Training Data Development for Mapping Global Impervious Surfaces Using Landsat Data

Jacqueline T. Phillips[1,3], Sarah E. Smith[1,3], Panshi Wang[2], Pui-Yu Ling[2], James C. Tilton[1], Bin Tan[1, 4], Chengquan Huang[2], Eric Brown de Colstoun[1], and Robert E. Wolfe[1]

[1] NASA Goddard Space Flight Center
[2] Department of Geographical Sciences University of Maryland
[3] USRA Goddard Earth Science Technology & Research
[4] Sigma Space Corporation

Introduction

The goal of the Global Land Survey-Impervious Mapping Project is to use Global Land Survey Landsat data to map and monitor global urbanization at 30 m resolution. Following a similar approach developed by DeFries et al. (1998, 1999) for AVHRR and MODIS data, and Yang et al. (2003) for the National Land Cover Dataset, we have produced a very large training data set based on interpreted high resolution commercial satellite data. This training set was derived using image segmentation tools and analysis and is used with a regression tree approach to produce the *first* continental-scale percent impervious cover data sets for the globe for 2010.

Step 1: Pre-Processing

1) Obtain High Resolution Imagery from the NGA
Scenes that were cloud free and from 2009 – 2011 were selected.

2) Review Imagery Order in ArcGIS 10.X
Scenes were organized and visually assessed for cloud cover, seasonality, sun angle, etc. using ArcGIS.

3) Re-Project & Subset
Images were then reprojected and orthorectified. Three 2500 x 2500 pixel subsets were created for each image.

Imagery Processed

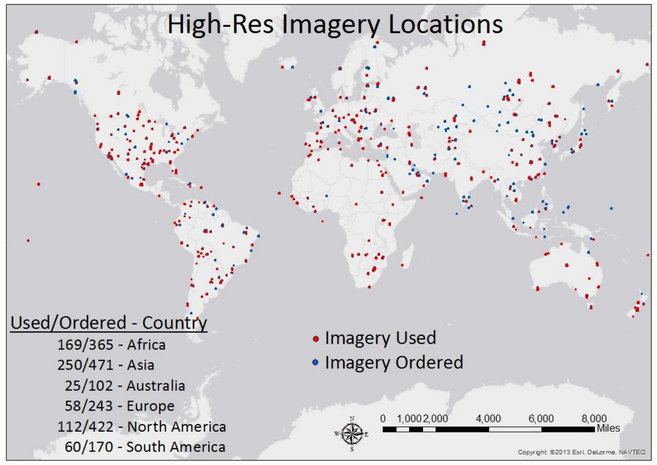

Figure 1. Map of training data locations depicting images ordered and images used.

Step 2: Training Process

HSegLearn was developed based on the HSeg software package Tilton et al. (2012). HSegLearn was built specifically for this project and provides a tool with which an analyst can label a particular ground cover class based on the hierarchical set of image segmentations produced by HSeg. For more information, please refer to the poster *HSegLearn – A Tool for Computer-Assisted Ground Reference Data Development* by James C. Tilton.

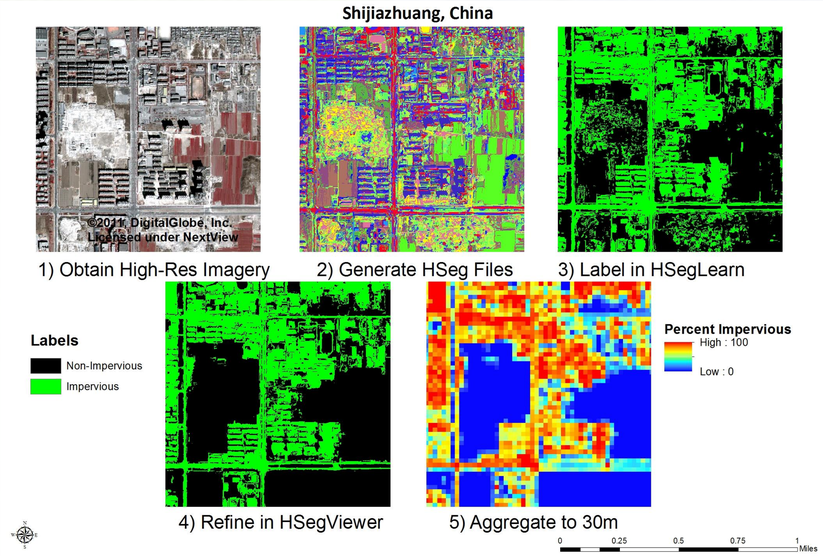

Figure 2. Steps in the training data process

Steps to Obtaining a Training Mask (Fig. 2)

1. Obtain reprojected and orthorectified high-resolution subset.
2. Generate HSeg files containing segmentation details.
3. Label impervious and non-impervious surfaces in HSegLearn.
4. Examine in HSegViewer to minimize commission or omission errors.
5. Aggregate the final training mask to percent impervious cover per pixel at the Landsat scale.

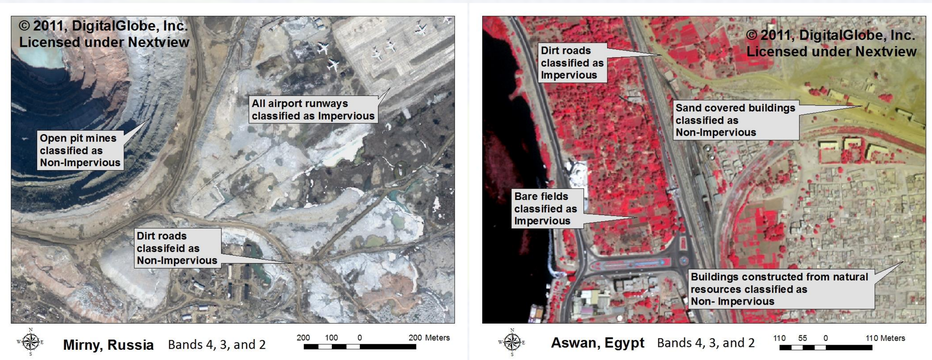

Figure 3. Maps of training data rules.

How impervious and non-impervious surfaces were labeled.

Impervious
- Man-made objects
- Ocean ports and docks
- Historical ruins

Non-Impervious
- Areas destroyed or overgrown
- Dirt roads and quarries
- Transient development

Step 3: Quality Assurance

Images were visually assessed and ranked based on the quality of the HSeg processing. Input from project analysts were examined as a group. This ranking was later used in the application of the regression tree. Figure 4 shows a few examples of these rankings.

- Imaged ranked 1 or 2 were used.
- Images ranked 3 or 4 were discarded.

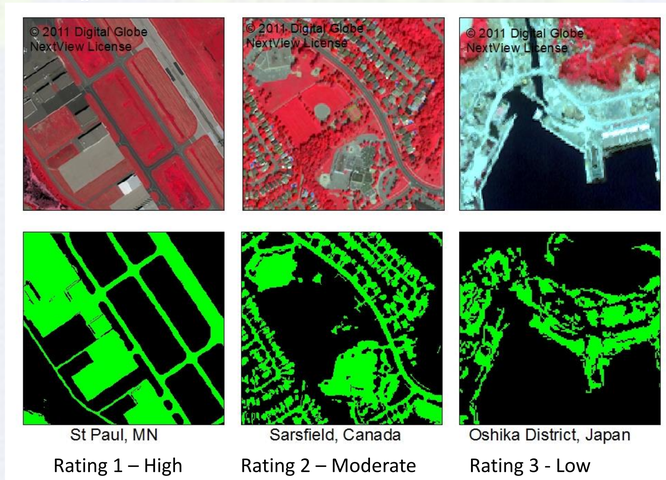

Figure 4. Examples of confidence rankings from 1 – 3. Imagery RGB 4, 3, 2.

Next steps in quality assurance:

- Verify accuracy of the cloud mask.
- Inspect for residual errors in orthorectification.

Training Data Summary

Continent	30m Training Pixels
Europe	2,816,916
North America	5,099,105
South America	3,571,483
Africa	3,603,191
Asia	9,505,845
Australia/Oceania	1,694,548
Total	26,291,088

Each of the ~26.3M training data pixels input into the regression tree is comprised of nine variables.

- Dependent variable: Percent impervious cover
- Independent variables: Surface reflectance for five Landsat bands (less Band 1) and vegetation indices (Kauth-Thomas and NDVI)

The training data are being used to produce the final continental scale products on a continent by continent basis, as well as validation of the final products.

References
DeFries, R.S., M. Hansen, J.R.G. Townshend, and R. Sohlberg (1998). Global land cover classifications at 8 km spatial resolution: the use of training data derived from Landsat imagery in decision tree classifiers. *Int. J. Remote Sens.*, 19:3141-3168.
DeFries, R.S., J.R.G. Townshend, and M. Hansen (1999). Continuous fields of vegetation characteristics at the global scale at 1km resolution. *J. Geophys. Res.*, 104:16911-16925.
Yang, L., C. Huang, C.G. Homer, B.K. Wylie, and M.J. Coan (2003). An approach for mapping large-area impervious surfaces: Synergistic use of Landsat 7 ETM+ and high spatial resolution imagery. *Canadian Journal of Remote Sensing*, *29*, 230-240.
Tilton, J. C. , Y. Tarabalka, P. M. Montesano and E. Gofman, "Best merge region growing with integrated non-adjacent region object aggregation," *IEEE Trans. Geosci. Remote Sens.*, 50(11), pp. 4454-4467, 2012.

NASA LCLUC Science Team Meeting, April 23-25, 2014, Rockville, MD