Research Thrust R3
Image and Data Information Management

A. Overview
B. R3 Fundamental Research
   R3a. Data Compression
   R3b. Multimedia Networks
   R3c. Scalable Computing
   R3d. Image Databases and Metadata
C. Impact and Timeline
D. I-PLUS

Diverse Problems – Similar Solutions

James Modestino
David Kaeli
A. Overview
Scope

- Acquisition, transmission, storage, processing, retrieval, dissemination, display and visualization of sensor data
- Development of a distributed collaborative research environment
- Provide remote Web access to I-PLUS resources
R3 Goals

**L3 Engineered System**
- Provide repository for tools and Solutionware
- Administer CenSSIS information management
- Provide Web access to CenSSIS resources

**L2 Enabling Technologies**
- Provide compression, processing, transmission and storage resources
- Provide communications/networking support infrastructure and collaborative work environment

**L1 Fundamental Science**
- Develop efficient, flexible compression approaches
- Develop efficient/reliable transport schemes
- Develop new collaborative work environments
- Develop specific scalable computing approaches
- Develop image database, archiving/search schemes

**E Education**
- Enable distributed education component
- Enhance engineering curriculum
Typical Localized Probing & Mosaicing (LPM) Application

- Multiple Overlapping Optical Images
- Sensor Data

<table>
<thead>
<tr>
<th>Number of Images</th>
<th>Image Size in Pixels</th>
<th>Number of Bits/Pixels</th>
<th>Total Size of Data Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^4$</td>
<td>1K x 1K</td>
<td>24</td>
<td>30 GBytes</td>
</tr>
</tbody>
</table>

- Processing/Storage/Transmission Problem
- Transmission Time 40 min @ 100 Mb/s
- Compression Reduces to < 0.5 min
- Layered Progressive Scheme Useful
Typical Multi-View Tomography (MVT) Application

- Single Quadrature Microscope View of Mouse Egg
- Multiple Views per Slice and Multiple Slices per Reconstruction
- Multiple Reconstructions for Moving Objects

NU/MGH Data Set

<table>
<thead>
<tr>
<th>Number of Views/Slices</th>
<th>Image Size in Pixels/View</th>
<th>Number of Bits/Pixels</th>
<th>Number of Slices</th>
<th>Total Size of Data Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1K x 1K</td>
<td>24</td>
<td>100</td>
<td>3 GBytes</td>
</tr>
</tbody>
</table>

- Transmission Time 4 min @ 100 Mb/s
- For 100 Reconstructions is 400 min
- Compression Reduces to < 1 min
- Utilize 3-D Volumetric Compression
- Multiresolution Layered Scheme Useful
Remote Hyperspectral Sensing of Coastal Regions
UPRM Data Set

Typical Multi-Spectral Discrimination (MSD) Application

- Multiple Wavelengths
- Multiple Images
- Sensor Data

<table>
<thead>
<tr>
<th>Number of Images/Wavelength</th>
<th>Image Size In Pixels/View</th>
<th>Number of Bits/Pixels</th>
<th>Number of Wavelengths</th>
<th>Total Size Of Data Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^3$</td>
<td>1K x 1K</td>
<td>16</td>
<td>$10^3$</td>
<td>20 TBytes</td>
</tr>
</tbody>
</table>

- Transmission Time 56 hr @ 100 Mb/s
- Compression Reduces to < 30 min
- Exploit Correlation Across Wavelengths
- Layered Progressive Scheme Useful
Typical LPM/MSD Application

- **Focused or Pulsed Probe**
- **Focused or Gated Detector**

- **Wide Band Probe**
- **Narrow Band Detectors**

Multiple Overlapping Temporal Sequence of Multi-Spectral Images

- **Photomosaic of the Curved Human Retina RPI Data Set**

- **Sensor Data**

<table>
<thead>
<tr>
<th>Image Size In Pixels</th>
<th>Number of Bits/Pixels</th>
<th>Number of Channels</th>
<th>Frame Rate</th>
<th>Length of Procedure</th>
<th>Total Size Of Data Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>1K x 1K</td>
<td>12</td>
<td>2</td>
<td>30/sec</td>
<td>1-2 hrs</td>
<td>300-600 GBytes</td>
</tr>
</tbody>
</table>

- **Real-Time Transmission Requires 720Mb/s**
- **Compression Reduces to < 6 Mb/s**
- **Transmission Time 6.6 hr @ 100 Mb/s**
- **Layered Transport Scheme Useful**
How Can We Provide Adequate Communications/Networking Infrastructure?
B. Fundamental Research
Goal: Develop information management concepts, tools and infrastructure to support basic CenSSIS scientific mission
# The R3 Research Team

## R3a/b

### Data Compression/Multimedia Networks

<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>James Modestino</td>
<td>Leader, RPI</td>
</tr>
<tr>
<td>John Proakis</td>
<td>Co-Leader, NU</td>
</tr>
<tr>
<td>Shiv Kalyanaraman</td>
<td>RPI</td>
</tr>
<tr>
<td>Masoud Salehi</td>
<td>NU</td>
</tr>
<tr>
<td>John Woods</td>
<td>RPI</td>
</tr>
<tr>
<td>Gary Saulnier</td>
<td>RPI</td>
</tr>
<tr>
<td>Greg Suski</td>
<td>LLNL</td>
</tr>
</tbody>
</table>

## R3c/d

### Scalable Computation/Database and Metadata

<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>David Kaeli</td>
<td>Leader, NU</td>
</tr>
<tr>
<td>Robert Futrelle</td>
<td>Co-Leader, NU</td>
</tr>
<tr>
<td>Jose Cruz-Rivera</td>
<td>UPRM</td>
</tr>
<tr>
<td>Roscoe Giles</td>
<td>BU</td>
</tr>
<tr>
<td>Claudio Rebbi</td>
<td>BU</td>
</tr>
<tr>
<td>Elias Manolakos</td>
<td>NU</td>
</tr>
<tr>
<td>Waleed Meleis</td>
<td>NU</td>
</tr>
</tbody>
</table>
R3a/b Image and Data Management Issues

Issues and Requirements Depend on Sensor Modality Type and Communications/Storage Infrastructure
Multi-resolution, or layered, coding approaches
- Model-based compression schemes
- Feature-based compression schemes
- Multi-objective compression schemes
- Data fusion in compressed domain
- Joint Source-Channel Coding (JSCC) schemes
- Applications built on enhanced TCP/IP networking protocols
- Utilize existing lower-layer networking infrastructure

Barrier 3
Lack of Robust, Physics-Based Recognition and Sensor Fusion Techniques

Barrier 6
Lack of Rapid Processing and Management of Large Image Databases
R3a Compression Research Issues

Existing Approaches
- General Purpose
- Intended for Human Evaluation
- Based on HVS Distortion Criterion
- Generally Single Resolution Lossy Compression
- Compression Efficiency Saturated at < 100:1

CenSSIS Approach
- Physics-Based
- Intended for Specific Imaging Functions
- Based on Function-Specific Distortion Criterion
- Provides Progressive Lossy-Lossless Compression
- Compression Efficiency Improvements by 10-100

- Will Leverage Existing Approaches Where and When Possible
- Significant Improvements in Compression Efficiency Critical to Success of CenSSIS Scientific Mission
Physics-Motivated Region-Based Coding

Original-Healthy

Segmentation

Regions
- Macula/Fovea Region-Center
- Optical Disc-Left
- Vascular Structure
- Background

ROI Coding
- Macula Region-Near Lossless
- Optical Disc-Fine
- Vascular Structure-Medium
- Background-Coarse
Multiresolution Approach

- Coarse segmentation at lowest band
- Use physics-motivated segmentation
- Successive refinement with higher bands
- Differential ROI coding of regions
- Coding approach tailored to region type
- Exploit correlation across scales
- Progressive Lossy-Lossless encoding

- **Dyadic 3-Level Wavelet Subband Decomposition**

**Schematic Representation**

- **L** - Low Frequency Component
- **H** - High Frequency Component

- **LL-LL**
- **LH-LL**
- **LL-LH**
- **LH-LH**
- **H-L**
- **L-H**
- **H-H**

- **Lowband**
- **HalfBand**
- **Fullband**
Operational Use

Presence of Pathology Obscures Underlying Region Types
How to Reliably Extract Important Regions with Pathology
How to Exploit the Physical Manifestations of Disease Pathologies
Investigate and assess compression, transmission and storage requirements associated with initial level L2 testbeds

Develop preliminary modality-specific compression Approaches:
- LPM
- MVT
- MSD

Provide preliminary extensible collaborative work environment based on IP telephony model

Develop transition plan to Internet-2 and identify and formulate middleware and QoS research issues

Identify and demonstrate promising shared applications such as Virtual Lab
Components of R3 Research Thrust

R3 Image and Data Information Management

R3a Data Compression
- Representation
- Compression
- Transmission
- Storage

R3b Multimedia Networks
- Distribution
- Transport
- Collaboration
- Web Access

R3c Scalable Computation
- Computing
- Storage
- Scalability
- Solutionware

R3d Image Databases and Metadata
- Archiving
- Image Formatting
- Image Browsing
- Data Migration
### R3c. How Do We Meet the Real-Time Processing Requirements of Hyperspectral Imaging?

<table>
<thead>
<tr>
<th></th>
<th>Today</th>
<th>Future</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Image Resolution</strong></td>
<td>1,000 x 1,000</td>
<td>10,000 x 10,000</td>
</tr>
<tr>
<td><strong>Dynamic Range</strong></td>
<td>10 bits / pixel</td>
<td>12 bits / pixel</td>
</tr>
<tr>
<td><strong>Scanning Wavelengths</strong></td>
<td>200 samples</td>
<td>1000 samples</td>
</tr>
<tr>
<td><strong>Frame rate</strong></td>
<td>30 fps</td>
<td>30 fps</td>
</tr>
<tr>
<td><strong>Required Bandwidth</strong></td>
<td>60 Gbits/s</td>
<td>36 Tbits/s</td>
</tr>
</tbody>
</table>

- 4096-node system
- Enables spatial and spectral slicing
- UPRM will develop parallelized versions of CenSSIS algorithms for focal plane processing
- End-to-end hardware/software sensor design

**Georgia Tech/UPRM SIMPil Architecture**

**SIMD Focal Plane Architecture**

**ACU**

**Image Resolution**

Today: 1,000 x 1,000
Future: 10,000 x 10,000

**Dynamic Range**

Today: 10 bits / pixel
Future: 12 bits / pixel

**Scanning Wavelengths**

Today: 200 samples
Future: 1000 samples

**Frame rate**

Today: 30 fps
Future: 30 fps

**Required Bandwidth**

Today: 60 Gbits/s
Future: 36 Tbits/s
R3c. How can CenSSIS Applications Scale Up to Use Real Data Sets?

- Exploit NSF Partnerships for Advanced Computational Infrastructure (PACI) resources
- Construct Beowulf-class clusters to provide low-cost, scalable processing (BioMed 8-node cluster and NSF MRI 32-node cluster)
- Utilize local NSF PACI/GRID computational resources at Boston University (Mariner Center)
- Develop parallel programming toolsets around MPI and OpenMP standards

CCS Center for Computational Science
SCV Scientific Computing and Visualization Group
How Do We Store/Organize CenSSIS Information to Enable Rapid Assessment and Dissemination?
## Database (Oracle 8i)

<table>
<thead>
<tr>
<th>Image ID</th>
<th>Image type</th>
<th>Image format</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPM099</td>
<td>Hyperspectral</td>
<td>?</td>
</tr>
<tr>
<td>MVT002</td>
<td>ECG</td>
<td>?</td>
</tr>
<tr>
<td>MSD321</td>
<td>Diffusive wave</td>
<td>?</td>
</tr>
<tr>
<td>MVT061</td>
<td>CAT scan</td>
<td>?</td>
</tr>
</tbody>
</table>

### Metadata - eXtensible Markup Language (XML)

- Sensor
- Manufacturer
- Calibration date
- Related Publications
- Serial #
- Compression algorithm
- Software version
- HLLs supported

### Additional Information

- *TBs of RAID storage*
- *Replication and distribution*

### Image Format Standards

- The data about the data
- Contextual browsing

\[ y = C(x, z) + n \]
Need for Software Engineering standards
- Standard programming interfaces (e.g., APIs)
- Version and release control
- Testing standards

Enable rapid prototyping of new algorithms

Develop parallel implementations for compute-bound algorithms and models

Provide libraries of Solutionware toolboxes, validated in our L2 testbeds

Leverage PACI and industrial-partner software development tools

R3d. How do we develop the subsurface imaging toolbox of the future?
R3c/d First Year Anticipated Results

- Complete benchmarking of SIMPiL focal plane architecture and applications
- Set up CenSSIS web, image storage and database servers
- Design and implement CenSSIS image database system
- Develop CenSSIS image format standards
- Provide preliminary specification for CenSSIS metadata databases
- Develop CenSSIS software library control system
- Complete prototype of JavaPorts programming toolset
C. R3 Impact and Timeline
R3 Impact on Education

- Deliver experiments and testbeds to the classroom
  - Immediate practice of classroom theory
  - Remote virtual laboratory*
- New Course modules
  - Digital Image and Video Processing (ECSE-6630)
  - Digital Picture Processing (ECSE-6640)
  - Digital Computer Architecture (ECE-3391)
- Collaborative work environment
  - Distance education and joint seminars
- Allow students to explore using CenSSIS Solutionware and image databases
  - Develop computer-based analysis skills applied to subsurface sensing and imaging**

*Already being piloted in RPI Nuclear Navy program
**Already being piloted in NU GE1103 course
R3 Technology: Real-time Hyperspectral Imaging

**Barrier 5**
Lack of Optimal End to End Sensor Design Methods

**Barrier 7**
Lack of Validated, Integrated Processing and Computation Tools
**R3 Technology: Scalable Inverse Solutions**

**y = C(x, z) + n**

for (i = 0; i < n, i++)

**Inverse ECG speedup**

<table>
<thead>
<tr>
<th>Nodes</th>
<th>Speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>10</td>
<td>8</td>
</tr>
</tbody>
</table>

**Linear speedup**

**Solutionware**

**Barrier 7** Lack of Validated, Integrated Processing and Computation Tools
R3 Technology: Commercial Databases, Compression and Multimedia Networks

Barrier 6: Lack of Rapid Processing and Management of Large Image Databases
<table>
<thead>
<tr>
<th>Year</th>
<th>Compression, Transmission and Storage</th>
<th>Multimedia Networking and Collaborative Work Environments</th>
<th>Focal-plane architectures and algorithms</th>
<th>Image databases and metadata</th>
<th>Solutionware</th>
<th>Education</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Yr 1</strong></td>
<td>Assessment of L2 needs</td>
<td>Deliver image DB</td>
<td>Simulation of SIMPiL</td>
<td>Deliver image DB</td>
<td>Construct library system</td>
<td><strong>Yr 1</strong></td>
</tr>
<tr>
<td><strong>Yr 2</strong></td>
<td>New modality-specific compression approaches</td>
<td>Develop web-based query capabilities</td>
<td>Development of parallel hyperspectral applications</td>
<td>Develop web-based query capabilities</td>
<td>Develop web-based rapid prototyping library system</td>
<td><strong>Yr 2</strong></td>
</tr>
<tr>
<td><strong>Yr 3</strong></td>
<td>Continued development for new L2/L3 approaches</td>
<td>Maintenance and upgrades to image database</td>
<td>Evaluation of performance on SIMPiL hardware</td>
<td>Development/implementation of XML-based metadata database</td>
<td>Continued support and test of evolving Solutionware libraries</td>
<td><strong>Yr 3</strong></td>
</tr>
<tr>
<td><strong>Yr 4</strong></td>
<td></td>
<td></td>
<td>Continued Investigation of Transport and QoS Issues</td>
<td>Contextual browsing</td>
<td></td>
<td><strong>Yr 4</strong></td>
</tr>
<tr>
<td><strong>Yr 5</strong></td>
<td></td>
<td></td>
<td>Continued Development Based on Internet Evolution</td>
<td></td>
<td></td>
<td><strong>Yr 5</strong></td>
</tr>
</tbody>
</table>
D. I-PLUS
I-PLUS

- Develops a unified problem-solving platform
- Includes techniques, tools and infrastructure
- Combines the capabilities of the L2 testbeds and Systems Integration
- Allows CenSSIS to address new sets of real-world problems
  - Leveraging the combined resources, techniques and lessons learned from multiple sensing and imaging modalities
  - Provides a natural path for technology transfer to outreach collaborators and industry partners
I-PLUS - Integrated Probes to Look Under Surfaces

**Systems Integration**

D. Kaeli, NU
Leader

---

**CenSSIS Software**

- **Resource Center**
  - CenSSIS Software Engineer
  - Software libraries
  - Programming Tools and Parallel Systems Management

**CenSSIS Image and Data Information Systems**

- Database Administrator
  - Access to parallel processing resources
  - Website Administration