Building Models for Solving General Inverse Problems

Harold Trease
Sensor and Decision Analytics Group
Computational Science and Mathematics Divison
Computational Information and Science Directorate
Pacific Northwest National Laboratory

Abstract

- In this case study we make use of a combination of inverse methods, forward simulations and uncertainty quantification to develop a method for characterizing a source or media based on known sensor data. In principal this is a fairly general concept, where based on known sensor data and constraints we iteratively define a model for projecting back to an unknown source through an unknown media, such that we can then define and run a forward simulation with initial conditions, boundary conditions and model closure assumptions to produce synthetic sensor data. Then based on the comparison of actual sensor data vs.. synthetic sensor data we can refine the inverse model and/or modify the constraints and iterate the process. In this presentation we will also discuss some of the underlying mathematical and computational considerations for solving the general inverse problem. Examples of the application of this process will be shown in the context of non-proliferation, treaty verification programs and image processing.

Sensor ←→ Media ←→ Source

- Image processing
- Information processing
- Modeling and simulation: forward and inverse

- Sensor data explosion: 1000 X sensor data to ~infinite unstructured streaming data

- Decisions: Sensor ←→ Media ←→ Source

Application Areas

- Biology (PNNL's Data Intensive Computing Initiative and NIH)
  - Ion mass spectrometry
  - Computed tomography (lungs and hearts)
- Subsurface transport (OFS / ASCR / SciDAC / ITAPS)
  - Migration of heavy metal waste
  - Carbon sequestration
- Border Control (DHS)
  - Passive and radiography analysis of shipping cargo containers
- Atmospheric and aquatic plume detection and analysis (NNSA)
  - Chemical identification, chemical processes (Hyperspectral analysis)
  - Chemical detection network
- Seismic monitoring (NNSA)
  - Nuclear explosion monitoring (explosion vs. earthquake)
- Information processing (IC)
- Standoff detection of explosives (PNNL/DHS Initiative)

Inverse Problems and Inverse Solutions

- Sensor data is given
  - Sparse sensor data (hyperspectral, VACIS)
  - Dense sensor data (biology, videos)
- The general inverse problem has no unique solution
  - Ill-posed problems
  - Ill-conditioned problems
- Forward simulations map parameter space one point at a time
- Requires domain specific knowledge to constrain solution space

Sources of Uncertainty and Errors

- Sensors
  - Signal-to-noise
  - Response functions
- Numerical approximations
  - Approximating physical system using systems of PDEs
- Numerical integration error
  - Roundoff, precision, truncation, closure
- Database uncertainty

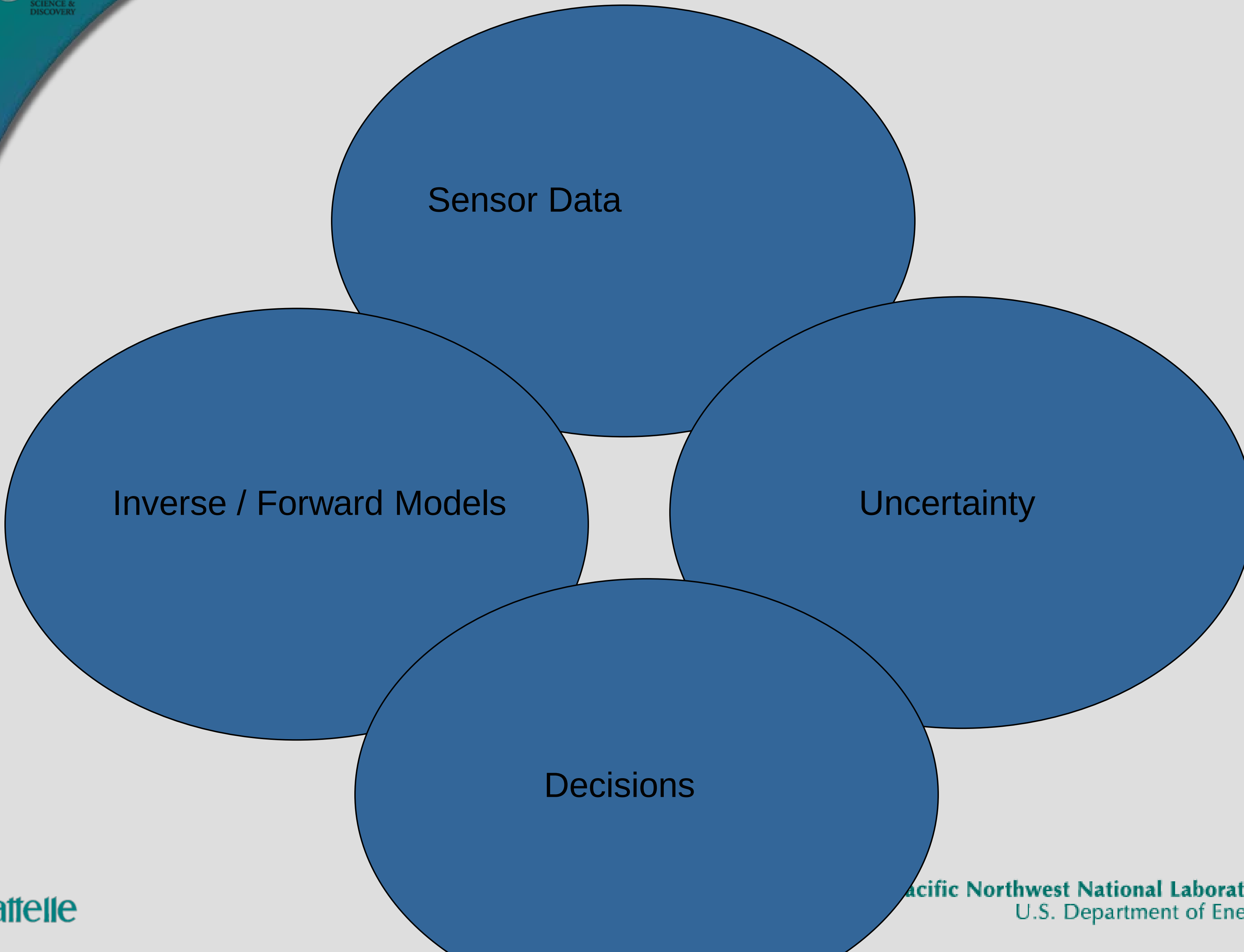
Inverse Solution Methods

- Mapping parameter space using forward solutions
  - Deterministic methods (PDEs and ray tracing)
  - Monte Carlo methods
- Populate a covariance matrix to indicate how everything changes with respect to everything
  - How complete does an approximate inverse have to be?
- Sensor data ➔ Inverse Models ➔ Forward Models
  - S = G (U)
  - Sensor data = Models (PDEs, initial conditions, boundary conditions, …)
- [Classifier: X(Images) ➔ Y(Benchmarks)]

Two Scenarios of Known vs. Unknown Data

- Sensor Data ←→Media ←→Source
  - Unknown(media), Known(source, sensor data): Cargo analysis, detection of explosives, biology
  - Unknown(source, media), Known(sensor data): plume detection and characterization, explosion monitoring

Decisions: Sensor ←→Media←→Source



Hyperspectral: Sensor ←→Media←→Source

Sensor Data:
- Hyperspectral Images

Inverse ←→ Forward Models:
- Plume profile (diffusion, hydro, gravity, bouyency, etc.)
- Plume growth (time-dependence)
- Scene Geometry

Uncertainty:
- Parameter Uncertainty

Decisions:
- Chemical signatures
- Chemical process
- Plume location
- Plume fate and transport

Seismic: Sensor ←→ Media ←→ Source

Sensor Data:

- Sonograms

Inverse ←→ Forward Models:

- Time-reversed wave propagation
- Substructure media
- Substructure geometry

Uncertainty:

- Parameter Uncertanity
- Media Uncertainty

Decisions:

- Earthquake vs. Explosion ?
- Where, how deep, how big ?

Image reconstruction: Sensor ←→Media←→Source

Sensor Data:

- MRI/NMR, CT, ET, Confocal

Inverse ←→ Forward Models:

- Back projection into datacube

Uncertainty:

-Parameter Uncertanity

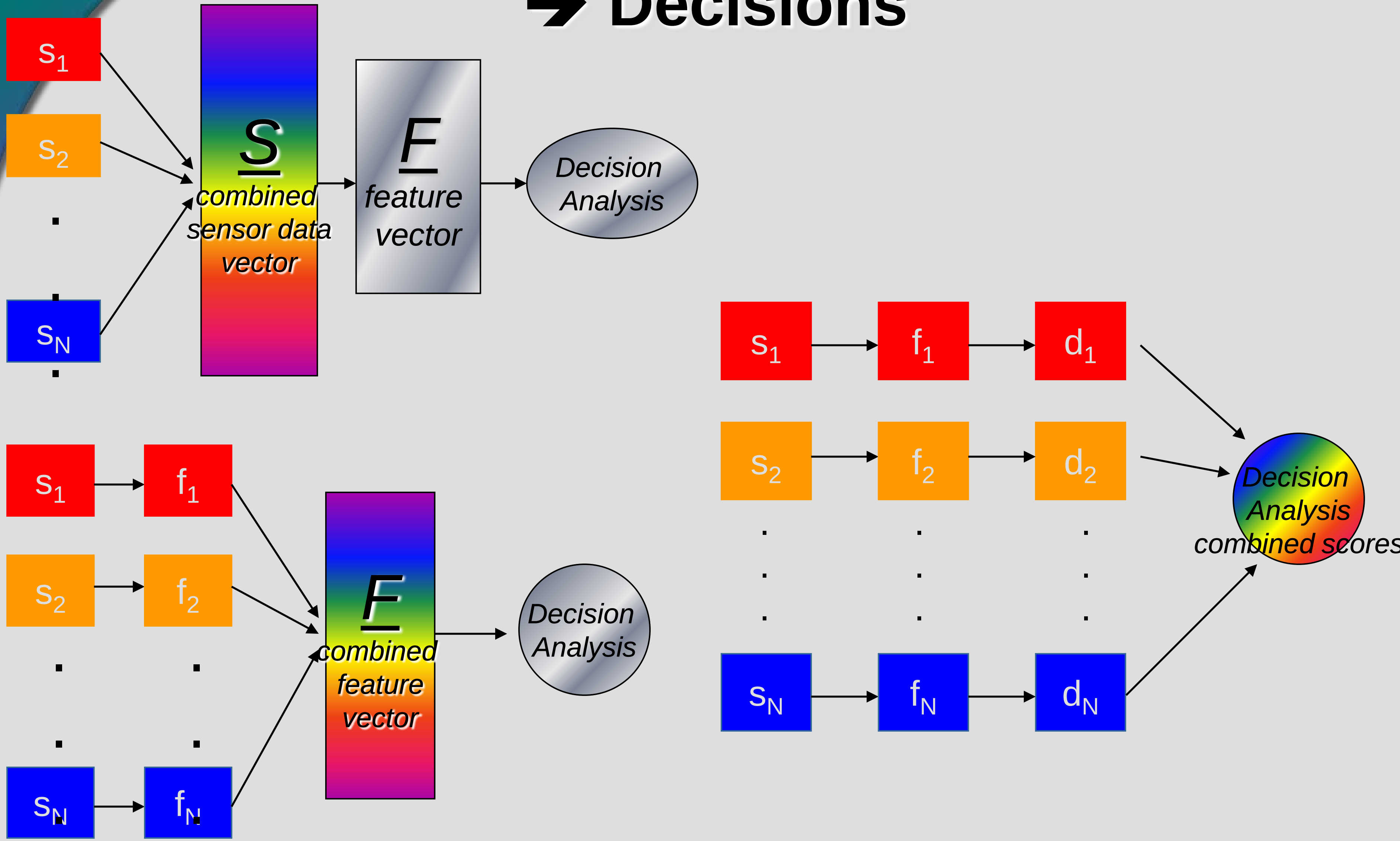
- Media Uncertainty

Decisions:

- Object extraction

- Media characterization

Multi-Sensor Integration: Raw Sensor Data ➔ Decisions



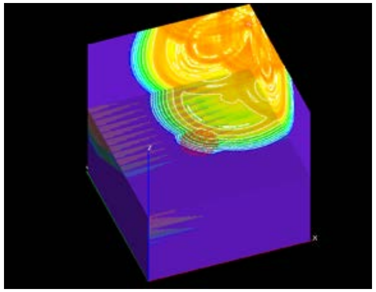
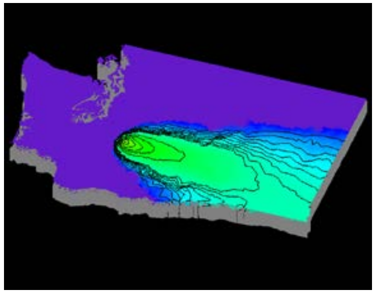
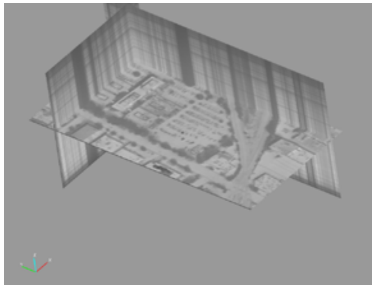
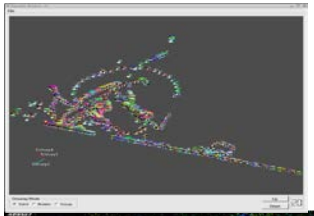
Sensor ←→ Media ←→ Source

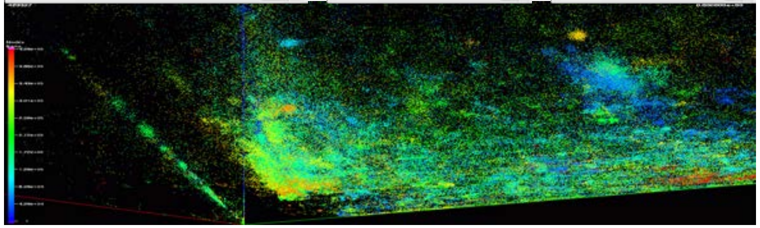
- Image processing
- Information processing
- Modeling and Simulation

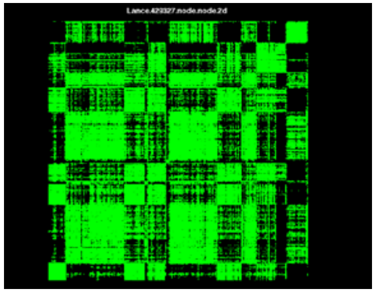
P3D: Computational Physics and Information Modeling, Simulation and Prediction Framework

- **Applications:**
  - CFD, CMM, CEM
  - Modeling, simulation and prediction of coupled continuum and discrete information
  - Image processing: Large volumes of static images and streaming video databases
  - Computational mathematics framework
- **Capability:**
  - Coupled continuum physics modeling and simulation, including: hydrodynamics, structural mechanics, transport phenomena, electromagnetics problems using finite-volume integration techniques
  - Solves coupled continuum and discrete problems
  - Partitions and solves large graph problems
  - Determines and tracks the principal information flow directions and trends
  - The P3D environment is useful in looking for and discovering “special” cases and counter examples in mathematical theories
- **Algorithms**:
  - High-fidelity geometry and mesh generation
  - Generate large N-dimensional meshes
  - Solves coupled discrete or continuum process(es)
  - High-performance, parallel implementation. Scalable from laptops to super-clusters.
- **P3D Codes:** DDV/DDATK, OSO, NWGrid/NWPhys, GMV
- **Authors:** Harold Trease, Et al.

Examples:
- Wave propagation
- Modeling of plumes
- Hyperspectral imaging processing
- PCA clustering of images & video
- Large-scale graph analysis

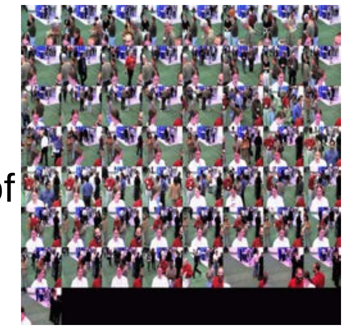







High-Performance Video Analysis: Surveillance, Forensics, Biometrics

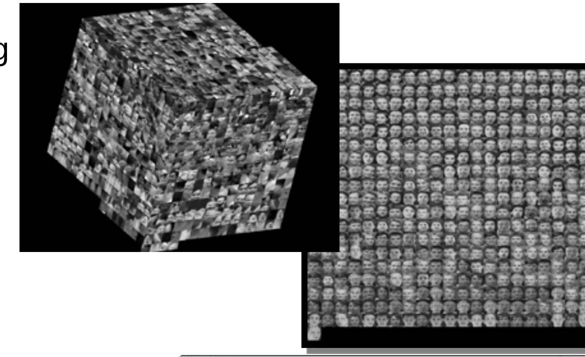
- **Applications:**
  - Video surveillance, forensics and biometrics
  - Analyzing shopper's patterns
- **Capability:** Multi/Many cameras, lots of data [demonstrated 1 DVD/sec, ~120,000 frames/sec, 41.6Gbytes/sec]
  - Have I seen this person?
  - Where and when?
  - Whom were they with?
  - What were they doing (possible intent)?
- **Algorithms**:
  - Information, statistical and (invarent) geometry algorithms
  - Face extraction and recognition
  - Tracking in space and time
  - 3-D geometry reconstructions of faces and scenes
  - High-performance, parallel implementation. Scalable from laptops to super-clusters.
- **P3D Codes:** DDV/DDATK, OSO, NWGrid/NWPhys, GMV
- **Authors:** Harold Trease, Robert Farber, Ryan Mooney, Tim Carlson, Et al.
- **Data Sources:** SC2005 videos

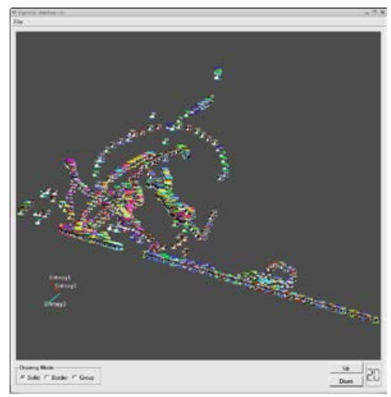


Streaming video

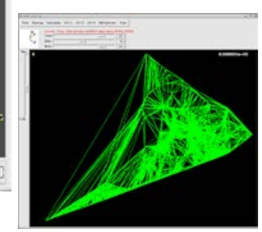
➔

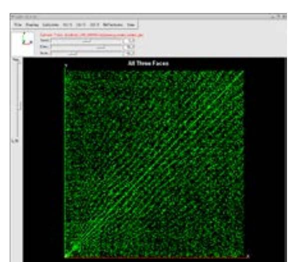
Face database





Building Social Network Graphs From Face Data

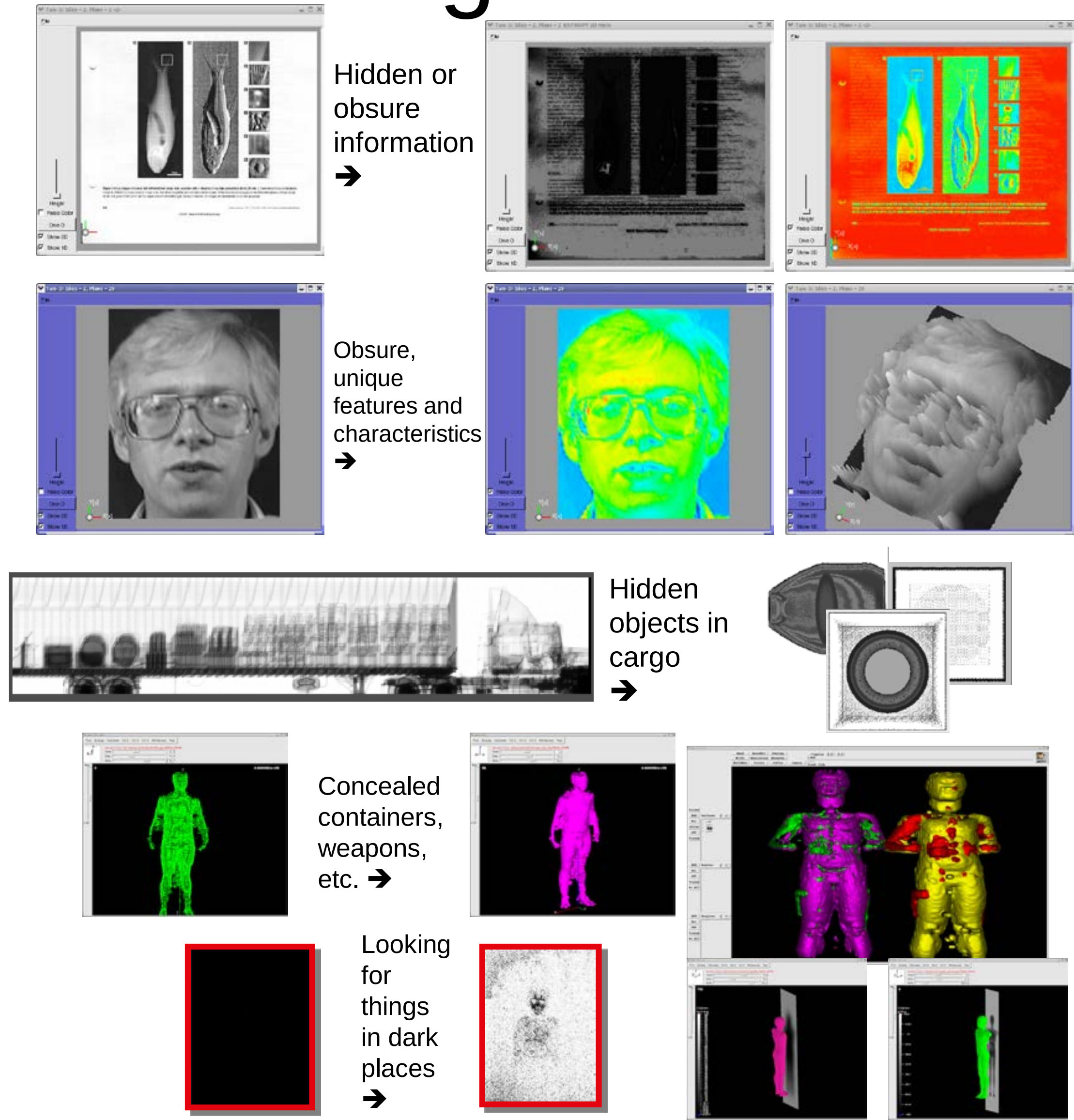




Partitioning face based graphs to discover relationships

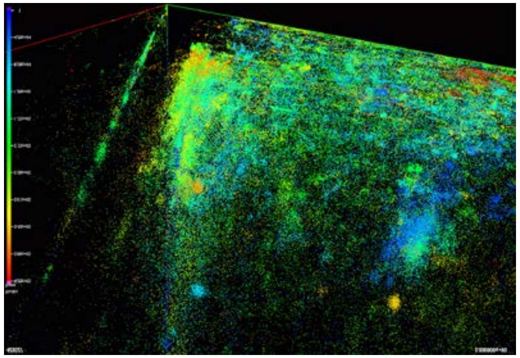
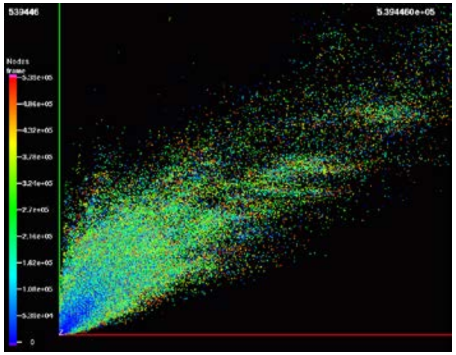
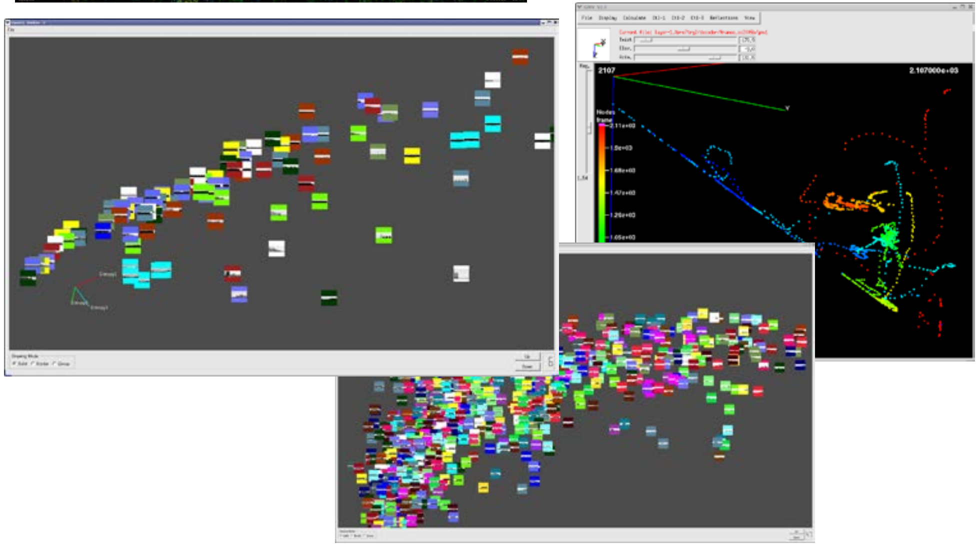
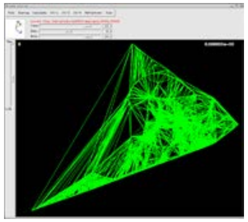
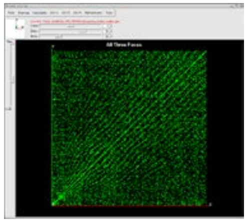
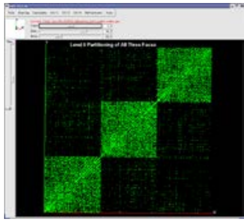
Seeing and Finding the Unseen in Static and Video Image Data

- **Applications:**
  - Detecting anomalies, outliers, fakes, watermarks, etc.
- **Capability:**
  - High-performance, parallel anomaly detection
  - Large databases and multi-stream video data
- **Algorithms**:
  - Transformations of image data into interesting spaces
  - Information regression and prediction
- **P3D Codes:** DDV/DDATK, OSO, GMV
- **Authors:** Harold Trease, John Fowler, Lynn Trease
- **Data Sources:** X-ray, VACIS, Intellifit, Safeview, Internet faces



Classification, Characterization and Clustering of High-Dimensional Data

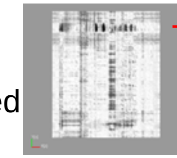
- **Applications:**
  - Static image and video data analysis
  - Border control (looking for drugs, people, etc. in commerical shipping cargo)
  - Organizing desktops and disk drive images
- **Capability:**
  - Find interesting patterns and clusters in high-dimensional data.
  - Predict the principal information flow paths to follow trends
  - Incorporate conditional dependence and independence using PDFs
  - Multi-INT, multi-sensor data fusion
- **Algorithms**:
  - Clusters data by using signatures of high-dimensional data, represented and manipulated as large sparse graphs
  - Classification, characterization, conditional dependence/independence algorithms uses the measure of the “distance” between PDF’s
  - High-performance, parallel implementation. Scalable from laptops to super-clusters.
- **P3D Codes:** DDV/DDATK, OSO, NWGrid/NWPhys, GMV
- **Authors:** Harold Trease, John Fowler, Lynn Trease, Robert Farber
- **Data Sources:** SC2005 videos, Discovery Channel, VACIS

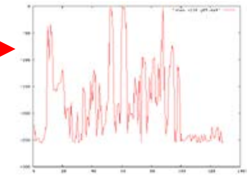
4-D (Spectral/Spatial/Time) Hyperspectral Image Processing and Analysis

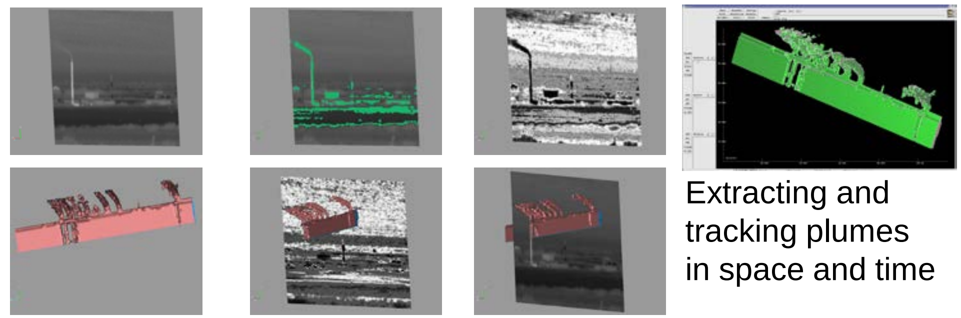
- **Applications:** Remote sensing, tracking and targeting
  - Chemical plume detection/tracking/prediction
  - Structural reconstruction and identification
- **Capability:**
  - Chemical end-member extraction
  - Plume extraction and tracking (space/time)
  - Plume modeling and simulation in space/time
- **Algorithms**:
  - Unique transformations based on:
    - Information content
    - Statistical quantities (PDFs)
    - Geometric invariants
  - Algorithms represented (4-D) data in the form of “datacubes”
  - High-performance, parallel implementation. Scalable from laptops to super-clusters.
- **P3D Codes:** DDV/DDATK, OSO, NWGrid/NWPhys, GMV
- **Authors:** Harold Trease, John Fowler, Lynn Trease
- **Data Sources:** Hyperspectral (128 infrared bands)

Image data ➔ Transformed data



To ➔ Spectra





Extracting and tracking plumes in space and time

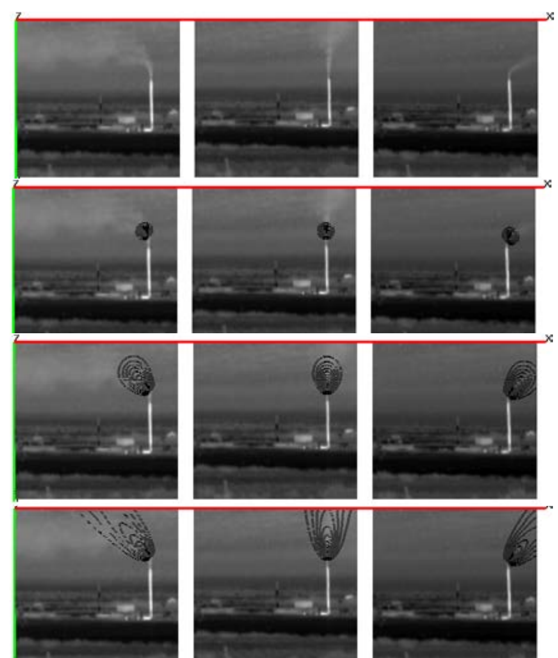
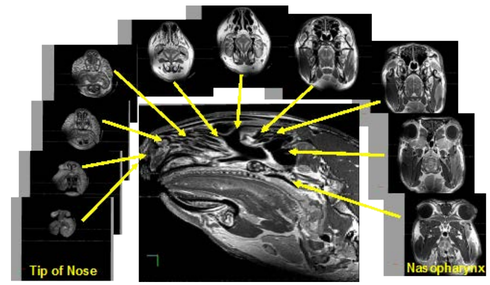
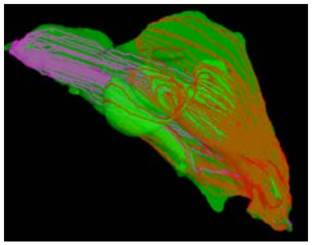


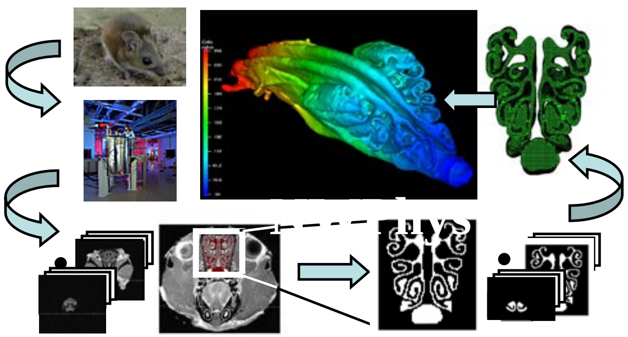
Image data 🡐 Geometry Models 🡐 Physical Models 🡐 Time-dependent simulations

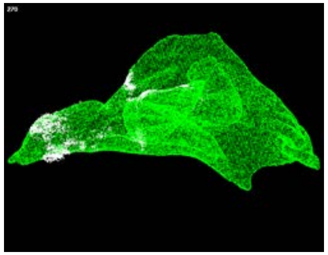
Computational Biology: The Virtual Respiratory Tract

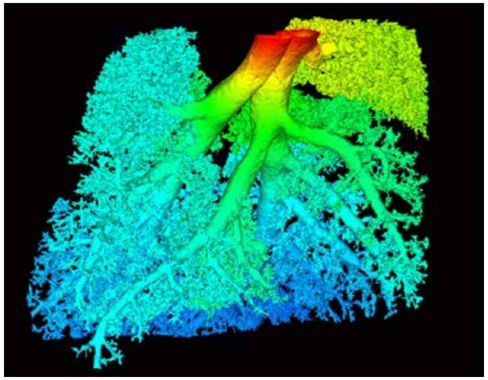
- **Applications:**
  - Bioterrorism related to the inhalation of aerosols
  - Pollution, chemicals, respirator design
  - NIH health related biomedical applications (animal ➔ human studies)
- **Capability:**
  - Image processing, segmentation and feature extraction using NMR/MRI and CT scans.
  - Particle dynamics and chemical reactive transport
  - Coupled hydrodynamic and material response
  - High-fidelity, geometry produces quantitative surface area and volume calculations
- **Algorithms**:
  - Finite volume integration
  - Unstructured boundary-fitted / volume-filling meshes
  - Hydro, structural mechanics, reaction/diffusion bio-physics models.
  - High-performance, parallel implementation. Scalable from laptops to super-clusters.
- **P3D Codes:** DDV, OSO, NWGrid/NWPhys, GMV
- **Authors:** Harold Trease, John Fowler, Lynn Trease
- **Data Sources:** NMR/MRI and CT





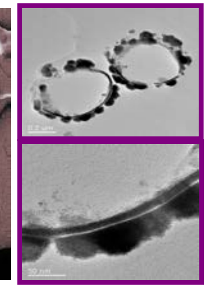




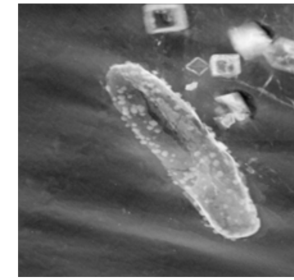


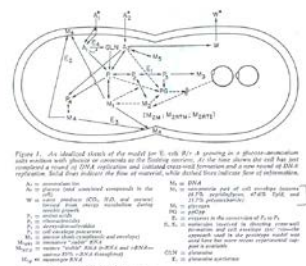
Computational Biology: The Virtual Microbial Cell Simulator in P3D

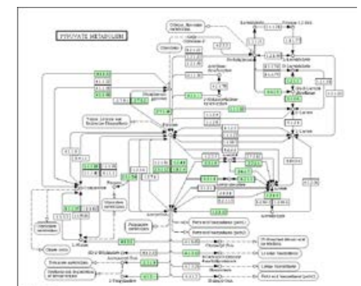
- **Applications:**
  - Bioremediation of heavy metal, radioactive waste for environmental cleanup
  - Bioterrorism related to the inhalation of aerosolized microbes
- **Capability:**
  - Image process based on electron tomography, confocal microscopy, and NMR.
  - Incorporates multiple spatial scales from single cells through communities.
  - Flux based, genome-scale metabolic pathway descriptions of the production and use of energy within and between cells.
- **Algorithms**:
  - Finite volume spatial integration coupled to Global Flux Balance for transport
  - Unstructured boundary-fitted / volume-filling meshes
  - Reaction/Diffusion transport models
  - High-performance, parallel implementation. Scalable from laptops to super-clusters.
- **P3D Codes:** DDV, OSO, NWGrid/NWPhys, GMV
- **Authors:** Harold Trease, TSTT group
- **Data Sources:** Electron microscopy/tomography



Electron tomography ➔ Geometry models







$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C_1 \xrightarrow{k_2} E + P$$

$$S + C_1 \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} C_2 \xrightarrow{k_4} C_1 + P$$

$$\frac{ds}{dt} + \vec{u} \cdot \vec{\nabla} s = -\vec{\nabla} \cdot D_s \vec{\nabla} s - k_1 s e + k_{-1} c_1 - k_3 s c_1 + k_{-3} c_2$$

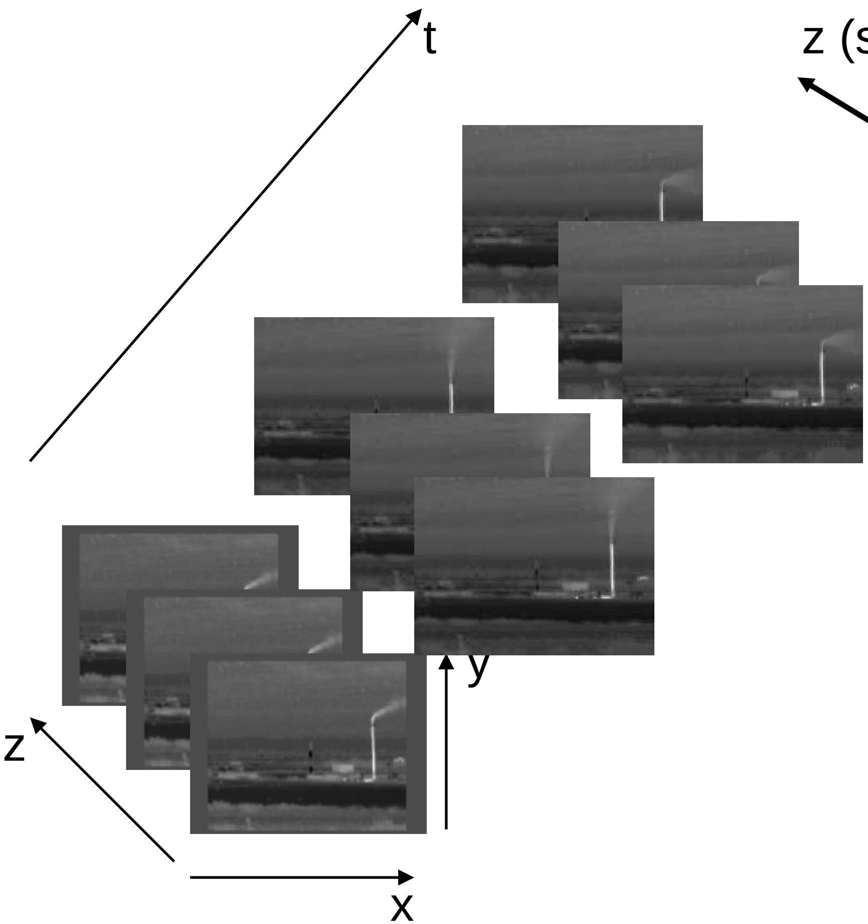
$$\frac{dc_1}{dt} + \vec{u} \cdot \vec{\nabla} c_1 = -\vec{\nabla} \cdot D_{c_1} \vec{\nabla} c_1 + k_1 s e - (k_{-1} + k_2) c_1 - k_3 s c_1 + (k_4 + k_{-3}) c_2$$

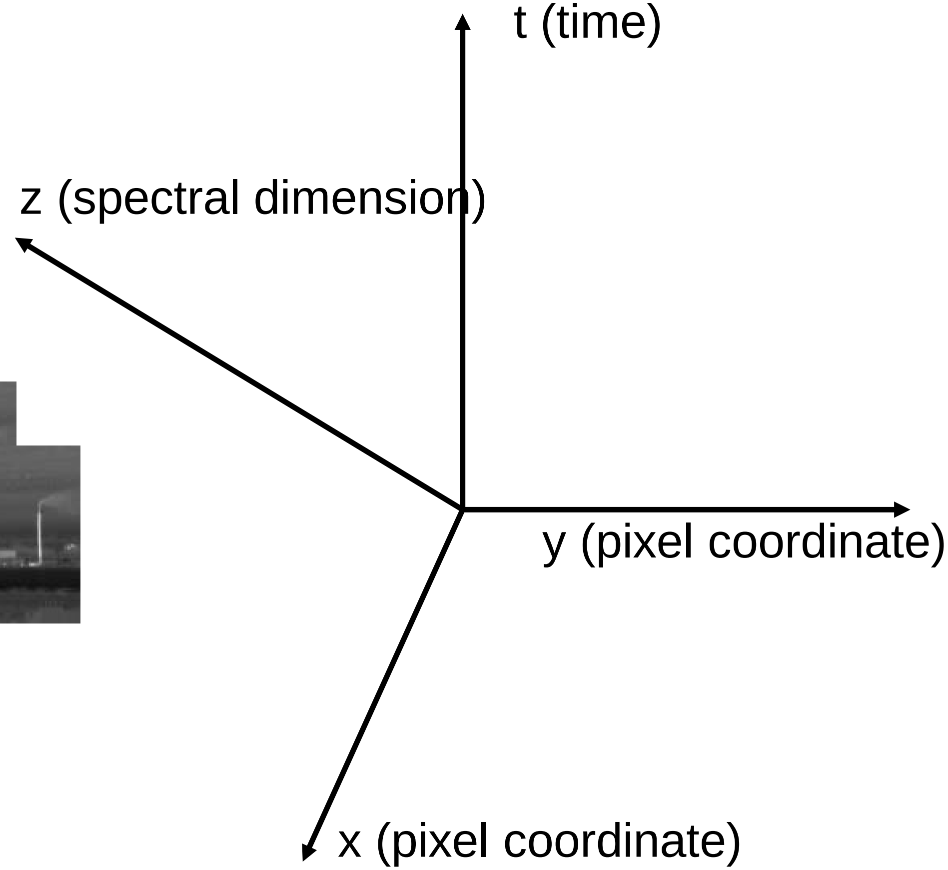
$$\frac{dc_2}{dt} + \vec{u} \cdot \vec{\nabla} c_2 = -\vec{\nabla} \cdot D_{c_2} \vec{\nabla} c_2 + k_3 s c_1 - (k_4 + k_{-3}) c_2$$

$$e + c_1 + c_2 = e_0$$

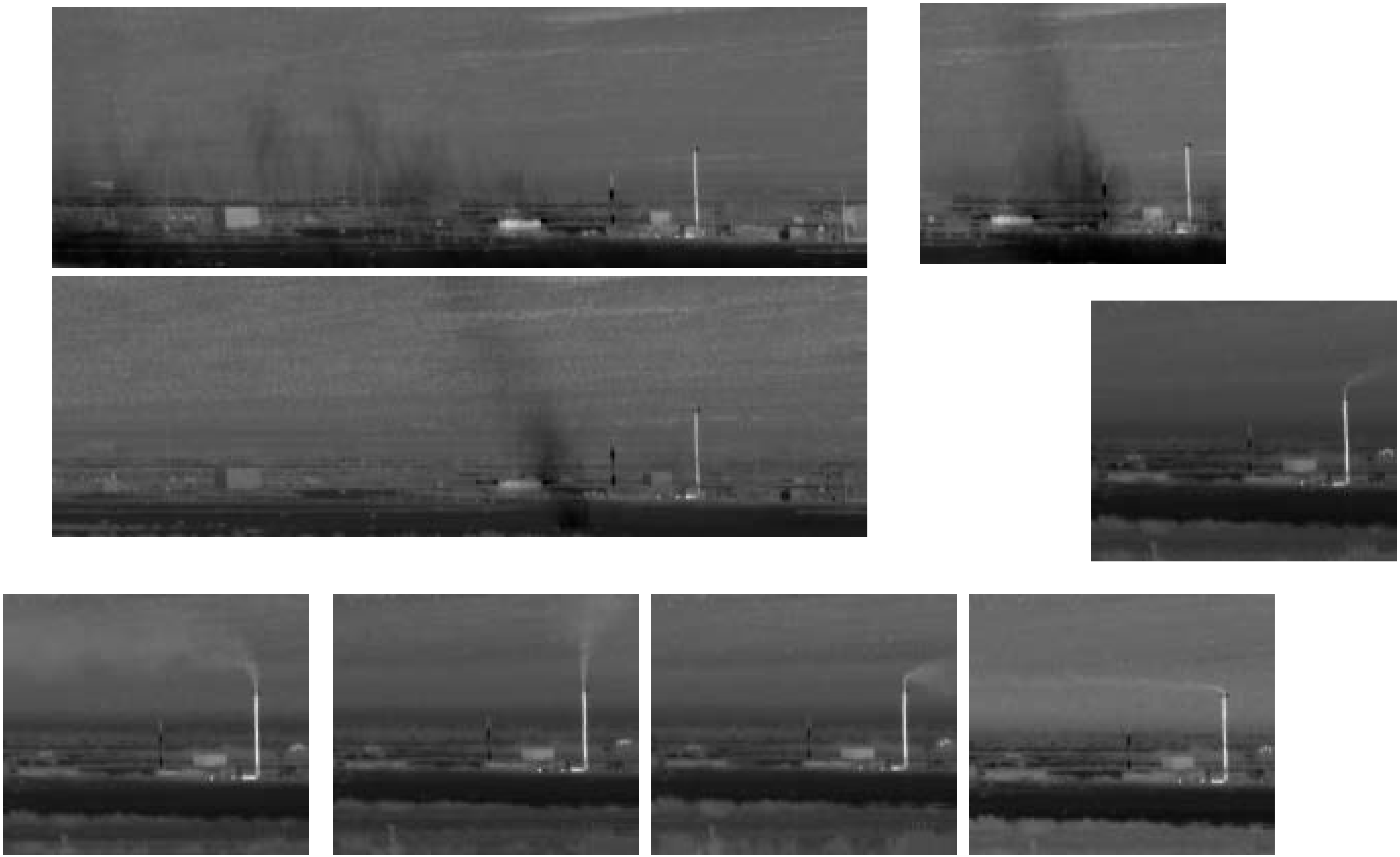
Biological models ➔Biologist's models ➔Mathematical Models ➔Simulations

4-D Hyperspectral Cubes



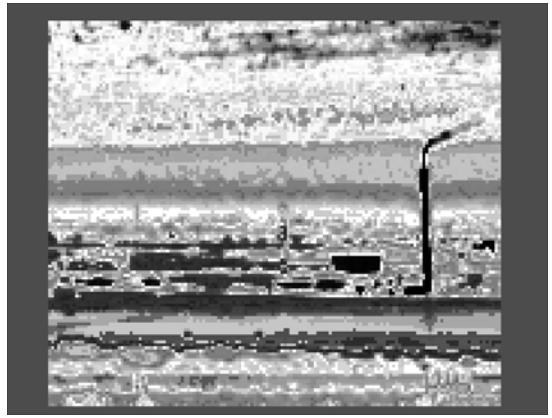


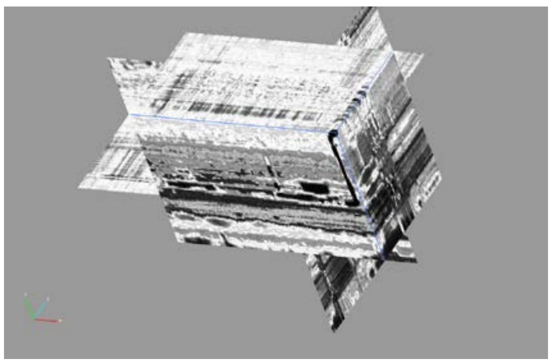
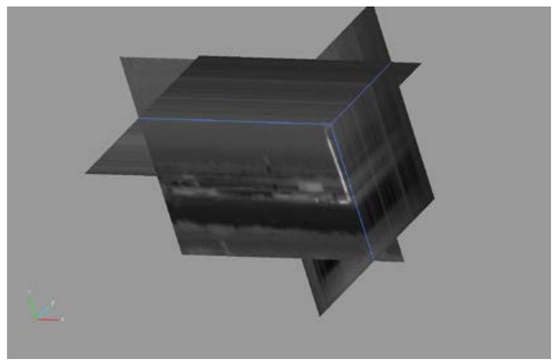
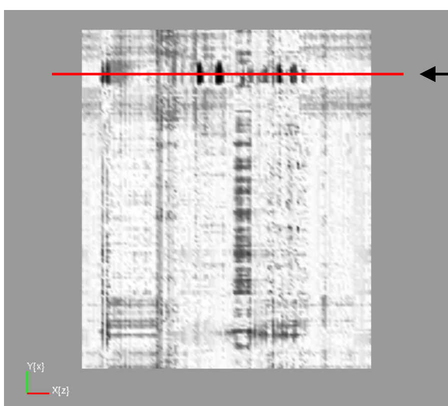
Possible Plumes



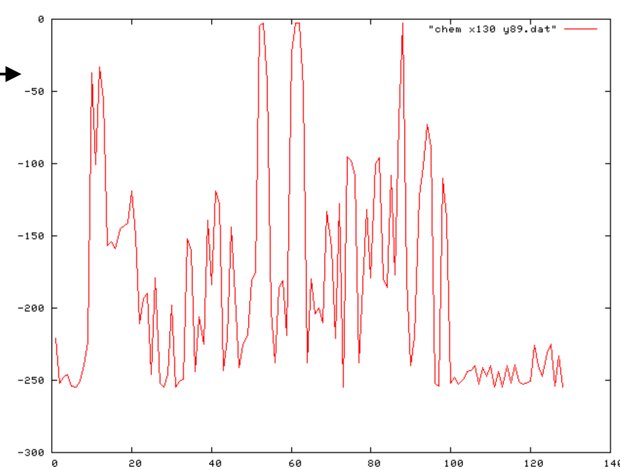
Stabilize decompose1.2

The process of using DDATK to determine possible plume location and chemical signature

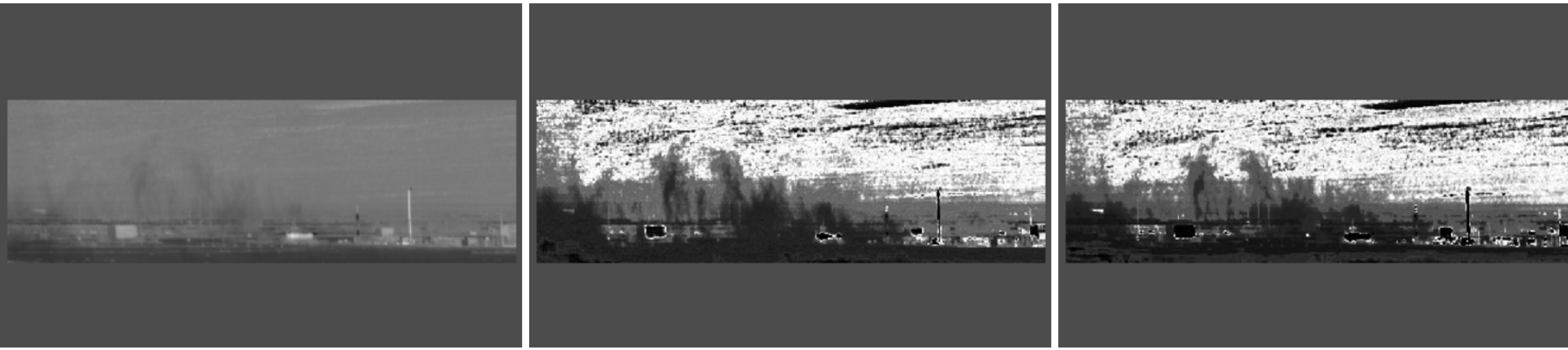



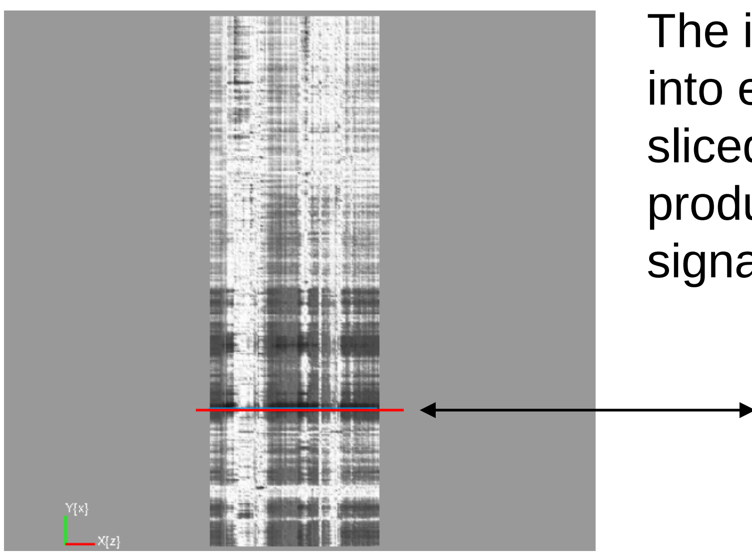




The image is transformed into entropy space and sliced across the bands to produce a chemical signature

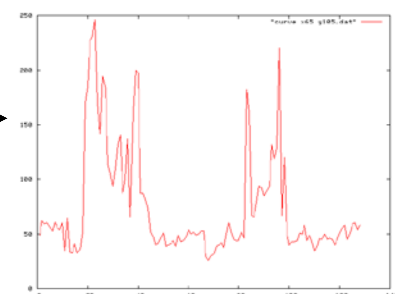


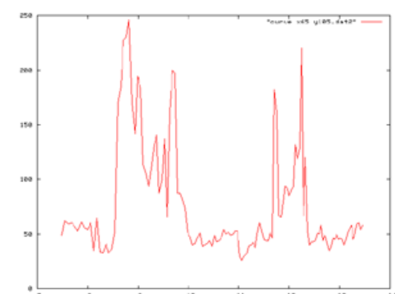
The process of using DDATK to determine possible plume location and chemical signature



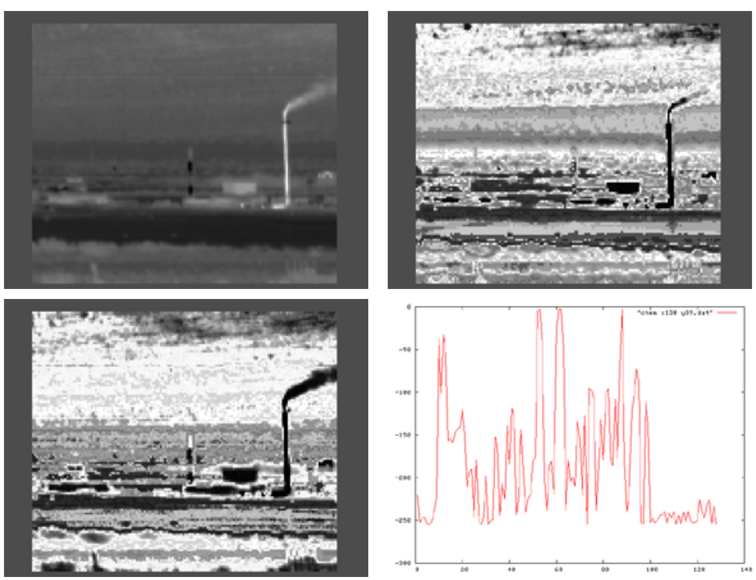


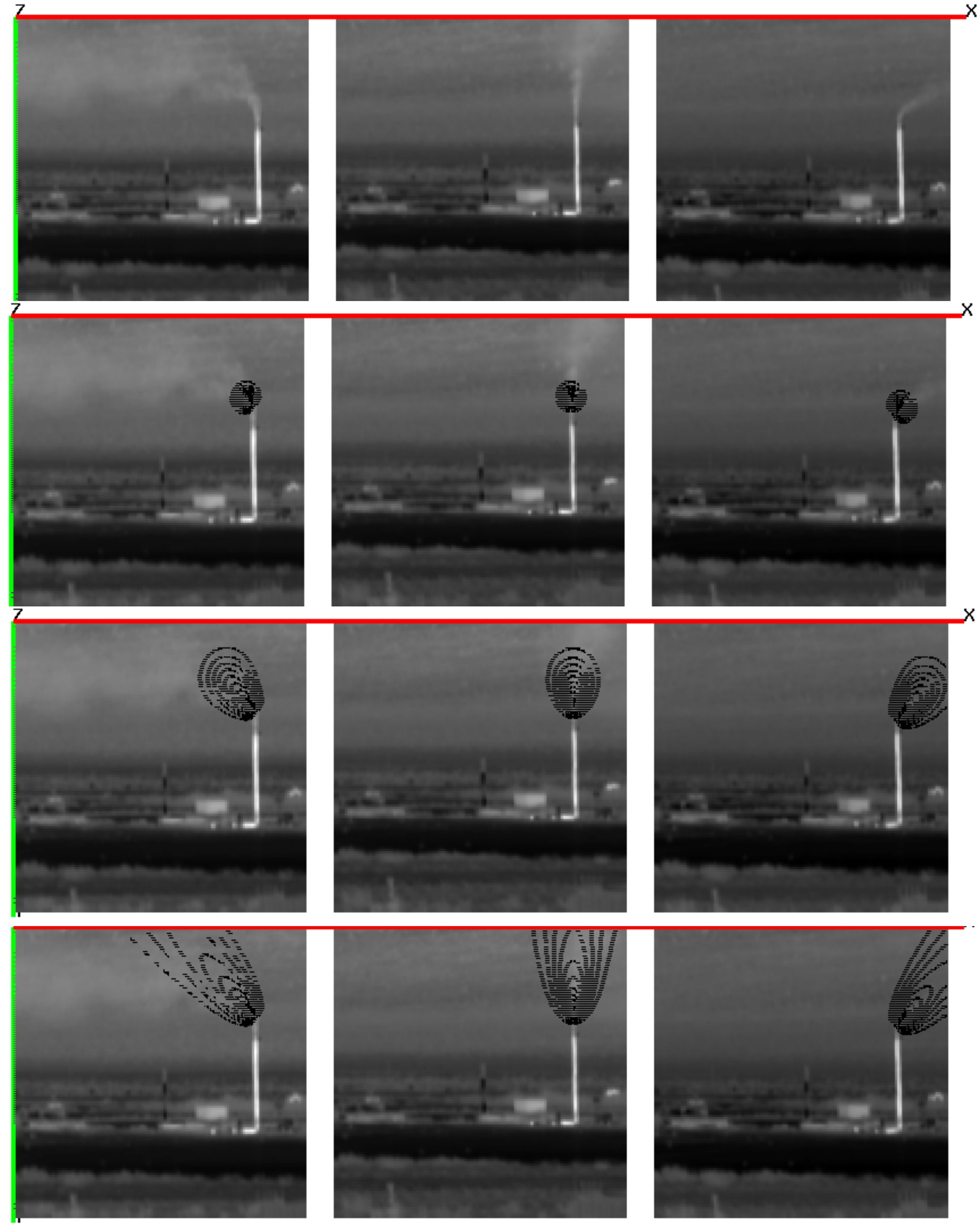
The image is transformed into entropy space and sliced across the bands to produce a chemical signature





Using image analysis to determine a possible source location, plume chemistry, geometry and assumptions about the environmental conditions (wind, temperature, etc.) we then perform simulations senerios to start to determine uncertainty map for the problem. Using the uncertainties we can update the forward model and iterate this process.





Current VACIS Processing

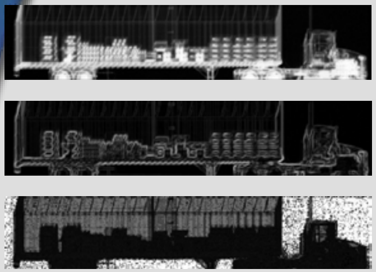
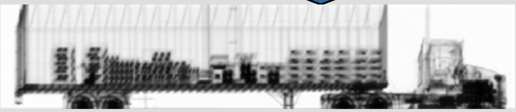


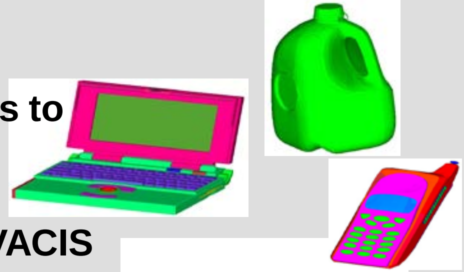
Image filters applied

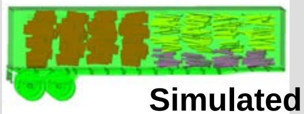


Truck cab and wheels cropped

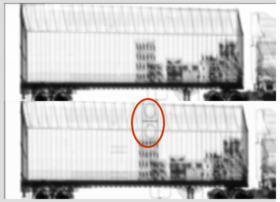
Cargo extracted

Generate geometries to simulate objects to imbed in VACIS images

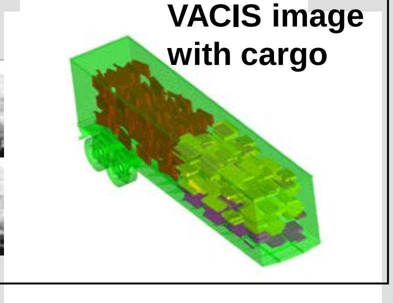




Insert Object in VACIS Image

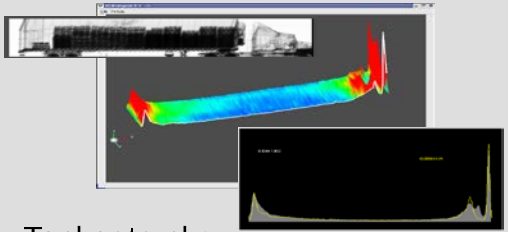


Simulated VACIS image with cargo

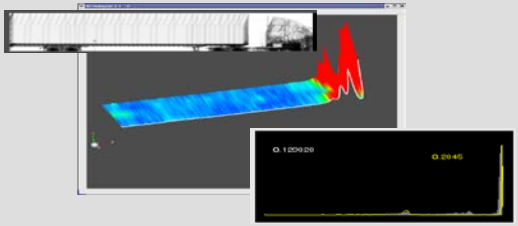


Entropy used to find signature of truck type and cargo

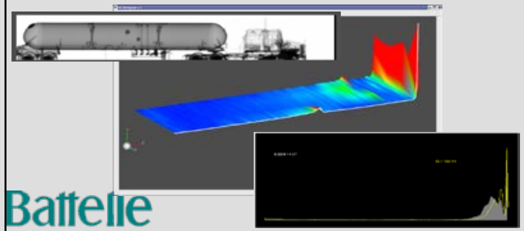
Loaded semi-trucks



Empty semi-trucks



Tanker trucks



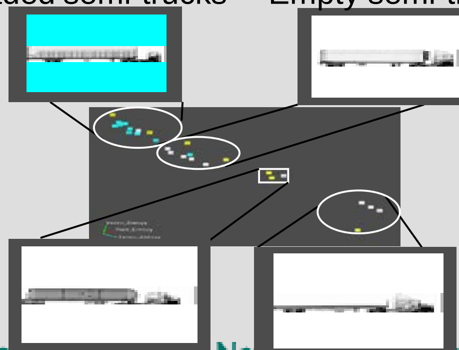
Flatbed trucks

Entropy used to classify and sort truck types and cargo

Loaded semi-trucks

Empty semi-trucks

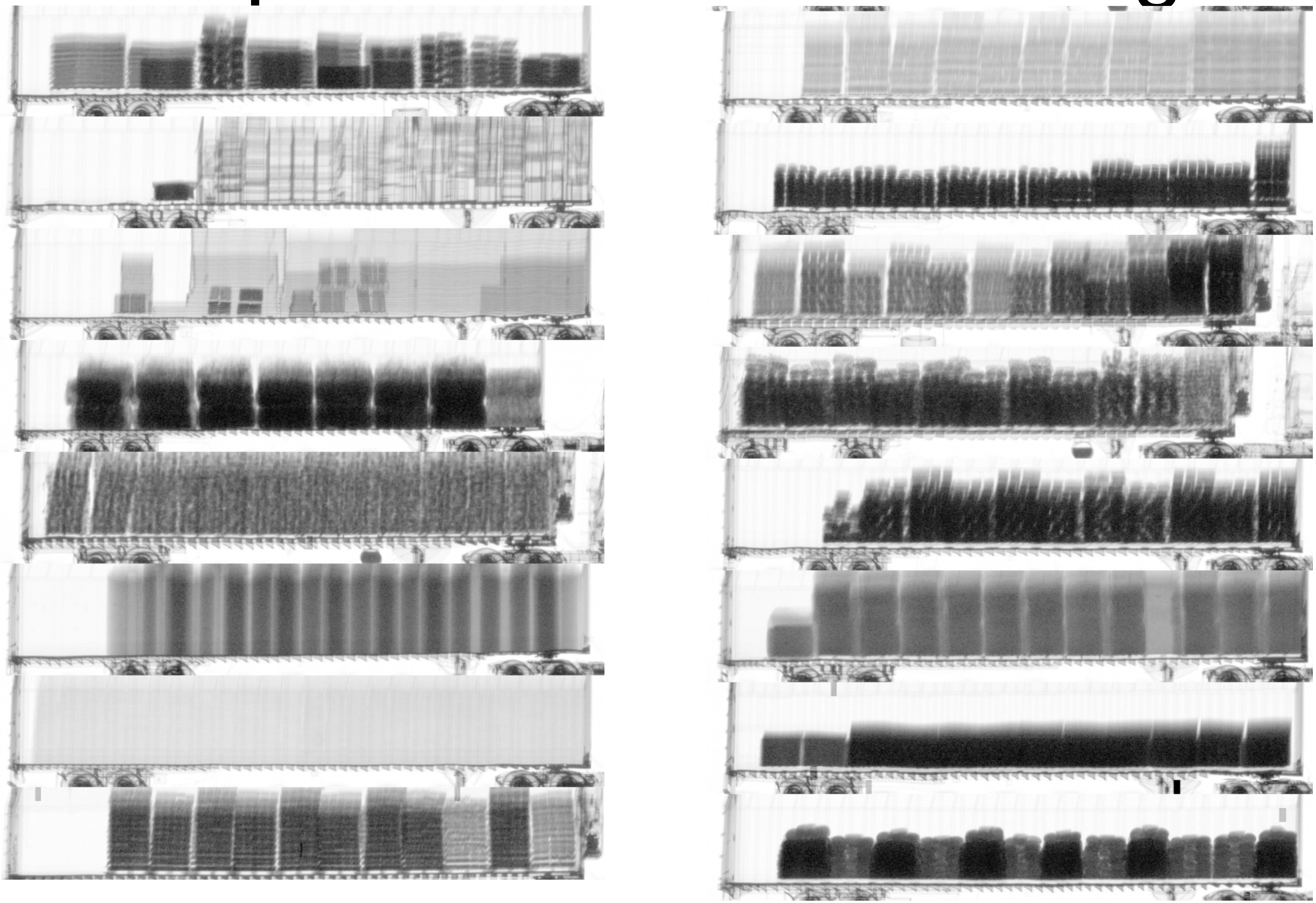
Reference trucks found in cluster



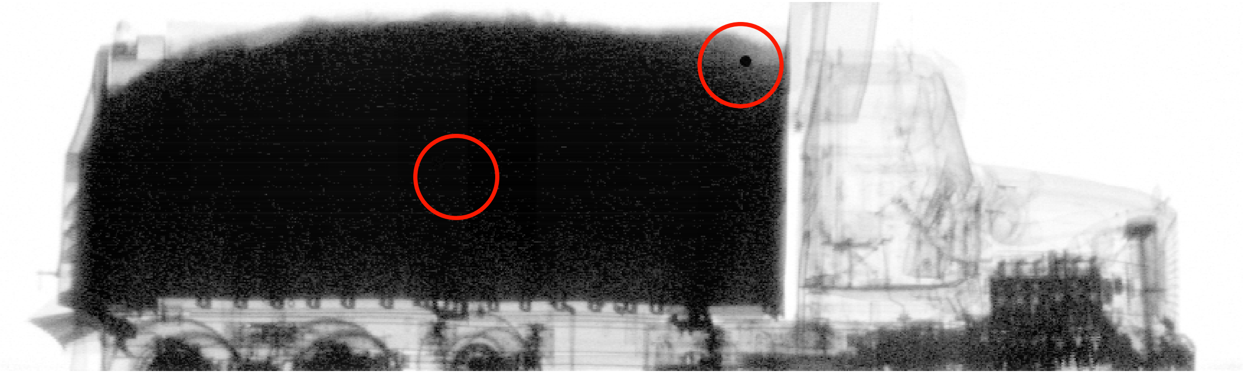
Tanker trucks

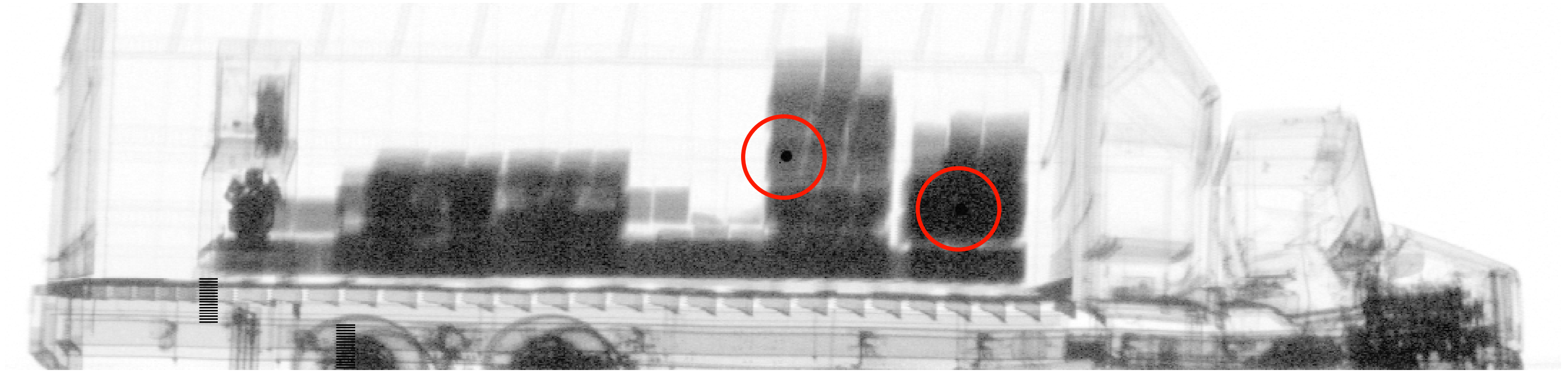
Flatbed trucks

Sample of semi-trailer images



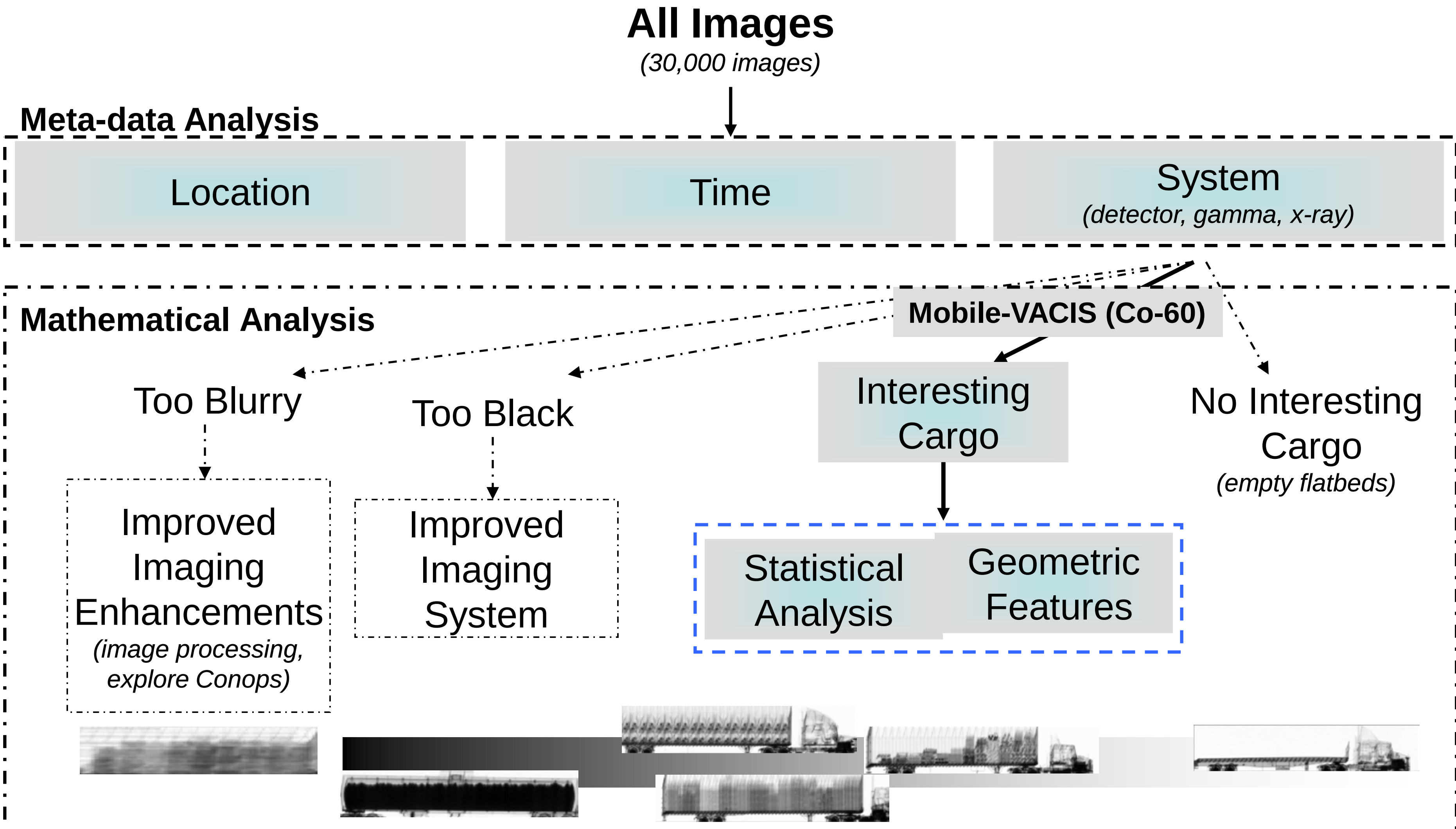
Benchmarks for Passive Detectors



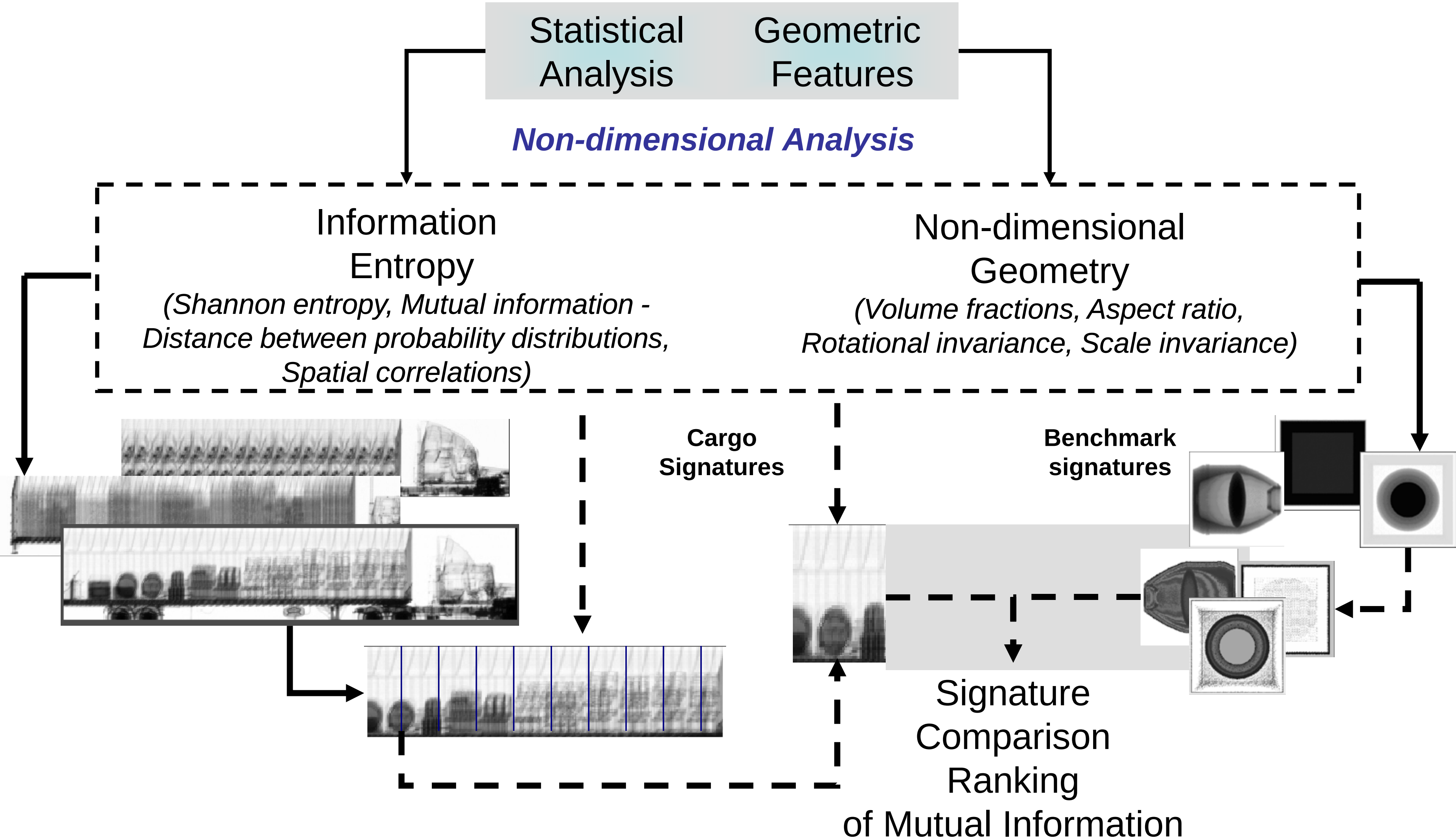


- Location in the image will affect detectability

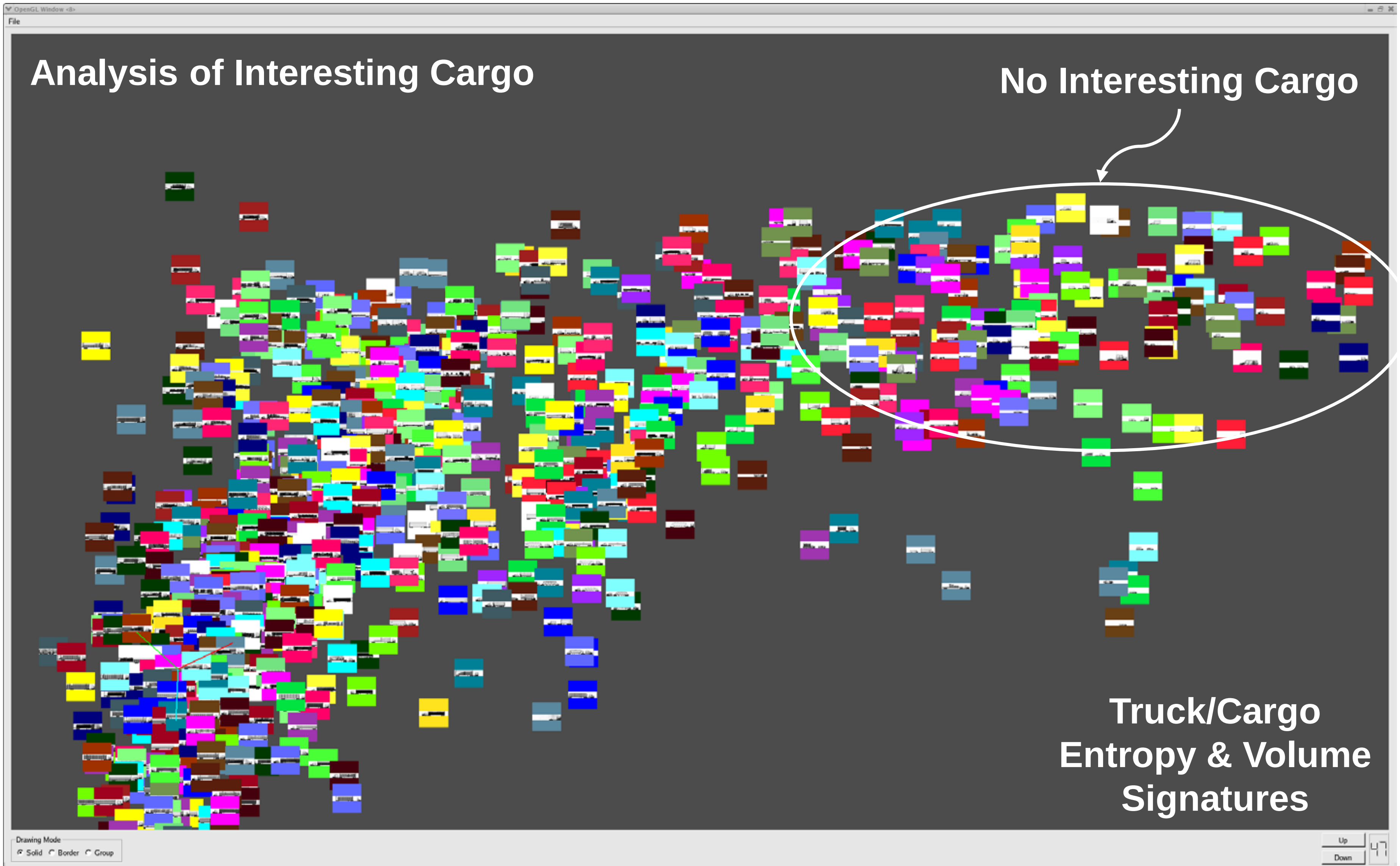
Categorization Framework
(Automated Clustering, Sorting and Classifying Images Using Metadata, Information Physics and Geometry)



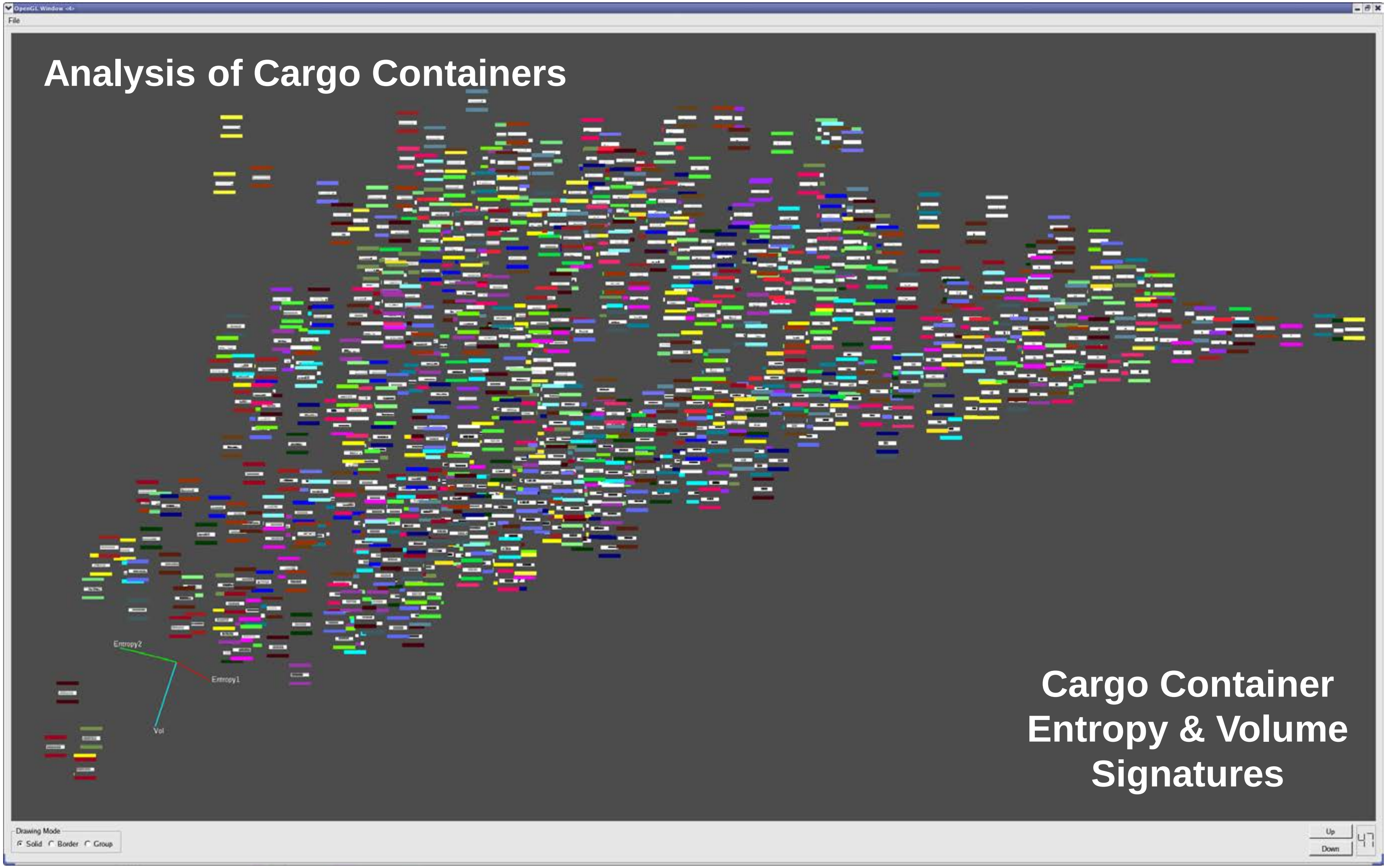
Mathematical Analysis of Interesting Cargo



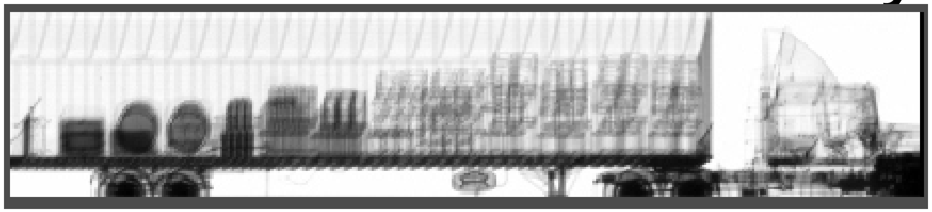
Subclass of Real Images Analysis
Mobile VACIS (Co-60)



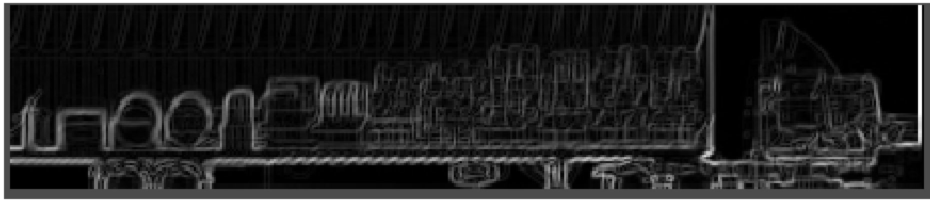
Subclass of Real Images Analysis- Mobile VACIS (Co-60)



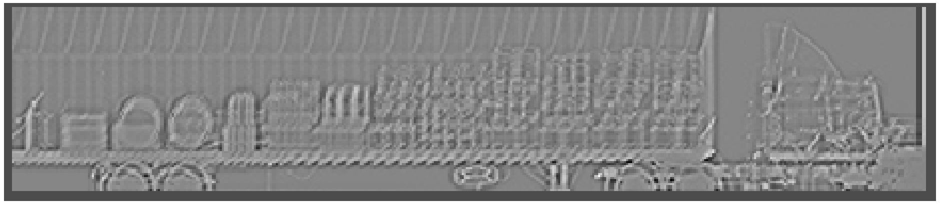
Non-Dimensional Geometry and Statistical Analysis



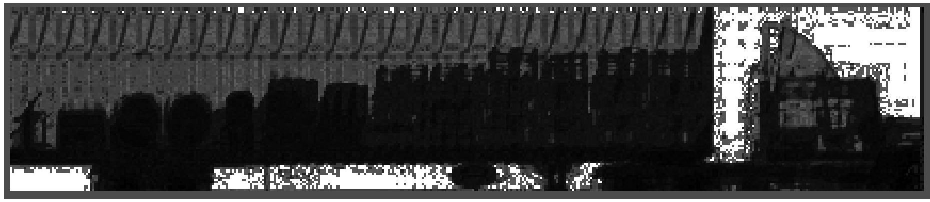
Original Cargo Truck



Sobel Edge Detection Algorithm

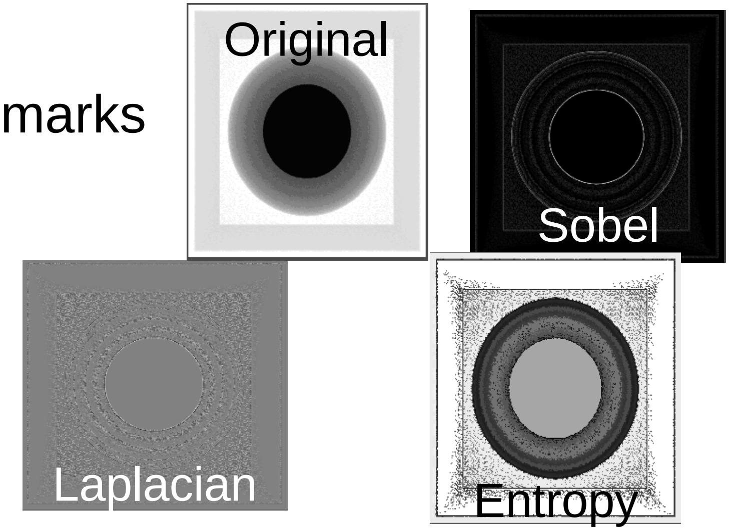


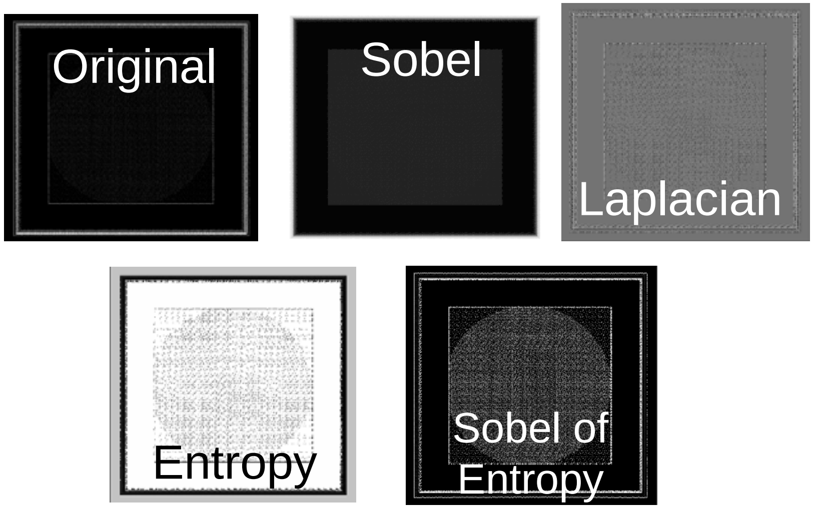
Laplacian Edge Detection Algorithm



Shannon Entropy Image

Benchmarks





Summary and Conclusions

- Sensor data explosion (volume not content)
  - 1000 X data can be serious
- Computational Capability
  - Commodity processors for getting things done
  - IBM Cells for data preprocessing and decomposition
  - Cray XMT (Eldorado) for data analysis (graphs, database searches)
- Decisions: Sensor ←→ Media ←→ Source