Large Scale Analytics in the Era of Abundant Data

John M. Irvine
Draper Laboratory

Presented at the
ENVI Analytics Symposium
August 25-26, 2015
Acknowledgments
This research was supported by the Office of the Secretary of Defense (OSD) under the Human, Social, Cultural and Behavioral (HSCB) Modeling Program, through the Office of Naval Research (ONR) contract N00014-12-C-0053.

Research Team:
Draper: John M. Irvine, Jennessa Kimball, Janet Lepanto, Natasha Markuzon, John Regan, Edward Vaisman, Richard Wood, Mon Young,
MIT: Fotini Christia, Roger Petersen

Imagery: Commercial satellite imagery provided by NGA

Disclaimer:
The views expressed here are those of the authors and do not represent the positions of OSD, ONR, Draper Laboratory, or MIT.
### The Changing World of Earth Imaging

<table>
<thead>
<tr>
<th>Year</th>
<th>Satellite</th>
<th>GSD</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1959-1963</td>
<td>KH-1: 7.5m</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1960</td>
<td>KH-4: 1.8m</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1963</td>
<td>KH-4: 1.8m</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td>GeoEye - 2008</td>
<td></td>
<td>Panchromatic 90 cm 2 M</td>
</tr>
<tr>
<td>2013</td>
<td>Skybox</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>Planet Labs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>OmniEarth</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Growth in Volume of Data

- **1959-1963:** 866,041 frames
- **GeoEye - 2008:** 350,000 km²/day MSI
- **Worldview 3 - 2014:** 680,000 km² per day

### Key Satellites and Dates

<table>
<thead>
<tr>
<th>Date</th>
<th>Satellite</th>
<th>Resolution</th>
<th>GSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oct 24, 1946</td>
<td>V-2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aug 1960</td>
<td>Corona</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jul 1972</td>
<td>LANDSAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sep 1999</td>
<td>IKONOS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oct 2001</td>
<td>Quickbird</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007 …</td>
<td>Worldview 1,2,3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sep 2008</td>
<td>GeoEye</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nov 2013</td>
<td>Skybox</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jan 2014</td>
<td>Planet Labs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td>OmniEarth</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
By 2020, foreign competition likely will strengthen. France, Germany, India, Israel, Japan, and South Korea all should have mature commercial programs for optical, sub-meter imagery. Operators in other countries could also impact the market. Nothing can be done to slow this technology development because the United States does not control it.

*Alternative Futures: United States Commercial Satellite Imagery in 2020*

*Robert A. Weber and Kevin M. O’Connell*
New Analytic Methods for Commercial Imagery

- **Abundant imagery available:**
  - Repeated imaging of targets
  - Softcopy data in near real time
  - Easy access to reference data
    - Historical imagery
    - Geospatial information

- **New approaches to analysis:**
  - Automated analytics
  - Sampling methods & statistical characterization of phenomena
  - Merging with other geospatial information, e.g. social media
“… Wall Street getting in on this technology … UBS Investment Research issued its earnings preview for Wal-Mart's second quarter … publicly revealed that UBS had used satellite services … to gather the comings and goings of the parking lots at Wal-Mart stores…”

New Big Brother: Market-Moving Satellite Images

Satellite surveillance has found a new home in business

Eamon Javers | @EamonJavers
“Data mined from satellite imagery could serve as an early indicator of socially disruptive events like epidemics, especially in countries with limited surveillance resources.”

Satellite Imagery Analysis: What Can Hospital Parking Lots Tell Us about a Disease Outbreak?

Patrick Butler & Naren Ramakrishnan, Virginia Tech, Elaine O. Nsoesie & John S. Brownstein, Harvard Medical School and Boston Children’s Hospital
National Imagery Interpretability Ratings Scale (NIIRS) is a user-oriented measure of image interpretability

- Based on analytic tasks that can be done with an image
- Highly correlated with sensor parameters: GSD, relative edge response, contrast, noise
“This is just the beginning,” said Mr. Crawford, who also previously was a senior vice president of Climate Corp., a San Francisco company that uses weather data to help farmers predict their crop yield. Monsanto bought the company for more than $1 billion in 2013.

Startups Mine Market-Moving Data From Fields, Parking Lots—Even Shadows Firms Seek Insights on Business Outlook for Investors Seeking an Edge

Bradley Hope, The Wall Street Journal

- Early analysis indicates area under cultivation
  - Requires accurate crop classification
- Production depends on yield, driven by numerous factors
  - Weather
  - Irrigation (if applicable)
  - Genotype
  - Crop health
The Chernobyl reactor in the USSR exploded two months after SPOT’s launch, giving news organizations worldwide their best overhead view of the scene, and a way to “penetrate Soviet secrecy.”


“Wildfires cause extensive damage… modeling the probabilistic risk of a current fire becoming large and dangerous, based upon fire, weather and surrounding(s) extracted from remote [sensing data]”

Data Driven Approach to Estimating Fire Danger from Satellite Images and Weather Information
Natasha Markuzon and Stephan Kolitz
IEEE Applied Imagery Pattern Recognition Workshop, 2009
What useful indicators about society can be derived from commercial satellite imagery?

- Model and correlate the relationship between survey data and features derived from satellite imagery
  - Political and economic theories drive model formulation
- Characterize and quantify strengths and limitations of approach
  - Validate models by testing with sequestered data
  - Apply approach to new geographic regions and assess results

Project addresses novel acquisition of data

- Reduced risks and costs for data collection
- Data collection for remote areas is possible
- Enables data collection over large regions
What can remotely sensed imagery tell us about societies?

- **Food production:**
  - Agricultural analysis & crop estimation
- **Land use:**
  - Characterizing urban areas
- **Economic Well-being**
  - Relative affluence
  - Distribution of wealth
- **Governance**
  - Centrality & decision authority
  - Social Capital

Diagram:

- **OBSERVED**
  - Physical Structure
  - Social & Institutional Structure

- **INFERRED**
Indicators of Economic Conditions

Boston Area
- Somerville (Ball Square)

Rio de Janeiro
- Rocinha
- Zona Sul

http://persquaremile.com/2012/05/24/income-inequality-seen-from-space/
Afghanistan: Imagery & Survey Data

All imagery data provided by NGA

Imagery Data
Extensive commercial imagery
- Approx. 500+ images
- Panchromatic
- Multispectral

Surveys of 500 villages (10 districts)

- Male head of household
- Male focus group

Surveys conducted in 2007, 2009, and 2011

- Development interventions in 50% of villages

Surveys address economic, governmental, and societal issues
Question: Protector

Who do you think will protect the people in your village when they require protection against war or invasion by insurgents?

(a) Villagers  (b) Government

Classification Results

Correctly Classified  78 %
Incorrectly Classified  22 %
Exploration of Key Factors

![Graph showing the relationship between Edge Density and Tree Density with data points labeled as Government and Village.](image-url)
Question: Community Project

We want to ask you, do you have plans to jointly work with other people of the village on a project for the village next year?

(a) Yes  (b) No

Classification Results

Correctly Classified  77%
Incorrectly Classified  23%
Question: Visit Hajj

- We want to ask you, do you have plans to Visit to the Hajj next year?
  (a) No  (b) Yes

Classification Results

Correctly Classified 70%
Incorrectly Classified 30%
“Ties to Own Identity Group” are lower in urban regions

Responses related to urbanization and economic well being

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Coefficient</th>
<th>t-statistic</th>
<th>Significance level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>3.86</td>
<td>6.49</td>
<td>0.00</td>
</tr>
<tr>
<td>Spatial Correlation</td>
<td>2.96</td>
<td>2.56</td>
<td>0.05</td>
</tr>
<tr>
<td>Local Texture</td>
<td>-63.64</td>
<td>-2.28</td>
<td>0.07</td>
</tr>
</tbody>
</table>

$R^2 = 0.77$
“From a remote sensing perspective, informal settlements share unique spatial characteristics that distinguish them from other types of structures ... These spatial characteristics are often captured in high spatial resolution satellite imagery.”

Transformation of remote sensing
- Abundant imagery data
- Analytic resources are the primary constraint

Need to focus on automated, statistical methods
- Broad, shallow analysis
- Leverage extensive spatial and temporal sampling
- Fusion with other sources (geospatial reference data, social media)

New and novel analysis becomes possible
- Monitor and understand workings of the society
- Commercial applications
Questions?