Visual Data Mining of Remote Sensing Data

Jürgen Symanzik*,
Utah State University, Logan, UT

*e-mail: symanzik@sunfs.math.usu.edu
WWW: http://www.math.usu.edu/~symanzik

with Louise Griffith, Rob Gillies, Di Cook, Jim Majure, Nicholas Lewin-Koh,
Noel Cressie, Sigbert Klinke
Contents

- Visual Data Mining
  - Definitions
  - Software
  - Techniques

- Examples

- Live Demos
Data Mining

Ed Wegman:

“Data Mining is Exploratory Data Analysis with Little or No Human Interaction using Computationally Feasible Techniques, i.e., the Attempt to find Interesting Structure unknown a priori”
Working Definition:

- Find structure (cluster, unusual observations) in large and not necessarily homogeneous data sets based on human perception using graphical methods and user interaction
- Goal or expected outcome of exploration usually unknown in advance
First uses of the term:


Software: XGobi/GGobi
Swayne, Cook and Buja

• Interactive environment for exploring multivariate data
  * Linked views allow “linked brushing”
  * Univariate, Bivariate and Multivariate views of the data
  * Grand tour
  * Wide variety of methods
  * Open source
  * Free

• Caveats
  * XGobi only on UNIX and Linux platforms
  * GGobi also available for PCs but not yet fully developed
Software: ArcView

• Desktop GIS with wide Range of Viewing and Data Manipulation Functions
  – Editing Features
  – Query Operations
  – Map Display
  – Interactive Interface
  – High Level Internal Scripting Language
Software: ExplorN/CrystalVision
Wegman, Luo, Carr

- Interactive environment for exploring multivariate data (similar to XGobi/GGobi)
  - Advanced Parallel Coordinates Displays
  - 3D Surfaces
  - Stereoscopic Displays

- Caveats
  - ExplorN only on SGI platforms
  - CrystalVision available for PCs but not yet fully developed
  - No interface to GIS software
Tools: Linked Brushing

XGobi

ArcView
Tools: Parallel Coordinate Plots

ExplorN
Tools: Scatterplot Matrix & Density Plot

ExplorN
Tools: Grand Tour

– Continuous random sequence of projections from n dimensions into 2 (or more) dimensions.
Examples

– Historical Examples
– Vermont/New Hampshire:
  » Quarry, Water, Clouds
– Atlanta
  » City, Forest
– California (Imperial Valley)
  » Fields
– North Africa
  » Desert
Stat Graphics & Remote Sensing

- Klein, Moreira (1994): Agricultural Region in Brazil
- Scott (1986): Agricultural Scene on 5 Days
- Salch, Scott (1997): 3 Groups of Farm Crops
- Carr (1991): Nevada Test Site
ArcView/XGobi/XploRe & Remote Sensing

- Symanzik, Majure, Cook (1996, 1997)
- Cook, Majure, Symanzik, Cressie (1996)
- Symanzik, Cook, Klinke, Lewin (1998)
- Symanzik, Griffiths, Gillies (2000)
The Vermont/New Hampshire Data

- Landsat Thematic Mapper (TM) data
- 6 Spectral Bands
- Outstanding Water Body is Connecticut River
Field/Quarry/Clouds ??
Water = Water ??
The Atlanta Data

- **NOAA-14 Satellite** (National Oceanic and Atmospheric Administration)
- **AVHRR Sensor** (Advanced Very High Resolution Radiometer):
  - Band 1: Red
  - Band 2: Near Infrared
  - Band 3: Mid Infrared
  - Band 4: Long Infrared
  - Band 5: (Very) Long Infrared
- **Data from “NASA’s Project Atlanta”**
- **18 Days from Jan 1997 to Dec 1997**
- **Resolution**: 1 km x 1 km per Pixel
- **Main Study Area**: 70 km x 46 km
Some Definitions

- Normalized Difference Vegetation Index:
  \[ NDVI = \frac{\text{Band 2} - \text{Band 1}}{\text{Band 2} + \text{Band 1}} \]

- \( NDVI \sim 0.8 \) for Highly Vegetated Surfaces
- \( NDVI \sim 0.1 \) for Bare Soil
- Surface Radiant Temperature \( T_0 \): Band 4
- Surface Moisture Availability \( M_0 \)
NS001-TMS derived $T_o$-NDVI scatterplot (gray spectral scaling) at a 5 meter spatial resolution for a 7 x 3 km area of the Mahantango Watershed, Pennsylvania. 18 July 1990, 1145 LST. Isopleths representing moisture availability index, $M_o$ are overlaid with the legend, $o = 0.0$ (‘warm’ edge), $\hat{o} = 0.2$, $\square = 0.4$, $\Delta = 0.6$, $\nabla = 0.8$, and $\times = 1.0$ (cold edge).
The Main Study Area
Two Months

August

December
The City

December

August
Clouds in August

August
2 Pixels of Interest

December

Linked

August
The Imperial Valley/CA Data

- Landsat-4 Thematic Mapper (TM) Data
- December 12, 1982
- 7 Spectral Bands
- 124 Fields with known Crop Information
Ground Truth
Alfalfa x 2
Live Demo

GGobi

CrystalVision
Result CrystalVision
Overall Conclusion

- Visual approach effective to see unexpected structure in data.
- Combination of different techniques most effective.
- Can be used for almost all types of RS data.
Future Work (1)

- Enhance new software (GGobi, CrystalVision) to operate in a linked environment with GIS software.
- Allow access to databases.
Future Work (2)

- Use 3D environment (CAVE, MiniCAVE) for visualization and visual data mining.
Contact

- Jürgen Symanzik
  - symanzik@sunfs.math.usu.edu
- Website
  - www.math.usu.edu/~symanzik