MASH UP OF VHR IMAGE CORNER DENSITY WITH KILOMETRIC POPULATION DENSITY FOR THE AUTOMATIC DETECTION OF BUILT-UP EXTENT

Lionel Gueguen, Image Mining Labs, DigitalGlobe
Motivation

- Fast growth of population:
  - 1 additional billion of people every 12 years (first billion attained after 18000 years);
  - 95% of the growth takes place in developing countries;
  - 52% / 48% of population live in urban/rural environment, with an increasing rate of urban immigration;

- Impact on built-up development:
  - Rapid urban growth with environmental (water access, waste management...) and socio-economic implications (social services deterioration, poverty, ...);
  - Lack of consistent information about human settlements in the developing countries;

- Collect of built-up area informations for sustainable growth:
  - Current global estimations, based on 500m-1km remote sensing data, miss rural settlement clusters;
  - Information extracted from VHR imagery allows the detailed description and monitoring of built-up in urban/rural areas at a global scale;

Outline

• Automatic extraction of built-up extent from VHR at country scale:
  - Extraction of corner densities with textural measurements (PanTex);
  - Local model training with low resolution a priori knowledge;

• Joint statistical model:
  - Statistical link between PanTex distribution and kilometric population density;
  - Expectation-maximization local model fitting;

• Hi-res Urban Globe (HUG):
  - cloud computing (Hadoop) environment;
  - Japan case;
Corner density extraction from VHR: PanTex

- Textural Haralick-contrast aggregated along various directions (PanTex) is highly correlated to a corner density:
  - High discrimination power for built-up vs. non-built-up classification;
  - Highly dependent on local conditions of acquisitions (angles, atmospheric visibility and building pattern);
  - One classification model fails over many images;

VHR Optical input → PanTex measurements → Built-up extent
Low Resolution A Priori Knowledge

• Low resolution a priori knowledge:
  - Consistent information representation;
  - Global scale availability, no missing part;

• Landscan Kilometric Population:
  - Product of Oak Ridge National Labs;
  - multi-variable dasymetric disaggregation of census;
  - use of low resolution remote sensing images;
  - People counts per 1km cells;

• Natural link between built-up extent (BUE) and population density:
  - Exploit the link to adapt local classification model of PanTex values;
Statistical Link Between BU and Pop density

- We assume the availability of Built-up extent given at high resolution (<10 meters);
- We observe the statistical link between kilometric BU extent density and kilometric population density:
  - BU density increases linearly with pop density for small values;
  - After a population break point, the BU density saturates to its maximum;
- Model:

\[
\pi_{BU}(y \mid \lambda, s) = s(1 - e^{-u/s})
\]
Empirical analysis of PanTex measurements:
- Distribution computed in a urban area;
- Opposed to a distribution computed in agricultural, forest area.

Model:
- Mixture of two unimodal distributions (NBU vs. BU);
- Distribution shapes suggest the use of parametric Gamma distribution;
- Classification decision can be taken with the knowledge of both distributions (decision theory);

Title: Gamma distribution
url:http://en.wikipedia.org/wiki/Gamma_distribution

Optimal Classification Threshold
Mash-up Statistical Model

• Mixture of Gamma distributions over PanTex values:

\[
p(x \mid \pi_{BU}, k_{BU}, \theta_{BU}, \pi_{NBU}, k_{NBU}, \theta_{NBU}) = \\
\pi_{BU} \cdot p_\Gamma(x \mid k_{BU}, \theta_{BU}) + \pi_{NBU} \cdot p_\Gamma(x \mid k_{NBU}, \theta_{NBU})
\]

BU density  
Gamma dis. of PanTex in BU  
1 - BU density  
Gamma dis. of PanTex in Non BU

• Make use of population density to determine the BU density of the mixture model:

\[
\pi_{BU}(y \mid \lambda, s) = s(1 - e^{-y/\lambda})
\]

• Parameter estimation: break point, saturation, Gamma parameters
Parameter estimation by Expectation-Maximization

• Parameters resolution:
  - Break point, saturation are estimated in cells of ~ 20 km;
  - PanTex distribution parameters are estimated in the 1km population cells;
  - Given the parameters, we have one classification model per 1km cell;

• Spatial smoothness constraint:
  - Gamma distribution parameters form a multi-dimensional image at 1km resolution;
  - Apply low pass filtering at each algorithm iteration to ensure spatial smoothness of the final solution;

• Expectation-Maximization equations described in detailed in the paper;
Hi-res Urban Globe

• Mash-up implemented in a cloud computing framework:
  - Hadoop environment;
• Built-up extraction:
  - PanTex computed from 2-5 m multispectral imagery (Quickbird, WV-02, GeoEye-1, Ikonos);
  - Values spatial aggregated and locally classified to produce a Built-up extent at 8 meters;
• Enables Geospatial Big Data queries;
Hadoop Implementation

• Cloud computing architecture:
  - High scalability of storage capacity and of compute power;
  - Seamless portability of algorithms;
  - Scale up from single servers to thousands of machines, each offering local computation and storage;

• Hadoop Map-Reduce:
  - Highly robust to failures;
  - Optimized exploitation of cloud architecture;
  - Distributed computation;

![Hadoop Implementation Diagram]

- Map: Individual stripes
- Reduce: Built-up extent mosaic
Japan case

- Collection of Quickbird and WorldView-2;
- Area: 377,944km²
  - Equivalent to 1500 individual stripes;
- Cloud computing resources:
  - 18 nodes with 10 cpus;
  - 48GB RAM memory per node;
- HUG-BUE extraction time:
  - less than 2h00;
  - 17km²/min/cpu
Mosaick of Multispectral Inputs
HUG – Built-Up Extent
Local Classification of Pantex: Robust to Acquisition Variations
Comparison to kilometric built-up extent from MODIS land classification

HUG-BUE reproduces the MODIS built-up extent and add all the small clusters of buildings
Big Data Query

• What is the percent of built-up coverage in the country Japan?

<table>
<thead>
<tr>
<th>Modis - Land cover</th>
<th>HUG - BUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1%</td>
<td>5.5%</td>
</tr>
</tbody>
</table>

• The kilometric layer is missing ~25% of the built-up coverage being of the coarse resolution;

• The high resolution and the global coverage of Japan allows to get a tighter estimate of building stock;
Conclusion

• Extraction built-up extent from VHR multi-spectral imagery;
• Mash-up of kilometric population density with corner density;
• Local and adaptive classification;
• Cloud computing/Hadoop Map-Reduce for scalable and robust computation/storage;
• Hi-res Urban Globe – Built-Up Extent system:
  - Scalable built-up extraction at country, continental, global scale;
  - Support to Big Data queries;
Perspective: Large scale
Building footprint extraction
www.digitalglobe.com