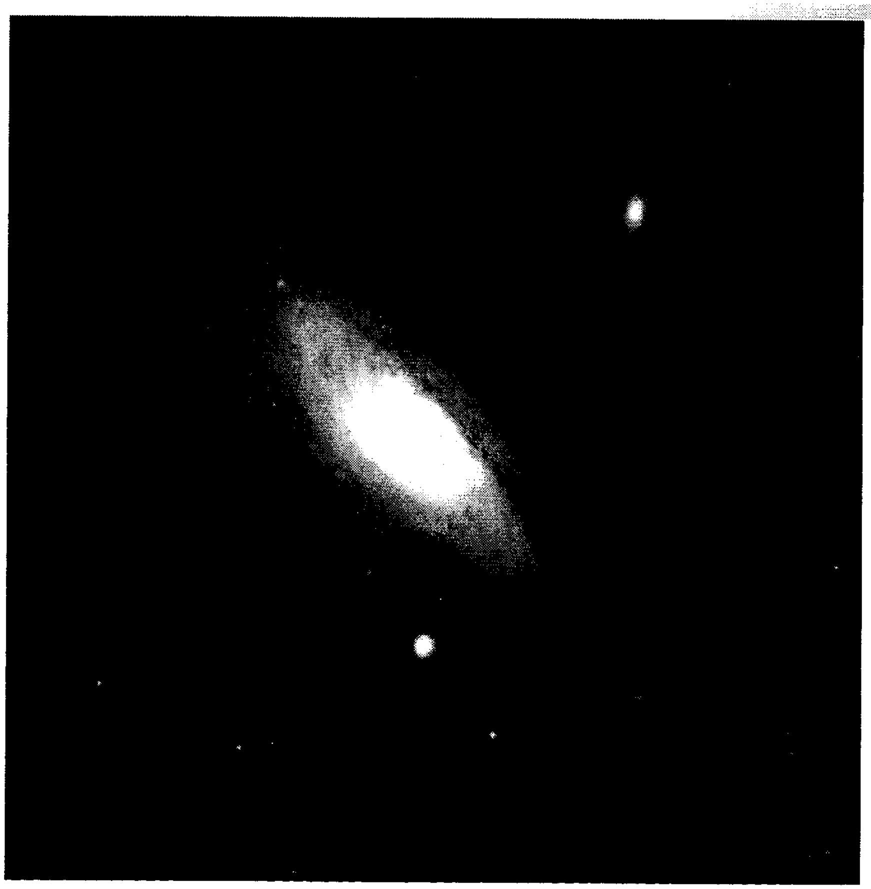
Access and Visualization Using Clusters and Other Parallel Computers



Daniel S. Katz

Parallel Applications Technologies Group Supervisor

Attila Bergou, Bruce Berriman, Gary Block, Jim Collier, Dave Curkendall, John Good, Laura Husman, Joe Jacob, Anastasia Laity, Peggy Li, Craig Miller, Lucian Plesea, Tom Prince, Herb Siegel, Roy Williams

Abstract: (to be deleted)

- JPL's Parallel Applications Technologies Group has been exploring the issues of data access and visualization of very large data sets over the past 10 or so years. This work has used a number of types of parallel computers, and today includes the use of commodity clusters. This talk will highlight some of the applications and tools we have developed, including how they use parallel computing resources, and specifically how we are using modern clusters. Our applications focus on NASA's needs; thus our data sets are usually related to Earth and Space Science, including data delivered from instruments in space, and data produced by telescopes on the ground.

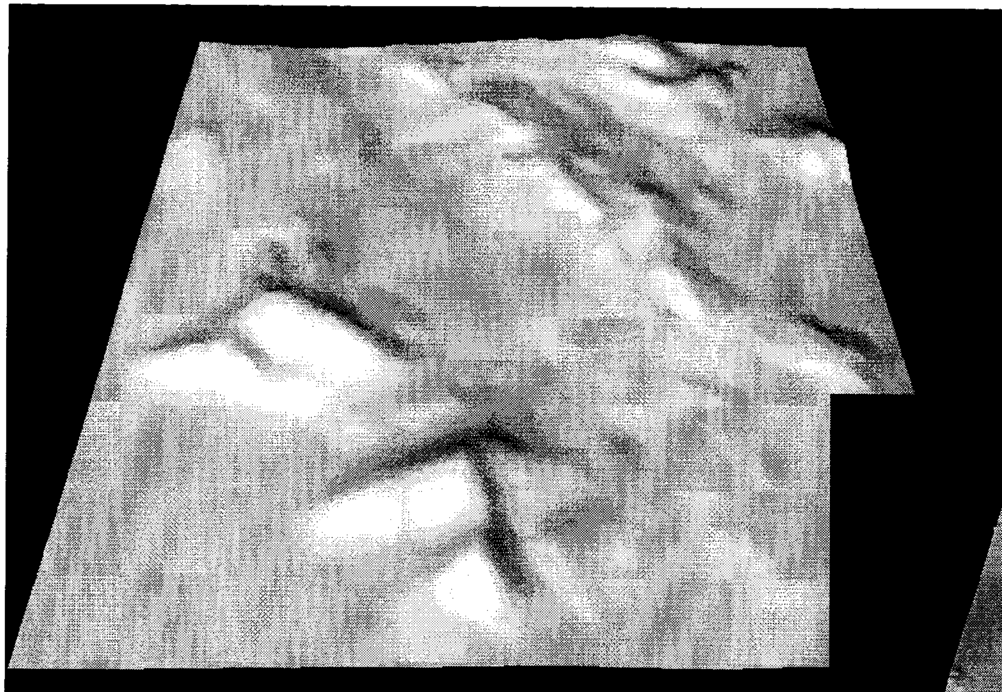
JPL's Parallel Applications Technologies Group

- PAT Group's goal: Helping JPL and Caltech scientists deal with large data sets
- The Problem:
  - Amount of available data is ever expanding
    - Many common data sets are now each 5-10 TB
  - Gaining knowledge from data is still hard
  - One very effective method is to visualize the data
- This talk will cover general methods for accessing and visualizing data, as well as highlighting specific examples

Visualizing Local Data

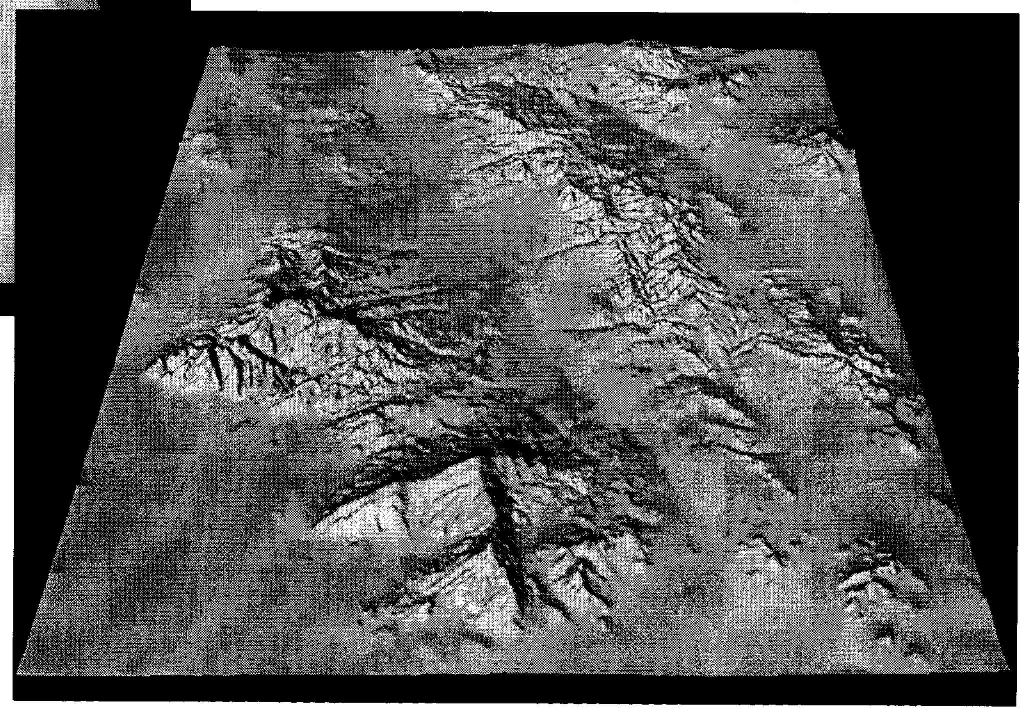
- Assume a very large data set exists on a local supercomputer
  - It's on the supercomputer because that's where it was generated
- Example: synthetically-enhanced Martian terrain data
  - Single Processor Algorithm by Bob Gaskell (JPL)
  - Parallelized by Richard Chen, Craig Miller, Herb Siegel
  - Used as a component of Terrain and Environmental Data Server (TEDS)
- TEDS is a 24/7 service for those doing simulations involving Mars terrain
  - Includes terrain storage, generation, enhancement, and access
    - Storage and access includes:
      - Measured terrain (Mars Yard, field sites, etc.)
      - Other modeled terrain (Mars sites, etc.)
  - Interfaces with simulation tools
  - Uses parallel computing as needed, specifically for generating terrain and enhancing terrain

Ideal Enhancement

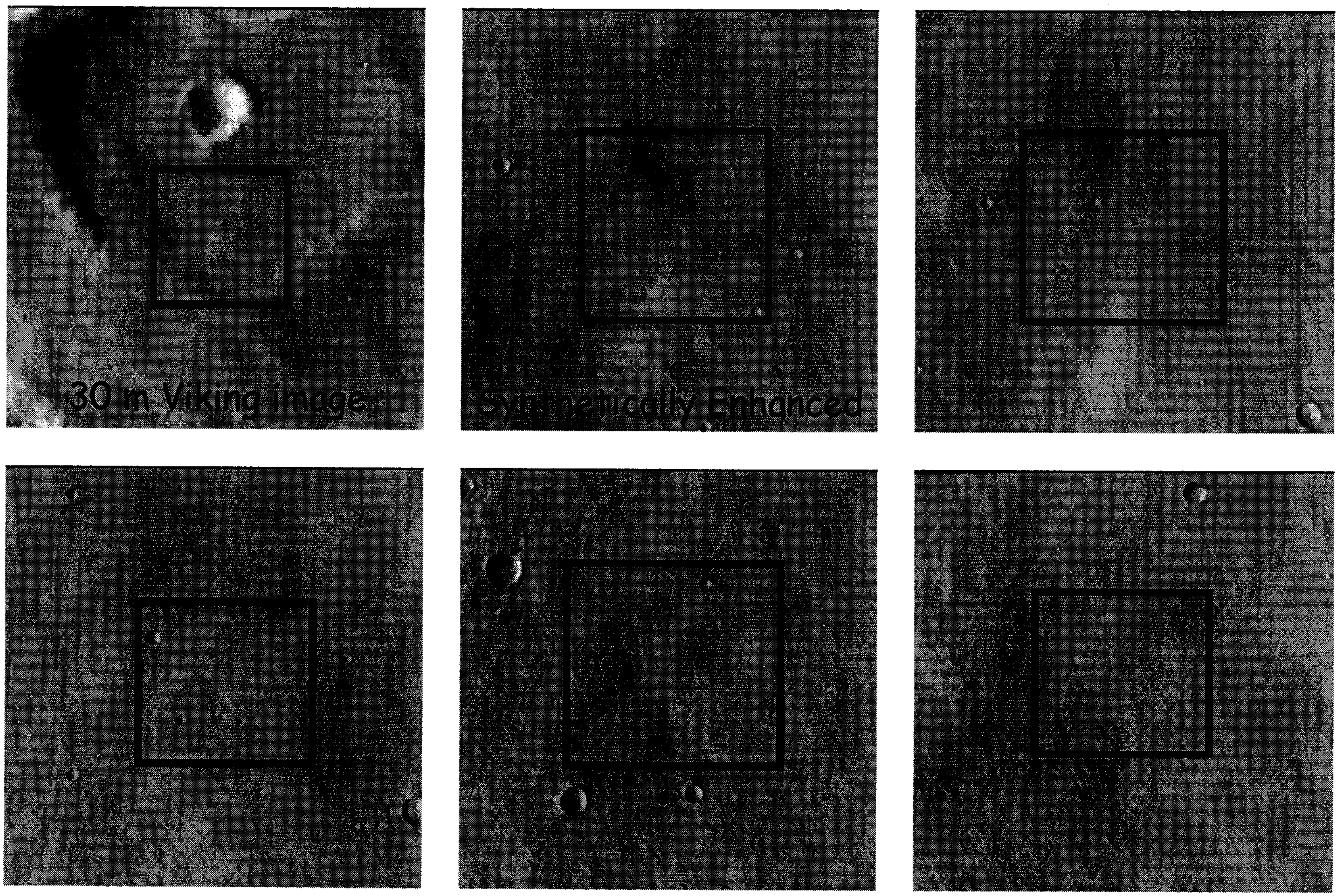


Input: Coarse Data (i.e., MOLA)

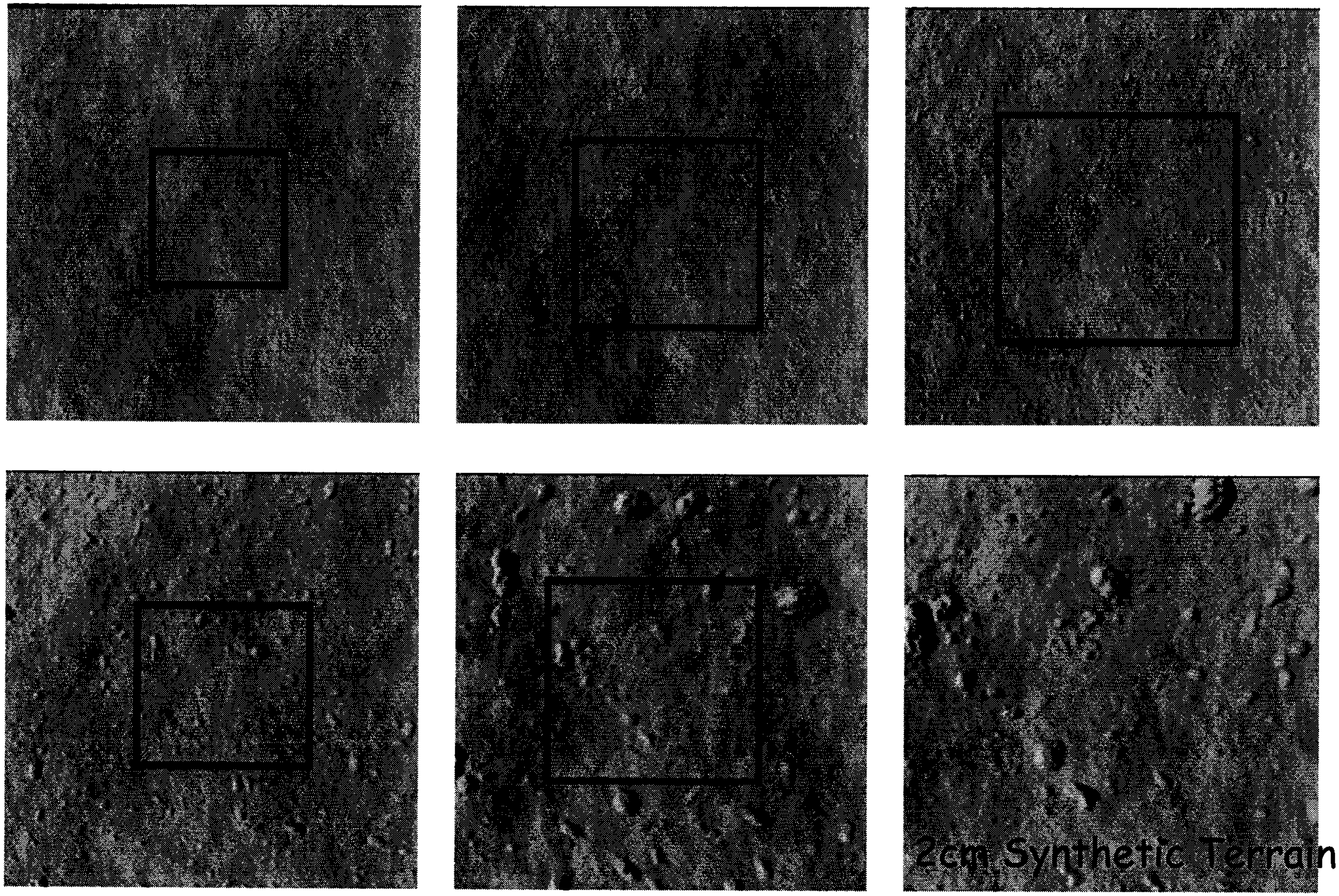
Output: Synthetically Enhanced



Synthetic Terrain: Starting with what we know...



...and adding detail



A Sample Synthetic Terrain

QuickTime™ and a YUV420 codec decompressor are needed to see this picture.

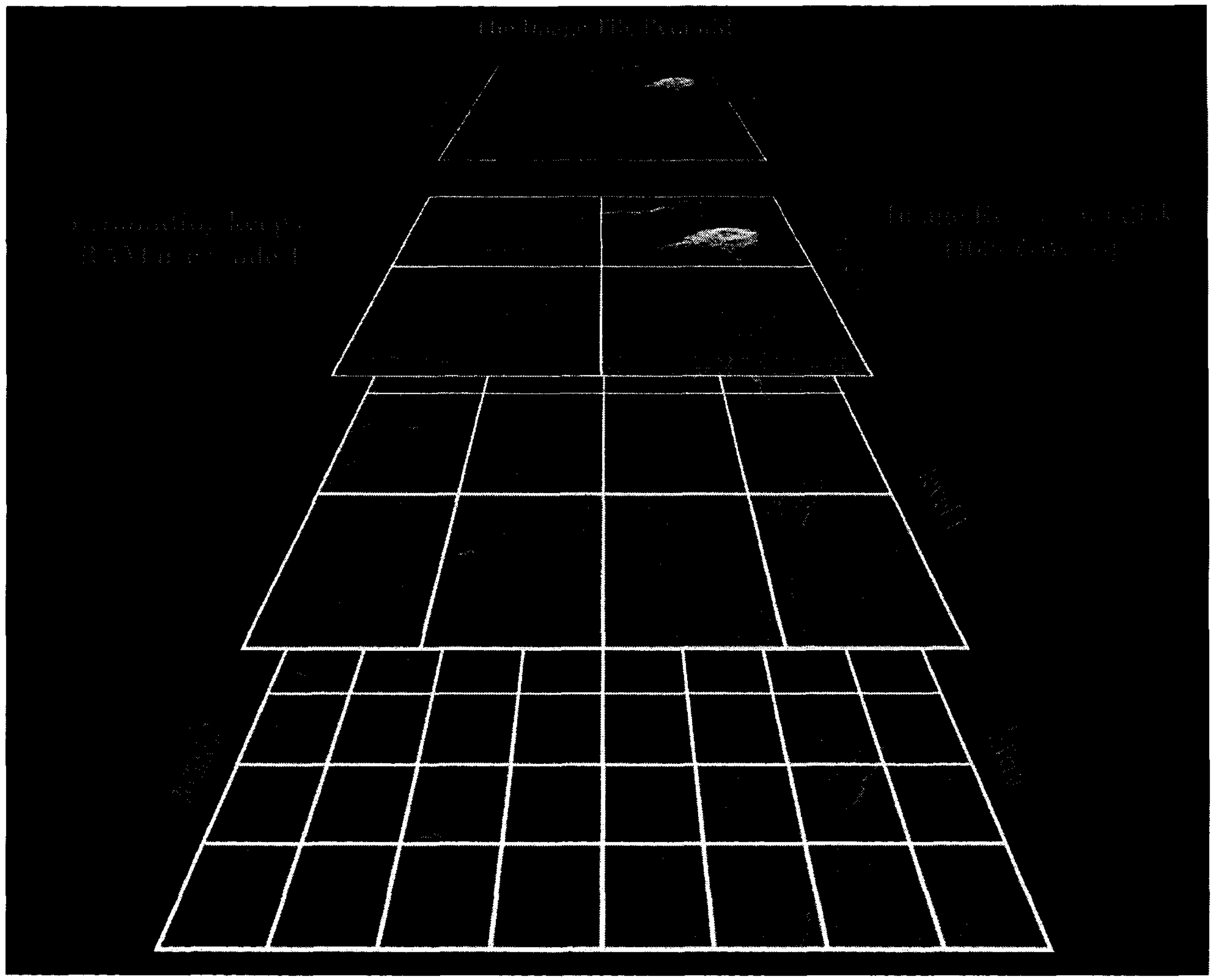
Digital Light Table (DLT)

- Previous animation built using Digital Light Table (by Herb Siegel and Craig Miller)
- Built for fast visual interactive access to very large data sets, including terrain data w/ elevation
- Originally built for viewing JERS-1 Amazon mosaic (mosaic created by Paul Siqueira and Bruce Chapman), sponsored by HPCC ESS
- Uses graphics hardware for fast pan and zoom

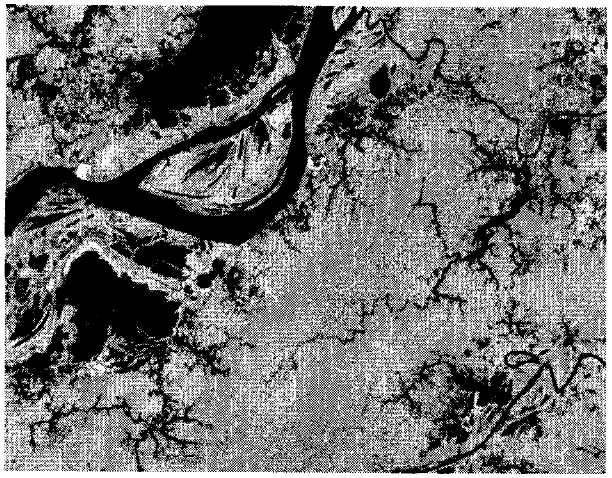
DLT Disk Storage

To permit smooth pan and zoom, the DLT needs to be able to quickly access data at the appropriate resolution

- The DLT uses the Image Tile Pyramid is to store the input data on disk:
- The original data is tiled.
- Each level has tiles 1/4 the resolution of prev. level
- This tiling **allows the DLT to smoothly pan and zoom** by only using the proper level's data in order to keep the output screen(s) updated



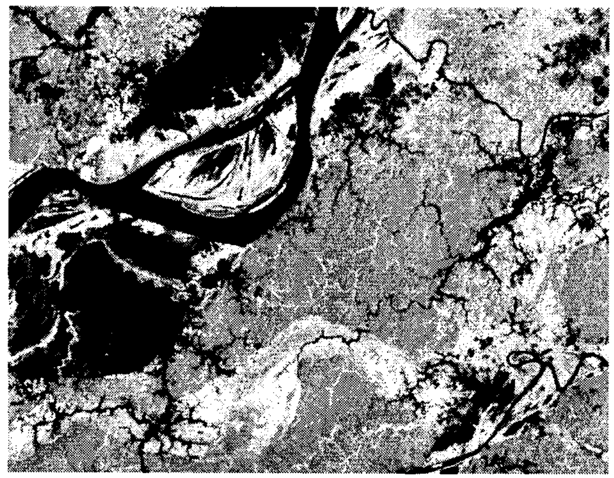
DLT's Change Detection Capability

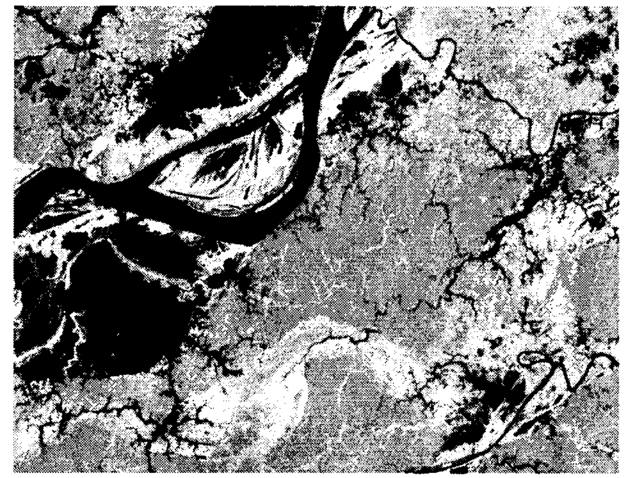


Dry season image

Mapped to blue

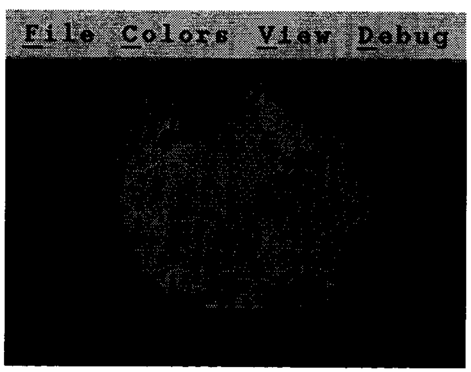
Multi-season image, where
Black = covered with water in both seasons
Grey = not covered with water in either season
Yellow = inundated in wet season
Blue = marshy in dry season



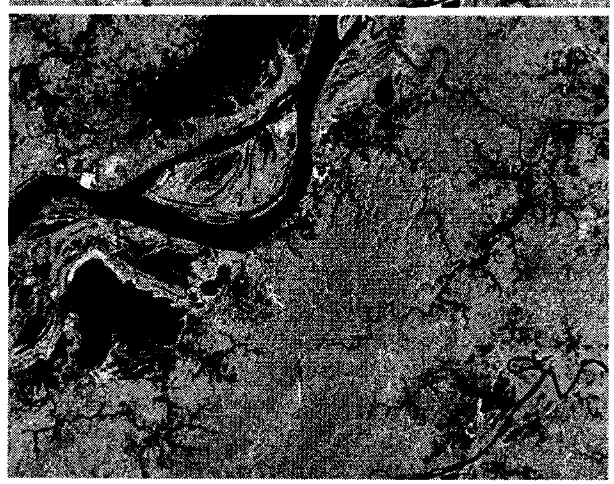


Wet season image

Mapped to green and red (yellow)



Viewer GUI tool, showing the JERS-1 Amazon SAR mosaic



Alternating the wet and dry Images - flash capability

DLT Zoom and Pan Examples

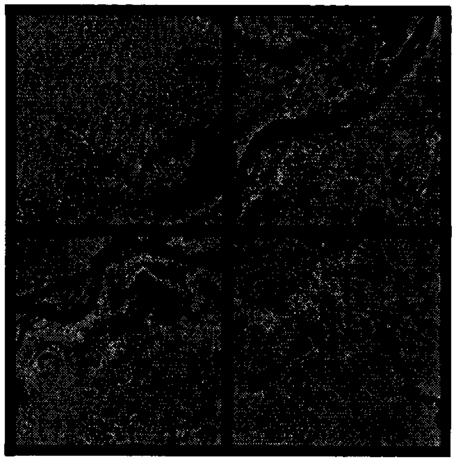
QuickTime™ and a YUV420 codec decompressor are needed to see this picture.

QuickTime™ and a YUV420 codec decompressor are needed to see this picture.

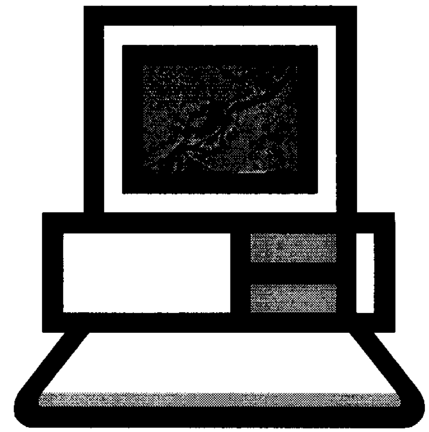
Key:
Black = covered with water in both seasons
Grey = not covered with water in either season
Yellow = inundated in wet season
Blue = marshy in dry season

DLT Views and Collaboration

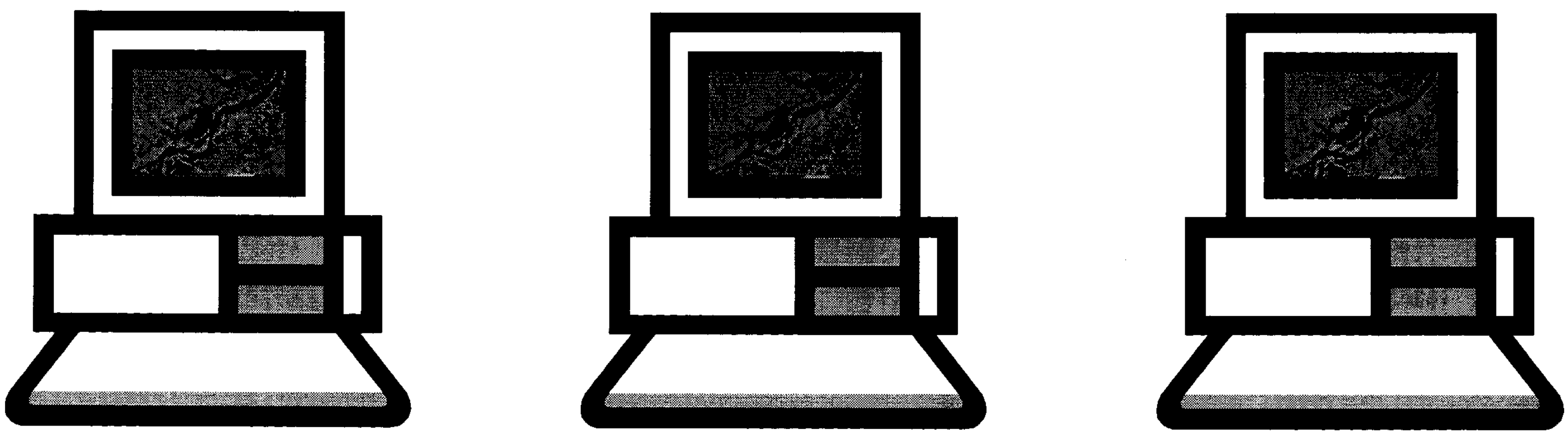
Three Ways to Use the DLT:



Multi-Screen Auditorium



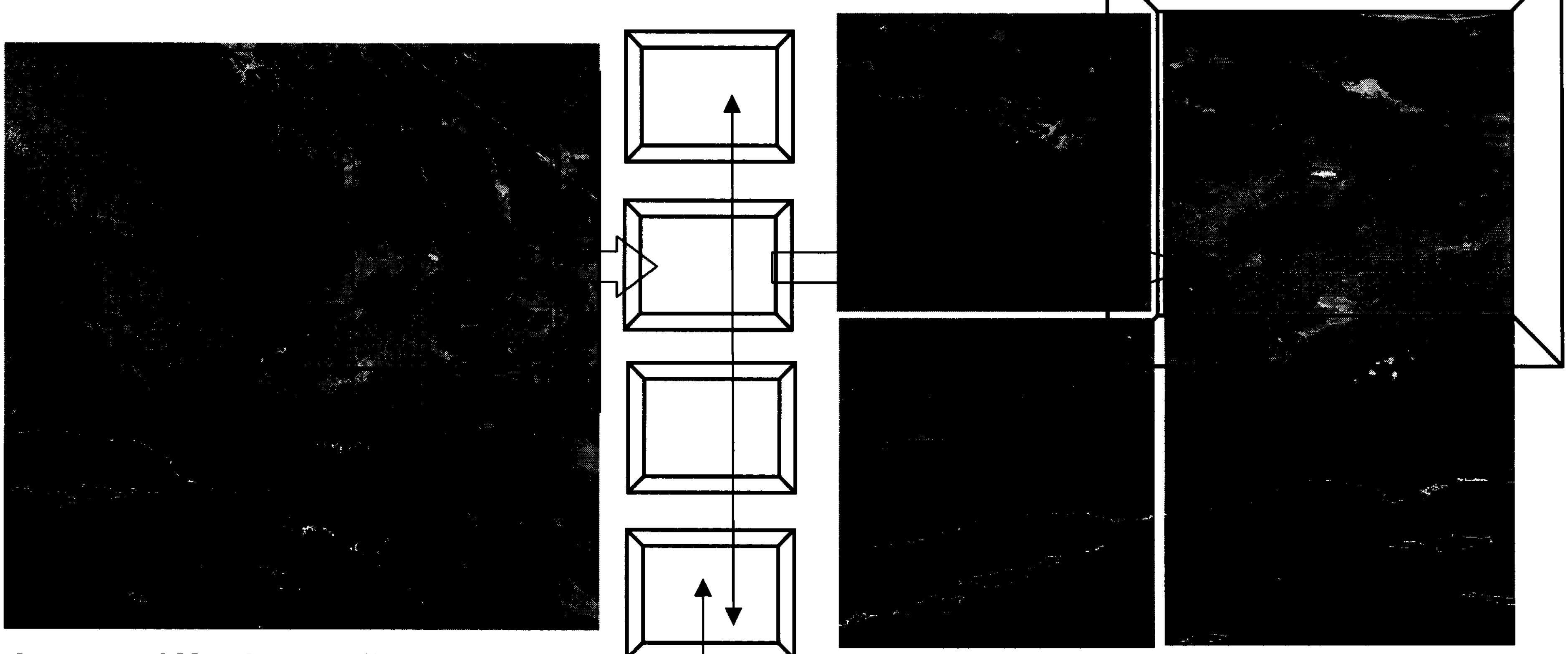
One Scientist



Multiple Scientists Discussing an Image (Anyone Can Control the DLT, Other Displays are Mirrored)

DLT Architecture

For multiple displays, each graphic engine selects the tiles and performs the rendering for its display:

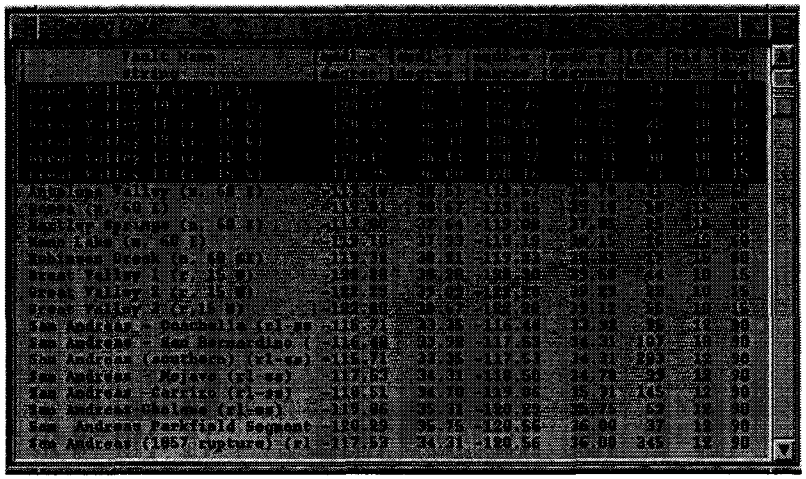


Displays

Image and Map layers, tiled at multiple resolutions: Appropriate tiles are selected & then read by graphic engines.

Parallel Graphic Engines: Commands & live data sent to any engine are automatically shared by all graphic engines.

Correspondence of Image Data and Metadata



California fault database

California Landsat with fault segment overlay

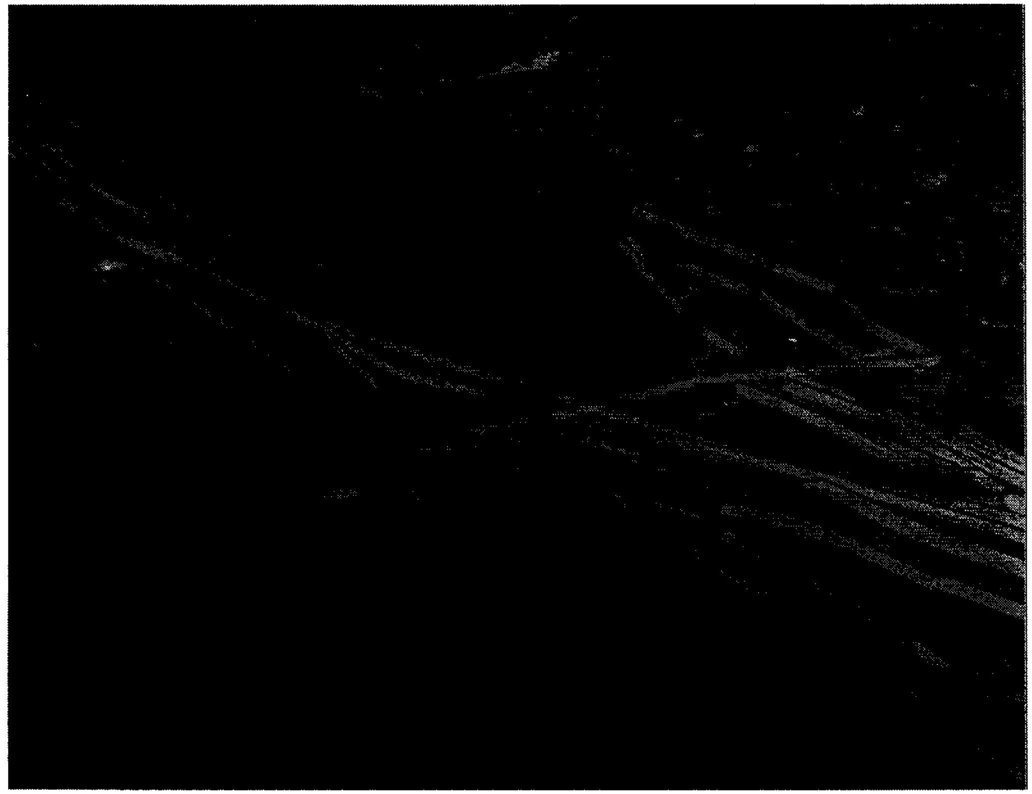
Information in a catalog is tied to associated locations in an image

Image to Catalog: User may select a region and see the catalog entries for those objects in that region highlighted both in the image and in the catalog window

Catalog to Image: User may select a catalog entry and see that object highlighted in the image or jump to the position of that object in the image

DLT -> MSLT

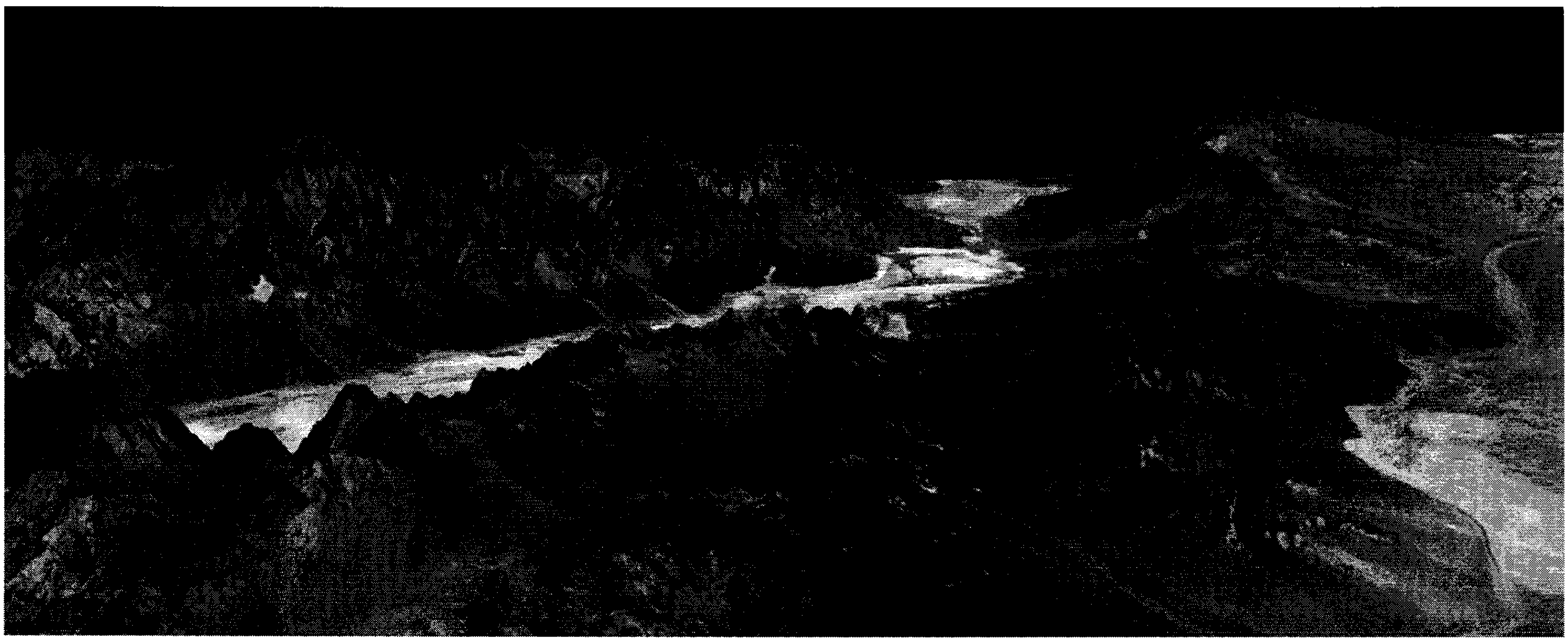
- DLT changing into Multi-Surface Light Table (MSLT), sponsored by NASA ESTO-CT
- Allows viewing of multiple surfaces, using transparency
- This permits fault segments to be displayed, under the land
- Can use color in different ways, such as to show dip angle



California Fault Database, from California Geological Survey
(fault colors represent the dip angles with green for 90 degree and blue for 0 degree)

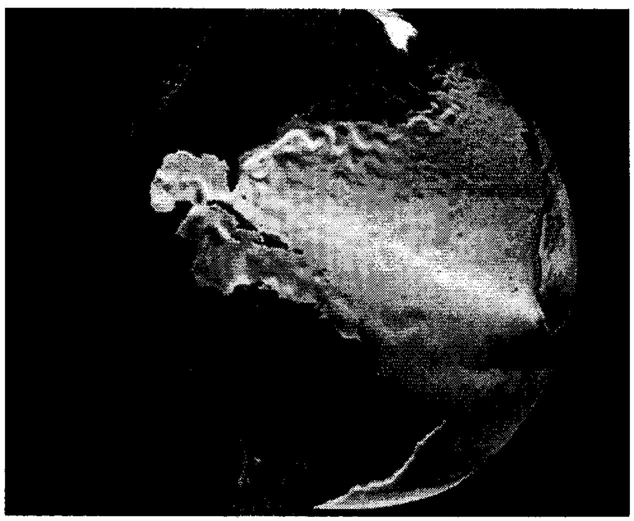
RIVA: Another Terrain Data Visualization Tool

- Peggy Li's Remote Interactive Visualization and Analysis (RIVA) System
- Scalable, parallel software rendering for 3-D planetary data, sponsored by internal funds and HPCC-ESS

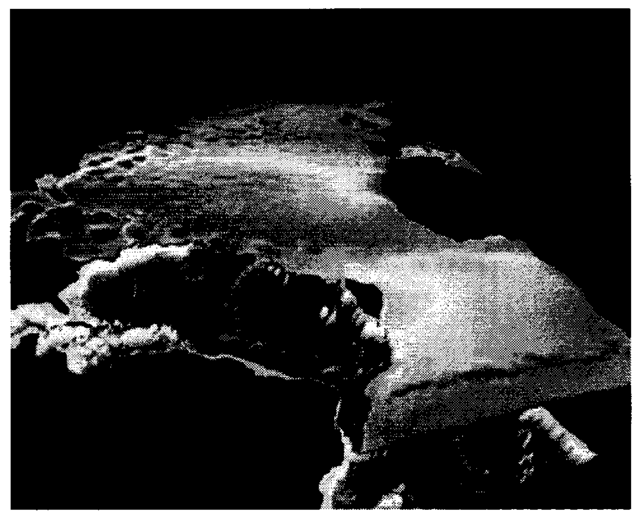


RIVA Features

- Spherical data representation capable of global data sets and regional terrain data sets
- Scalable to large number of processors, large input data sets and large output images
- Generate high-resolution still images and animations (IMAX and HD formats)
- Support out-of-core rendering for data sets bigger than the physical memory



Ocean surface temperature overlay on top of a global earth image



Ocean surface temperature separated from the ocean bottom topo map using zbuffer

RIVA Features (2)

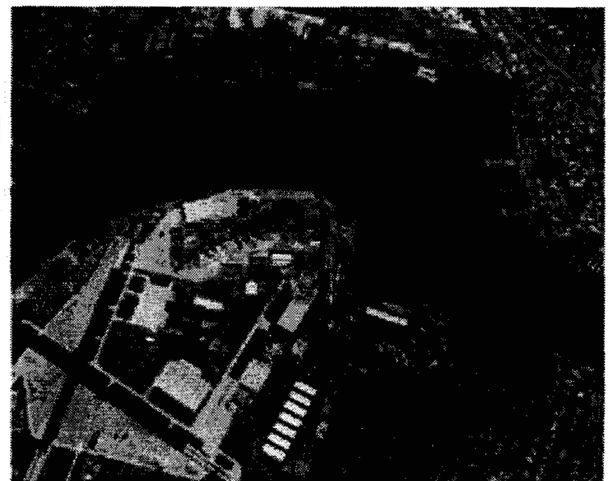
- Multiple surface rendering with different resolution, different format, and different coverage, compositing surface using zbuffer or alpha-blending
- Distributed and interactive data exploration and visualization
- Animation of time-varying simulation data set using out-of-core rendering technique
- Batch mode movie production

(a)

30 meter Landsat, bands 7, 4, and 2

(b)

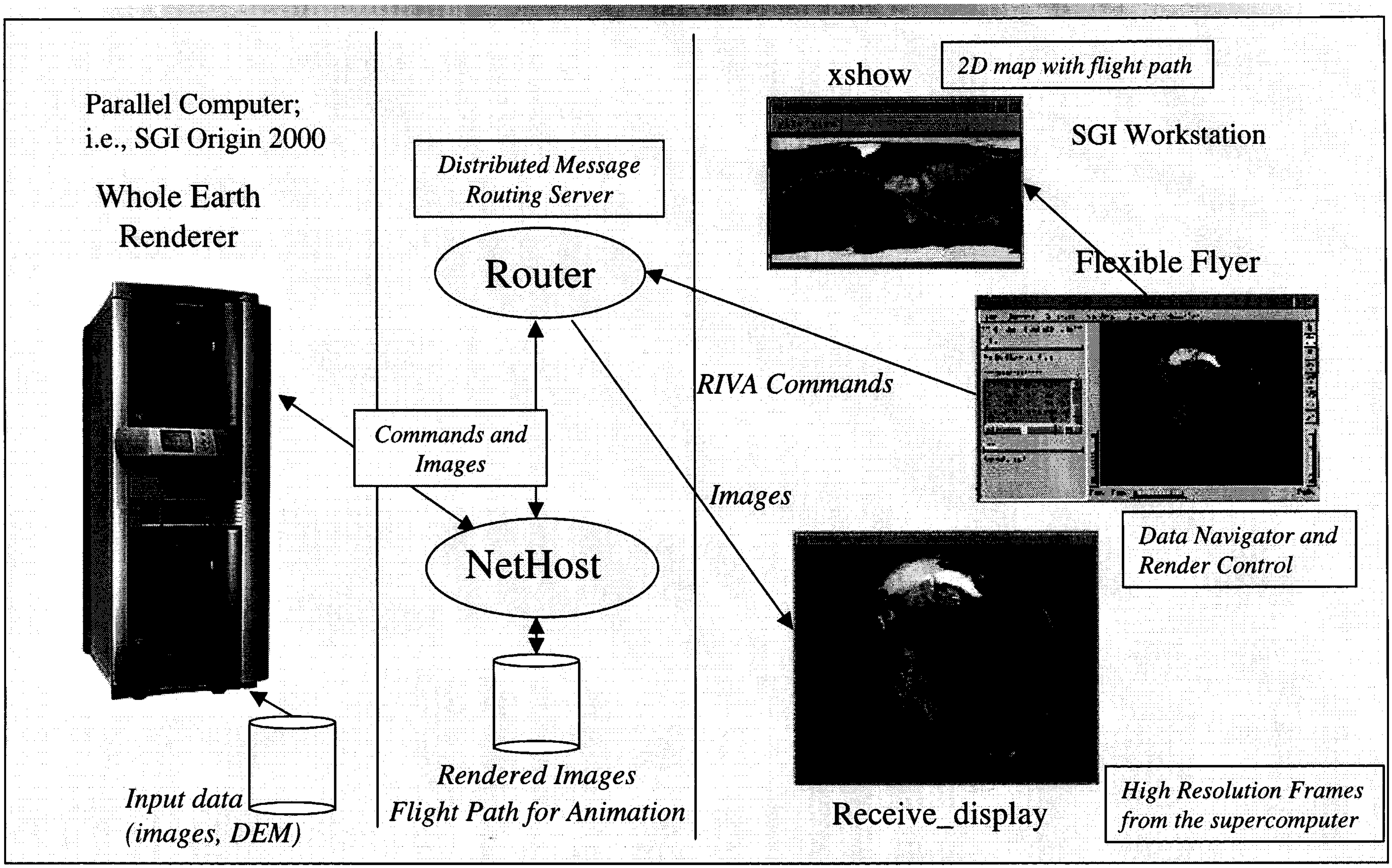
Two images blended with the opacity of 2.25 meter image set to 0.58



(c)

2.25 meter grayscale satellite image

RIVA System Architecture



RIVA Movie

A Fly-over movie using four datasets: a southern California map, a 154 meter resolution Southern California image and a 30 meter LandSAT image of San Diego, ending with 2.25 meter resolution data of San Diego.

Sharing Your Data Visually

- The obvious choice today:
  - Deliver images through a web browser
  - Allow users to decide what images to build
- Example: Lucian Plesea's MAPUS (http://mapus.jpl.nasa.gov/)
  - Starting with a 180 GB mosaic of 30 m Landsat data, including:
    - Access to all 6 bands
    - Digital Elevation Model (DEM)
    - Can overlay political boundaries, roads,rivers…
  - GIS Web Map Server (WMS)
  - Original mosaic sponsored by Air Force, web server sponsored by Digital Earth Project