan to create data warehouse

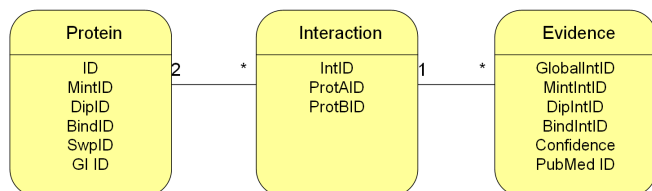


Fig. 3. Global schema for data warehouse

efficient to process a query. I use materialized view for efficiency and to make the query planning easier.

As a first step to create a data warehouse I made local copies of each dataset. The process consisted of downloading the information from the internet in flat file format (tab or pipe delimited) or XML format, parsing the protein-protein interaction information from the file and then putting that information into tables for each dataset. After inspecting the information contained in each dataset I created a general protein-protein interaction database schema for use as a global schema.

The global schema has a table for proteins, a table for interactions and a table for experimental evidence of the interaction. All information from the sources is retained within the global schema, including the original source ID's which help track the origin of the data.

Each dataset contains at least one protein identifier from general protein databases. MINT has SwissProt [7] ID's associated with its proteins, BIND has GenBank [8] ID's associated with its proteins, and DIP has SwissProt ID's and GenBank ID's associated with its proteins. Using these protein identifiers, it is possible to find where the proteins from the three datasets overlap. Using joins in MySQL, I found all the overlaps and put the proteins into a table, "globalproteins", such that no protein is repeated, each has a new unique global identifier and each protein has a "mintid", "dipid", and "bindid" field, which is null if the protein is not from that database and not null if it

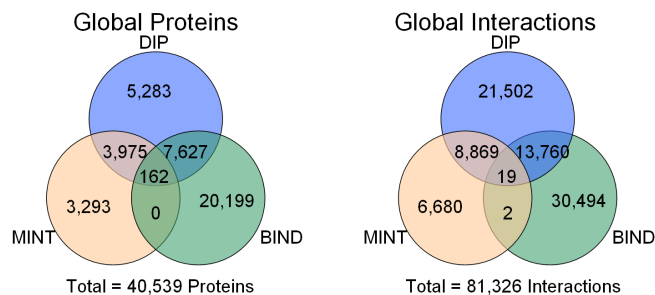


Fig. 4. Protein and interaction overlap

comes from that dataset. I found that when joining tables on fields that are indexed the MySQL performance is acceptable.

Once all the proteins are mapped to a unique global ID it is easy to tell where the interactions in the datasets overlap. The proteins involved in an interaction are mapped to their unique global ID and if two interactions involve the same two proteins then they are considered the same. Using this logic a MySQL query can put all the interactions into the table "globalints", where each is given a unique global interaction ID and the proteins are stored as their global ID. In order to define this transformation the query requires an OR in the WHERE clause, in order to check if ((interaction1.protein1 = interaction2.protein1 AND interaction1.protein2 = interaction2.protein2) OR (interaction1.protein2 = interaction2.protein1 AND interaction1.protein1 = interaction2.protein2)). I found that MySQL currently does not use indexes with the OR in the WHERE clause and is very inefficient, so I had to use two queries and UNION them as a workaround.

With the proteins and interactions integrated all that is left is the experimental evidence for the interactions. Here is where I found some problems with biological data integration. I found that each dataset has different evidence fields for the protein-protein interactions, data is missing from many entries, and sometimes an interaction is listed multiple times with slightly different changes in experimental variables. These factors made it hard to find where the evidence overlapped and I ended up including all evidence for an interaction without integrating it.

After implementing the data warehouse using global-as-view I investigated how local-as-view would have worked for these particular datasets. I found that the queries to add datasources would be less complex and already existing definitions would require no updating. In comparison, global-as-view definitions are very complex, even with only three datasets, and require updating if a new source is added. On the other hand, with global-as-view it is very easy to answer a global query while local-as-view requires difficult query planning.

III. VISUALIZATION

The data contained within the data warehouse could be useful to many biologists but an easy way to access the

Interface

[Data Source](#) | [Index Structure](#) | [Tool](#) | [Data Integration](#) | [Information Integration \(BN\)](#) | [Information Integrati](#)

Global schema name:

Tables:

☐ foo
☐ globalbindevidence
☐ globaldipevidence
☐ globalints
☐ globalmintevidence
☐ globalproteins
☐ globaltax
☒ intersectdipbind
☐ intersectmintdip
☐ intersectmintdipbind

Global-as-view transformations:

```
GlobalProteins(mintid, dipid, swpid, giid) =
select MintProteins.id, DipProteins.id, DipProteins.swpid, DipProteins.giid
from DipProteins left join MintProteins on(swpid)
union
select MintProteins.id, DipProteins.id, MintProteins.swpid, DipProteins.giid
from MintProteins left join MintProteins on(swpid) where DipProteins.id is
NULL;
```

Fig. 5. Creating a global view with the general interface

data is necessary. To do this I created two types of web based interfaces. The first is an interface specific to the data warehouse that I made. It allows the user to query for a certain protein using an assortment of ID's, including taxonomy, SwissProt, GenBank, etc. A protein search can also be constrained to proteins that are in a certain combination of source databases. For example a search can be constrained to all proteins in MINT AND DIP AND BIND or any other boolean combination of the three. This constraint can be used in conjunction with the ID search or separately. Then a list of proteins matching the search criteria are displayed. From here the user may select a protein to get more information about it and a list of the proteins it interacts with. From the protein page an interacting protein can be selected to find out more about that protein or an interaction can be selected to display all available evidence that the two proteins interact.

The second interface is more general. The idea is for this interface to be a place to manage global schemas and local sources and to choose which sources to integrate and how to integrate them. It allows the user to add a data source by indicating a name and which tables are included. It also allows data integration by defining global schemas in the same manner as the sources, by giving a name, which tables to use and the transformations (GAV and/or LAV) that define the global schema. This is a work in progress and in the future would hopefully allow querying across any of the sources or integrated views. Currently it implements the specialized query interface described previously for querying the global view.

Considering that biologists are not just interested in single proteins, the question arises of how to visualize networks of protein-protein interactions in a useful way. McCloskey et al. [9] have found 121 proteins that may suppress programmed cell death. They would like to find other proteins that may be involved in cell death. The problem is that experiments to find these proteins are time consuming and expensive, so it is impossible to test all possible proteins. Instead they depend on data driven research to

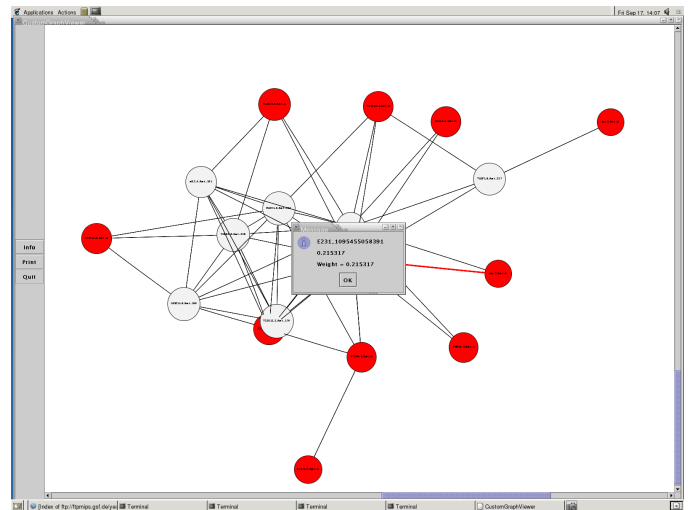


Fig. 6. Confidence information for an interaction within the network

help decide which proteins might be interesting to test experimentally. They would like to integrate protein-protein interaction data from multiple sources to create a network of interacting proteins. This network can then be used to discover untested proteins that cluster with the known proteins. In order to help with this I added interactivity to a graph viewing program. I made a program to display a graph of protein-protein interactions where a protein is shown as a node and two nodes have an edge if they interact. Then the nodes are colored according to whether they are known or not. Displaying the graph can help detect clusters and also allows interactivity. By clicking on an edge you can find more information about which sources provide evidence for the interaction and clicking on a node will take you to WormBase, a specific biological database for the worm *C. Elegans*, and look up the corresponding protein.

IV. CONCLUSION

As a result of integrating the three biological datasets MINT, DIP, and BIND, I created a materialized data warehouse using global-as-view containing 40,539 proteins and 81,326 interactions. This data warehouse has an interface to help biologists find interesting proteins and protein-protein interaction information. Some of the main points that were discovered from this research are the following. The transformations for global-as-view are quite complex and the complexity increases with the number of sources being integrated but the global query is straight forward to answer. With local-as-view it is easier to add a source but answering global queries is more difficult. The MySQL database engine is currently not efficient enough for biological data integration as shown by the problems experienced with OR's in the where clause of a SELECT and not having index structures on sub-queries. Also, experimental data is difficult to integrate because it contains many variables that differ slightly between experiments and the same variables are not used between datasets.

V. ACKNOWLEDGEMENTS

I would like to thank my mentor Vebjorn Ljosa and my advisor Ambuj Singh.

This research was supported in part by the National Science Foundation under grant EIA-0080134 and by the National Science Foundation Integrative Graduate Education and Research Traineeship (IGERT) program under grant no. 022713.

REFERENCES

- [1] Eric M. Phizicky and Stanley Fields, "Protein-protein interactions: methods for detection and analysis," *Microbiol.Rev.*, vol. 59, no. 1, pp. 94–123, mar 1995.
- [2] Andreas Zanzoni, Luisa Montecchi-Palazii, Michele Quondam, Gabrielle Ausiello, Manuela Helmer-Citterich, and Gianni Cesareni, "MINT: a molecular interaction database," *FEBS Letters*, vol. 513, no. 1, pp. 135–140, feb 2002.
- [3] Lukasz Salwinski, Christopher S. Miller, Adam J. Smith, Frank K. Pettit, James U. Bowie, and David Eisenberg, "The database of interacting proteins: 2004 update," *Nucleic Acid Research*, vol. 32, no. 1, pp. D449–51, sep 2004.
- [4] Gary D. Bader, Doron Betel, and Christopher W.V. Hogue, "BIND: the biomolecular interaction network database," *Nucleic Acids Research*, vol. 31, no. 1, pp. 248–50, Jan. 2003.
- [5] Sudarshan Chawathe, Hector Garcia-Molina, Joachim Hammer, Kelly Ireland, Yannis Papakonstantinou, Jeffrey D. Ullman, and Jennifer Widom, "The TSIMMIS project: Integration of heterogeneous information sources," in *16th Meeting of the Information Processing Society of Japan*, Tokyo, Japan, 1994, pp. 7–18.
- [6] Alon Y. Levy, Anand Rajaraman, and Joann J. Ordille, "Querying heterogeneous information sources using source descriptions," in *Proceedings of the Twenty-second International Conference on Very Large Databases*, Bombay, India, 1996, pp. 251–262, VLDB Endowment, Saratoga, Calif.
- [7] Amos Bairoch and Rolf Apweiler, "The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000," *Nucleic Acids Research*, vol. 28, no. 1, pp. 45–48, 2000.
- [8] Dennis A. Benson, Ilene Karsch-Mizrachi, David J. Lipman, James Ostell, Barbara A. Rapp, and David A. Wheeler, "GenBank," *Nucleic Acids Research*, vol. 28, no. 1, pp. 15–18, 2000.
- [9] McCloskey et al. Personal communication.



Kevin Hawkins is an undergraduate at UCSB working towards his BS in Computer Science with an expected graduation date of Spring 2005. His graduate mentor is Vebjorn Ljosa and his advisor is Ambuj Singh.

MATConcat: An Application for Exploring Concatenative Sound Synthesis Using MATLAB

Bob L. Sturm

IGERT Fellow

Electrical and Computer Engineering

Abstract—The author has developed an application in MATLAB implementing concatenative sound synthesis (CSS) using feature matching. CSS is a process of combining short pieces of recorded sound to construct new sonic forms. Historically, CSS was developed for text-to-speech synthesis, but recently it has been explored as a musical sound synthesis method. The results have been called ‘musaics,’ the sonic analogue to mosaics made from small pieces of colored tile. Though this MATLAB application is less sophisticated than other CSS algorithms, it is meant to be a free and open application for demonstrating and experimenting with the process. The author has used this application to create many interesting and entertaining sound examples, as well as compositions. The application, and many sound examples, are available at <http://www.mat.ucsb.edu/~b.sturm>.

Index Terms—Concatenative sound synthesis, feature extraction, feature matching, composition

I. INTRODUCTION

A method exists in the synthesis of computer speech, called concatenative synthesis [1]. This technique, developed in the early sixties, is used mostly for text-to-speech synthesis. A computer segments written text into elementary spoken units that are synthesized using a large database of sampled speech sounds, like “ae”, “oo”, “sh”. These components are pieced together to obtain a synthesis of the text. These methods have recently been applied to creating “audio mosaics,” or “musaics” [2], [3], [4], but instead of using written text, they use recorded sound. As in mosaicing, a ‘target’ sound is approximated by small sound samples from a ‘corpus.’

Creative application of concatenative sound synthesis (CSS) has been minimal, and software for exploring it is not available. The author thus created an application to explore this technique for composing music. *MATConcat* is an implementation of CSS in MATLAB, a powerful but slow mathematics software language. With *MATConcat* a sound or composition can be concatenatively synthesized from audio segments in a database of any size. CSS provides many interesting and unique possibilities for sound design and electroacoustic composition. I have used it to create several intriguing sound examples, as well as electroacoustic compositions. These demonstrate the potential of this technique for sound synthesis.

II. ALGORITHM

The techniques used in *MATConcat* are very simple. Figure 1 displays the algorithm used. An analysis of the target, or the sound being approximated, produces feature vectors for ‘frames’ taken by sliding a user-specified win-

dow across the audio by a constant hop size. A six-element feature vector is created for each frame of the sound. Table I shows the current dimensions of the feature vector and interpretations of each component.

Feature Measure	Meaning of Feature
Number of Zero Crossings	General noisiness, existence of transients
Root Mean Square (RMS)	Mean acoustic energy (loudness)
Spectral Centroid	Mean frequency of total spectral energy distribution
Spectral Drop-off	Frequency below which 85% of energy exists
Harmonicity	Deviation from harmonic (integral) spectra
Pitch	Estimate of fundamental frequency

TABLE I

CURRENT FEATURE VECTOR ELEMENTS

The analysis database of sound used for the synthesis is called the corpus, which can be several seconds to hours long. The analyzed sound being approximated is called the target. Iterating through the frames of the target analysis, optimal matches are found in the corpus database using specified matching parameters and thresholds. Other options can be specified, like forcing a match or selecting one at random, or extending the previous match if none is found. These numerous matching criteria and synthesis options creates many different possibilities.

III. EXAMPLES

Several intriguing sound examples have been created so far. The dramatic percussion crescendi from Gustav Mahler’s second symphony have been synthesized using corpora of monkey and animal sound effects, a Muslim Imam chanting the Koran, an hour of vocal music by John Cage, three hours of nostalgic Lawrence Welk, and all four string quartets of Arnold Schoenberg.

The example using the monkey vocalizations, shown in Figure 2, is particularly amazing. In this example the RMS and spectral roll-off feature elements are matched. The slowly building crescendo is ‘aped’ by the monkeys, creating a sense of increasing hysteria. At the climax the dominant gorillas grunt as lesser monkeys cower in submission.

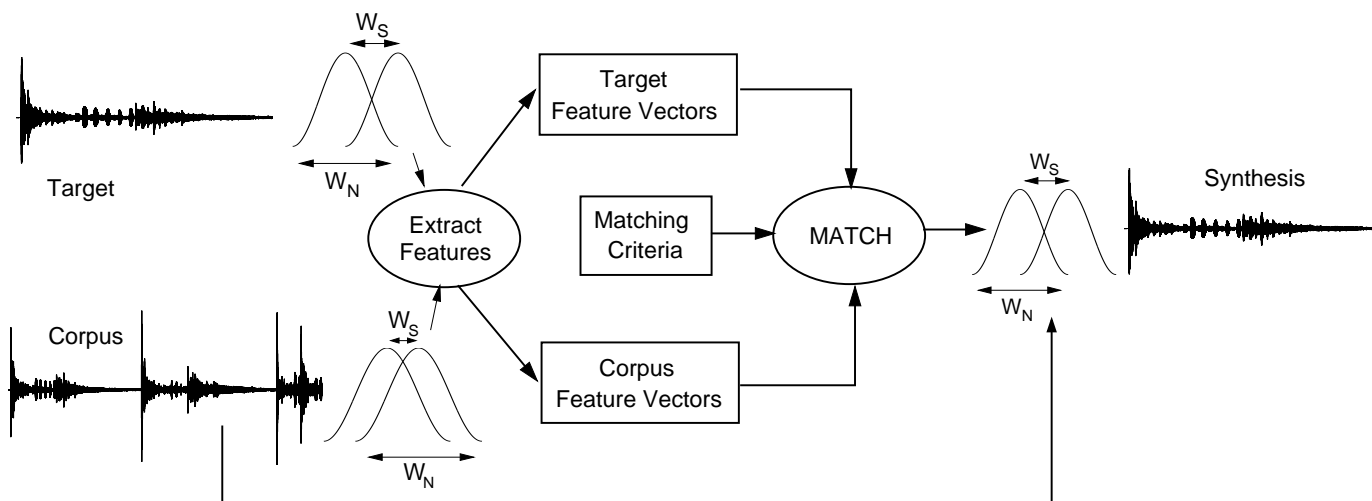


Fig. 1. Algorithm of MATConcat. A target sound is analyzed by a constant hop-size window, creating for each frame a unique feature vector that identifies its characteristics (figure 1). The best match for each frame in the target is found in the collection of sounds called the corpus—analyzed in the same manner—based on criteria specified by the user.

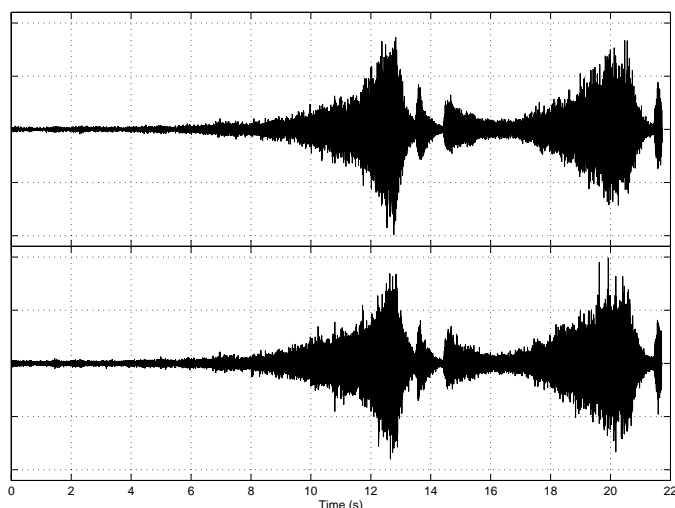


Fig. 2. Mahler's crescendi performed by London Symphony Orchestra (Gilbert Kaplan, cond.) (top), and performed by ensemble of monkeys (bottom).

Synthesizing the same target using the same matching criteria, but from a corpus of John Cage's vocal music, creates an entirely different sonic experience. The impressions of Mahler's crescendi remain however.

IV. CONCLUSION

Through the sound examples and compositions created, *MATConcat* demonstrates that this relatively simple implementation of CSS creates effective and intriguing sound and music material. *MATConcat* serves well as a massive sample-mill, grinding sound into minuscule pieces for reconstitution into familiar forms. Surely with machine listening and score analysis, other interesting possibilities will emerge; but currently this implementation of CSS is far from being exhausted.

As it stands *MATConcat* is prototype software. To take advantage of the numerous possibilities of CSS, this application will be ported to C++. Many improvements can be made including increased speed, increasing the dimensions of the feature vector, and expanding the list of synthesis options. Future work will implement the features of the MPEG-7 audio framework standard [5], and working toward a real-time implementation. These extensions will further open up the interesting avenues for creative CSS.

MATConcat, and many sound examples, are available for free at <http://www.mat.ucsb.edu/~b.sturm>.

REFERENCES

- [1] A. Hunt and A. Black, "Unit selection in a concatenative speech synthesis system using a large speech database," *ICASSP*, vol. 1, no. 1, pp. 373–376, 1996.
- [2] A. Lazier and P. Cook, "Mosievius: Feature driven interactive audio mosaicing," in *Proc. of the 6th Int. Conference on Digital Audio Effects (DAFx-03)*, 2003.
- [3] D. Schwarz, "A system for data-driven concatenative sound synthesis," in *Proc. of the COST G-6 Conference on Digital Audio Effects (DAFX-00)*, 2000.
- [4] A. Zils and F. Pachet, "Musical mosaicing," in *Proc. of the COST G-6 Conference on Digital Audio Effects (DAFX-01)*. University of Limerick, 2001.
- [5] J. Martínez, "Mpeg-7 overview," <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm>, 2003.



Bob L. Sturm has presented his research and music all around the world. In 1999 he earned an M.A. from Stanford University in Computer Music Technology, and in 2004 he earned an M.S. in Multimedia Engineering from UCSB. He is currently working on a PhD in Electrical and Computer Engineering at UCSB. His IGERT advisors are Stephen Pope and Dr. Curtis Roads.

A Real-Time Application Demonstrating the Interaction of Atmosphere and Ocean Using Sound and Image

Bob L. Sturm
IGERT Fellow

Electrical and Computer Engineering

August Black
IGERT Fellow

Media Arts and Technology

Abstract—Using sound and image our application *Pacific Pulse* presents a truly unique way of perceiving the interaction between atmosphere and ocean along the Pacific coast of the United States. Accompanying animated satellite imagery in three wavelengths are sonifications of data collected by ocean buoys along the Pacific Coast of the United States. This novel interface takes advantage of both the visual and auditory senses for presenting a large mass of data. Some of the phenomena that can be heard are the low rumbles of swell energy, the high-frequency effects of afternoon winds, and the gradual passing of wave-trains that originate from distant storms. This is an exciting multimedia application for use in science education and public outreach. So far it has been implemented in a kiosk environment in the Institute for Genetic Medicine at the University of Southern California, and the physics building at the University of California, Santa Barbara.

Index Terms—Sonification, visualization, data display, physical oceanography

I. INTRODUCTION

SINCE 1975 the Coastal Data Information Program (CDIP), within the Scripps Institution of Oceanography (SIO) in Southern California, has measured, disseminated and archived coastal environment data for use by coastal engineers, planners, scientists, mariners, and the military [1]. CDIP operates approximately twenty off-shore and near-shore buoys that monitor ocean conditions including wave height, period, and direction, air and sea temperature, and wind velocity. The multidimensionality of the data and its cyclic wave nature are inviting for new methods for data display, such as sonification.

Sonification, or auditory display, is a parametric representation of data using sound, vis-à-vis a visual or graphic representation [2]. Two remarkable examples of sonification are the Geiger counter, and the heart rate monitor. A Geiger counter gives a qualitative impression of proximate radioactivity; a click is produced for every detected radioactive decay. One can listen to the clicks while walking with the device and note the relative danger in the vicinity. Used in a hospital, a heart rate monitor provides continuous feedback about a patient's status without requiring focused attention or the use of vision. One need not even be in the same room to hear a heart stop beating.

Since 2002 one author (Sturm) has experimented with using sonification for presenting the data measured by deep-water ocean buoys [3], [4]. The results have been very effective for illustrating concepts of physical oceanography, and demonstrating the use of sound for interpreting

data. Combining the sonifications with visual representations of the data is expected to increase the effectiveness of the display.

II. APPLICATION

Pacific Pulse highlights the dynamic conditions and evolution of the mighty Pacific Ocean using sound and image. Datasets measured by buoys along the Pacific coast of the United States are downloaded from CDIP every thirty minutes and turned into sound, providing an aural representation of ocean conditions along the coast. Satellite images in three wavelengths are downloaded every thirty minutes from the National Oceanic and Atmospheric Administration (NOAA) [5] and displayed synchronized with the sound. All data used in *Pacific Pulse* is in the public domain and is accessible from the World Wide Web.

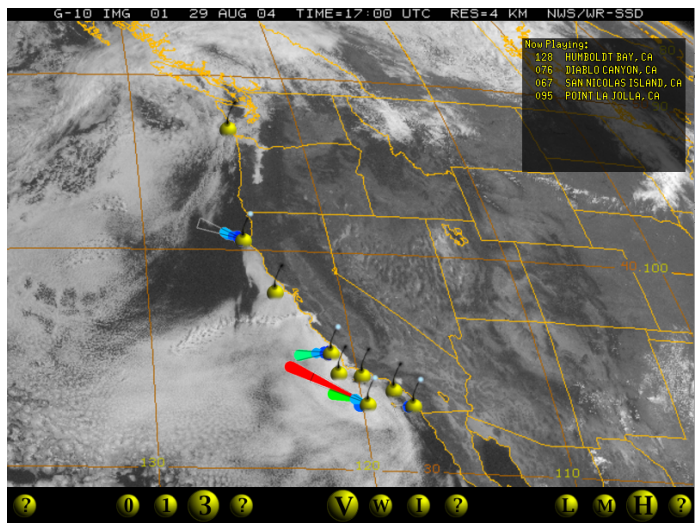


Fig. 1. A screen grab of *Pacific Pulse*. Buoys are shown superimposed on satellite images. Users can click on buoys to hear the most recent data sonifications. Activated buoys show the most recent directional spectrum distribution.

The sonifications are generated from spectral analyses of the buoy time-series data. Each buoy measures accelerations in the vertical and horizontal planes. A spectral analysis of this data reveals the distributions of energies among frequencies in the bandwidth 0.025 – 0.58 Hz, as well as their direction of origin. The sonification maps the

buoy frequencies onto audible frequencies and performs additive synthesis to generate a time-domain signal from any number of spectral records. A much more detailed description of the process is given in [4].

The application, seen in figure 1, has several elements. At the bottom of the screen are buttons that control what is seen and heard. Superimposed on the satellite imagery are the buoys available for data display. A list of “currently playing” buoys is located at the top-right. The buttons marked by a “?” provide information about, from left to right, the application, datasets, the visuals, and the sonifications. Three durations of data are available: three days, one day, and 2 hours, from the most recent time. Three bandwidths of satellite image data are available: visible wavelengths, water vapor, and infrared. Additionally three different sonification mappings can be chosen, each emphasizing a different part of the ocean wave spectra, and each having a different sound quality.

The user interacts with *Pacific Pulse* by using a mouse or touch-screen to click on buttons and buoys. Clicking on a buoy enables its sonification to play synchronized with the satellite imagery. Any number of buoys can be heard together. With three different data periods, one can easily hear the decay of wave-trains over several days, and the effects of afternoon winds over hours. Superimposed on an activated buoy is a flower plot showing the most recent directional distribution of wave energies measured. *Pacific Pulse* runs continuously and updates itself every thirty minutes. Since the conditions along the coast are constantly changing, each visit to *Pacific Pulse* provides a different visual and aural experience.

This application is programmed in C++ using open source audio and graphic libraries. It has been successfully tested on Linux and OS X operating systems.

III. CONCLUSION

Pacific Pulse provides an exciting interactive environment for observing the interaction between the atmosphere and ocean, and could certainly be used by students to learn about physical oceanography. In the past the sonifications have been found to be very useful for pedagogical purposes. On numerous occasions the author (Sturm) have presented them to college students in oceanography and marine science classes using a lecture that introduces principles of physical oceanography with visual and auditory displays of the data. At the end the students are asked to mimic sonifications of various phenomena, and describe why they sound the way they do. Motivated by the sonifications and their novelty most students appear more engaged with the material. A more formal study will be conducted to quantitatively assess improvements in learning the material and interpreting the data.

Pacific Pulse is currently being demonstrated at the Institute for Genetic Medicine at University of Southern California, and in the physics building at University of California, Santa Barbara. In these roles it not only serves as a tool with which a person can observe the current state of and interactions between the atmosphere and Pacific

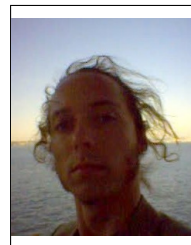
Ocean, it also serves as an attractive illustration of using multimedia to convey complex concepts and display multidimensional data.

REFERENCES

- [1] CDIP, “Coastal Data Information Program Homepage,” <http://cdip.ucsd.edu>, 2004.
- [2] G. Kramer et al., “Sonification report: Status of the field and research agenda, prepared for the national science foundation by members of the international community for auditory display,” Available online at: <http://www.icad.org>, 1999.
- [3] B. L. Sturm, “Music from the Ocean, multimedia CD-ROM,” Composerscientistrecordings: <http://www.composerscientist.com>, 2002.
- [4] B. L. Sturm, “Pulse of an ocean: Sonification of ocean buoy data,” *Leonardo Journal of Arts and Sciences*, forthcoming in 2005.
- [5] NOAA, “National Oceanic and Atmospheric Administration Homepage,” <http://www.nwsa.noaa.gov>, 2004.



Bob L. Sturm worked at CDIP for 16 months as a “data janitor,” constantly cleaning, shuffling, and getting personal with buoy data. His present research includes scientific sonification, science and media arts pedagogy, concatenative synthesis, and composition. Sturm has earned an M.A. from Stanford University in Computer Music Technology, and an M.S. in Multimedia Engineering from UCSB. His IGERT advisors are Stephen Pope and Dr. Curtis Roads.



August Black has an awful habit of calling himself an artist. Previously, this has meant making marks on paper and later on canvas. Now, this means almost anything concerning material, concept, and form. His research is based in the overlap of media, focusing mostly on the kinds of audiences that are created and induced by emerging conventions of observation and involvement.

Notes

Notes
