Understanding vegetation indices' efficiency in land use change detection on Minnesota Urban Vegetation

Ali Hossaini

GIS 5577
Spring 2021

Abstract

In recent studies, many approaches have been exploited for automatic land use change detection, specifically vegetation extraction, from satellite imagery. With the completion of the a vegetation land use dataset, planners have the ability to map changes in urban growth and development in a geographic information system (GIS) database and can better visualize development trends and anticipate future growth needs. The homogeneity of features and their reflectance in specific bands provide a great potential to utilize image indices as an efficient approach toward improving vegetation extraction and monitoring land use change within time periods. In this project, I examined vegetation spectral indices to extract green spaces' polygons and compared the result with spatial historical land use tabular data in PostGIS geodatabase management system. That is to say, this project is an attempt to visualize the vegetation areas from above space scene in satellite imagery to improve database design in monitoring vegetation land use change. This is done by employing vegetation feature indices on a landscape classification of vegetation areas in Seninel-2 satellite images. By doing so, this project found a considerable decrease in total area of green spaces in the area of study, and more general approved the analytical capabilities of spatial database design in land use change detection in urban regions.

Introduction

Many studies noted the importance of satellite systems in Remote Sensing (RS) which has continually been rising in response to the increasing demand on urban planning, emergency management, change detection, crop estimation and so on [1,2]. The goal of RS is to extract information and identify interested targets in satellite imagery to complete image understanding [3,4]. Some of these studies have primarily focused on land use change detection from a RS imagery which is a challenging but important research topic [5,6,7]. However, monitoring the change in land use of agricultural lands in current urban areas during time periods have been less considered, regarding the land signature in image bands of such areas.

In raster imagery, vegetation bodies appear as green polygon structures and sometimes do not appear having distinct lines with neighboring lands [7,8]. Depending on the geographic location, they often have homogenous square-shaped or circular-shaped profile in the images and can be seen directly as collinear distinct edges of field borders which is highlighted in their reflectance in specific bands[7]. For instance, Waldner[9] examined a cropland mapping survey in Brazil and found out a land-cover class was assigned to most of the large fields clearly identifiable from the built-up surrounding. Agricultural areas show high intensities in the near infrared band [3,10] and make it possible to use image indices to distinct vegetation features.

There are different approaches to vegetation extraction methods such as classification-based methods which can be classified into supervised and unsupervised classification methods [2]. Masek [11] et al. established a technique based on a normalized difference vegetation index (NDVI) to distinguish built-up land features among croplands, through unsupervised classification. Then, Zha [12,13] et al. developed the novel normalized difference built-up index (NDBI) to automate the mapping process of built-up areas and to accurately extract urban features by arithmetically manipulating NDBI and NDVI. However, this method was ineffective because the extracted features did not distinguish vegetation areas from bare land successfully.

To map urban areas, Xu [14] developed the index-based built-up index. This index is derived from three different indices, namely, the soil adjusted vegetation index (SAVI), modified normalized built-up index, and NDBI. Although the results showed that these indices are automated, transferable, and efficient in built-up extraction in urban areas [17], finding an efficient accurate vegetation index in urban areas is still a research gap.

Furthermore, several previous studies focused on spectral indices for Landsat TM imagery [18]. Despite the effectiveness of these indices on such satellite data, they may be difficult to apply to novel imagery with high spatial resolution [19]. For instance, Sentinel-2 satellite constellation has a good potential for vegetation extraction, providing imagery data in 13 spectral bands with high temporal resolution (5 days) and high spatial resolutions (10-60 m) [20]. Thus, utilizing vegetation indices based on land spectral signature in agricultural areas on Sentinel-2 time-series images would be a promising approach toward monitoring change in those areas. The result of this extraction provided a polygon shapefile of green spaces which is loaded to PostGIS to be analyzed with historical land use database and give us the total area of change in vegetation land use.

Database description

This project includes a geodatabase of 4 spatial tables. The conceptual design of the database is based on finding land features and land use type among tabular data and vegetation index output raster. The goal is to monitor vegetation land use change in the urban area of interest. Thus, one table is needed to be loaded to the database to show historical land use of the urban region. Another one is the county borders which is necessary to clip the study area. The database also contains a table of vegetation index polygon extraction result. Finally, the fourth table stores the analytical operations about the area statistics (i.e. NDVI Mean, Max, Min, and total sum of area).

Data

The “Historical Generalized Land Use” dataset encompasses the area of study for this project which is the seven county Twin Cities (Minneapolis and St. Paul) Metropolitan Area in Minnesota (i.e. comprises Anoka, Carver, Dakota, Hennepin, Ramsey, Scott and Washington counties). The dataset was developed by the Metropolitan Council, a regional governmental organization that deals, in part, with regional issues and long-range planning for the Twin Cities area. The data were interpreted from 1984, 1990, 1997, 2000, 2005, 2010 and 2016 air photos and other source data, with additional assistance from county parcel data and assessor's information.

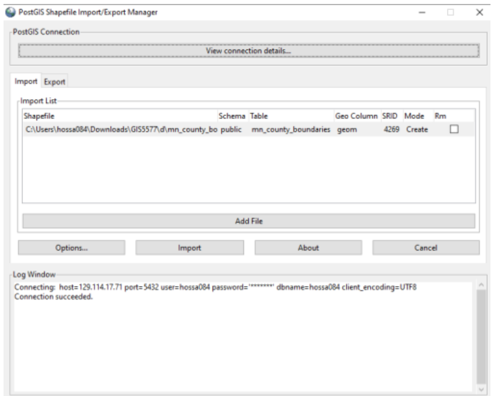
“County Boundaries” dataset represents the county boundaries, as recognized by the Minnesota Department of Transportation. There are 87 counties in Minnesota which this project focuses on the seven county Twin Cities and clips the area of interest from this dataset.

Another dataset is the “NDVI mask” of vegetation indices on Sentinel-2 imagery. This data provides vegetation raster polygon for the region which is generated from RGB and NIR spectral bands of a time-series image for 2-month temporal resolution (mostly growing season in July and August) in 2019 with high spatial resolution (10 m). Thus, utilizing vegetation indices and different Sentinel-2 time-series images, this dataset allows to track land use change.

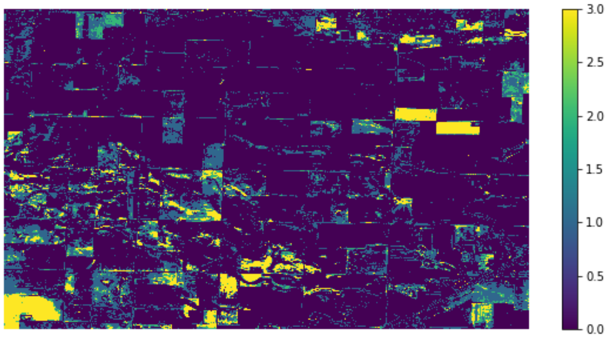
The last dataset is the “NDVI_statistics” which is created in the database to store the result of spatial analyses in the database. It includes statistics about NDVI, the total area of vegetation polygons and total area of land use change.

Methods

To insert the datasets to PostGIS database, I loaded the shapefile of Historical Generalized Land Use dataset through Command Prompt. I set the SRID field according to the metadata for the shapefiles. Following the same method, I loaded the NDVI mask dataset. To load County Boundaries, I tried a different method to load the file as a database file (.dbf) through PostGIS import manager. As I mentioned earlier, there is also NDVI_statistics dataset which is defined and created in the database and the fields were filled with the result of statistics analyses.



To extract NDVI mask from the satellite imagery, I rewrote a part of my python code for my graduate program thesis in Jupyter notebook. In this code, I used several raster libraries' (such as rasterio, geopandas, scikit-learn, and scikit-image) tools to export vegetation polygons as a shapefile. More documentation is available in Jupyter file in the class project repository.



Data analysis

The following SQL query statements are the ones dealing with raster data and shapefiles as well as joining to tabular data that could be performed on this database:

1) Write the select or create statement that results in a new table NDVI_statitics that contains the following fields

2) If you use the count of NDVI pixels as a proxy for area, which year has the most vegetation area per pixel?

3) Write the SQL query that shows the total area of vegetation land in northern half of Ramsey county based on spatial extent of the NDVI mask.

4) Write the SQL query that identifies all vegetation land use in 2010 that are within 5 kilometers of Ramsey county centroid.

5) Write the SQL code that creates a histogram of the vegetation land use in Ramsey county based on NDVI_mask. A histogram is each unique pixel type and total number of pixels.

6) Write the SQL code that reclassifies vegetation land use in Twin-cities so that Parks & Recreation Areas, Vacant/Agricultural, Farmsteads in years before 2000 and Park, Recreational, or Preserve,Golf Course, Agricultural in years after 2000 (LUSE_code= 7,8,10,70,100,173) would take a unique LUSE_code.

7) Which vegetation types (LUSE_code= 7,8,10,70,100,173) decreased from 2010 to 2016 in Ramsey County.

Challenges

One major challenge is about comparative analysis of the Historical Generalized Land Use dataset in geodatabase. It is important to understand the changes between land use inventory years from 1984 and 2016 and how to compare recent land use dataset from NDVI mask to historical data. In general, over the land use years, more detailed land use information has been captured and changes in data collection methodology effects the ability to interpret land use changes and trends in land consumption.

This challenge demonstrates itself in different image resolution of the satellite imagery (10 m) with the resolution of air photos in Historical Generalized Land Use dataset which differs based on the years. With the improved data in terms of resolution, individual properties were reviewed to assess the extent of development. According to the Metropolitan Council, in most cases, if properties under 5 acres were assessed to be at least 75% developed, then the entire property was classified as a developed land use (not 'Undeveloped'). As a result of these realignments and development assessments, changes in land use between early land use years (1984-1997) and more recent years (2000-2016) will exist in the data that do NOT necessarily represent actual land use change. These occurrences can be found throughout the region which skew the result of comparative analysis.

The other challenge is with a range of vegetation land use types that have vegetation coverage in the Historical Generalized Land Use dataset while the vegetation mask just extracts polygons with a specified NDVI threshold. The following table shows the vegetation attributes in Historical Generalized Land Use dataset metadata. I learned PostGIS database is flexible with this issue.

LUSE1984, LUSE1990 and LUSE1997: 1984, 1990 and 1997 land use codes (2 digit integer field type)	LUSE2000, LUSE2005 and LUSE2010: 2000, 2005, 2010 and 2016 land use codes (3 character text field type).
0 = No Data (for 1984 and 1990 only)	**100 = Agricultural**
1 = Single Family Residential	**111 = Farmstead**
2 = Multi-Family Residential	112 = Seasonal/Vacation
3 = Commercial	113 = Single Family Detached
4 = Industrial	114 = Single Family Attached
5 = Public Semi-Public	115 = Multifamily
6 = Airports	116 = Manufactured Housing Parks
7 = Parks & Recreation Areas	120 = Retail and Other Commercial
8 = Vacant/Agricultural	130 = Office
9 = Major Four Lane Highways	141 = Mixed Use Residential
10 = Open Water Bodies	142 = Mixed Use Industrial
11 = Farmsteads	143 = Mixed Use Commercial and Other
12 = Extractive (1997 only)	151 = Industrial and Utility
41 = Industrial Parks not Developed	153 = Extractive
51 = Public & Semi-Public Vacant	160 = Institutional
54 = Public Industrial (1997 only)	**170 = Park, Recreational, or Preserve**
	173 = Golf Course
	201 = Major Highway
	202 = Railway
	203 = Airport
	210 = Undeveloped
	220 = Water

Solution

For the comparative analysis challenge, I looked for detailed category definitions between 1984 and 2016, and I referred to the Attribute Accuracy and the Data Quality section of the metadata. I found the photo resolution was 0.6-meter resolution which is much more accurate than the NDVI mask of the satellite imagery with 10-meter resolution.

Considering second challenge and PostGIS capabilities, I used where clause to identify all vegetation land use types in one category in the Historical Generalized Land Use dataset. for instance, `WHERE LUSE_code IN (7,8,11).` Or I could also use `ST_Reclass` order in PostGIS to classify NDVI values of the mask dataset based on vegetation fields in historical dataset.

References

1. E. Karaman, U. Çinar et al. "A new algorithm for automatic road network extraction in multispectral satellite images", Proceedings of the 4th GEOBIA, May 7-9, 2012.
2. Weixing Wang, Nan Yang et al. "A review of road extraction from remote sensing images", Journal of Traffic
and Transportation Engineering (English Edition), Volume 3, Issue 3, June 2016
3. S. Movaghati, A. Moghaddamjoo and A. Tavakoli, "Road Extraction From Satellite Images Using Particle Filtering and Extended Kalman Filtering," in IEEE Transactions on Geoscience and Remote Sensing, vol. 48, no. 7, pp. 2807-2817, July 2010, doi: 10.1109/TGRS.2010.2041783.
4. Gruen Armin, Haihong Li "Road extraction from aerial and satellite images by dynamic programming." ISPRS Journal of Photogrammetry and Remote Sensing, 50 (4) (1995)
5. Sujatha, C., Selvathi, D. "Connected component-based technique for automatic extraction of road centerline in high resolution satellite images." J Image Video Proc. 2015, 8 (2015). https://doi.org/10.1186/s13640-015-0062-9
6. E. Christophe and J. Inglada, "Robust Road Extraction for High Resolution Satellite Images," 2007 IEEE International Conference on Image Processing, San Antonio, TX, 2007, pp. V - 437-V - 440, doi: 10.1109/ICIP.2007.4379859.
7. Uwe BACHER and Helmut MAYER "Automatic road extraction from IRS satellite images in agricultural and desert areas", Institute for Photogrammetry and Cartography, Bundeswehr University Munich, 2004
8. Fortier, M., D. Ziou and C. Armenakis. "Survey of Work on Road Extraction in Aerial and Satellite Images." (2002).Ali Hossaini GEOG 8292
Dec. 23, 2020 Project Report
9. François Waldner, Diego De Abelleyra, (2016) "Towards a set of agrosystem-specific cropland mapping methods to address the global cropland diversity", International Journal of Remote Sensing, 37:14, 3196-3231, DOI: 10.1080/01431161.2016.1194545
10. Xia, N., Wang, Y., Xu, H. et al. Demarcation of Prime Farmland Protection Areas around a Metropolis Based on High-Resolution Satellite Imagery. Sci Rep 6, 37634 (2016). https://doi.org/10.1038/srep37634
11. Masek et al., 2000 "Dynamics of urban growth in the Washington DC metropolitan area, 1973–1996, from Landsat observations" Int. J. Remote Sens., 21 (2000), pp. 3473-3486
12. Zha et al., 2001 "An effective approach to automatically extract urban land-use from TM imagery" J. Rem. Sens. Beijing, 7 (1) (2001), pp. 37-40
13. Zha et al., 2003 "Use of normalized difference built-up index in automatically mapping urban areas from TM imagery" Int. J. Remote Sens., 24 (3) (2003), pp. 583-594
14. Xu, 2008 "A new index for delineating built-up land features in satellite imagery" Int. J. Remote Sens., 29 (2008), pp. 4269-4276
17. Dwijendra Pandey, K.C. Tiwari, "Extraction of urban built-up surfaces and its subclasses using existing built up indices with separability analysis of spectrally mixed classes in AVIRIS-NG imagery", Advances in Space
Research, Volume 66, Issue 8, 2020
18. Pankaj Pratap Singh & R.D. Garg (2014) A two-stage framework for road extraction from high-resolution satellite images by using prominent features of impervious surfaces, International Journal of Remote Sensing, 35:24, 8074-8107, DOI: 10.1080/01431161.2014.978956
19. He et al., 2010 "Improving the normalized difference built-up index to map urban built-up areas using a semiautomatic segmentation approach" Remote Sens. Lett., 1 (4) (2010), pp. 213-221
20. Lebourgeois, V., Dupuy, S., Vintrou, É., Ameline, M., Butler, S., & Bégué, A. (2017). "A combined random forest and OBIA classification scheme for mapping smallholder agriculture at different nomenclature levels using multisource data (simulated Sentinel-2 time series, VHRS and DEM)". Remote Sensing, 9(3), 259.