A NOVEL 3D INTELLIGENT FUZZY ALGORITHM BASED ON MINKOWSKI-CLUSTERING

Shima Toori[1], Ali Esmaeily[2]
[1]M.Sc. in GIS Engineering, Graduate University of Advanced Technology, Kerman, Iran
Email: st69626@yahoo.com
[2]Assistant professor, Dept. of Remote Sensing Engineering, Graduate University of Advanced Technology, Kerman, Iran
Email: aliesmaeily@kgut.ac.ir

KEY WORDS: Fuzzy, 3D NDVI, Minkowski Clustering, Algorithm, Distance

ABSTRACT:

Assessing and monitoring the state of the earth surface is a key requirement for global change research. In this paper, we propose a new consensus fuzzy clustering algorithm that is based on the Minkowski distance. This research concentrates on Tehran's vegetation mass and its changes during 29 years using remote sensing technology. The main purpose of this research is to evaluate the changes in vegetation mass using a new process by combination of intelligent NDVI fuzzy clustering and Minkowski distance operation. The dataset includes the images of Landsat8 and Landsat TM, from 1989 to 2016. For each year three images of three continuous days were used to identify vegetation impact and recovery. The result was a 3D NDVI image, with one dimension for each day NDVI. The next step was the classification procedure which is a complicated process of categorizing pixels into a finite number of separate classes, based on their data values. If a pixel satisfies a certain set of standards, the pixel is allocated to the class that corresponds to those criteria. This method is less sensitive to noise and can integrate solutions from multiple samples of data or attributes for processing data in the processing industry. The result was a fuzzy one dimensional image. This image was also computed for the next 28 years. The classification was done in both specified urban and natural park areas of Tehran. Experiments showed that our method worked better in classifying image pixels in comparison with the standard classification methods.

1. Introduction

Using high resolution remote sensing datasets in order to detect vegetation change offers an access to monitoring vegetation change dynamics and to some degree plant diversity. Vegetation mapping also presents valuable facts for comprehending the natural and man-made environments through measuring vegetation cover from local to global scales at a certain time point or over a continuous period.

Based on recent researches, half of the world's population lives in cities, and in the recent past, the main raise of the urban population took place in underdeveloped and developing countries. Cities are the most obvious human environment, and now almost 50% of global populations live in cities (Willem et al., 2006). From an ecological point of view, urbanization has a large impact on the ecosystem as it always accompanied by land use changes, altering its combination and structure, and consequently influencing ecosystem processes and functioning (Alberti, 2005). Land use and land cover changes affects the original matter and energy cycles of the ecosystem (Wackernagel and Yount, 1998; Pielke et al., 1999; Imhoff et al., 2000). In the direction of urbanization, the transmutation of land to urban use mainly decreases photosynthesis of the ecosystem in regions with productive forests (Wear and Greis et al., 2001; Nizeyaimana et al., 2001).

The world is encountering urban extend and experiencing rapid population growth, which is transforming regional natural landscapes. For instance, large patches of fertile forests or cropland tend to be fragmentized due to quick urbanization (Imhoff et al., 1997; Folke et al., 1997). Loss of cropland in some areas has an important affect on climate conditions, food security, and the environment (Yu et al., 2006). Scientists have done many researches about the effect of urbanization on matter cycles, energy flow and ecosystem service (Schimel et al., 2000).

Nowadays remote sensing has become a useful alternative to collect spatial information of vegetation dispersion during a period of time, by means of studying definite spatial and temporal resolution. Regional land reclamation programmers are progressively combining remotely sensed imagery for monitoring current and historical vegetation dynamics. Many researchers used Spectral vegetation index data to monitor vegetation, government and minerals industry (Willem et al.,2006). AlBakri and Taylor (2003) used the time series data including NOAA AVHRR images in order to monitor vegetation conditions for 15 years and they find that NDVI is a proper index to monitor vegetation conditions.

Specifically, this research focuses on Tehran's vegetation mass and its changes within 29years with the use of Remote sensing technology.

2. Study Area

The study area is Tehran province, located in North of Iran, at 35 40 N Latitude and 51 26 E Longitude, which is the capital of Iran (Figure1).

Figure1: location of study area, Tehran

Tehran's population was about 800,000. It had increased to 3 million by 1966 and by 1986 migrants brought the population to 6million. Today, the metropolitan region has more than 10 million residents, more than the sum of the Iran's next five major metropolitan provinces combined. This explosive growth has public health and environmental consequences, including air pollution and water impurity and the loss of arable land. Figure2 and Figure3 illustrate Tehran transportation network and Tehran green areas respectively.

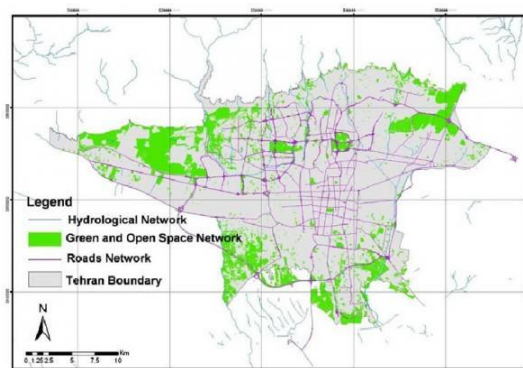Figure2: Tehran Transportation network

Figure3: Tehran green areas

The NASA's Landsat 5 satellite with Thematic Mapper sensor, acquired an images of Tehran on August 2, 1989 (figure4 top), and another one on July 19, 2009 (figure4 bottom). In both images, vegetation appears bright green, urban areas range in color from gray to black, and unproductive areas appear brown. Whereas non-urbanized areas border the earlier image, urbanization fills almost the entire frame of the later image. Major roadways crisscrossing the city in 1989 remain visible in 2009, but many additional roadways have been added, particularly in the north.

Figure4: False color images of Tehran acquired on August 2, 1989 (top), and on July 19, 2009 (bottom)

3. METHODS

The main purpose of this paper is to measure the changes in vegetation mass. By the use of different high resolution remote sensing Landsat8 and LandsatTM datasets from three continuous days, vegetation impact and recovery can be identified for years 1988 to 2016. For this purpose the first step is Classification. Compared to previous attempts, the suggested algorithm of this research differs in the similarity matrix construction. It does not use the resampling technique, but it uses the Minkowski distance to measure the input data. Each pixel belongs to the class with shortest distance. The new intelligent algorithm evaluates the accuracy using distribution. In low preciseness conditions, the intelligent algorithm changes the class center. Depending on the number of pixels in each class, if we do not reach the desired accuracy, the cycle is repeated 100 times. The final respond optimized by the algorithm will be considered.
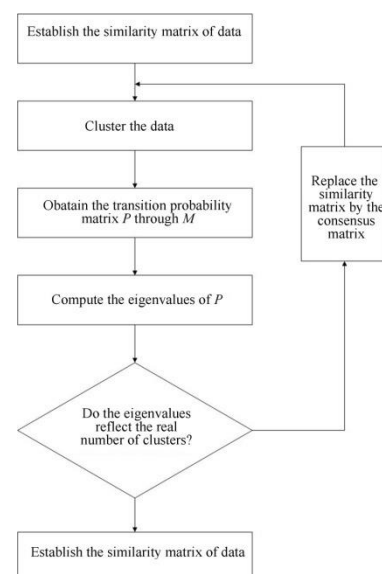

Figure5: Flow graph of iterative Minkowski clustering

The Landsat images received on 9 June 1988 to 11 June 1988 - three continuous days- with 30 meter resolution, processed and the NDVI factor computed using pixel values of bands 3 to 5. No significant deviation must be observed in NDVI values, because of the three day period. This was repeated for the years 1989 to 2016. The outcome of this step is a 3D NDVI image, with one dimension for each day NDVI. Classification is the next step. Classification is a complex process of sorting pixels into a finite number of individual categories of data, based on their data values. We used an intelligent Minkowski weighted distance algorithm in order to identify the four NDVI class centers. If a pixel satisfies a certain set of standards, the pixel belongs to the class that corresponds to those criteria. The Minkowski distance is the distance between the two measured NDVI dimensions (De-Gang X, 2017).

The similarity matrix does not adopt one specific Minkowski distance, but makes use of the different forms of Minkowski distances by adjusting the parameters in the equation. In this paper, we use the Minkowski distance to construct the similarity matrix:

$$M_P(x,y) = \left(\sum_{i=1}^{n} |x_i - y_i|^P \right)^{1/P} \tag{1}$$

$$S_{M_P}(x,y) = \mu \, exp\left(-\mu M_P(x,y)\right) \tag{2}$$

μ values can be selected from{0.1, 0.2, 0.5, 0.8, 0.9}. In case P=1, the first equation is the Manhattan distance, which reflects the sum of the absolute value of the difference between pixel i and pixel j. If P=2, formula (1) is the shortest distance between pixel i and pixel j, namely, the diagonal distance. If P $\rightarrow \infty$, then equation (1) is the Chebyshev distance, which reflects the maximum deviation between i and j in a special dimension. Additionally, P could have a value that is less than 1, and the categorizing method, by taking different values of P, will provide different outcomes. μ is an adjustable element. Because there are two variables, the distance equation can reflect similar information for different aspects by adapting the parameters. Minkowski distances can define the similarity of the special data in advance. We can use this equation to establish different similarity matrices and acquire the real similarity information picked from different clustering effects. This method is proper for the idea of consensus clustering, namely, selecting the appropriate matrix from the different results.

In order to obtain the clusters:

1. Different similarity matrices based on the Minkowski distance with different parameters (p and μ) must be established.

2. One of the above matrices must be categorized with the clustering algorithms and the clustering numbers (n numbers), and obtain the $3 \times n$ matrices.

3. The consensus similarity matrix M has been obtained by analyzing the $3 \times n$ matrices (which reflect the information of the clustering).

4. The transition probability matrix have been obtained.

5. The transition probability matrices have been obtained corresponding to the other similarity matrices by using the above methods.

6. The number of clusters must be obtained by analyzing the Eigen values of the different transition probability matrices.

4. RESULTS AND DISCUSSION

The main goal of this study is to evaluate the changes in vegetation mass. The research was carried out in Tehran. A number of varied high resolution remote sensing Landsat TM and Landsat8 datasets from three continuous days were used to identify vegetation impact and retrieval from 1989 to 2016. The Landsat images received on 9 June 1988 to 11 June 1988 - three consecutive days- with 30-meter resolution, processed and the NDVI factor calculated using pixel values of bands 3to5.
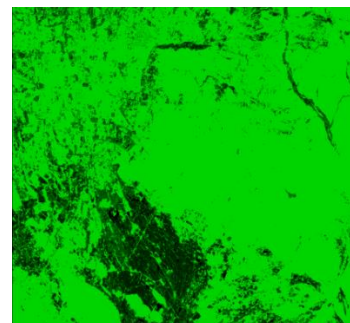
Because of the 3day period duration of the proposed algorithm, no significant changes must be observed in NDVI values. We used an intelligent Minkowski distance algorithm in order to identify the four NDVI class centers. If a pixel satisfies a certain set of criteria, the pixel is allocated to the corresponding class to those criteria.

Three calculated NDVI values must be clustered for each pixel of the 3D NDVI image and as they were for three consecutive days, there should not be any significant alters in their classes. The intelligent algorithm decides the possible NDVI class for each pixel of 3D NDVI image. The four class names were determined by the class center NDVI value. Every pixel with significant deviation between their three measured NDVI values is categorized as unclassified.
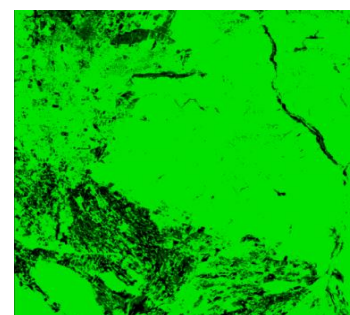
The result was a fuzzy one dimensional image. This image was also computed for the next 26 years, the classification *was done* in both specified urban and natural park areas of Tehran that showed detailed measurable vegetation, re-growth and reduction.

Intelligent algorithm automatically accomplished classification in four classes, which Class 1 covers large and healthy plants and Class 4 determines poor vegetation areas.
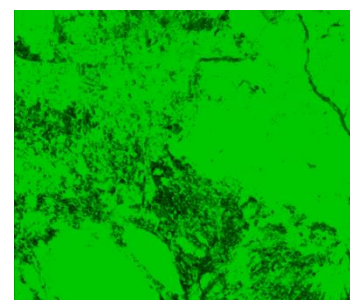
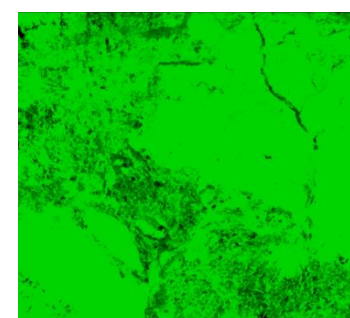|	
⬛	Unclassified
🟩	Class1
🟩	Class2
🟩	Class3
🟩	Class4

1988

2000

2006

2016

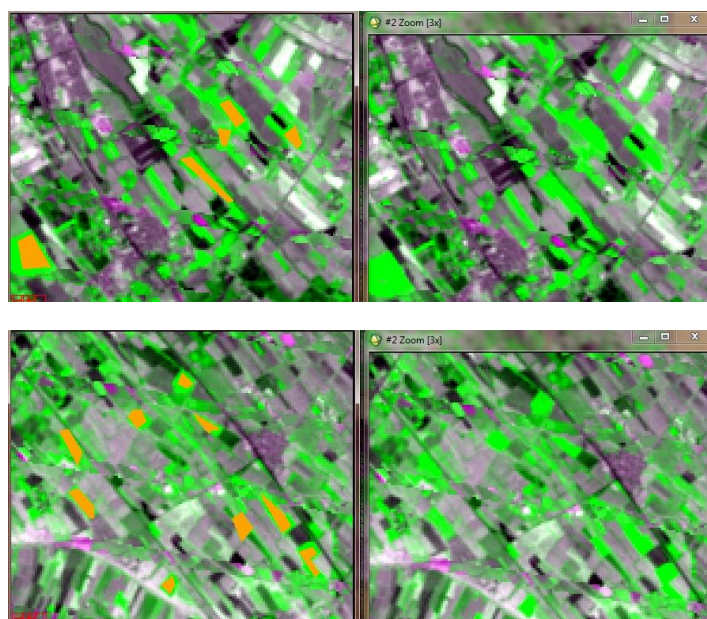Figure6: The final output of algorithms, fuzzy images

Figure7: fuzzy Minkowski classification test pixels

For evaluation of the accuracy of output, a number of test pixels were selected (Figure7). The pixels were categorized in two separate sets; one related to urban vegetation and the second related to the pixels of the natural vegetation, urban parks and plants. The table1&2 describe that the algorithm has performed very successful in detecting intervals.

Test	Total Pix					
1988-DAY1	948	841	76	12	5	14
%	-	88.71%	8.01%	1.2%	0.52%	1.47%
1988-DAY3	659	614	13	9	16	7
Urban	-	93.17%	1.97%	1.36%	2.42%	1.06%
2000- DAY3	1473	1341	94	4	13	21
%	-	91.03%	6.38%	0.27%	0.88%	1.42%
2006-DAY2	1649	1498	104	17	20	11
%	-	90.84%	6.3%	1.03%	1.21%	0.66%
2016-DAY1	993	941	24	6	15	7
%	-	94.76%	2.41%	0.6%	1.51%	0.7%
2016-DAY3	878	814	29	11	8	16
Urban	-	92.71%	3.3%	1.25%	0.91%	1.82%
Table1: Test pixels						

Area	1988	2000	2006	2016
	4.97%	5.64%	5.66%	1.53%
	7.12%	8.48%	6.89%	4.84%
	9.34%	8.51%	9.14%	8.48%
	8.39%	9.14%	7.67%	5.34%
	68.82%	68.23%	70.64%	79.81%
Table2:Test pixels result				

5. CONCLUSION

The aim of this study is to focus on Tehran's vegetation mass and its changes within 29 years using mathematics and remote sensing technology, a key that extends possible data archives from present time to over several decades back. According to the previous studies, the combination of a 3D fuzzy NDVI clustering and weighted Minkowski distance operators have never been used on images bands, in order to detect changed pixels. Our proposed algorithm, repeated for the same three days of years 1989 to 2016. The outcome of this step is a 3D NDVI image, with one dimension for each day NDVI. The next step was classification. The four class names were determined by the class center NDVI value. Also pixels with significant deviation between their three measured NDVI values were classified as unknown.

The outcome was a fuzzy one dimensional image, which was also computed for the remained 28years, the classification was done in both specified urban and natural park areas of Tehran that denoted detailed measurable vegetation, re-growth and reduction. The results shows that the method of this paper works better in classifying image pixels compared to some of the standard classification methods. The results of this study emphasize the need for more attention to changes and vegetation reduction in that areas. In order to improve the results obtained and continue working in this direction, it is suggested that future studies can be done by adding parameters such as the change of seasons.

REFERENCES

Al-Bakri JT and Taylor JC (2003) Application of NOAA AVHRR for monitoring vegetation conditions and biomass in Jordan. *Journal of Arid Environments* 54: 579–593

Alberti, M., 2005. The effects of urban patterns on ecosystem function. *International Regional Sciences*. Rev. 28, 168–192.

De-Gang X., Pan-Lei Z., Chun-Hua Y., Wei-Hua G., Minkowski-distance-based Consensus Clustering Algorithm, *International Journal of Automation and Computing*, 14(1), February 2017, 33-44

Folke, C., Jannson, A., Larsson, J., Costanza, R., 1997. Ecosystem appropriation by cities. *AMBIO 26*, 167–172.

Imhoff, M.L., Tucker, C.J., Lawrence, W.T., Stutzer, D.C., 2000. The use of multisource satellite and geospatial data to study the effect of urbanization on primary productivity in the United States. *IEEE Trans. Geoscience Remote sensing* 38 (6), 2549–2556.

Imhoff, M.L., Lawrence, W.T., Stutzer, D.C., Elvidge, C.D., 1997. Using nighttime DMSP/OLS images of city lights to estimate the impact of urban land use on soil resources in the US. *Remote Sensing Environment*. 59, 105–117.

Imhoff, M.L., Bounoua, L., DeFries, R., Lawrence, W.T., Stutzer, D., Tucker, C.J., Ricketts, T., 2004. The consequences of urban land transformation on net primary productivity in the United States. *Remote Sensing Environment*. 89, 434–443.

Nizeyaimana, E., Petersen, G.W., Imhoff, M.L., Sinclair, H., Waltman, S., Reed- Margetan, D.S., Levine, E.R., Russo, P., 2001. Assessing the impact of land conversion to urban use on soils with different productivity levels in the USA. *Soil Science Soc*. 65, 391–402.

Pielke, S.R.A., Walko, R.L., Steyaert, L.T., Vidale, P.L., Liston, G.E., Lyons, W.A., Chase, T.N., 1999. The influence of anthropogenic landscape changes on weather in south Florida. Mon. *Weather Rev*. 127, 1663–1673.

Schimel, D., Enting, I.G., Heimann, M., 1995. CO2 and the carbon cycle. In: Climate Change 1994, *Intergovernmental Panel on Climate Change*, Cambridge UniversityPress, Cambridge.

Schimel, D., Melillo, J., Tian, H.Q., 2000. Contribution of increasing CO2 and climate to carbon storage by ecosystems in the United States. *Science*, 287 (5460), 2004– 2006.

Wackernagel, W., Yount, D., 1998. The ecological footprint: an indicator of progress toward ecological sustainability. *Environment Monitoring Assessment*. 51, 511–529.

Wear, D. N., Greis, J. G., 2001. The Southern Forest Resource Assessment Draft Summary Report. http://www.srs.fs.fed.us/sustain.

Willem JD, Barron J, Stuart E and Stefanie M (2006) Multi-sensor NDVI data continuity: Uncertainties and implications for vegetation monitoring applications. *Remote Sensing of Environment* 100: 67–81

Yu, D.Y., Pan, Y.Z., Liu, X., Wang, Y.Y., Zhu, W.Q., 2006. Ecological capital measurement by remotely sensed data for Huahou and its socio-economic application. *Plant Ecology*. 30 (3), 404–413 (in Chinese, with English abstract)