Ground-Truth Collection System for Nuclear Proliferation Activities

Oscar Alejandro Pérez Cruz
Research Alliance in Math and Science
http://sites.google.com/a/g.ornl.gov/oscar-perez_cruz/
perezcruz.oscar@gmail.com

Advisor: Dr. Ranga Raju Vatsavai
Collaborators: Soumya De / Regina Ferrell
Oak Ridge National Laboratory
r7v@ornl.gov / ferrellrk@ornl.gov

Problem Background

For the last three decades satellites have produced large collections of geospatial images with a variety of spectral, spatial, and temporal resolutions. Automated analysis is a problem, because

- Extensive database of high resolution satellite imagery
- Limited number of image analysts and software tools

This has caused an increased difficulty in the ability to process and comprehend these images. A possible solution to this problem is the creation of an

- Image Analysis System
- Data Management System

Research Objectives

- Broad objective
  - Create a system that can assist the image analyst in the identification of different region types from
    - ✓ Structures
    - ✓ Activities

    → of the region
  - Aim: Develop a system that
    - ✓ Search
    - ✓ Identify

    → potential nuclear proliferation activities
- Summer research focus
  - Create a large ground-truth dataset
    - ✓ Train machine learning algorithms to analyze images
  - Develop an interactive ground-truth collection system containing
    - ✓ Spatiotemporal database management system
    - ✓ Image feature extraction and analysis system

Research Background

Proliferation: spreading of a resource
Nuclear proliferation: spreading of nuclear weapons

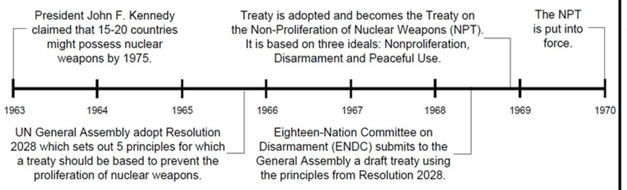
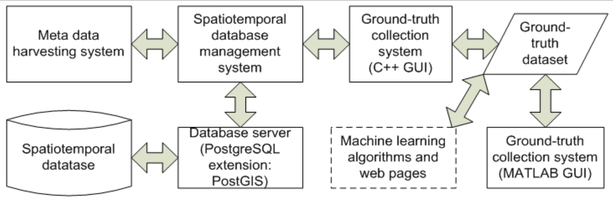


Fig. Nuclear Non-Proliferation Treaty (NPT) creation summary timeline.

Ground-truthing: term widely used in many remote sensing supervised classification task to denote the process of assigning labels for each image sample

- Used to relate image data to real features and physical structures on the ground
- Enables calibration of remote-sensing data
- Aids in the interpretation and analysis
- Helps in the initial supervised classification of an image by
  - Providing training sites for remote sensing software to construct a model (estimating parameters)
  - Helping with atmospheric correction by identifying objects in satellite images

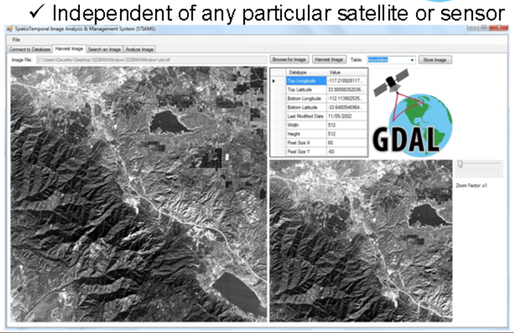
System Architecture



Meta Data Harvesting System

Meta data: data that provide information about data (location/size/date image taken)
Harvesting: process of extracting the meta data from the ancillary data sources

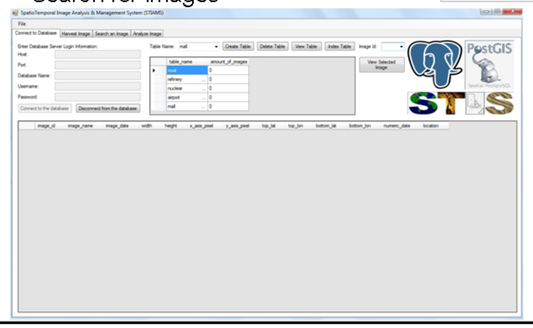
- Features
  - Uses GDAL open source library
    - ✓ Supports 50/60 image formats
    - ✓ Independent of any particular satellite or sensor



Spatiotemporal Database Management System

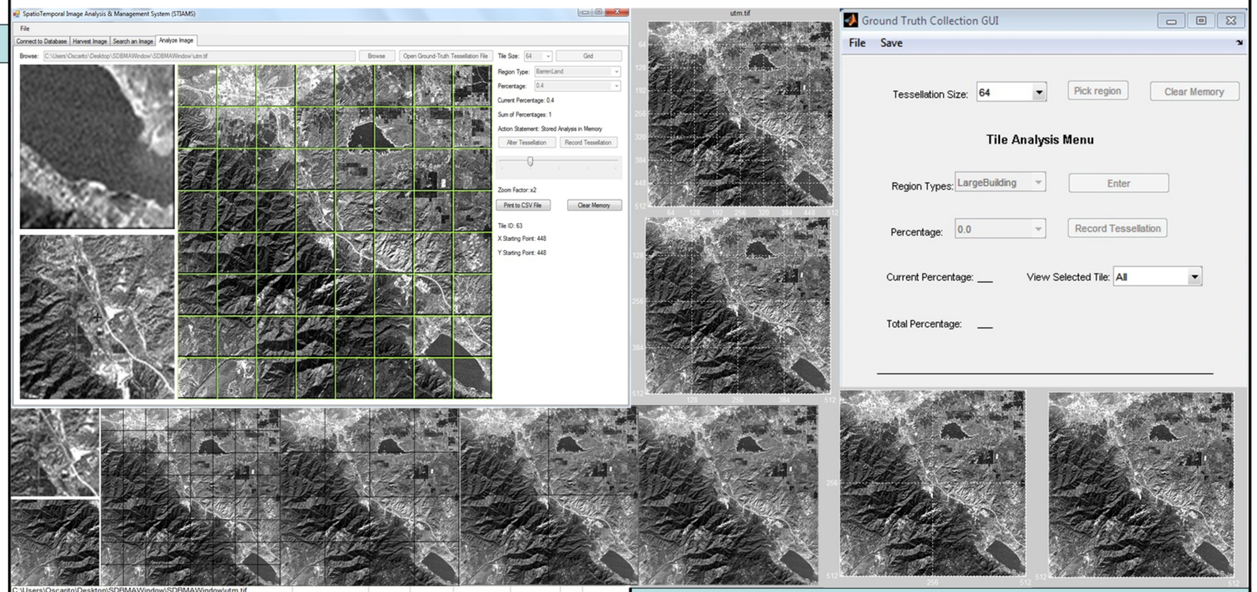
Spatiotemporal database: large collection of interrelated data which contains object that contain objects with an identifier, spatial attribute and a timestamp

- Spatial attribute
  - Point: entity with a location
  - Line: network of entities
  - Region: entity with a large area
- Capabilities
  - Connect to database
  - Create, delete, view, and index table
  - Harvest meta data to insert in table
  - Search for images



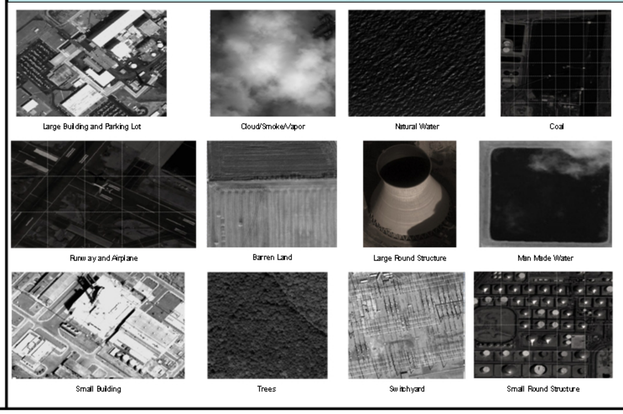
Ground-truth Collection System

- Enables the user to view satellite imagery
- Allows user to label regions with class label
- Can be used with either the MATLAB or C++ GUI
- Capabilities
  - Divides the image in a fixed tile size
    - ✓ Sizes: 64, 128, 256 and 512 pixels
  - Enables user to
    - ✓ Identify a percentage a class within the tiles
    - ✓ Add more region type classes
    - ✓ View ground-truth dataset and edit it through another analysis
    - ✓ View only selected region type on analyzed tiles



X-coord	Y-coord	Width	Height	LargeBuilding	SmallBuilding	LargeRoundStructure	SmallRoundStructure	Switchyard	ParkingLots	ManMadeWater	NaturalWater	Trees	BarrenLand
64													
0	0	64	64	0	0	0	0	0	0	0	0	0	1
64	0	64	64	0	0	0	0	0	0	0	0	0	1
128	0	64	64	0	0	0	0	0	0	0	0	0	1
192	0	64	64	0	0	0	0	0	0	0	0	0	1
256	0	64	64	0	0	0	0	0	0	0	0	0	1
320	0	64	64	0	0	0	0	0	0	0	0	0	1
384	0	64	64	0	0	0	0	0	0	0	0	0	1
448	0	64	64	0	0	0	0	0	0	0	0	0	1
0	64	64	64	0	0	0	0	0	0	0	0	0	1
64	64	64	64	0	0	0	0	0	0	0	0	0	1
128	64	64	64	0	0	0	0	0	0	0	0	0	1
192	64	64	64	0	0	0	0	0	0	0	0	0	1
256	64	64	64	0	0	0	0	0	0	0	0.4	0	0.6
320	64	64	64	0	0	0	0	0	0	0	0.3	0	0.7
384	64	64	64	0	0	0	0	0	0	0	0	0	1
448	64	64	64	0	0	0	0	0	0	0	0	0	1
0	128	64	64	0	0	0	0	0	0	0	0	0	1
64	128	64	64	0	0	0	0	0	0	0	0	0	1
128	128	64	64	0	0	0	0	0	0	0	0	0	1
192	128	64	64	0	0	0	0	0	0	0	0	0	1
256	128	64	64	0	0	0	0	0	0	0	0.1	0	0.9
320	128	64	64	0	0	0	0	0	0	0	0.1	0	0.9
384	128	64	64	0	0	0	0	0	0	0	0	0	1
448	128	64	64	0	0	0	0	0	0	0	0	0	1
0	192	64	64	0	0	0	0	0	0	0	0	0	1
64	192	64	64	0	0	0	0	0	0	0	0	0	1
128	192	64	64	0	0	0	0	0	0	0	0	0	1
192	192	64	64	0	0	0	0	0	0	0	0	0	1
256	192	64	64	0	0	0	0	0	0	0	0	0	1
320	192	64	64	0	0	0	0	0	0	0	0	0	1
384	192	64	64	0	0	0	0	0	0	0	0	0	1
448	192	64	64	0	0	0	0	0	0	0	0	0	1

Region Types



Summary and Future Direction

System will enable an easier data management of the geospatial library.

- Stage of development
  - Continued expansion of geospatial library dataset
  - Machine learning algorithms for proliferation detection using labeled tiles
  - Other ground-truth collection system component upgrades
- Future extension for the system
  - Image classifier system
  - Other kinds of search criteria like
    - ✓ Cloud coverage percentage
    - ✓ Forestation percentage
    - ✓ Population growth
    - ✓ Type of resolution

References

- Gleason, S., Ferrell, R., Cheriyadat, A., Vatsavai, R., De, S. (2010), *Semantic Information Extraction from Multispectral Geospatial Imagery Via a Flexible Framework*. IEEE Geoscience and Remote Sensing Symposium (IGARSS, 2010).
- Doyle, J. (2008). *Nuclear Safeguards, Security, and Nonproliferation.* Butterworth-Heinemann.
- PostgreSQL 8.3.3 Documentation, The PostgreSQL Global Development Group.
- Neufeld, K. (2009). PostGIS 1.4.0rc2 Manual.
- Geospatial Data Abstraction Layer (GDAL), GDAL/OGR Project Management Committee, http://www.gdal.org/.

*This Research Alliance in Math and Science program is sponsored by the Office of Advanced Scientific Computing Research, U.S. Department of Energy. *This work was performed at the Oak Ridge National Laboratory, which is managed by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725. *This work has been authored by a contractor of the U.S. Government, accordingly the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes. I would like to thank Dr. Ranga Raju Vatsavai for the opportunity to work on this project. I finally appreciate the support of my colleagues Soumya De and Regina Ferrell.