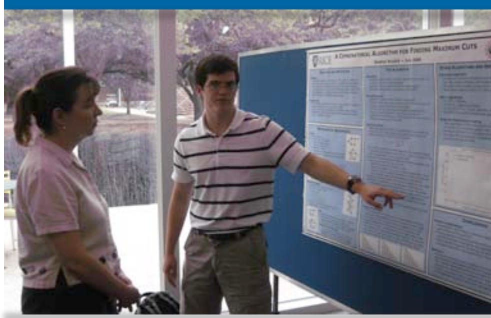
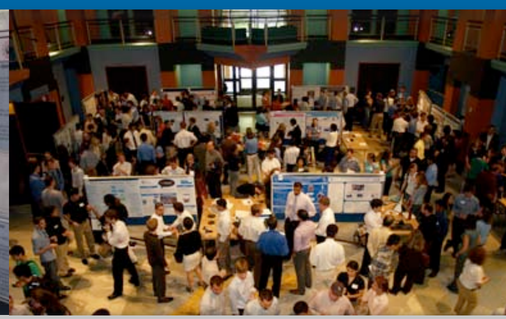
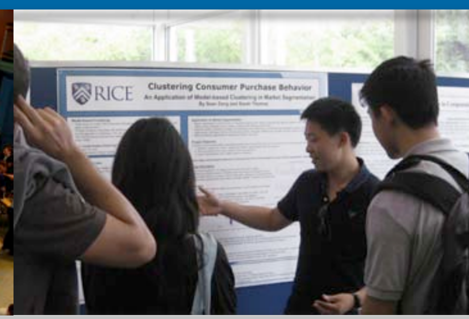
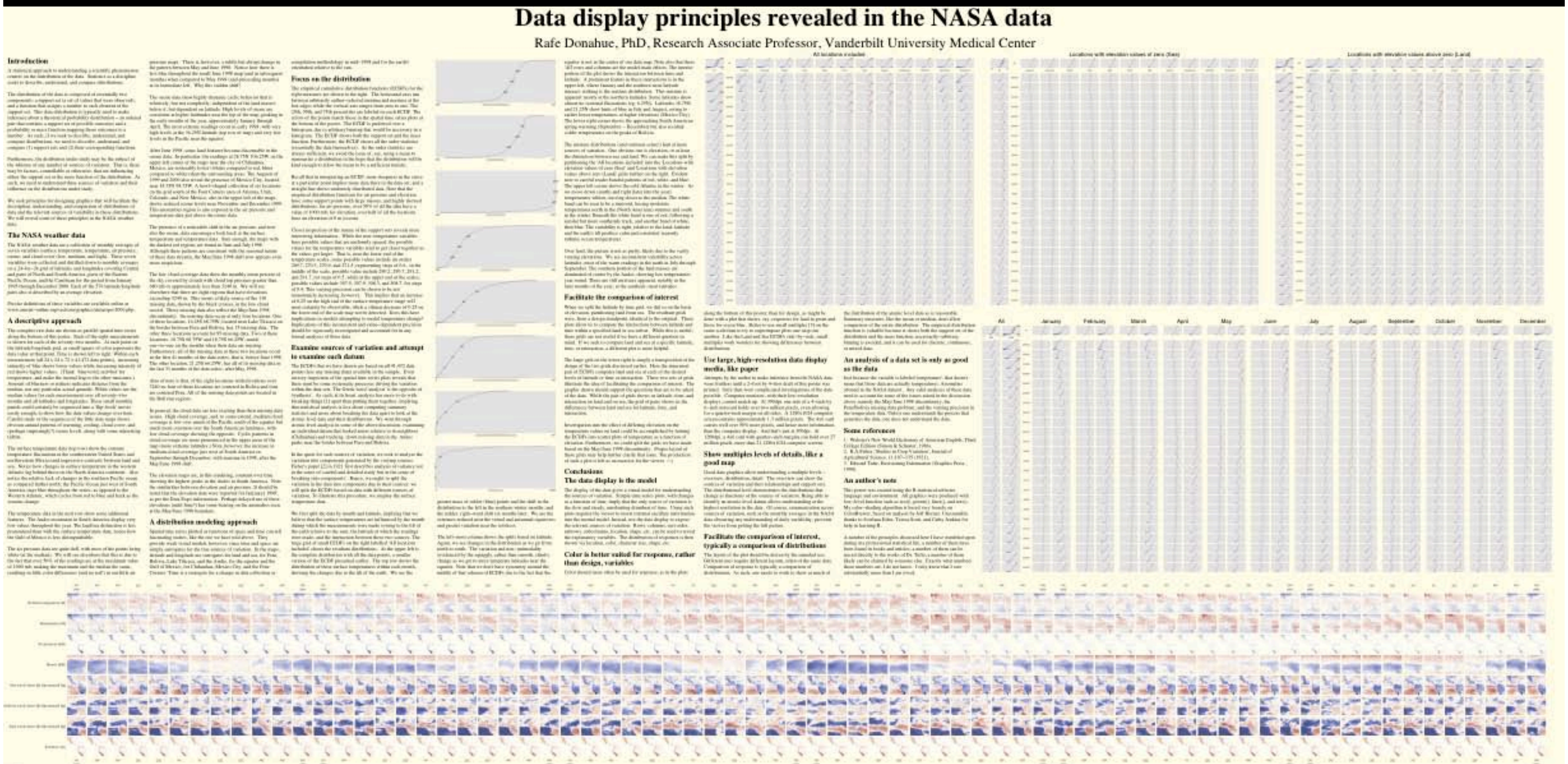
How to Prepare and Present a Poster for STAT 405

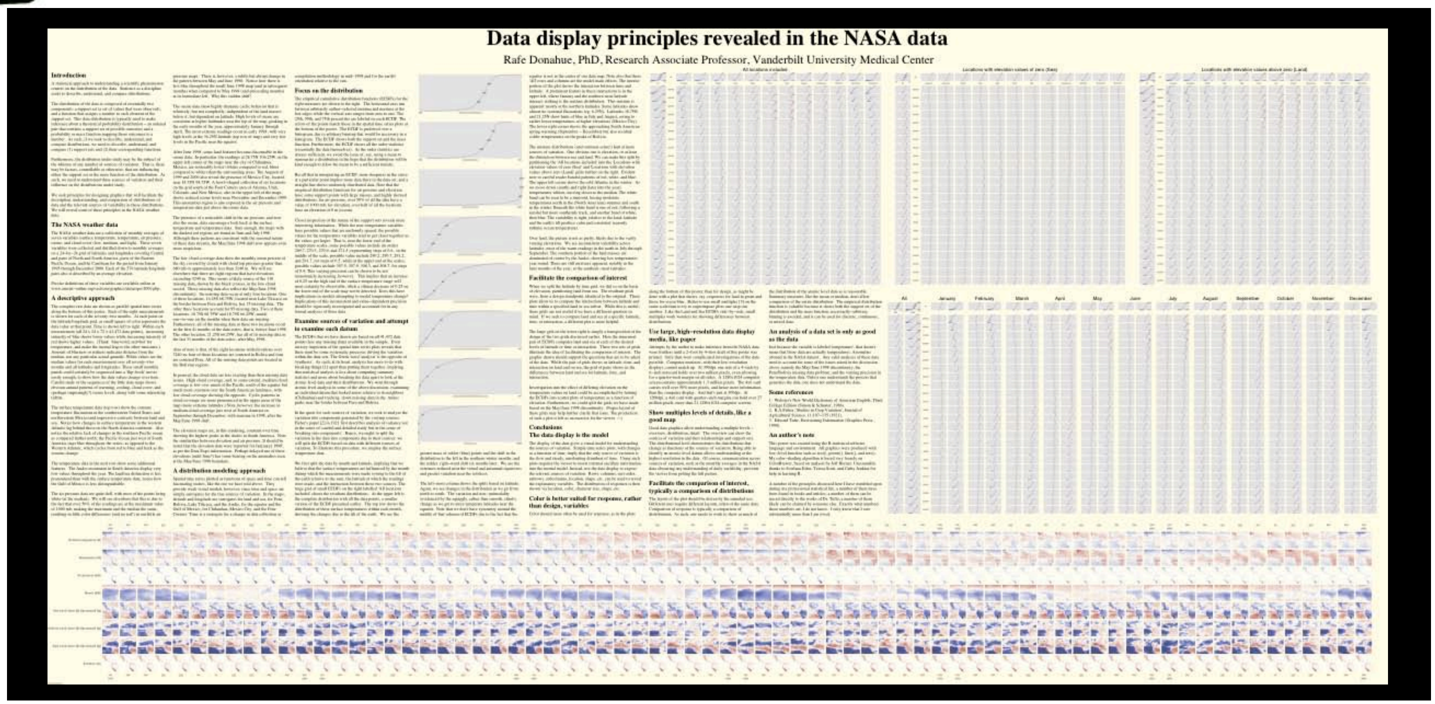
Tracy Volz, PhD
Rice Center for Engineering Leadership
tmvolz@rice.edu

http://math.rice.edu/VIGRE/Images/2009FallPosterAlbum/2009FallPhoto.html

How would you react if you encountered a poster that looked like this?



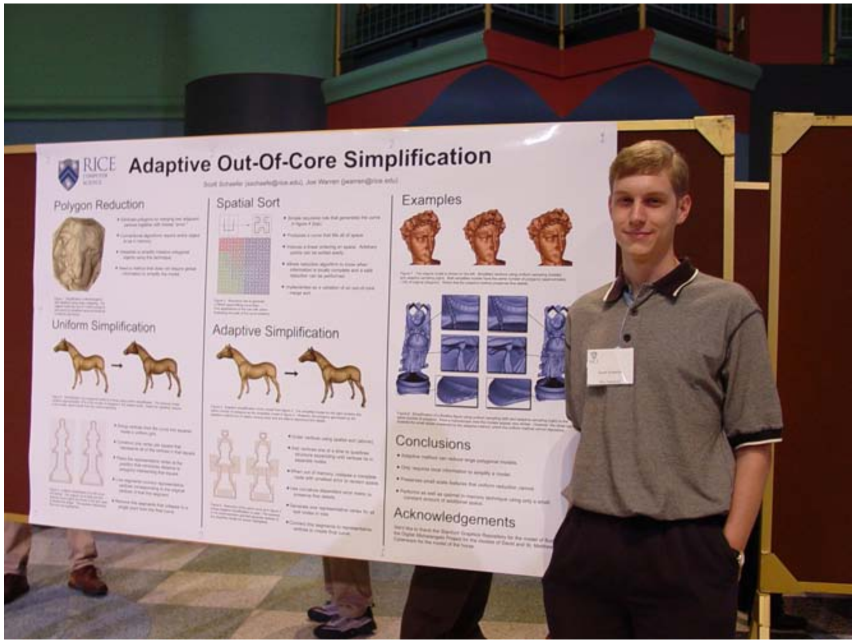
http://stat-computing.org/dataexpo/2006/entries.html



http://stat-computing.org/dataexpo/2006/entries.html

Today's plan

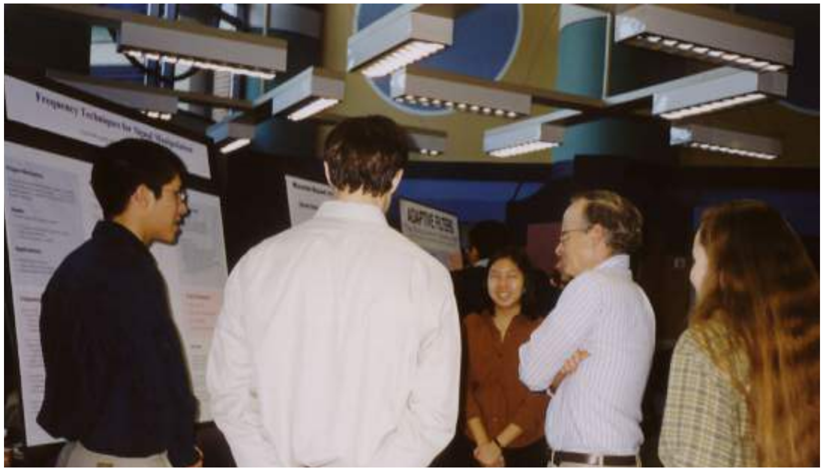
- Review design principles
- Critique posters
- Present poster with confidence



Posters present a challenge

- Audiences
  - Make decisions quickly
  - Come and go
- Posters must be lean and clean
  - Attractive
  - Accessible
  - Comprehensible
- Posters must stand alone

Design for the audience



Why are they interested?
What are their backgrounds?
How will they benefit?
What are their questions?

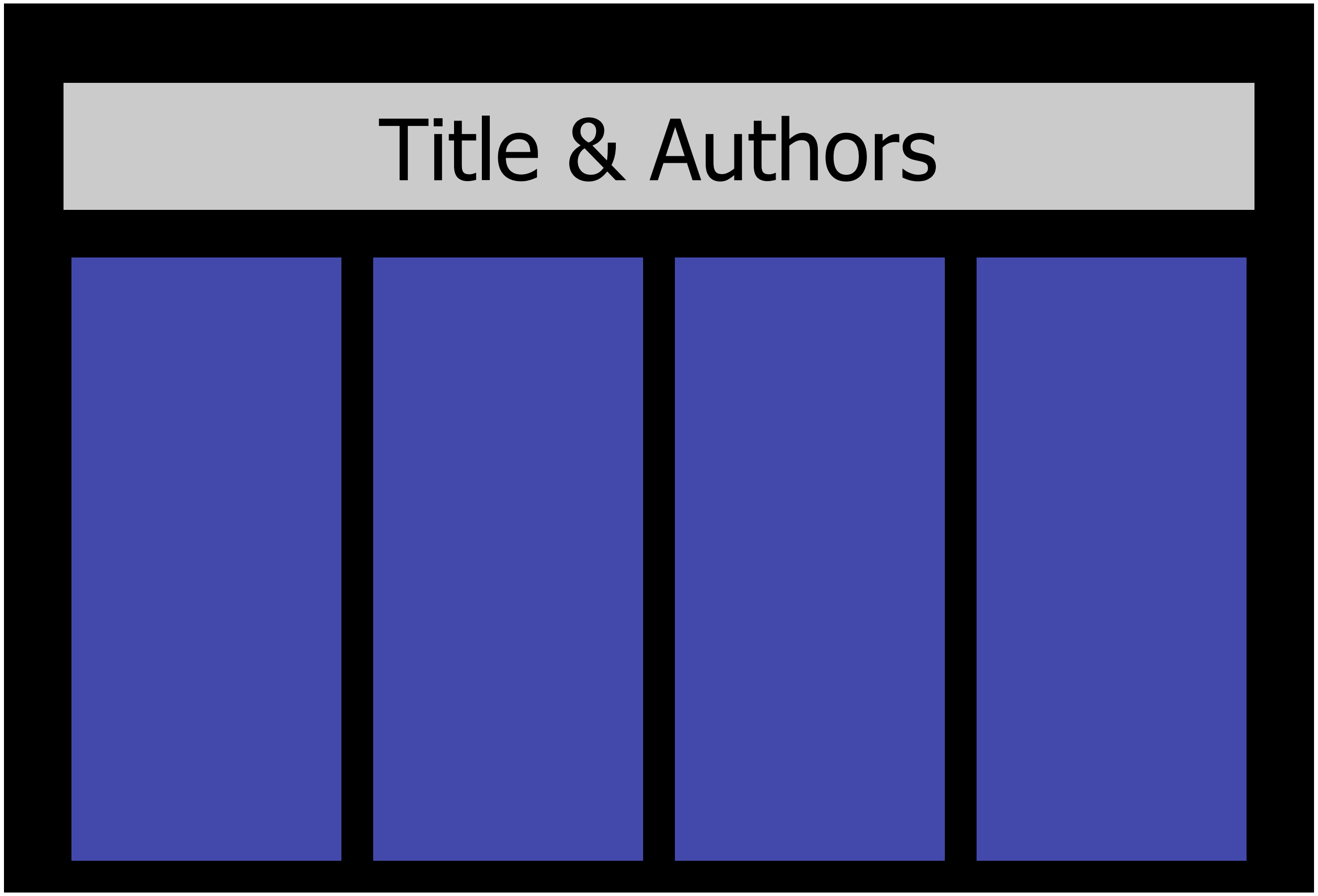
What’s your story?

- Why is your data analysis interesting?
- What data are you exploring? What variables are you analyzing?
- What’s specific question(s) do you answer?
- What’s your approach?
- What are your assumptions?
- What are your key findings?
- How can your work be extended or applied?

Visualize

"the story"

L-to-R flow in vertical columns



Flows top to bottom, left to right

Suggestions for a CEO:

A Statistical Analysis for Optimizing a Company's Resources

Daniel Zales, Chao Pan, Mary Horn

Introduction:

CEOs struggle to optimize a companies resources. Using data from a local company that creates eLearning training modules for the oil industry, we sought to provide insight into three different sectors of the company:
- **Marketing and Advertising**
- **Course Design**
- **Product Management**

Yearly Data (445720 Total Observations)

Data Source: CEO of an eLearning company

Attempts_2007:2010 – 4 sets, 163013 observations – UNITSTATUS_ID, ATTEMPT_ID, QNUM, QCORRECT, and DATE
Unistatus_2007:2010 – 7 data sets, 298340 observations – UNITSTATUS_ID, UNIT_ID, and USER_ID
Units – 1 data set, 1332 observations – UNIT_ID, TITLE, CREATED, and ACTIVE
Users – 1 data set, 10228 observations – USER_ID, CLIENT_ID, and DEMO
Clients – 1 data set, 145 observations – CLIENT_ID, CERTIFICATE, CREATION_DATE and CLIENT
Data sets were joined together by color coded variables.

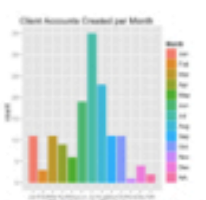
Variable Definitions:

UNITSTATUS_ID – string for each lesson license
UNIT_ID – string to identify lesson module
USER_ID – string to identify user
CLIENT_ID – string to identify client
CERTIFICATE – what certificate was issued
CREATION_DATE – day account was made
QCORRECT – # of answers that were correct

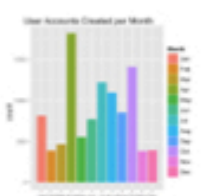
DATE – day of attempt (ymd_hm)
TITLE – name of lesson module
CREATED – date of unit creation
ACTIVE – T/F if unit is still sold
DEMO – if account used the demo
CLIENT – Company
QNUM – # of questions

Findings for Marketing Manager
- Users create accounts on job hunting season.
- Clients open accounts before new workers arrive

Yearly Account Creation Cycle



Most clients (companies) create accounts in the summer

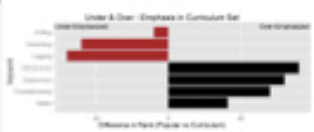


Creation of accounts for employees of clients

- Marketing efforts should be increased in the months leading up to summer so the company can best capitalize on this opportunity.
- Marketing likely does not see a positive return on investment during November.

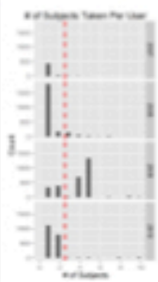
Findings for Course Designer
- Users take a longer time to complete popular units
- Users who attempt the test more than 3 times do not take it seriously

Units Over- and Under- emphasized

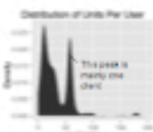


- Keywords are extracted from popular units and curriculum separately
- The difference shows that high demands in advanced skills such as Drilling, Cementing, Logging courses are not met
- The over-provision of supporting courses- such as Maintenance, Troubleshooting, Safety and Assessment courses- is not necessary.

Diverse Units, Focused Subjects

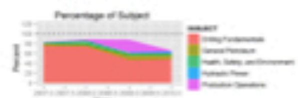


- Users on average take more than 20 units, a number that is usually less than or equal to 2 total courses.
- This shows a **high correlation** within units.

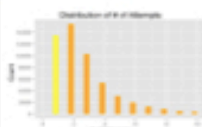


Popular Units and Subjects

- **Popular** units are defined as units with more than 1000 users. Units in the same subject generally cluster together



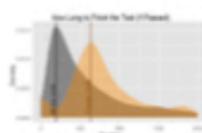
More Attempts, Less Variation



Users usually make more than one attempt to pass the test.

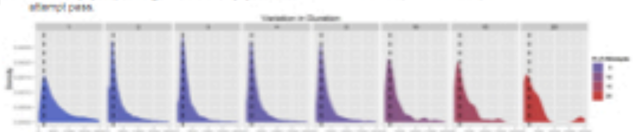
- The mean of the duration between "# of attempts" does not vary much.
- Variance of duration grows smaller from 1-attempt pass to 3-attempt pass, but larger from 3-attempt pass to 20-attempt pass.

More Popular, Longer Duration



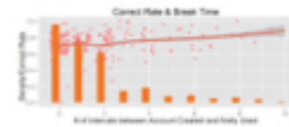
Popular units have more rigorous tests than the average test.

- Clear difference in duration. Popular units are usually time-consuming with a lower number of questions correct (not shown here)



Findings for Product Manager
- Longer break equates to a higher score
- Some exams need improvement as they are too hard

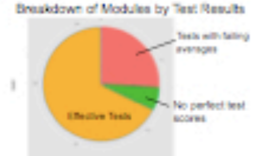
Longer Break Time Shows Higher Scores



- Users who take longer to start courses after signing up perform better
- Users who start the course immediately tend to be too ambitious to perform well.

Effective and Ineffective Exams

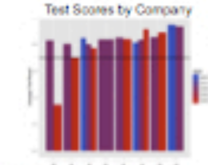
Breakdown of Modules by Test Results



- 18.6% of active modules have a failing (scores < 70%) average on their tests, indicating some tests are too hard
- 4.5% of active modules have never had a perfect score, indicating an error on the answer key
- 77% of active modules have tests that appear to be highly effective

Success of Major Clients

Test Scores by Company



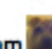
- Company names have been redacted at compan's request
- Black line shows the passing score
- Provides an interesting insight into company culture
- **Company A** was intimately involved with Deepwater Horizon Blowout. Notice it did not require its employees to pass the tests.

Suggestions for Managers:

Marketing:
- Start advertisement & promotion before summer to win during job hunting season
- Focus more on big clients, rather than individual users.

Design:
- Provide an automatic recommendation for users, due to specific units' high correlation.
- Update the curriculum set, add the hot, remove the redundant.
- Add penalty to more than three attempts to discourage taking multiple tests.

Product:
- Make hardest tests easier
- Make sure answer keys are correct
- Worry when a company performs really poorly

Team 

L-to-R flow in rows

ClimaTears Goggle System

Anastasia Alex, Austin Edwards, J. Daniel Hays, Michelle Kerkstra, Amanda Shih
Department of Bioengineering, Rice University
climatears@gmail.com

BCM Baylor College of Medicine
Department of Opthalmology
Cullen Eye Institute

Dry Eye Testing Lacks Standardization

- 4.8 million Americans affected by Dry Eye (DE) disease [1]
- DE caused by tear production deficiency or excessive tear evaporation [2]
- DE research limited by lack of:
  - Concensus on diagnostic criteria and tests that define DE
  - Control of ambient conditions affecting DE test results
- Low-humidity environment induces DE symtoms useful to DE research
- Current method:
  - Costs up to $2 million
  - Has limited locations
  - Exposes entire body to adverse conditions

Mission: To create a low-cost, portable device that controls the relative humidity around the eye while recording blink rate, providing a standardized environment for clinical and diagnostic DE testing.

Design Criteria	
Control Relative Humidity (RH)	15 - 40% ± 3%
Monitor Blink Rate	errors ≤ 5%
Manufacturing Cost	< $1,000
Goggle Weight	< 0.45 kg
Conditioning Pack Weight	< 2 kg
Set-up Time	< 10 min

CTGS Achieves Desiccating Environment and Increases Dry Eye Severity

Device Evaluation

Tests were performed to evaluate the ability of the CTGS to create a localized desiccating environment.

Desired RH Maintained within 3% RH Tolerance

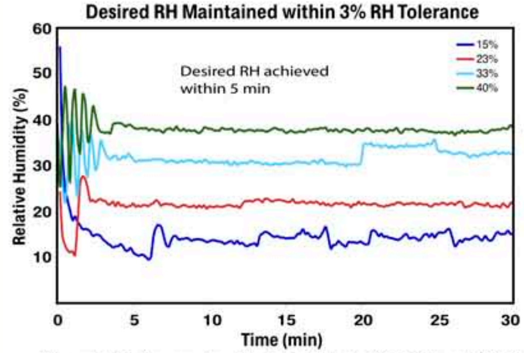


Figure 1. RH in the goggles with desired RH set at 15, 23, 33, and 40% RH.

Clinical Evaluation

Ocular tests before and after CTGS use:
- Visual acuity
- Fluorescein tear film breakup time
- **Corneal topography (SRI) [Fig. 2]**
- Kinetic topography (tear film stability)
- Tear meniscus measurement (OCT)
- Manual tear collection (tear composition)
- **Blink rate at t = 0, 45, 90 min [Fig. 3]**

Testing conditions:
- 5 normal patients, 5 DE patients
- Goggles worn for 90 minutes
- RH set to 23%

Potential confounding factors:
- Small sample size (n<30)
- Age variability amongst subjects
- Whether or not DE subject was on treatment plan at time of test

CTGS Use Increases Eye Surface Irregularity

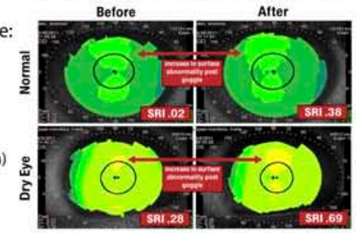


Figure 2. Surface Regularity Index (SRI) before & after CTGS use.

CTGS Use Increases Blink Rate in DE Patients

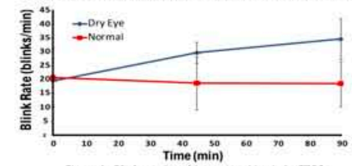


Figure 3. Blink rate in subjects over 90 min in CTGS

Overall Design

ClimaTears Goggle System (CTGS) Controls Relative Humidity

Conclusions and Summary

The ClimaTears Goggle System enables:
- Standardization of testing conditions for both clinical and research applications
- Automatic recording of periocular temperature, relative humidity, airflow and blink rate
- Control of relative humidity between 15 and 40% ± 3%
- Manufacturing cost of <$700 per device compared to >$1 million for the leading competitor's product

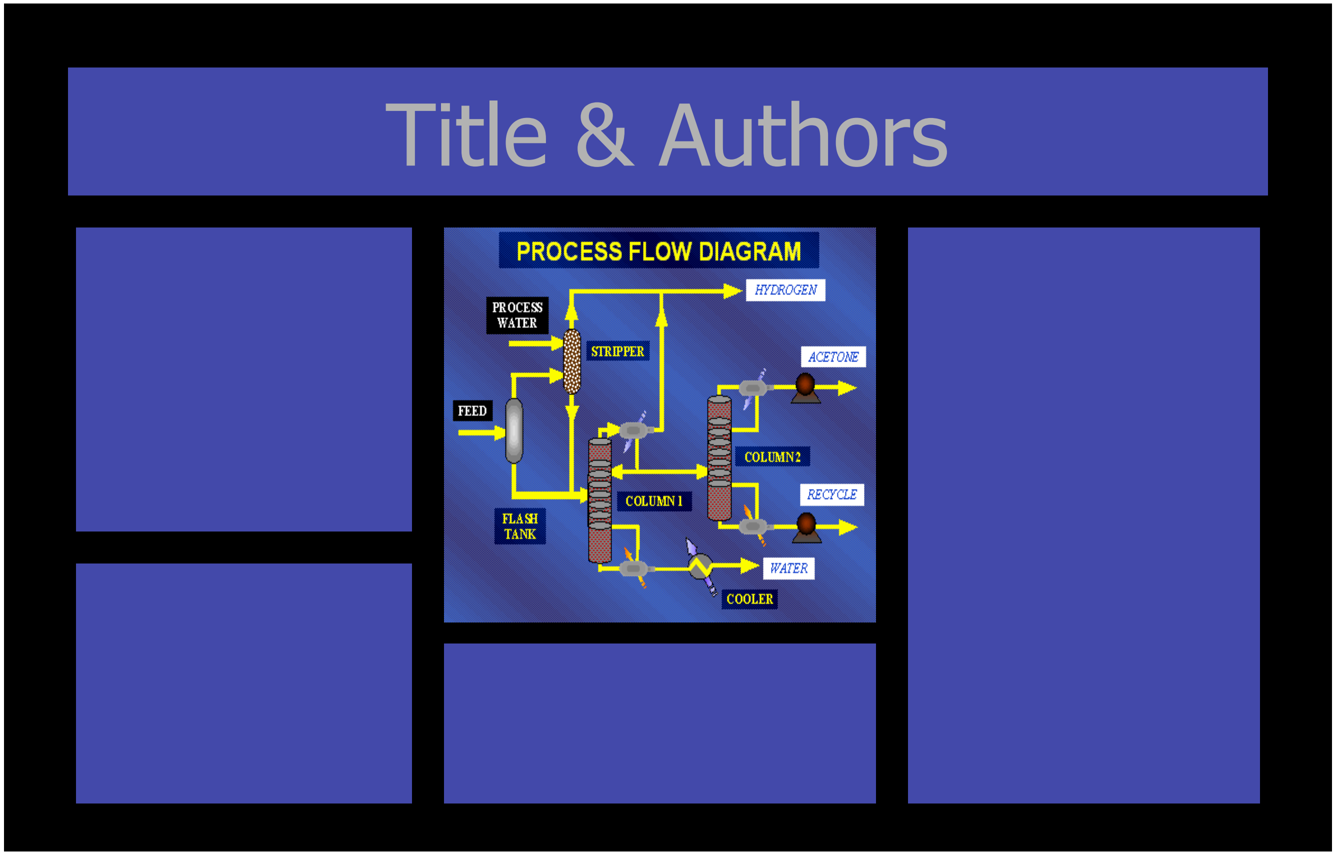
A clinical pilot study using the CTGS showed signi cant increases in DE symptoms, indicating potential use of the device as both a diagnostic and research tool.

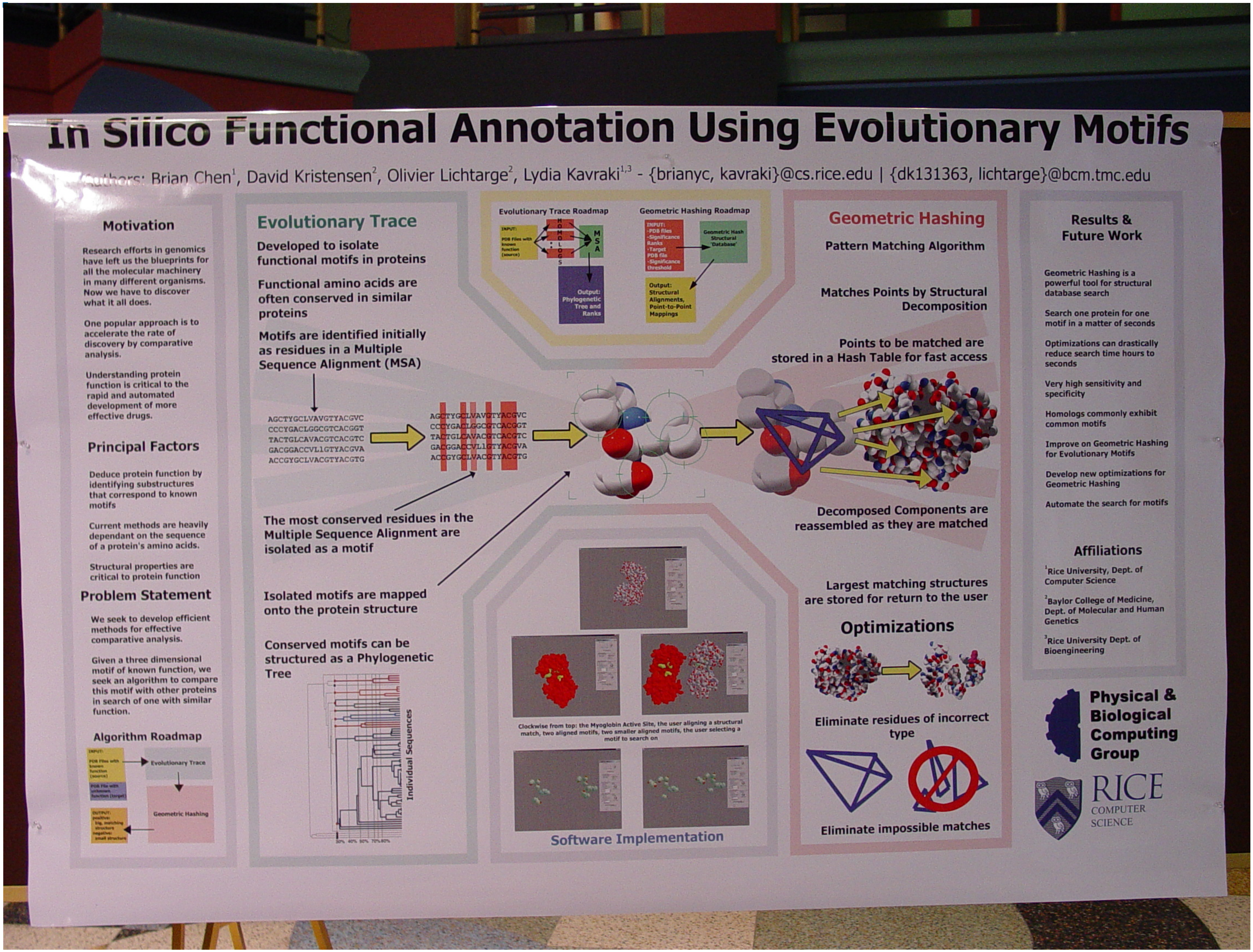
References & Acknowledgements

1. "Definition and Classification of Dry Eye Disease: Report of the Definition and Classification Subcommittee of the International Dry Eye WorkShop (2007)." The Ocular Surface 5.2 (2007): 75-92. The Ocular Surface: A Review Journal for Clinicians and Researchers. 2007.
2. Dry Eye Disease (DED) Report. Rep. Custom Business Development & Management Technology, 25 Nov. 2009. Web. 19 Oct. 2010.

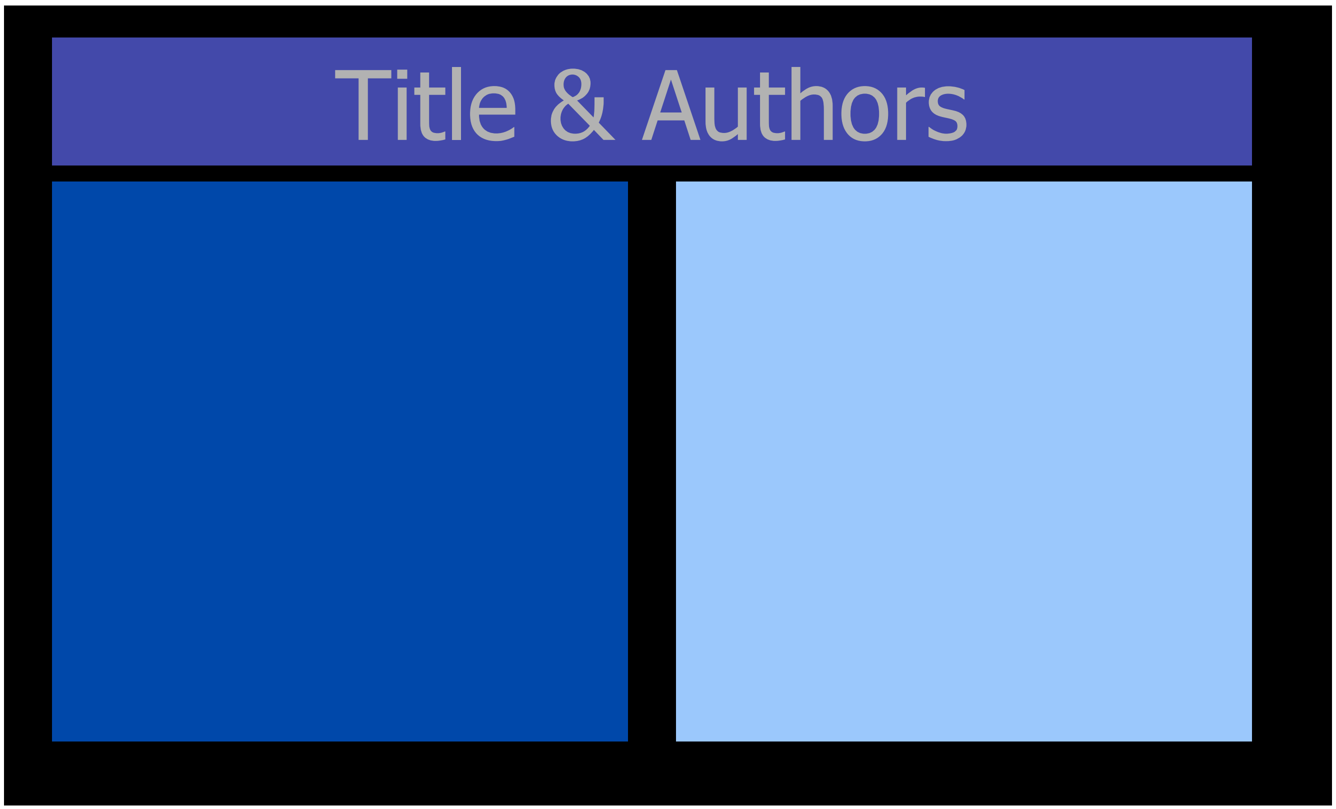
Many thanks to Dr. Stephen Pflugfelder, Dr. Cintia De Paiva, Dr. Maria Oden, Dr. Renata Ramos and Dr. Gary Woods for their tremendous support. This project was made possible by the Oshman Engineering Design Kitchen and the Cullen Eye Institute at Baylor College of Medicine

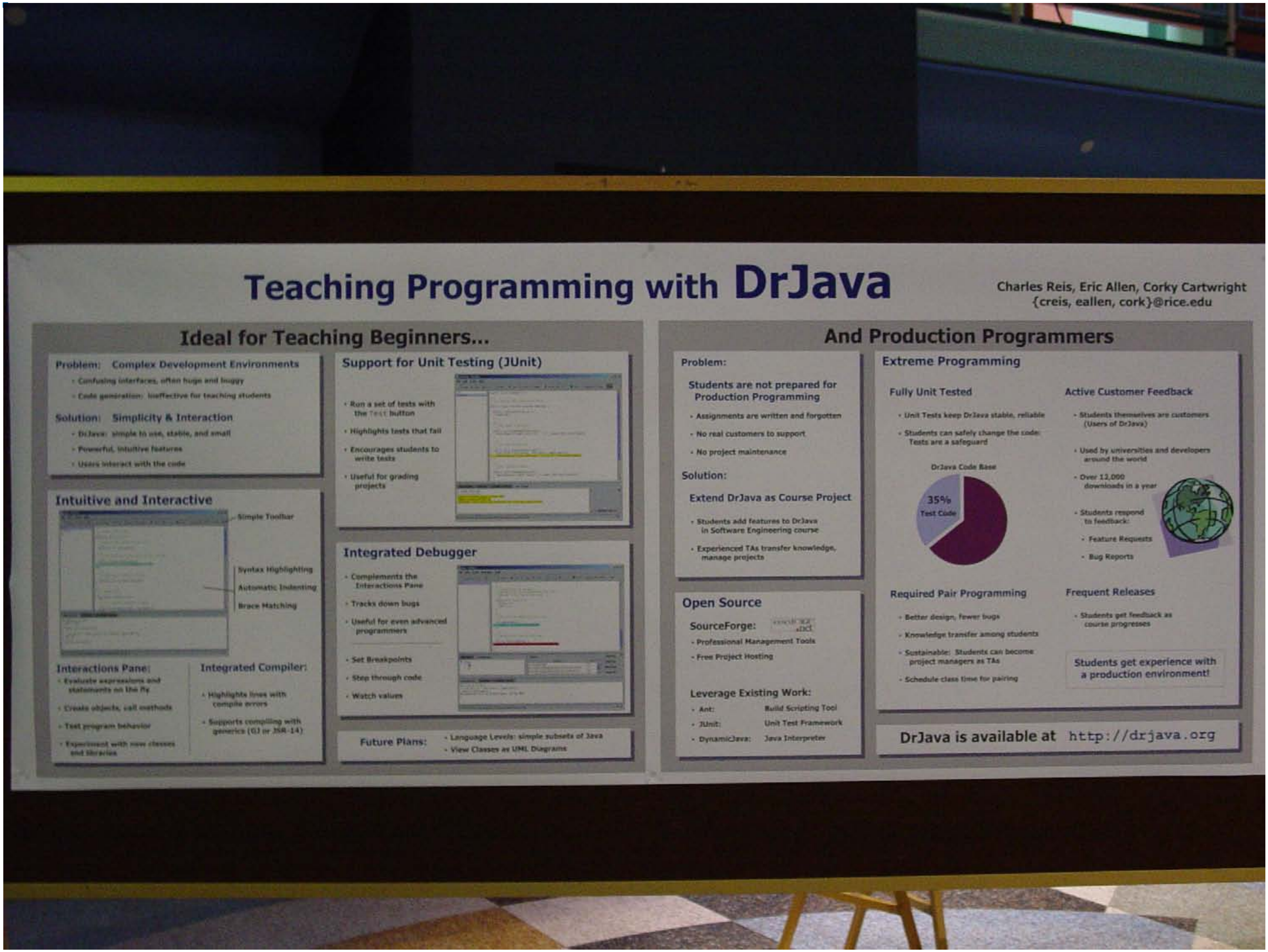
Centered image & explanations




In Silico Functional Annotation Using Evolutionary Motifs
Authors: Brian Chen[1], David Kristensen[2], Olivier Lichtarge[2], Lydia Kavraki[1,3] - {brianyc, kavraki}@cs.rice.edu | {dk131363, lichtarge}@bcm.tmc.edu
Motivation
Research efforts in genomics have left us the blueprints for all the molecular machinery in many different organisms. Now we have to discover what it all does.
One popular approach is to accelerate the rate of discovery by comparative analysis.
Understanding protein function is critical to the rapid and automated development of more effective drugs.
Principal Factors
Deduce protein function by identifying substructures that correspond to known motifs
Current methods are heavily dependant on the sequence of a protein's amino acids.
Structural properties are critical to protein function
Problem Statement
We seek to develop efficient methods for effective comparative analysis.
Given a three dimensional motif of known function, we seek an algorithm to compare this motif with other proteins in search of one with similar function.
Algorithm Roadmap
Evolutionary Trace
Geometric Hashing
Evolutionary Trace
Developed to isolate functional motifs in proteins
Functional amino acids are often conserved in similar proteins
Motifs are identified initially as residues in a Multiple Sequence Alignment (MSA)
AGCTYGCLVAVGTYACGVC
CCCYGACLGGCGTCACGGT
TACTGLCAVACGTCACGTC
GACGGACCVL1GTYACGVA
ACCGYGCLVACGTYACGTG
The most conserved residues in the Multiple Sequence Alignment are isolated as a motif
Isolated motifs are mapped onto the protein structure
Conserved motifs can be structured as a Phylogenetic Tree
Individual Sequences
Evolutionary Trace Roadmap
Output: Phylogenetic Tree and Ranks
Geometric Hashing Roadmap
Output: Structural Alignments, Point-to-Point Mappings
Clockwise from top: the Myoglobin Active Site, the user aligning a structural match, two aligned motifs, two smaller aligned motifs, the user selecting a motif to search on
Software Implementation
Geometric Hashing
Pattern Matching Algorithm
Matches Points by Structural Decomposition
Points to be matched are stored in a Hash Table for fast access
Decomposed Components are reassembled as they are matched
Largest matching structures are stored for return to the user
Optimizations
Eliminate residues of incorrect type
Eliminate impossible matches
Results & Future Work
Geometric Hashing is a powerful tool for structural database search
Search one protein for one motif in a matter of seconds
Optimizations can drastically reduce search time hours to seconds
Very high sensitivity and specificity
Homologs commonly exhibit common motifs
Improve on Geometric Hashing for Evolutionary Motifs
Develop new optimizations for Geometric Hashing
Automate the search for motifs
Affiliations
[1]Rice University, Dept. of Computer Science
[2]Baylor College of Medicine, Dept. of Molecular and Human Genetics
[3]Rice University Dept. of Bioengineering
Physical & Biological Computing Group
RICE COMPUTER SCIENCE

Two fields in contrast





Teaching Programming with DrJava
Charles Reis, Eric Allen, Corky Cartwright
{creis, eallen, cork}@rice.edu
Ideal for Teaching Beginners...
Problem: Complex Development Environments
- Confusing interfaces, often huge and buggy
- Code generation: ineffective for teaching students
Solution: Simplicity & Interaction
- DrJava: simple to use, stable, and small
- Powerful, intuitive features
- Users interact with the code
Intuitive and Interactive
Simple Toolbar
Syntax Highlighting
Automatic Indenting
Brace Matching
Interactions Pane:
- Evaluate expressions and statements on the fly
- Create objects, call methods
- Test program behavior
- Experiment with new classes and libraries
Integrated Compiler:
- Highlights lines with compile errors
- Supports compiling with generics (GJ or JSR-14)
Support for Unit Testing (JUnit)
- Run a set of tests with the Test button
- Highlights tests that fail
- Encourages students to write tests
- Useful for grading projects
Integrated Debugger
- Complements the Interactions Pane
- Tracks down bugs
- Useful for even advanced programmers
- Set Breakpoints
- Step through code
- Watch values
Future Plans:
- Language Levels: simple subsets of Java
- View Classes as UML Diagrams
And Production Programmers
Problem:
Students are not prepared for Production Programming
- Assignments are written and forgotten
- No real customers to support
- No project maintenance
Solution:
Extend DrJava as Course Project
- Students add features to DrJava in Software Engineering course
- Experienced TAs transfer knowledge, manage projects
Open Source
SourceForge:
- Professional Management Tools
- Free Project Hosting
Leverage Existing Work:
- Ant: Build Scripting Tool
- JUnit: Unit Test Framework
- DynamicJava: Java Interpreter
Extreme Programming
Fully Unit Tested
- Unit Tests keep DrJava stable, reliable
- Students can safely change the code: Tests are a safeguard
DrJava Code Base
35%
Test Code
Active Customer Feedback
- Students themselves are customers (Users of DrJava)
- Used by universities and developers around the world
- Over 12,000 downloads in a year
- Students respond to feedback:
- Feature Requests
- Bug Reports
Required Pair Programming
- Better design, fewer bugs
- Knowledge transfer among students
- Sustainable: Students can become project managers as TAs
- Schedule class time for pairing
Frequent Releases
- Students get feedback as course progresses
Students get experience with a production environment!
DrJava is available at http://drjava.org

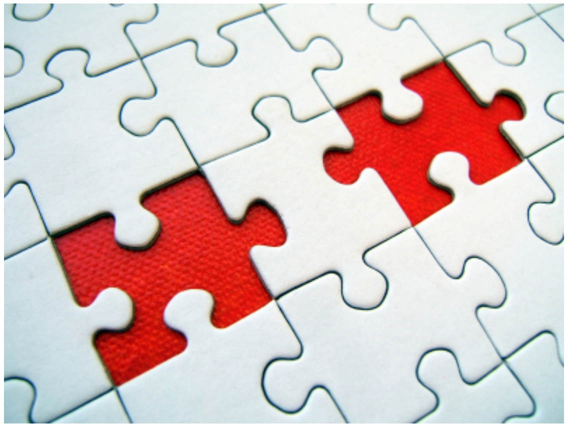
Make the story accessible

- Use informative headings
  - **Generic**: Results
  - **Topic-oriented**: Houston geography and crime
  - **Informative:** Crime widely dispersed across Houston
- Reduce jargon

Arrangement

Contrast

Hierarchy



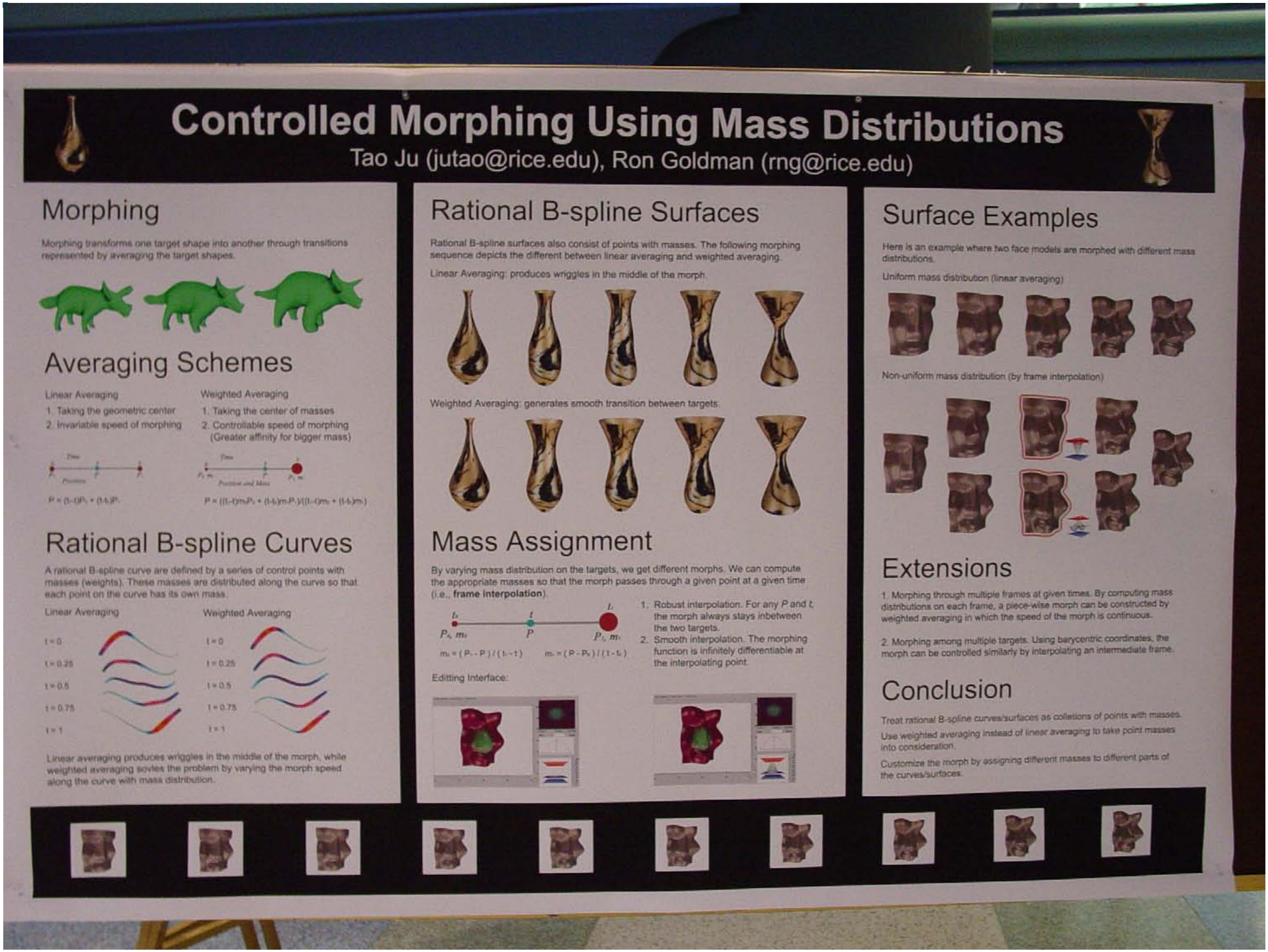
White space

Istockphoto.com

Contrast

- Size
- Shape
- Color
- Shade
- Proximity

Istockphoto.com


Controlled Morphing Using Mass Distributions
Tao Ju (jutao@rice.edu), Ron Goldman (rng@rice.edu)
Morphing
Morphing transforms one target shape into another through transitions represented by averaging the target shapes.
Averaging Schemes
Linear Averaging
1. Taking the geometric center
2. Invariable speed of morphing
Weighted Averaging
1. Taking the center of masses
2. Controllable speed of morphing (Greater affinity for bigger mass)
Rational B-spline Curves
A rational B-spline curve are defined by a series of control points with masses (weights). These masses are distributed along the curve so that each point on the curve has its own mass.
Linear Averaging
Weighted Averaging
t = 0
t = 0.25
t = 0.5
t = 0.75
t = 1
Linear averaging produces wriggles in the middle of the morph, while weighted averaging sovles the problem by varying the morph speed along the curve with mass distribution.
Rational B-spline Surfaces
Rational B-spline surfaces also consist of points with masses. The following morphing sequence depicts the different between linear averaging and weighted averaging.
Linear Averaging: produces wriggles in the middle of the morph.
Weighted Averaging: generates smooth transition between targets.
Mass Assignment
By varying mass distribution on the targets, we get different morphs. We can compute the appropriate masses so that the morph passes through a given point at a given time (i.e., frame interpolation).
1. Robust interpolation. For any P and t, the morph always stays inbetween the two targets.
2. Smooth interpolation. The morphing function is infinitely differentiable at the interpolating point.
Editting Interface:
Surface Examples
Here is an example where two face models are morphed with different mass distributions.
Uniform mass distribution (linear averaging)
Non-uniform mass distribution (by frame interpolation)
Extensions
1. Morphing through multiple frames at given times. By computing mass distributions on each frame, a piece-wise morph can be constructed by weighted averaging in which the speed of the morph is continuous.
2. Morphing among multiple targets. Using barycentric coordinates, the morph can be controlled similarly by interpolating an intermediate frame.
Conclusion
Treat rational B-spline curves/surfaces as colletions of points with masses.
Use weighted averaging instead of linear averaging to take point masses into consideration.
Customize the morph by assigning different masses to different parts of the curves/surfaces.

Hierarchy

Enables audience to discern relationships

Istockphoto.com

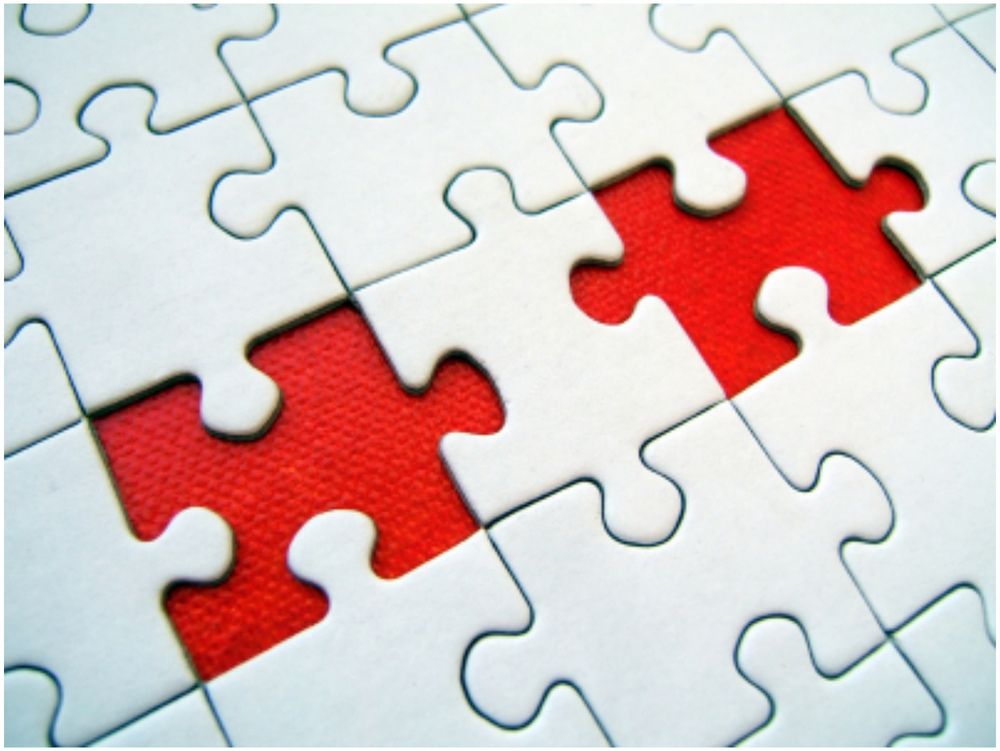
Bulleted List Layout

- This is a primary bullet.
  - This is a secondary bullet.
    - This is subordinate to the bullets above it.
    - This is subordinate to the bullets above it.
- This is a primary bullet.
  - This is a secondary bullet.

Bulleted List Layout

- This is a primary bullet.
  - This is a secondary bullet.
    - This is subordinate to the bullets above it.
    - This is subordinate to the bullets above it.
- This is a primary bullet.
  - This is a secondary bullet.

White space



Directs gaze

Istockphoto.com

Adaptive Out-Of-Core Simplification

RICE COMPUTER SCIENCE

Scott Schaefer (sschaefe@rice.edu), Joe Warren (jwarren@rice.edu)

Polygon Reduction

Figure 1. Simplification of Michelangelo's Saint Matthew using vertex collapsing. The original model has over 6.7 million polygons and cannot be simplified using conventional in memory techniques.

- Eliminate polygons by merging two adjacent vertices together with lowest "error."
- Conventional algorithms require entire object to be in memory.
- Infeasible to simplify massive polygonal objects using this technique.
- Need a method that does not require global information to simplify the model.

Uniform Simplification

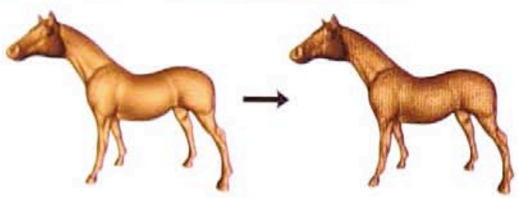


Figure 2. Simplification of a polygonal model of a horse using uniform simplification. The reduced model contains approximately 12% of the number of polygons in the original model. Notice the "gridding" present in the model, which results from the uniform sampling.

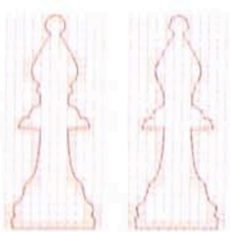


Figure 3. Uniform simplification of a 2D curve of a bishop. The original curve (left) and the reduced curve (right) are shown in the grid used to simplify the shape. The squares intersecting the curve are highlighted.

- Group vertices from the curve into squares inside a uniform grid.
- Construct one vertex per square that represents all of the vertices in that square.
- Place the representative vertex at the position that minimizes distance to polygons intersecting that square.
- Line segments connect representative vertices corresponding to the original vertices of that line segment.
- Remove line segments that collapse to a single point from the final curve.

Spatial Sort

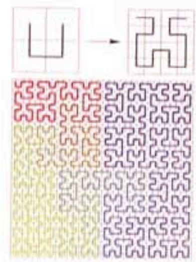


Figure 4. Recursive rule to generate a Hilbert space-filling curve (top). Five applications of the rule with colors illustrating the path of the curve (bottom).

- Simple recursive rule that generates the curve in figure 4 (top).
- Produces a curve that fills all of space.
- Induces a linear ordering on space. Arbitrary points can be sorted easily.
- Allows reduction algorithm to know when information is locally complete and a safe reduction can be performed.
- Implemented as a variation of an out-of-core merge sort.

Adaptive Simplification

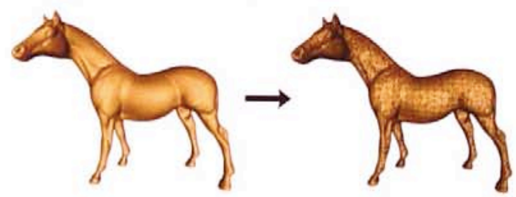


Figure 5. Adaptive simplification of the model from figure 2. The simplified model on the right contains the same number of polygons as the simplified model in figure 2. However, the polygons generated by the adaptive method are of widely varying sizes and are able to reproduce fine details.

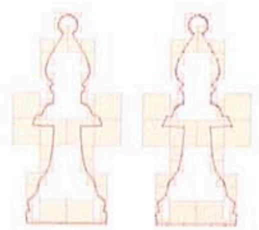


Figure 6. Reduction of the same curve as in figure 3 except adaptive simplification is used. The squares in the multi-resolution grid that generate vertices of the simplified model are shown highlighted.

- Order vertices using spatial sort (above).
- Add vertices one at a time to quadtree structure expanding until vertices lie in separate nodes.
- When out of memory, collapse a complete node with smallest error to reclaim space.
- Use curvature dependent error metric to preserve fine details.
- Generate one representative vertex for all leaf nodes in tree.
- Connect line segments to representative vertices to create final curve.

Examples

Figure 7. The original model is shown on the left. Simplified versions using uniform sampling (middle) and adaptive sampling (right). Both simplified models have the same number of polygons (approximately 1.5% of original polygons). Notice that the adaptive method preserves fine details.

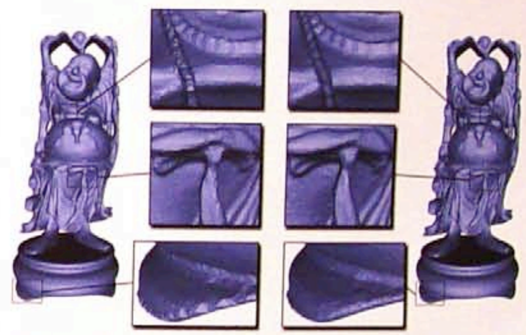


Figure 8. Simplification of a Buddha figure using uniform sampling (left) and adaptive sampling (right) to the same number of polygons. From a macroscopic view the models appear very similar. However, the close-ups illustrate the small details preserved by the adaptive method, which the uniform method cannot reproduce.

Conclusions

- Adaptive method can reduce large polygonal models.
- Only requires local information to simplify a model.
- Preserves small scale features that uniform reduction cannot.
- Performs as well as optimal in-memory technique using only a small, constant amount of additional space.

Acknowledgements

We'd like to thank the Stanford Graphics Repository for the model of Buddha, the Digital Michelangelo Project for the models of David and St. Matthew, and Cyberware for the model of the horse.

- ***Therapeutic Systems and Technologies*** –
  - Dorin Panescu, Refractec Inc., Irvine , CA, **Chair,** Cardiac Catheter Ablation
  - Jean-Yves Chapelon Ph.D., INSERM, Lyon, France, High Intensity Focused Ultrasound for Prostate Tumor Ablation
  - Rahul Mehra Ph.D., Medtronic, Inc. Minneapolis, MN, Cardioverter-defibrillators, Tachyarrhythmia Research
  - Tim McIntyre M.S.,Manager, St. Jude Medical, St Paul, MN, Medical Device Industry R&D and Management
  - John Pearce Ph.D., ECE Department, University of Texas, Austin, TX, Electromagnetics and Acoustics Applied to Medical Devices
  - Kouros Azar M.D., B.S.BME, Thousand Oaks, CA, Reconstructive Surgeon
  - Reese Terry M.S., Co-founder/Board Member Cyberonics, Inc., Houston, TX, Neurostimulation Devices

Therapeutic Systems & Technologies

Dorin Panescu, *Chair*
Refractec Inc., Irvine, CA
Cardiac Catheter Ablation

Jean-Yves Chapelon, Ph.D.
INSERM, Lyon, France
High Intensity Focused Ultrasound for Prostate Tumor Ablation

John Pearce, Ph.D.
University of Texas, Austin, TX
Electromagnetics and Acoustics Applied to Medical Devices

Rahul Mehra, Ph.D.
Medtronic, Inc. Minneapolis, MN
Cardioverter-defibrillators
Tachyarrhythmia Research

Kouros Azar, M.D., B.S.
BME, Thousand Oaks, CA
Reconstructive Surgeon

Tim McIntyre, M.S.
St. Jude Medical, St Paul, MN,
Medical Device Industry R&D and Management

Reese Terry, M.S.
Co-founder/Board Member
Cyberonics, Houston, TX
Neurostimulation Devices

Text

- Legible
- Large
- Succinct
- Parallel
- Conservative
- Consistent

Istockphoto.com

Sans serif font best

Good for posters	Good for print
E	E
Sans Serif	Serif ("tail")
Arial	Times New Roman
Helvetica	Courier
Calibri	Garamond

Large font sizes

Title ◄	80-120 pt
Headings ◄	36-54 pt
▪ Text ◄	24-36 pt
Captions and data labels ◄	18-20 pt

Fat text ▸lean text

The findings support the idea that there are gender differences in the mate selection process: women tend to choose less attractive mates while men tend to choose more attractive mates.

Gender differences in mate selection

- Women select less attractive mates
- Men select more attractive mates

Color

Istockphoto.com

Did Van Gogh Really Paint This ?
Frequency Analysis for Art Forensics

Don Johnson (dhj@rice.edu) Lu Sun (ls4@rice.edu) Zeting Liu (zbl1@rice.edu) Xiang Guo (xg1@rice.edu)

ABSTRACT

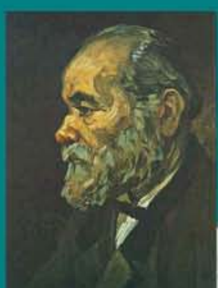
Frequency Distributions should match if two paintings are from the same canvas roll according to the **Thread Counting Algorithm** we employ.

BACKGROUND

- **Warp and Weft**
- **Canvas Texture Modeling**

$$c(\vec{x}) = p(\vec{x}) + \frac{A}{2}\cdot(2 + a_h \sin(2\pi \vec{f}_h \cdot \vec{x} + \phi_h) + a_v \sin(2\pi \vec{f}_v \cdot \vec{x} + \phi_v))$$

- **Weave Density**
- **Thread Count**
- **Frequency Analysis**

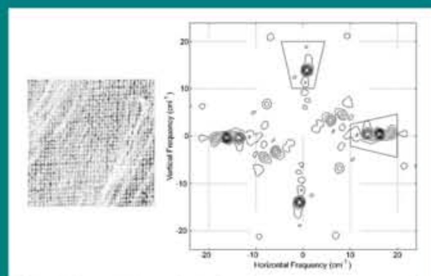


Van Gogh's *Portrait of an Old Man with a Beard (F205/JH971)*

X-ray image of F205 provided by the Van Gogh Museum, Amsterdam

METHODS

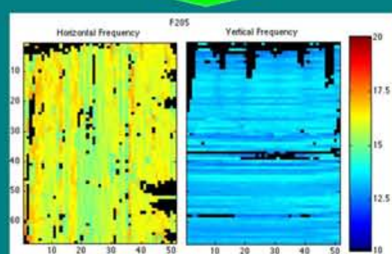
- **Raw Data: X-Ray Images**
- **Orthogonal Sinusoids Modeling**
- **Thread Count and Weave Density**
- **Short Time Fourier Analysis**
- **Spectrum of Paintings**
- **Frequency Analysis**
  - **Averages**
  - **Deviations**



1" × 1" swatch from the X-ray

Detailed spectra computed from a ½" × ½" square

The wedges indicate areas where weave-related spectral peaks are found



Spectrum of whole-painting (F205)

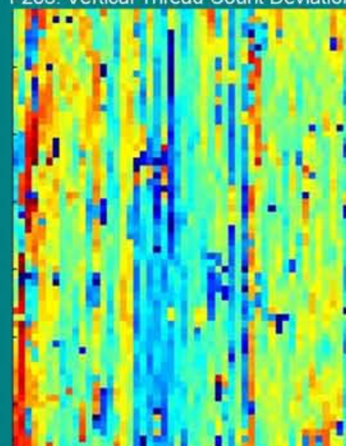
Horizontal and Vertical Frequencies

* Black spots indicates multiple or outrange frequency peaks

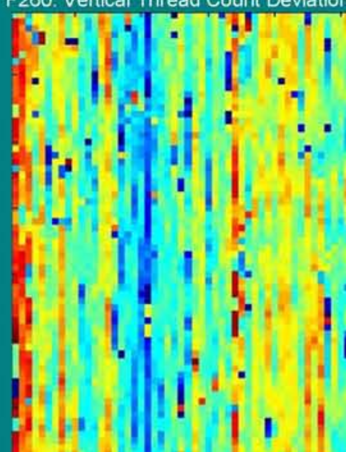
RESULTS

- **Critical Value**
- **Paintings Mapping Determination**

F205: Vertical Thread Count Deviations



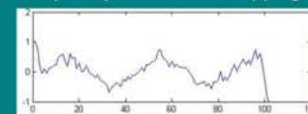
F260: Vertical Thread Count Deviations



CONCLUSIONS

Spectral techniques offer a better approach than manual methods in terms of efficiency and accuracy. Whole-painting analysis could better support the forensic evidence quantitatively for sequencing paintings.

Frequency Correlation Mapping Between F205 and F260



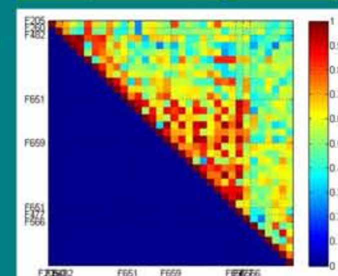
Sth. More here

Correlation in Vertical Direction

Correlation in Horizontal Direction

FUTURE WORK

- Orientation Issue in Modeling
- Multiple and Outrange Peak Frequencies

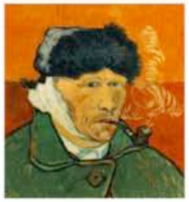


Sth. More here

Deviation Matching Results

ACKNOWLEDGEMENT

- A
- B
- C



Did Van Gogh Really Paint This ?
Frequency Analysis for Art Forensics

Don Johnson (dhj@rice.edu) Lu Sun (ls4@rice.edu) Zeting Liu (zbl1@rice.edu) Xiang Guo (xg1@rice.edu)

ABSTRACT

A canvas can be characterized by the vertical and horizontal weave densities while the actual painting serves as an additive signal that only distracts from the thread-counting process. The thread counting algorithm and the spectral techniques we employ here can analyze weave density for entire paintings with an accuracy comparable to human measurements more efficiently.

MOTIVATION

Van Gogh Museum of Amsterdam has a collection of artist's works and is looking for a more advanced analysis for sequencing paintings in addition to the traditional manual methods. Our whole-painting analysis shows that frequency distributions should match if two paintings are from the same canvas roll. This could better support the forensic evidence quantitatively when comparing two paintings.

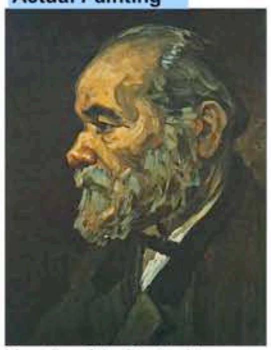
BACKGROUND

- Warp and Weft
- Vertical vs. Horizontal
- Canvas Texture Modeling

$$c(\vec{x}) = p(\vec{x})$$
$$+\frac{A}{2}\cdot(2+a_h\sin(2\pi\vec{f}_h\cdot\vec{x}+\phi_h)+a_v\sin(2\pi\vec{f}_v\cdot\vec{x}+\phi_v))$$

- Thread Count
- Weave Density
- Short-space Spectrum
- Whole Painting Analysis
- Correlation Determination
  - Averages
  - Deviations

Actual Painting



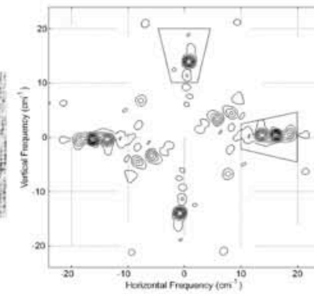
Van Gogh's *Portrait of an Old Man with a Beard (F205/JH971)*

Raw Data

X-ray image of F205 provided by the Van Gogh Museum, Amsterdam

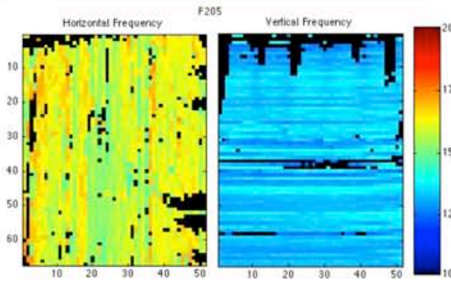
1" × 1" swatch from the X-ray

Detailed spectra computed from a ½" × ½" square



The wedges indicate areas where weave-related spectral peaks are found

Spectrum of whole-painting (F205) Horizontal and Vertical Frequencies

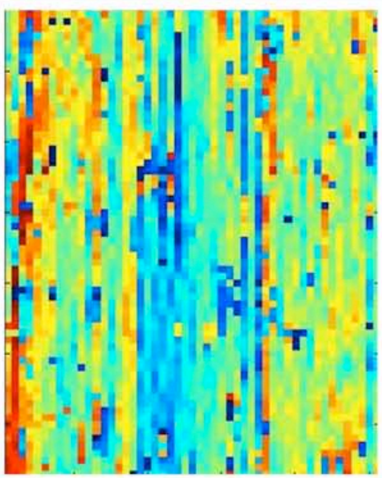


Black spots indicates multiple or outrange frequency peaks

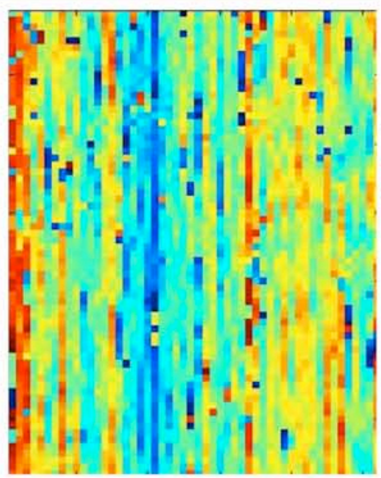
RESULTS

- Critical Values
- Paintings Mapping Analysis
  - Location
  - Peaks

We obtain the spectrum of another painting by going through the same process

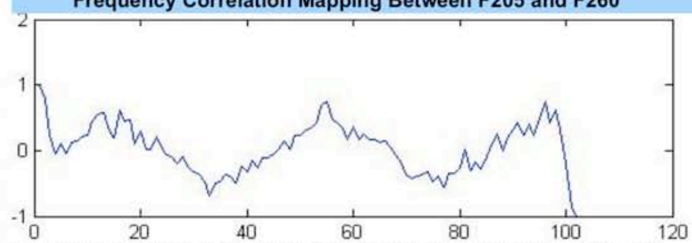


F205: Vertical Thread Count Deviations



F260: Vertical Thread Count Deviations

Frequency Correlation Mapping Between F205 and F260



Correlation in Horizontal Direction: Peak appears at 57th alignment

Clearly, our correlation analysis result is consistent with the matching between F205 and F260 as we can see from the two spectrums. And in fact, F205/F260 were painted by Van Gogh in the same month in 1885.

ACKNOWLEDGEMENT

- This poster would have not been possible without the help of Professor Don H. Johnson at ECE department of Rice University.
- The thread counting algorithm represents a collaboration between the Thread Count Automation Project directed by Professor Rick Johnson at Cornell University and the Van Gogh Museum, Amsterdam.
- Images in this poster are courtesy of Van Gogh Museum, Professor Don H. Johnson and Lu Sun.

Did Van Gogh Really Paint This?
Frequency Analysis for Art Forensics

Lucia Sun (ls4@rice.edu) Xiang (Jash) Guo (xg1@rice.edu) Zeting Liu (zbl1@rice.edu)

Motivation

The Van Gogh Museum of Amsterdam seeks a more accurate quantitative approach to sequencing its paintings.

Current Approach

Human measurement – weave threads are manually counted and compared. This approach is inefficient and may not be accurate.

New Approach

- **Use thread counting algorithm and spectral techniques to analyze weave density for entire paintings.**
  - **A canvas can be characterized by its vertical and horizontal weave densities.**
  - **In contrast, paint acts as an additive signal that only distracts from thread-counting process.**
- **Frequency distributions of weave densities should match if two paintings are from same canvas roll.**

Background

- **Warp and Weft**
- **Vertical vs. Horizontal**
- **Canvas Texture Modeling**

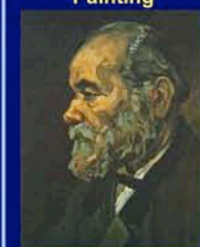
$$c(\vec{x}) = p(\vec{x}) + \frac{A}{2}\cdot\left(2 + a_h \sin(2\pi\vec{f}_h\cdot\vec{x} + \phi_h) + a_v \sin(2\pi\vec{f}_h\cdot\vec{x} + \phi_v)\right)$$

- **Thread Count**
- **Weave Density**
- **Short-space Spectrum**
- **Whole Painting Analysis**
- **Averages**
- **Deviations**
- **Correlation Determination**

Methods

- **Orthogonal Sinusoids Modeling of Raw Data**
- **Thread Count and Weave Density**
- **Short Time Fourier Analysis**
- **Spectra of Paintings and Frequency Analysis**

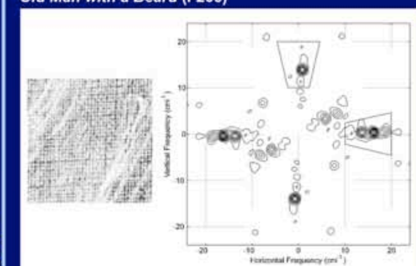
Painting



Van Gogh's *Portrait of an Old Man with a Beard* (F205)

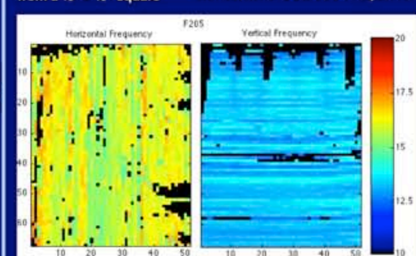
Raw Data

X-ray image of F205



▲ 1" × 1" swatch from X-ray image **▲ Wedges show where weave-related spectral peaks appear.**

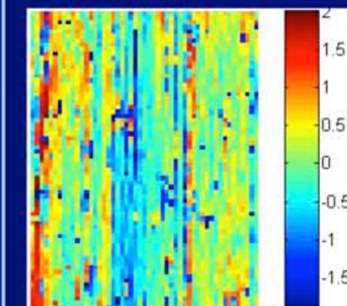
Detailed spectra computed from a ½" × ½" square **Spectrum of whole-painting (F205) Horizontal and Vertical Frequencies**



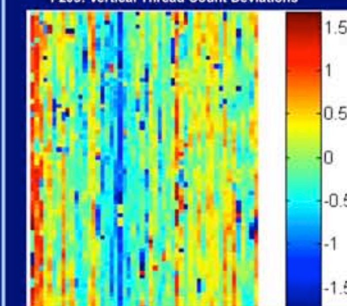
Black spots indicate multiple or outrange frequency peaks.

Results

- **Critical Value**
- **Locations & Peaks**



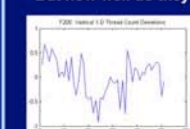
F205: Vertical Thread Count Deviations

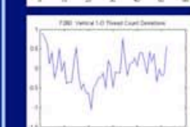


F260: Vertical Thread Count Deviations

▲ Two graphs above show vertical thread count deviation pattern of F205 and F260.

Visually, we can see the matching strips. But how well do they match?



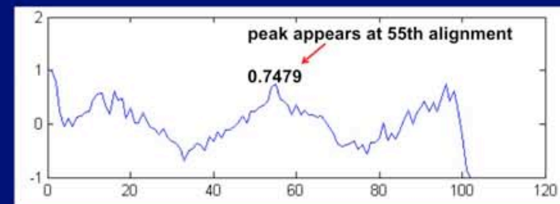


◄This graph shows the vertical 1-D thread count deviations of F205 and F260.

The 1-D thread count deviations are obtained by summing the column deviations of 2-D data.

We can clearly see the similarities of the vertical ones between F205 and F260.

How do they correlate?



▲Unbiased Correlation Coefficient between Vertical Thread Count Deviations of F205 & F260.

Peak (0.7479) appears at 55th alignment as two graphs are mapped to the matching alignment, or visually "best fit" together.

Conclusions

Spectral techniques offer a more efficient and accurate approach to analyzing and sequencing paintings than manual methods. Whole-painting analysis could provide quantitative support for forensic evidence.

Our correlation analysis result is consistent with the matching of F205 and F260, shown in two spectra on left.

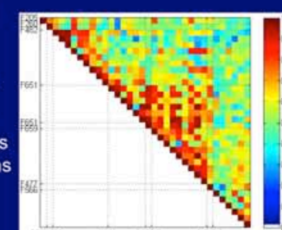
And in fact, F205 and F260 were painted by Van Gogh in the same month in 1885!

Future Work

- Elimination of NaN's in Thread Counting Densities
- Determining Critical Values
- Multiple & Outrange Peak Freq.

► Comparing weave density deviations of different paintings can generate cross correlations between these paintings.

Hot spots represent "good matches."



Deviations Matching Results

Acknowledgements

- We wish to thank Dr. Don H. Johnson of Rice University's ECE department.
- The thread counting algorithm is a collaboration between the Thread Count Automation Project directed by Professor Rick Johnson at Cornell University and Van Gogh Museum, Amsterdam.
- Images are courtesy of Van Gogh Museum, Dr. Don H. Johnson and Lucia Sun.

CAUSES OF DEATHS IN MEXICO 2008

Melissa Tsang, Ignacio Olivera, Eli Payne, & Jaime Ramos
Professor: Hadley Wickham TA: Garret Grolemund

Main Data from: http://sinais.salud.gob.mx
SINAIS: Sistema Nacional de Informacion en Salud

Motivation:
Motivated by the so called Mexican drug war, we started analyzing the data, looking for the most dangerous federal entities and municipalities. However, our analysis lead us to more results, by subdividing the causes of death and comparing death trends versus change in temperature throughout Mexico and comparing with 2005-2007.

Goals: To Study Deaths in Mexico

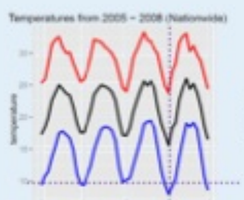
1. Death Trends Related to Temperature
2. External Causes of Death
3. Deaths of Foreigners

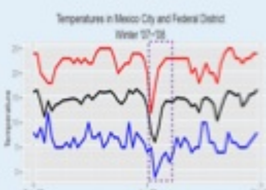
Death Trends Related to Temperature

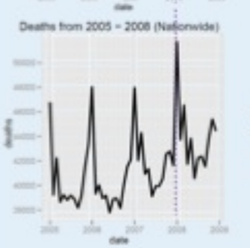
Nationwide and Mexico (state)-Federal District

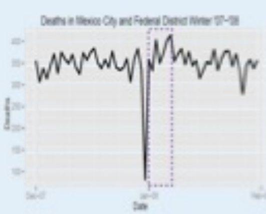
Nationwide Cold front 2008:

At both **National level** and **State level**, a decrease in temperature correlated with an increase in the number of deaths.









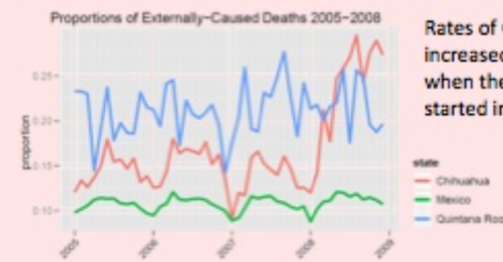
External Causes of Death

Percentage by State.

Chihuahua and Quintana Roo had the highest percentages of externally caused deaths.

Further analysis:
Quintana Roo had the highest rate of suicides at 0.040 (nearly 4 times the national average of 0.011) and **Chihuahua had the highest rate of homicides** at 0.131 (national 0.028).

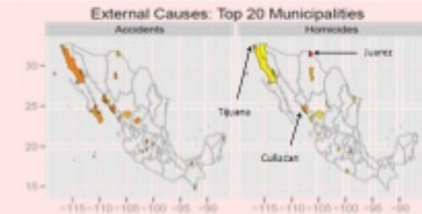
Time Line of Proportion for Chihuahua and Quintana Roo



Rates of Chihuahua increased dramatically when the drug war started in 2008.

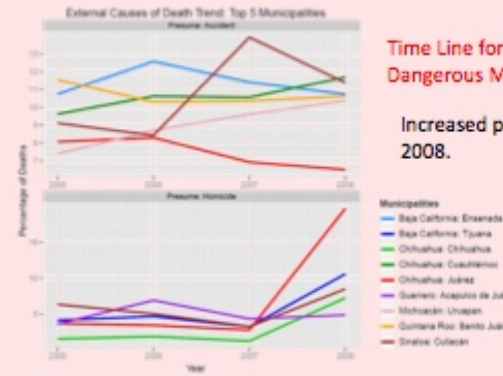
Focusing on Chihuahua

Juarez had the highest percentage of deaths by external causes followed by Chihuahua city.



Top 20 Most Dangerous Municipalities 2008

Juarez, Chihuahua.
Tijuana, Baja Cal.
Culiacan, Sinaloa.



Time Line for the Top 5 Most Dangerous Municipalities

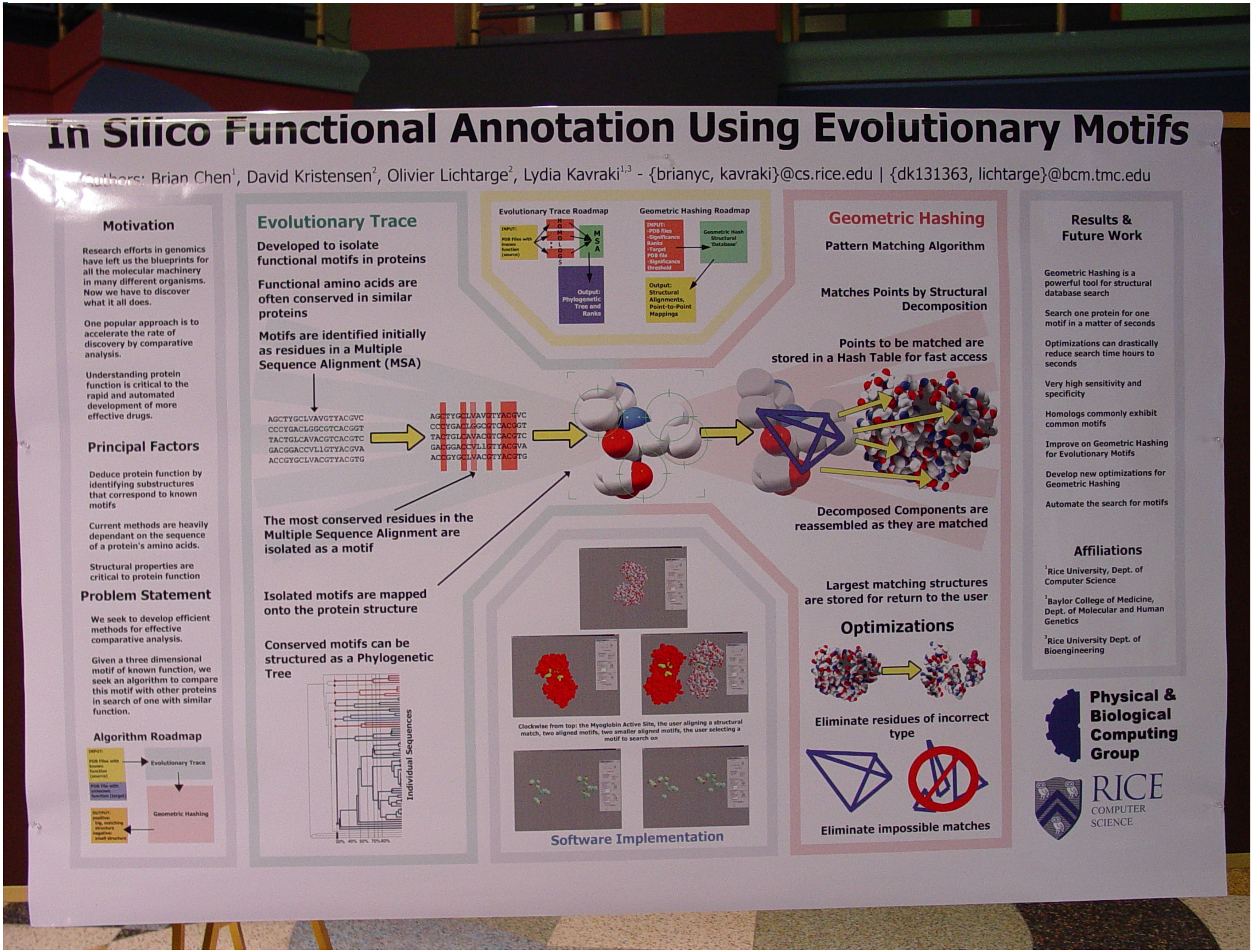
Increased percentages in 2008.

Deaths of Foreigners

Percentage by State.

Baja California Sur and Quintana Roo had the highest percentages of foreigner deaths.

Conclusion:
Deaths are negatively correlated with temperature at a national and State level.
Incidence of externally caused deaths in Chihuahua increased since the start of the drug war in 2008, mostly in the city of Juarez.
Most foreign deaths in Quintana Roo (Cancun) and Baja California (Tijuana).


In Silico Functional Annotation Using Evolutionary Motifs
Authors: Brian Chen[1], David Kristensen[2], Olivier Lichtarge[2], Lydia Kavraki[1,3] - {brianyc, kavraki}@cs.rice.edu | {dk131363, lichtarge}@bcm.tmc.edu
Motivation
Research efforts in genomics have left us the blueprints for all the molecular machinery in many different organisms. Now we have to discover what it all does.
One popular approach is to accelerate the rate of discovery by comparative analysis.
Understanding protein function is critical to the rapid and automated development of more effective drugs.
Principal Factors
Deduce protein function by identifying substructures that correspond to known motifs
Current methods are heavily dependant on the sequence of a protein's amino acids.
Structural properties are critical to protein function
Problem Statement
We seek to develop efficient methods for effective comparative analysis.
Given a three dimensional motif of known function, we seek an algorithm to compare this motif with other proteins in search of one with similar function.
Algorithm Roadmap
Evolutionary Trace
Geometric Hashing
Evolutionary Trace
Developed to isolate functional motifs in proteins
Functional amino acids are often conserved in similar proteins
Motifs are identified initially as residues in a Multiple Sequence Alignment (MSA)
AGCTYGCLVAVGTYACGVC
CCCYGACLGGCGTCACGGT
TACTGLCAVACGTCACGTC
GACGGACCVL1GTYACGVA
ACCGYGCLVACGTYACGTG
The most conserved residues in the Multiple Sequence Alignment are isolated as a motif
Isolated motifs are mapped onto the protein structure
Conserved motifs can be structured as a Phylogenetic Tree
Individual Sequences
Evolutionary Trace Roadmap
HOMOLOGS
MSA
Output: Phylogenetic Tree and Ranks
Geometric Hashing Roadmap
Geometric Hash Structural 'Database'
Output: Structural Alignments, Point-to-Point Mappings
Clockwise from top: the Myoglobin Active Site, the user aligning a structural match, two aligned motifs, two smaller aligned motifs, the user selecting a motif to search on
Software Implementation
Geometric Hashing
Pattern Matching Algorithm
Matches Points by Structural Decomposition
Points to be matched are stored in a Hash Table for fast access
Decomposed Components are reassembled as they are matched
Largest matching structures are stored for return to the user
Optimizations
Eliminate residues of incorrect type
Eliminate impossible matches
Results & Future Work
Geometric Hashing is a powerful tool for structural database search
Search one protein for one motif in a matter of seconds
Optimizations can drastically reduce search time hours to seconds
Very high sensitivity and specificity
Homologs commonly exhibit common motifs
Improve on Geometric Hashing for Evolutionary Motifs
Develop new optimizations for Geometric Hashing
Automate the search for motifs
Affiliations
[1]Rice University, Dept. of Computer Science
[2]Baylor College of Medicine, Dept. of Molecular and Human Genetics
[3]Rice University Dept. of Bioengineering
Physical & Biological Computing Group
RICE COMPUTER SCIENCE

Details matter!

- Apply consistent formatting
- Check grammar and spelling
- Cite references
- Give credit to others
- Include contact info

POST: A Secure, Resilient Cooperative Messaging System

http://freepastry.rice.edu/post

RICE

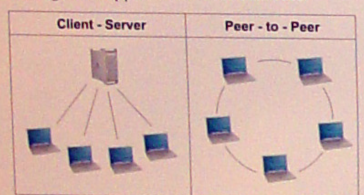
Motivation

Problem

- Current peer-to-peer (p2p) systems only used for illegal file sharing
  - Gnutella, KaZaA, Napster have provided the notion that p2p is not good for anything legal
- Are proposed p2p overlays mature enough to support collaborative applications?
  - High requirements of security
  - Existing p2p applications are simple
- Opportunity to improve existing collaborative applications (email, instant messaging)
  - Added robustness and resilience
  - Reduced cost
  - Increase security

General Solution

- Provide a generic, serverless collaborative platform, POST, based on p2p technologies
  - Create a middleware layer which enables the writing of collaborative applications
  - Use a p2p overlay, such as Pastry, for both data storage and application-level multicast



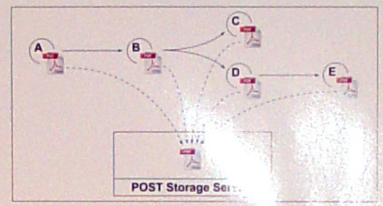
- Target applications inherit desirable properties from POST
  - Increased robustness and resilience from distributed nature of p2p
  - Reduced cost due to no dedicated servers
  - Better security from authenticated messages and default encryption of data

Architecture

- POST provides three primitives to applications written on top of it

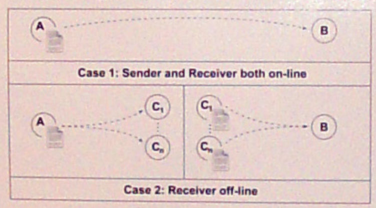
Single-Copy Data Storage

- Data is stored securely with multiple copies coalesced into one
  - In a collaborative system, sharing is common



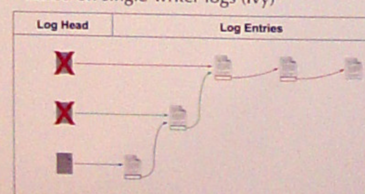
User Notification

- POST allows users to send application-specific notifications to others
  - Works regardless of recipient on or offline



User Specific Metadata

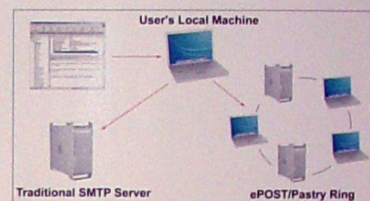
- POST provides user-specific metadata for each application it supports
  - Based on single-writer logs (Ivy)



Projected Applications

Email

- Email application ePOST compatible with existing clients and protocols
  - Users run local proxy
  - Messages broken into MIME components, each stored in Data Storage
  - Delivery using Notification service
  - Email folders represented using Metadata service



Other

- Instant messaging application imPOST
  - Uses Notification service for delivery and Metadata service for buddy lists
- Shared Calendaring application calPOST
  - Metadata service used to store appointments

Status

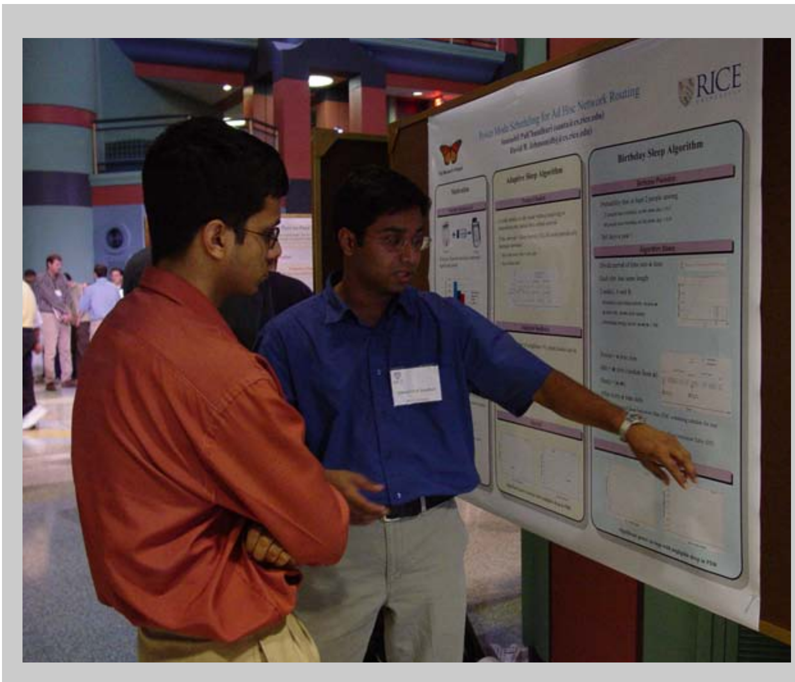
- POST implemented on top of FreePastry, PAST, and Scribe
- ePOST completely implemented with local IMAP and SMTP server
  - Efficiency and feasibility currently being studied within our group
  - Gaining experience with deployed p2p system
- imPOST completely implemented
  - Not yet integrated with existing IM protocols and clients

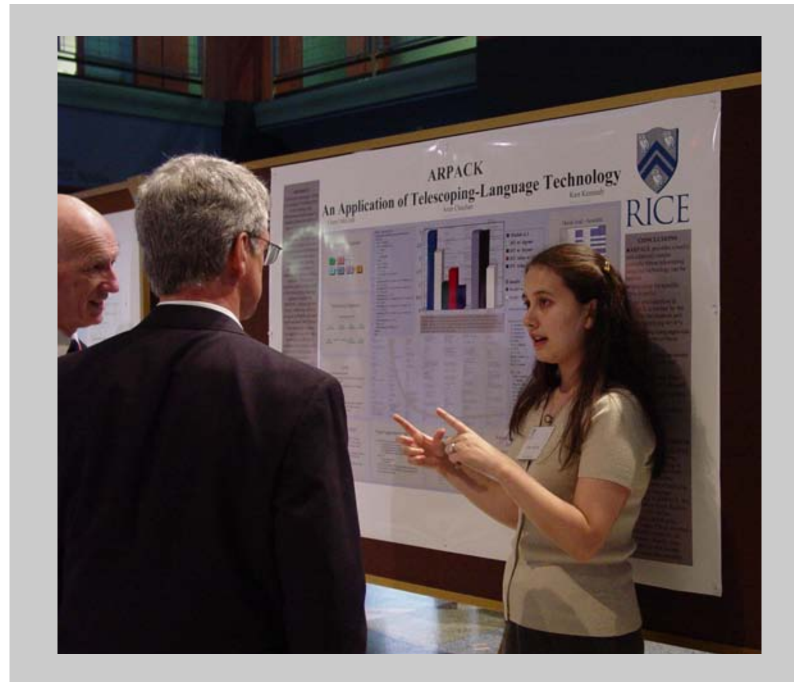
Present your poster with confidence

Connect with audience

- Greet people
- Prepare 30 sec, 90 sec, 3 min talks
- Summarize “gist” in 2-3 sentences
- Reinforce key points
- Be able to start spiel from any section
- Catch up newcomers quickly

Communicate with confidence





- Be enthusiastic
- Maintain eye contact
- Don’t block poster

- Integrate gestures
- Speak up
- Adjust pace

Wrap up

- Engage audience's interest
- Use images/figures/maps to make data accessible
- Analyze and interpret data
- Edit for brevity and precision