High-resolution land cover classification using low-resolution global map data

Mark J. Carlotto*, General Dynamics Advanced Information Systems

ABSTRACT

A fusion approach is described that combines texture features from high-resolution panchromatic imagery with land cover statistics derived from co-registered low-resolution global databases to obtain high-resolution land cover maps. The method does not require training data or any human intervention. We use an MxN Gabor filter bank consisting of M=16 oriented bandpass filters (0-180°) at N resolutions (3-24 meters/pixel). The size range of these spatial filters is consistent with the typical scale of manmade objects and patterns of cultural activity in imagery. Clustering reduces the complexity of the data by combining pixels that have similar texture into clusters (regions). Texture classification assigns a vector of class likelihoods to each cluster based on its textural properties. Classification is unsupervised and accomplished using a bank of texture anomaly detectors. Class likelihoods are modulated by land cover statistics derived from lower resolution global data over the scene. Preliminary results from a number of Quickbird scenes show our approach is able to classify general land cover features such as roads, built up area, forests, open areas, and bodies of water over a wide range of scenes.

Keywords: Gabor filters, land cover classification, texture features, image-map data fusion

1. INTRODUCTION

Land cover classification is important in mapping, navigation, automatic target recognition, and other applications. Historically much of the progress in land cover classification has been the result of exploiting spectral differences between land cover types. Sensors like Landsat were built specifically to support land use/cover classification and related applications. However the widespread availability of high resolution panchromatic imagery collected by commercial sensors has led to renewed interest in algorithms for panchromatic image classification.

We describe an algorithm that uses low-resolution terrain/feature databases to control a texture-based panchromatic image classifier. Our classifier operates on a multi-resolution texture representation of the image computed by a bank of Gabor filters (Section 2). Clustering reduces the complexity of the data by combining pixels that have similar texture into clusters (Section 3). Land cover classification is unsupervised and accomplished using a bank of texture anomaly detectors (Section 4). Class likelihoods are modulated based on land cover statistics derived from lower resolution global data over the scene. Preliminary results from a number of Quickbird pan image scenes that span a range of climates and seasons are presented along with conclusions in Section 5.

2. TEXTURE FEATURE EXTRACTION

Texture feature extraction starts with an MxN Gabor filter bank consisting of *M* oriented bandpass filters at *N* resolutions (bandpass frequencies), specifically *M*=16 filters (0-180°) at *N*=4 resolutions from 3-24 meters wide (at a typical image resolution of 0.67 meters/pixel imagery):

$$g_{x,y}(m,n) = \exp\left(-\frac{u_m^2+\gamma^2 v_m^2}{2\sigma_n^2}\right)\cos\left(\frac{2\pi u_m}{\lambda_n}\right)$$

* mark.carlotto@gd-ais.com

with

$$
\begin{aligned}
u_m &= x\cos\theta_m + y\sin\theta_m \\
v_m &= -x\sin\theta_m + y\cos\theta_m
\end{aligned}
$$

where x and y are image coordinates, and θ_m is the orientation of the m-th filter. The other variables are described by Serre and Riesenhuber [1]. The output from the filter bank is a set of MxN = 64 images

$$
f_{x,y}(m,n) = \left\| BRI_{x,y} * g_{x,y}(m,n) \right\|
$$

where BRI is the pan brightness image. The filter bank is implemented in the frequency domain with MN+1 FFTs per image.

Recognizing that roads, built up areas, and other kinds of structured features can appear at any orientation we rotate the Gabor features. If $p(n)$ is the direction with the maximum response at resolution n, the rotated features are

$$
h_{x,y}(m,n) = f_{x,y}\left(m + p(n) \bmod M, n\right).
$$

Linear features like roads have an anisotropic power spectral density [2] with most of the power in the same direction at each resolution. Rectilinear features and non-structured textures can have anisotropies in different directions at each resolution. We compute the dominant direction across resolution using a voting scheme. The voting method adds two votes per resolution in an M-element accumulator array. The first vote is in the direction with the largest response, the second vote in the direction with the next highest response, and so forth for a total of eight votes across all N=4 resolutions. The orientation m^* is the direction with the largest number of votes.

Instead of using the raw Gabor filter outputs we developed a set of physically-based texture features for classification. The first, linearity (LIN) responds to surface features oriented in a particular direction like roads, waterways, hedgerows, etc.

$$
LIN_{x,y}(n) = \tfrac{1}{3} \sum_{m=0,1,M-1} h_{x,y}(m,n)
$$

where

$$
h_{x,y}(m,n) = f_{x,y}\left(m + m^* \bmod M, n\right)
$$

The second, rectilinearity (REC) is designed to respond to features with a secondary orientation that is roughly perpendicular to the primary orientation like fields, buildings, vehicles, etc.

$$REC_{x,y}(n) = \tfrac{1}{3} \sum_{m=M/2-1,M/2,M/2+1} h_{x,y}(m,n)$$

The filters added on either side of the primary and secondary orientations provide a certain tolerance to linear features that are not perfectly straight and rectilinear features not exactly at right angles (e.g., which occurs when a rectangular feature is viewed off-nadir). A third nonstructured texture (TXT) feature was designed to respond to other textures like trees, rough, and dissected terrain

$$TXT_{x,y}(n) = \tfrac{1}{10}\left[\sum_{m=0}^{M-1} h_{x,y}(m,n) - 3\times LIN_{x,y}(n) - 3\times REC_{x,y}(n)\right]$$

Each of these features is computed at N=4 resolutions for a total of 13 features including brightness (BRI).

3. CLUSTERING

Large images (scenes) are divided into smaller sized tiles (typically 1024 x 1024 pixels) for processing. Texture features are clustered to reduce computational complexity for downstream processing from 10^6 pixels to about 10^2 texture clusters per tile. Tile clusters are then re-clustered with those from the other tile clusters to produce a set of super-clusters, which describe similar textures within the scene. The K-means algorithm is used both for tile and scene clustering. To promote the formation of compact clusters we divide each feature by its global mean:

$$\frac{\phi(i)}{E\left[\phi(i)\right]}$$

where the expectation is computed over the image, $\Phi = \left[\phi(i)\right]$ is a lexographic ordering of the 12 physically-based texture features plus brightness; i.e.,

$$\Phi_{x,y} = \begin{bmatrix} TXT_{x,y}(0) \\ REC_{x,y}(0) \\ LIN_{x,y}(0) \\ \vdots \\ TXT_{x,y}(3) \\ REC_{x,y}(3) \\ LIN_{x,y}(3) \\ BRI_{x,y} \end{bmatrix}$$

The number of clusters per tile and per scene are set by the user, typically K=256. After clustering the texture feature vector $\mathbf{\Phi}_k$ represents the average values of the 13 features within the k-th cluster.

4. LAND COVER CLASSIFIER

Global terrain and feature data over the scene controls the classification process. We use 30 arc-second (1 km) global AVHRR land cover class data [3] to estimate the relative frequency of classes within the scene being processed. The AVHRR data is based on a 14-class categorization scheme:

- Water
- Evergreen Needleleaf Forest
- Evergreen Broadleaf Forest
- Deciduous Needleleaf Forest
- Deciduous Broadleaf Forest
- Mixed Forest
- Woodland
- Wooded Grassland
- Closed Shrubland
- Open Shrubland
- Grasslands
- Cropland
- Barren
- Urban and Built-up

These classes are derived from spectral data and so do not extend well to texture classification. A smaller number of classes that can be computed from brightness and texture are used instead:

- Linear features
- Rectilinear features
- Non-structured features
- Open areas
- Water/shadow

Not surprisingly the first three classes have the same meaning as the three physically-based, texture-derived features described earlier – linearity, rectilinearity, and non-structured texture. Open areas can be understood as regions that lack texture. Water and shadows are those parts of open areas that relatively dark.

To detect textured regions we compute the following normalized statistics:

$$d_k^{LIN}(n) = \frac{LIN_k(n) - \mu_{LIN}(n)}{\sigma_{LIN}(n)}$$

$$d_k^{REC}(n) = \frac{REC_k(n) - \mu_{REC}(n)}{\sigma_{REC}(n)}$$

$$d_k^{TXT}(n) = \frac{TXT_k(n) - \mu_{TXT}(n)}{\sigma_{TXT}(n)}$$

where k is the cluster and n is the resolution. The pixels in each cluster are obtained from the cluster map $k(x,y)$, which is a by-product of the clustering process. The means and variances are computed as averages over all of the clusters in the scene

$$\mu_{LIN}(n) = E\left[LIN_k(n)\right] \quad \sigma^2_{LIN}(n) = E\left[\left(LIN_k(n)\right)^2\right] - \left(\mu_{LIN}(n)\right)^2$$

$$\mu_{REC}(n) = E\left[REC_k(n)\right] \quad \sigma^2_{REC}(n) = E\left[\left(REC_k(n)\right)^2\right] - \left(\mu_{REC}(n)\right)^2$$

$$\mu_{TXT}(n) = E\left[TXT_k(n)\right] \quad \sigma^2_{TXT}(n) = E\left[\left(TXT_k(n)\right)^2\right] - \left(\mu_{TXT}(n)\right)^2$$

For each cluster at each resolution we 1) find the class with the greatest statistical deviation, and 2) determine if the deviation exceeds a threshold. The threshold depends on the relative fraction of textured area, which is based, in turn, on the global landcover data over the scene. An empirically-derived mapping was developed for Quickbird (Table 1). The values depend primarily on resolution, although season, culture, population density, and other factors are also important.

Table 1 Mapping global landcover to texture fractions

Class	**Fraction**	**Class**	**Fraction**	**Class**	**Fraction**
Water	0	Evergreen Needleleaf Forest	0.8	Evergreen Broadleaf Forest	0.8
Deciduous Needleleaf Forest	0.8	Deciduous Broadleaf Forest	0.8	Mixed Forest	0.8
Woodland	0.8	Wooded Grassland	0.5	Closed Shrubland	0.3
Open Shrubland	0.3	Grassland	0.1	Cropland	0.2
Bare Ground	0.1	Urban and Built Up	0.9		

The initial textured/non-textured decision at resolution n is:

$$\omega_k(n) = \begin{cases} \operatorname{argmax}\left\{d_k^{LIN}(n), d_k^{REC}(n), d_k^{TXT}(n)\right\}, \mathit{if}\ \max\left\{d_k^{LIN}(n), d_k^{REC}(n), d_k^{TXT}(n)\right\} > t \\ OPEN, otherwise \end{cases}$$

The detection threshold satisfies

$$p_0 = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-x/2} dx$$

where p_0 is the fraction of the image that is textured, which is computed as an average of the global landcover data over the scene. The textured/non-textured decisions at the four resolutions are combined by picking the class that "fires" most; i.e., one of

{LINEAR,RECTILINEAR,NONSTRUCTURED,OPEN}

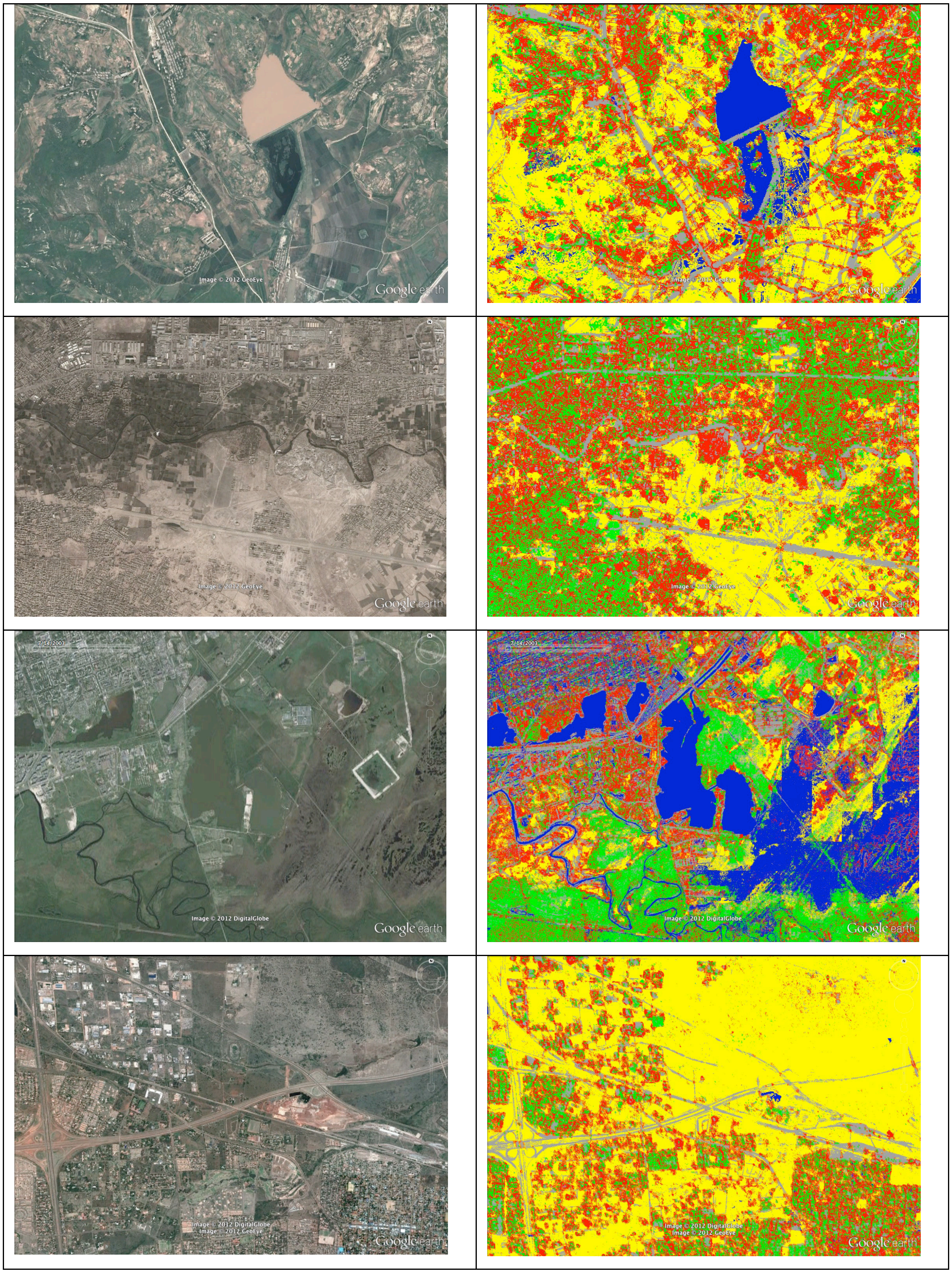

Figure 1 ALE landcover maps (right) – from top to bottom: northern Asia, southern Asia, northern Europe, and southern Africa. Reference images (left) courtesy Google Earth.

In the event of a tie, the class with the greatest deviation is selected. Water and shadows are detected from brightness. Using a similar idea,

$$\omega_k = \begin{cases} WATER/SHADOW, if \ \ BRI < b^* \\ \{LINEAR, RECTILINEAR, UNSTRUCTURED, OPEN\}, \ otherwise \end{cases}$$

The brightness threshold is the value that comes closest to satisfying

$$F(b^*) = \%\, Water$$

where F is the cumulative distribution of the brightness image histogram. The percentage of the scene that is water is estimated from GTOPO30 – 30 arc-second (1 km) global elevation data [4] as its resolution is better than the AVHRR land cover. If the elevation is less than or equal to zero or if the slope is zero, we infer water. The method tends to over-estimate the amount of water but can be refined using histogram analysis techniques [5].

5. RESULTS AND CONCLUSIONS

Figure 1 shows preliminary results from four scenes processed. The color scheme is:

- linear features – gray
- built up areas – red
- open areas – yellow
- non-structured features such as forests and rough terrain – green
- water/shadows – blue

In the first scene (northern Asia) acquired in winter there is good separation between built-up, forested areas, and water. Hedgerows between fields are extracted as linear features along with roads, rivers, and drainage features. Where this scene was acquired in winter with the sun low in the sky, the second scene (the southern Asia) was taken in the summer. This is a good test of the landcover classifier with respect to seasonal differences. Some built up areas that contain smaller structures like mud huts are misclassified as natural non-unstructured terrain. Neither of the previous scenes contains significant vegetation. A large portion of third scene (northern Europe) taken in summer is vegetated and so is another good test of the classifier. The results show very good discrimination between vegetation, built areas, roads, open areas, and water. The fourth scene is a fall image in southern Africa. Gaps in the road network occur here and in other scenes where roads are wider than the largest filter in the filter bank. In half-meter resolution imagery a six-lane divided highway (which is at least 6x16' wide) is wider than the passband of the largest Gabor filter (24 meters). A possible solution is to add another filter bank, which will increase computational complexity by 20%. Gaps also occur when the contrast between the road and surrounding terrain is low.

Since the classifier uses only brightness to discriminate water, there is confusion between water and shadows. A finer-grained method is required to distinguish the two. There is also some confusion between built up areas and natural terrain. Although one would think manmade and natural features are easily separated there are many situations where it is not the case. Examples include forests with tree shadows forming linear and rectilinear patterns, hamlets, groupings of mud huts, and other small structures in irregular (non-rectilinear) arrangements, and others. A possible solution is to modulate classification decisions in those areas, e.g., varying the range of scales (resolutions) over which manmade features are detected, based on cultural context. However the fact remains that surface features and lighting can and will conspire to create regular patterns that can be difficult for the texture classifier to distinguish from man-made activity, and vice versa.

REFERENCES

[1] T. Serre and M. Riesenhuber, "Realistic modeling of simple and complex cell tuning in the HMAX model, and implications for invariant object recognition in cortex," *Technical Report CBCL Paper 239 / AI Memo 2004-017*, Massachusetts Institute of Technology, Cambridge, MA, July 2004.

[2] Mark J. Carlotto, "Detecting Patterns of a Technological Intelligence in Remotely-Sensed Imagery," *J. British Interplanetary Soc*., Vol. 60, pp. 28-39, 2007.

[3] University of Maryland Department of Geography http://glcf.umiacs.umd.edu/data/landcover/

[4] NOAA National Geophysical Data Center http://www.ngdc.noaa.gov/mgg/topo/globe.html

[5] Mark J. Carlotto, "Histogram analysis using a scale-space approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 9, No. 1, Jan. 1987.