Using High Resolution Commercial Satellite Imagery to Quantify Spatial Features of Urban Areas and their Relationship to Quality of Life Indicators in Accra, Ghana

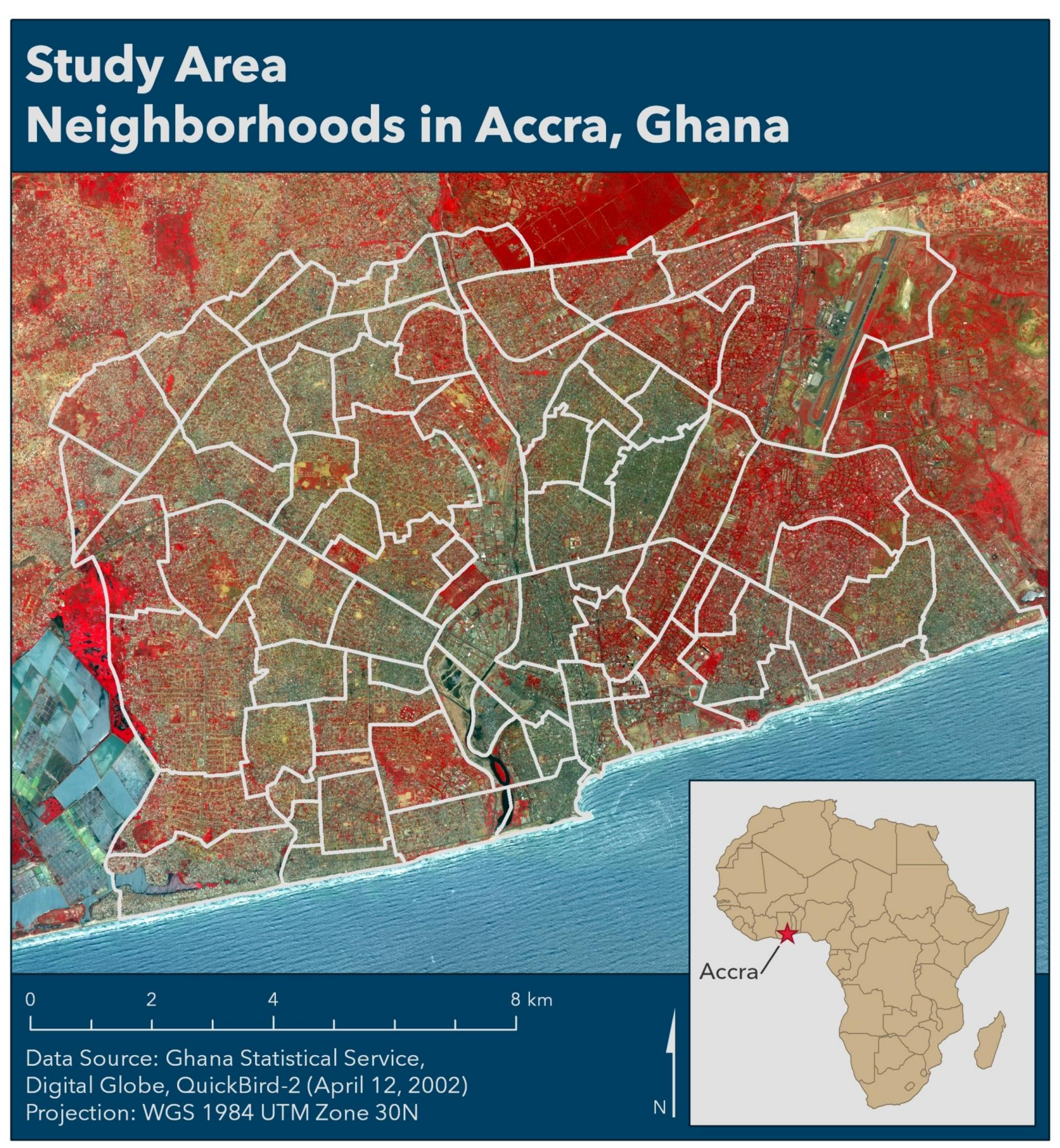
Avery Sandborn[1], Ryan Engstrom[1], and Qin Yu[1]
[1]The George Washington University, Department of Geography, Washington, DC

Poster: PA31B-4062

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Abstract

Mapping urban areas via satellite imagery is an important task for detecting and anticipating land cover and land use change at multiple scales. As developing countries experience substantial urban growth and expansion, remotely sensed based estimates of population and quality of life indicators can provide timely and spatially explicit information to researchers working to determine how cities are changing. In this study, we use commercial high spatial resolution satellite imagery, in combination with fine resolution census data, to determine if remotely sensed data is an effective and appropriate means of identifying spatial patterns of quality of life in Accra, Ghana. Traditionally, spectral characteristics are used on a per-pixel basis to determine land cover; however, in this study, we test a new methodology that quantifies spatial characteristics of urban areas using a variety of spatial features observed in the imagery. Spatial features, such as Line Support Regions (LSR) and PanTex, focus on extracting structural and textural patterns of built-up areas, such as roads, buildings, and building shadows. Information derived from aggregating the descriptive statistics of these spatial features at the neighborhood level are then compared to census derived quality of life indicators including information about housing and employment. The Normalized Difference Vegetation Index (NDVI) is also computed for comparative analysis between spatial and spectral features. Results indicate that the strongest correlations are between LSR and lighting source and literacy rates, PanTex and immigration and employment variables, and NDVI and housing and slum variables. Final results from this study will be used to determine if this methodology provides a new and improved way to measure a city structure in developing cities, and differentiate between residential and commercial land use zones, as well as formal and informal housing areas.



Data Acquisition

Satellite Imagery
- April 12, 2002 Quickbird-2 multispectral image (2.44m resolution)
- Covers approximately 80% of the Accra Metropolitan Assembly (AMA)
- Georeferenced using a second order polynomial transformation, 13 ground control points, and a digital elevation model (merged Cartosat, Spline, and ASTER terrain data)

Census Data
- 2000 Ghanaian census obtained from the Ghana Statistical Service (GSS)
- Contains detailed information about population, housing, education, and health
- Records compiled at the enumeration area (EA) level, equivalent to a US census tract and contain about 1000 people per EA
- EA units digitized into a GIS format and spatially georeferenced to the satellite imagery (Engstrom et al. 2011)
- EA units aggregated from 1724 EAs into 107 neighborhoods (Engstrom et al. 2013)
- Only neighborhoods that are fully covered by the Quickbird-2 image are included, reducing the number to 74 neighborhoods

Methodology

Block and Scale Size
Each spatial feature (LSR, PanTex, and NDVI) are computed with each combination of block and scale size (Figure 1).

Block size represents the pixel size at which the output feature will be aggregated to.
- Block Size 4 = 9.76m output
- Block Size 8 = 19.52m output

Scale size represents the area from which the spatial feature will extract contextual information from.
- Scale Size 8 = 19.52m
- Scale Size 16 = 39.04m
- Scale Size 32 = 78.08m

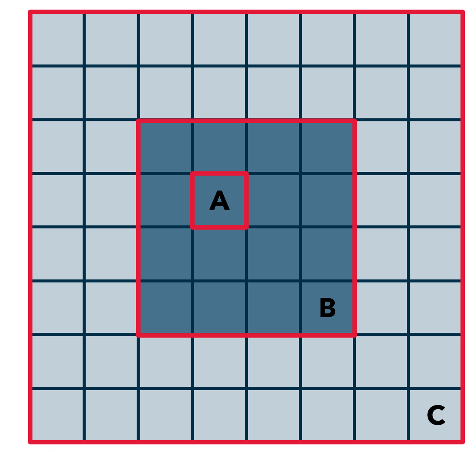


Figure 1: A) Pixel size 2.44m, B) block size 9.76m and C) scale size 19.52m

Line Support Regions (LSR)
LSR represent lines and their attributes extracted from the imagery (Figure 2).

Steps to Compute LSR
- Pixels are grouped and binned by gradient orientations
- Grouped pixels are taken as support regions
- Each support region represents a candidate area for a straight line
- Only areas with high gradient direction variation are extracted as line segments
- Line attributes can be calculated, such as location, orientation, length, and width

3 Features Returned
- Sum of line lengths
- Mean of line lengths
- Line Variance

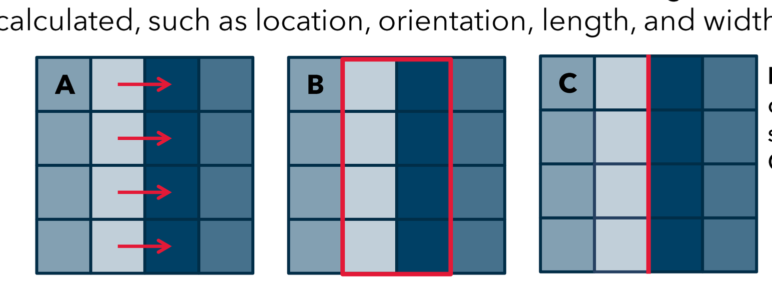


Figure 2: A) Gradient orientation, B) line support region, and C) line extraction

Output Interpretation
- Urban areas contain more frequent and longer line segments
- Rural areas contain shorter and more randomly distributed line segments
- Formal neighborhoods tend to have more uniform line segments
- Informal neighborhoods tend to have more inconsistent line patterns

PanTex
PanTex is a texture measure derived from the grey-level co-occurrence matrix (GLCM) which extracts built-up areas from panchromatic imagery (Figure 3). PanTex helps to extract buildings, while eliminating roads and other boundaries from the index.

Steps to Compute PanTex
- Compute GLCM contrast in multiple directions
- Take the minimum of all directional contrast
- Average operator produces undesirable edge effect

Output Interpretation
- High values indicate objects (buildings) are the size of the pixel block
- Low values indicate that objects are not the size of the pixel block (roads, fields)

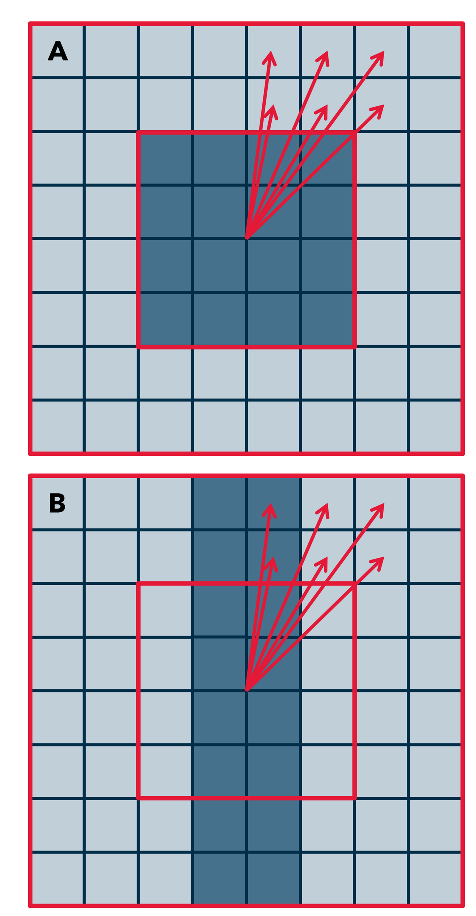


Figure 3: A) PanTex calculation for a building and B) PanTex calculation for a road

Normalized Difference Vegetation Index (NDVI)
NDVI is one of the best known and most widely used vegetation indices.

$$NDVI = \frac{(NIR-R)}{(NIR+R)}$$

Output Interpretation
- High values indicate the highest possible density of green vegetation and less built-up area
- Low values indicate less dense areas of vegetation and more built-up area

Zonal Statistics
Once all spatial features were computed, the descriptive statistics were calculated for each spatial feature by neighborhood. Specifically, zonal statistics summarize the values of a raster (average, standard deviation, and sum) within the zone of the neighborhood shapefile, and writes the data to a database file.

Correlation Coefficient
A correlation coefficient matrix was created to measure the strength and direction of the relationships that exist between the zonal statistics of the spatial features and the census data. Due to a fairly large dataset of 75 neighborhoods and an alpha significance level of 0.01, the critical value for correlation coefficients was found to be 0.3.

Output Interpretation
- Correlation coefficients greater than 0.5 = strong positive correlations
- Correlation coefficients less than -0.5 = strong negative correlations
- Correlation coefficients between 0.3 and 0.5 = weak positive correlations
- Correlation coefficients between -0.3 and -0.5 = weak negative correlations

Results

Table 1: Strongest correlation coefficients between line support regions and census data

Line Support Region (LSR)	Lighting Source Not Electric	Not Literate Age 15-24	Tenancy Rent-free	Population Density
Block 4, Scale 16, Mean of Line Length Average	-0.34	-0.27	-0.50	0.45
Block 4, Scale 16, Mean of Line Length Sum	-0.53	-0.53	-0.43	0.15
Block 4, Scale 16, Mean of Line Variance	-0.54	-0.53	-0.42	0.13
Block 4, Scale 16, St.Dev of Line Length Average	0.51	0.53	0.25	-0.02
Block 4, Scale 32, Mean of Line Length Average	-0.24	-0.18	-0.49	0.50
Block 4, Scale 32, Mean of Line Length Sum	-0.52	-0.48	-0.43	0.13
Block 4, Scale 32, Mean of Line Variance	-0.52	-0.49	-0.42	0.10
Block 4, Scale 32, St.Dev of Line Length Average	0.53	0.45	0.46	-0.24
Block 8, Scale 16, Mean of Line Length Average	-0.33	-0.28	-0.51	0.44
Block 8, Scale 16, Mean of Line Length Sum	-0.54	-0.54	-0.45	0.14
Block 8, Scale 16, Mean of Line Variance	-0.54	-0.56	-0.44	0.11
Block 8, Scale 16, St.Dev of Line Length Average	0.51	0.54	0.24	-0.02
Block 8, Scale 16, St.Dev of Line Variance	0.47	0.20	0.53	-0.43
Block 8, Scale 32, Mean of Line Length Average	-0.22	-0.18	-0.50	0.49
Block 8, Scale 32, Mean of Line Length Sum	-0.53	-0.50	-0.44	0.11
Block 8, Scale 32, Mean of Line Variance	-0.53	-0.52	-0.42	0.08
Block 8, Scale 32, St.Dev of Line Length Average	0.55	0.48	0.45	-0.23

LSR Results
- Higher population density in areas with longer lines (more built-up), lower population density in areas with shorter lines (less built-up)
- Higher illiteracy rates and less electricity in areas with varied line lengths (less built-up), and higher literacy rates and more electricity in areas with similar line lengths (more built-up)
- Block size was not a factor in correlation coefficients, however, scale size 8 did not have any strong correlations with census data

Strong Positive Correlation (Greater than 0.5)
Moderate Positive Correlation (0.4 to 0.5)
Weak Positive Correlation (0.3 to 0.4)
No Significant Correlation (-0.3 to 0.3)
Weak Negative Correlation (-0.3 to -0.4)
Moderate Negative Correlation (-0.4 to -0.5)
Strong Negative Correlation (Less than -0.5)

Table 2: Strongest correlation coefficients between PanTex and census data

PanTex	Informal Sector	Lighting Source Not Electric	Percent Born Outside of Neighborhood	Percent Who Were Not in Accra 5 Years Ago	Ga-Adangme Ethnic Group	Number of Dependents	Percent Who Do Not Work	Tenancy Rent-free
Block 4, Scale 8, Mean of PanTex Value	0.51	-0.02	-0.15	-0.22	-0.12	-0.10	-0.12	-0.55
Block 4, Scale 8, St.Dev of PanTex Value	0.15	0.13	0.44	0.19	-0.59	-0.45	-0.29	-0.30
Block 4, Scale 16, Mean of PanTex Value	0.48	0.01	-0.12	-0.19	-0.13	-0.13	-0.14	-0.52
Block 4, Scale 16, St.Dev of PanTex Value	-0.07	0.41	0.61	0.48	-0.52	-0.54	-0.49	-0.02
Block 4, Scale 32, St.Dev of PanTex Value	-0.12	0.60	0.64	0.62	-0.39	-0.56	-0.59	0.16
Block 8, Scale 8, Mean of PanTex Value	0.50	-0.01	-0.14	-0.21	-0.12	-0.11	-0.14	-0.54
Block 8, Scale 8, St.Dev of PanTex Value	0.08	0.21	0.47	0.24	-0.59	-0.47	-0.37	-0.27
Block 8, Scale 16, Mean of PanTex Value	0.48	0.00	-0.12	-0.19	-0.12	-0.13	-0.14	-0.52
Block 8, Scale 16, St.Dev of PanTex Value	-0.03	0.39	0.59	0.45	-0.52	-0.54	-0.47	-0.04
Block 8, Scale 32, St.Dev of PanTex Value	-0.11	0.60	0.63	0.62	-0.38	-0.56	-0.59	0.16

PanTex Results
- Immigrants tend to live in heterogeneous areas (mixed buildings and non-buildings), while non-immigrants live in homogeneous areas (mostly buildings)
- Higher unemployment rates and the informal employment sector tend to occur in homogeneous areas (mostly buildings) while lower unemployment rates and the formal sector tend to occur in heterogeneous areas (mixed buildings and non-buildings)
- Less electricity in heterogeneous areas (mixed buildings and non-buildings), and more electricity in homogeneous areas (mostly buildings)
- Block size was not a factor in correlation coefficients, however, scale size 32 did not appear to be as influential as scale sizes 8 and 16

NDVI Results
- There was not a significant differentiation between certain block sizes and scale sizes
- The mean of NDVI had correlation coefficients below ranging from -0.51 to -0.86 for eleven variables:

Biofuel, housing density, informal sector, no sewerage for liquid waste, no sewerage for solid waste, non-separate cooking space, not single family house type, percent of women not secondary educated, population density, slum index, and unimproved sanitation

Conclusions

- Although NDVI revealed strong correlations between more variables, both LSR and PanTex were able to reveal additional significant correlations that were not captured by NDVI
- Scale size appears to play a more important role than block size in the spatial features ability to extract information, although LSR favored scales of 8 and 16, while PanTex favored scales of 16 and 32

Future Work

- To capture more spatial properties of the urban areas, additional spatial features can be computed, including Fourier transform, gabor filters, histogram of oriented gradients (HOG), hough probabilistic transform, lacunarity, and local binary patterns
- Future work will also entail running a land cover classification algorithm on a virtual stack of all spatial features

Bibliography

- Burns, J., Hanson, A., & Riseman, E. (1986). Extracting Straight Lines. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-8 (4),* 425-455.
- Deering, D. W. (1978). *Rangeland reflectance characteristics measured by aircraft and spacecraft sensors*. Ph. D. Dissertation, Texas A & M University, College Station, TX, 338 pp.
- Engstrom, R., Ashcroft, E., Jewell, H., & Rain, D. (2011). Using Remotely Sensed Data to Map Variability in Health and Wealth Indicators in Accra, Ghana. *Joint Urban Remote Sensing Event (JURSE), 2011,* 145-148.
- Engstrom, R., Ofiesh, C., Rain, D., Jewell, H., & Weeks, J. (2013). Defining Neighborhood Boundaries for Urban Health Research in Developing Countries: A Case Study for Accra, Ghana. *Journal of Maps,* 1-7. DOI:10.1080/17445647.2013.765366.
- Graesser, J., Cheriyadat, A., Vatsavai, R. R., Chandola, V., Long, J., & Bright, E. (2012). Image Based Characterization of Formal and Informal Neighborhoods in an Urban Landscape. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 5 (4),* 1164-1176.
- Pesaresi, M., Gerhardinger, A., & Kayitakire, F. (2008). A Robust Built-Up Area Presence Index by Anisotropic Rotation-Invariant Textural Measure. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 1 (3),* 180-192.
- Unsalan, C. (2006). Gradient-Magnitude-Based Support Regions in Structural Land Use Classifications. *IEEE Geoscience and Remote Sensing Letters, 3 (4),* 546-550.

Acknowledgments

This research was funded by the NASA grant entitled "The Urban Transition in Ghana and Its Relation to Land Cover and Land Use Change Through Analysis of Multi-scale and Multi-temporal Satellite Image Data" Grant Number#: G00009708, NASA Award Number: NNX12AM87G.