Available Online at http://www.recentscientific.com

CODEN: IJRSFP (USA)

International Journal of Recent Scientific Research
Vol. 9, Issue, 4(L), pp. 26359-26364, April, 2018

ISSN: 0976-3031

International Journal of Recent Scientific Research

DOI: 10.24327/IJRSR

Research Article

HUMAN SETTLEMENT EXTRACTION AND POPULATION MAPPING USING MULTISPECTRAL REMOTE SENSING AND CENSUS DATA AT FARIDPUR DURGAPUR COMMUNITY DEVELOPMENT BLOCK

Suman Chatterjee*[1] and Kaniska Sarkar[2]

[1]Department of Geography, Jadavpur University, Kol-32
[2]Jadavpur University

DOI: http://dx.doi.org/10.24327/ijrsr.2018.0904.2044

ARTICLE INFO

Article History:
Received 19th January, 2018
Received in revised form 21st February, 2018
Accepted 05th March, 2018
Published online 28th April, 2018

Key Words:

Settlement Extraction, multi spectral remote sensing, Bayesian network Classifier & unsupervised technique, Census data Integration.

ABSTRACT

Population and settlement mapping is one of the most important tasks to do for the following purposes a) to assess the anthropogenic stress upon the environment, b) to assess the vulnerability of any community to environmental hazards and disasters, c) to estimate the degree and direction of urbanisation etc. Mixed, complex and confusing spectral characteristics and low spatial resolution are the limitation of Optical-Multispectral remote sensing with conventional classification techniques (Supervised and unsupervised classification) besides Hyperspectral, microwave/SAR (synthetic aperture radar) data provide promising result but often convey complicated analysis and cost-intensives. In this research an alternative methodology has been implemented where open access multispectral data has been used to extract the spatial distribution of settlement and afterward Census population data has been integrated with that spatial map to generate spatial population map. Bayesian network classifier integrated with unsupervised k means technique used to extract settlement later population data of census 2011 has been added according to area of the settlement cluster under a specific micro administrative unit (MAU). Current method has been compared with simple maximum likelihood classification (MLC) and produced better accuracy compared to MLC with overall classification accuracy of 97.83%.

INTRODUCTION

Obtaining information about human settlements using remotely sensed data has always been a challenge since the realm of Earth observation Began over 50 years prior (Pesaresi *et al.*, 2009). The most widely used conceptual paradigm in satellite based remotely sensed data analysis is Land use classification using the spectral information of the image. Yet mapping urban areas using these image classification algorithms brought unsatisfactory result due to spectral diversity of the settings of urban region and the fact that this diversity violates the fundamental assumption of spectral homogeneity upon which most classification algorithms are based (Small *et al.*, 2009). With advancement of remote sensing technologies it has been realised that traditional LULC paradigm which dominated the remote sensing methodological debate in the past 50 years, simplifies the data interpretation excessively and is not sufficient to produce greater amount of details to match the proficiency of these improved data (Pesaresi *et al.*, 2009). Mixed spectral characteristics, spectral and spatial complexity of the settlement built up leads to much misclassification all over the image and salt and peepers affect specially with settlement area. Unsupervised and even supervised classification often confuses the spectral signature of the settlement with other land-use and produces unsatisfactory result especially when the resolution is moderate to low. Post Classification Smoothening (Spatial Median Filter) to eliminate the salt and pepper affects leads to distortion of the features and drag far from the real world Scenario. Other than spectral classification and LULC, couple of other techniques emerged are Night time observations by Defence Meteorological Satellite Program (DMSP) a strong indicator of populated areas and population distribution; Synthetic aperture radar (SAR) measurements includes the three-dimensional characteristics of urban surfaces; thermal infrared data contain information about energy fluxes and local climatic conditions (Herold *et al.*, 2009). Built up area extraction using Normalised indices such as NDBI, IBI etc. are common in practice (Xu, 2007; Zha *et al.*, 2003). Indices like NDBI, BUI, or IBI can highlight the

Corresponding author:* **Suman Chatterjee
Department of Geography, Jadavpur University, Kol-32

built up features but cannot separate the settlement from the other built up areas and also contains noise from water body.

From the perspective of physical composition the built up system is heterogeneous which includes almost all materials in its unique structural manner; it includes both natural and anthropogenic elements and surface in a complex pattern. From the functional perspective it can be classified as Residential, commercial, Industrial, transportation etc. (Hu *et al*, 2016; Zhang and Lin, 2015). And often those elements are composed of same types of material. Beside that many different materials can be utilized for the same built-up element of the same settlement i.e., concrete, stone, clay tiles, plastic, corrugated metal, bitumen, grass, for building's roof and simultaneously the indistinguishable material for different elements i.e., the same stones for paved roads and building roofs (Pesaresi *et al*., 2009; Weng, 2014). Therefore often it is not feasible to differentiate natural and artificial if we only consider the spectral characteristics for example unpaved roads and bare soils, but also many roof materials and again bare soil or rocks. This situation is called mimetism between settlement and surrounding natural areas (Weng, 2014).

Image classification is the automated categorization of individual pixels or labelling them into certain classes (Campbell and Wynne, 2011). Satellite Image classification can be spectral based or pixel based, spatial based or object based (Lillesand *et al*, 2014). Spectral i.e. pixel or DN value based classification is most common for land use land cover classification (Aryaguna and Danoedoro, 2016; Lu and Weng, 2007). Maximum likelihood is the most widely used supervised classifier among all spectral based classification (Myint *et al*. 2011; Shengule, 2015; Sun *et al*, 2013) and also considered as most accurate (Shlien and Smith, 1975). Maximum likelihood is a well known parametric supervised probability based classification which assumes equal probabilities of classes and normal distribution of each spectral class. It evaluates both variance and covariance matrix of category spectral response (Otukei and Blaschke, 2010; Sun *et al*, 2013; Jaynes, 1974). Bayesian classification is also a supervised technique based on Bayes' theorem which accepts both continuous and discrete variables (Korb and Nicholson, 2010; Chepkochei, 2011) and considered as the extended approach of Maximum likelihood as Bayesian classification incorporates prior probability (Lillesand *et al*, 2014). Simple Bayesian analysis computes posterior probability with prior probability and likelihood function. Supervised learning is a vital part of machine learning and among them Bayesian network have been manifested to perform well (Korb and Nicholson, 2010). Although Bayesian Networks introduced (Pearl *et al*, 1988) in early 1988 it were not considered as classifiers until the discovery that Naive Bayesian Network, a type of probabilistic model effective for inference dealing with uncertainty (Yu *et al*. 2008; Solares and Sanz 2005). Since then researcher started to explore the possibilities of Bayesian network as classifiers. With different types of network learning algorithms Bayesian networks proven to be successful classifier e.g. the simplest among them is Naive Bayesian classifier and other modified networks such as tree-augmented naive Bayesian network (TAN), or the Bayesian network augmented naive Bayesian classifier (Campos *et al*. 2011). Numerous studies proved Bayesian Networks as powerful tools in remote sensing image classification and evaluated the efficiency of Bayesian network for image classification using multi-spectral (Park, 2004; Dlamini, 2010) and Hyperspectral images (Solares & Sanz, 2007). The basic difference between Bayesian estimation and maximum likelihood estimation is the inclusion of prior knowledge into account (Farrell and Ludwig, 2008). If the Bayes' estimation is based on uniform prior probability it is similar to Maximum likelihood. Having uniform probability equivalent to probability of constant function meaning there is no prior information (Jaynes, 1974).

Objectives: The Main Objectives of this work is a) to develop an alternative method for settlement extraction using freely available multispectral satellite data with satisfactory accuracy without or little misclassification using an approach combining Bayesian network classification and unsupervised classification, b) to compare with simple maximum likelihood classification and c) to integrate census population data with the settlement map to produce a spatial population map.

Study Area: All the MAUs (58) of Faridpur-Durgapur community development (CD) block under Burdwan District of West Bengal, India (Figure 1). The region has been chosen because the city is a developing city with moderate population. To assess the efficiency of the methodology the city has been chosen because of the mixed land-use pattern E.g. urban, rural, Residential, Commercial, industrial along with moving Infrastructures Roads, airport and other built up area, waste land, land spills can be seen which resulting complex spectral pattern all over the area.

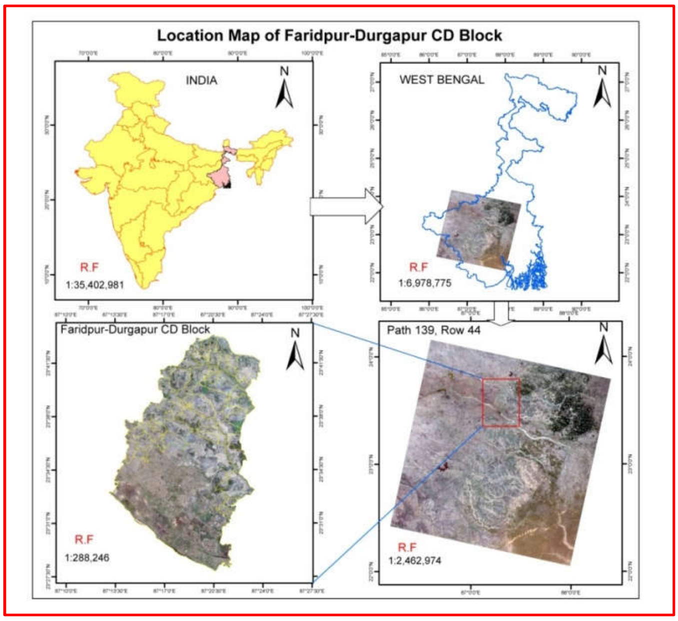

Figure 1 Location Map of the Faridpur-Durgapur CD Block

The city is a developing industrial hub inducing anthropogenic stress upon the environment therefore vulnerable to environmental hazard and disasters. The Block Covers an **Area** of 155.97 km^2, with Population (Census 2011): 115,924 and Population Density of 740/km^2. Co-ordinates: 23°40′04.8″N 87°18′34.2″E.

Data: Landsat OLI data (Path-139, Row-44) with 1.83 % cloud cover acquired on 04/03/2017 has been used for this purpose. Population Data: Primary Census Abstract from Census 2011.

Village/town level boundary have been obtained from Thana map and converted to vector format in ArcGIS. Village/town level total population data have been collected from primary census abstract published by the Office of the Registrar General & Census Commissioner, India.

METHODOLOGY

It is well understood if this part is described in two parts (a) Deriving settlement layer (first seven paragraphs) and (b) integrating population data to produce spatial population map (the last paragraph). Bayesian network classifier available in the Imagine Objective tool of Erdas Imagine has been used to construct the settlement probability layer later converted to binary layer using appropriate threshold. The Binary Image multiplied with individual bands and unsupervised k means classification has been performed finally recoded and excluded the non-settlement (Figure 2)**.** To integrate population data, settlement raster layer has been converted into polygon vector and assigned population proportional to the area of the polygon under any specific administrative units such as village, town or Municipal Corporation (Figure 4). Each step has been described in subsequently.

Satellite Image acquisition: LANDSAT OLI Image, WRS2 Path-139, Rows-44 acquired for the years 2017 acquired in winter season to avoid clouds.

Image Pre-processing: Individual bands (B2, B3, B4, B5, B6 and B7) were stacked and Atmospheric Correction has been performed Using ATCOR ad on tools of Erdas imagine. ATCOR converts raw image values (DNs) to surface reflectance (from solar reflective bands) by determining atmospheric parameters. It employs an atmospheric database which includes a wide range of pre-calculated radiative transfer runs for different weather conditions and sun angles. Corrected Image with spatial resolution of 30 meters has been merged with the panchromatic band (B8) with 15 meter resolution and pansharpned to increase the spatial resolution of the data to 15 meter. To enhance the image and to differentiate objects 3x3 edge enhancement filter ran.

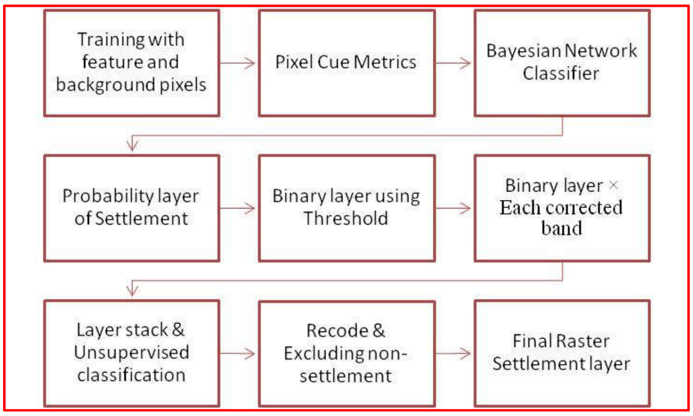

Figure 2 Deriving Settlement Map

Probabilistic classification and feature extraction using threshold: Only the first step of the "Imagine objective" process tree that is "Raster Pixel processor" has been used. Further steps i.e. Thresholding and extracting, Raster to vector conversion has been performed separately in Erdas imagine model maker and ArcGIS respectively. Threshold value has been carefully chosen after analysing the pixel probability layer scrupulously and comparing with the original Image and reference data such as Google earth. Raster pixel processor uses couple of pixel level cue to extract features for example DN value/colour/tone, texture and site/situation. Single feature probability (SFP) uses Bayesian Network classifier to computes the probability metric (a number between 0 and 1) for each pixel of the input image based on its pixel value and the given training samples. More a pixel is similar to the given feature training samples higher the probability is given. Low probability is given if a pixel value is similar to background training samples or it significantly differs from the feature training samples. The Bayesian network classifier has been trained by the pixels of the settlement and the background pixels. Both the feature pixels and the background have been chosen carefully (which should not overlap with each other) and were submitted to compute pixel cue metrics.

Conversion to Binary Image: The image has been transformed into a binary image using suitable threshold in the conditional statement resulting settlement with value "1" and others as "0".

Multiplication of Binary Image with each corrected Bands of the Satellite Image: The Binary Image then multiplied with the each band of the atmospherically corrected image which resulting each band having only the settlement covered area. Either

$\{Binary\ Image(1) \times Band\ n\} = Band\ n\ only\ Settlement$ Or $\{Binary\ Image\ (0) \times Band\ n\} = 0$ **Equation 1**

Unsupervised Classification: Afterward all the bands have been stacked into a composite image on which a K means unsupervised classification with 4 classes has been performed. Among 4 classes, two classes found to be as settlement class i.e. Settlement 1 and Settlement 2. Finally the output classified image has been re-coded to merge the settlement classes and to exclude the other built up areas.

Comparing Maximum likelihood supervised classification with present method: A conventional supervised classification using maximum likelihood has been also performed to compare with the current methodology. Initially classified the image with 9 classes 1) Dense vegetation 2) Waste land, 3) Agricultural Land, 4) Sand banks, 5) sparse vegetation, 6) Industrial & other built up, 7) Water body, 8) Barren land and 9) Settlement, latter all the classes have been recoded and merged to get two classes a) Settlement and b) Unclassified (Figure 3).

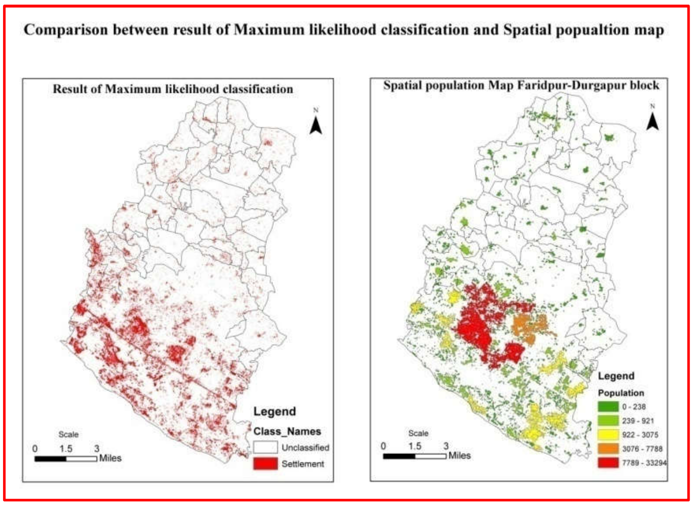

Figure 3 comparison MLC & Current Technique

Integration of Census Population Data with the Spatial Settlement map: Remotely sensed data can only provide the spatial distribution and extent of the human settlement. But to produce a spatial population map we must integrate the actual census data with the spatial settlement map (Figure 5). For that Village/Block level Census Data has been collected from the Primary census abstract 2011. The population data has been assigned to the map according to the area of any settlement cluster/polygon using a simple arithmetic equation (Equation 2).

Assigning population Proportional to the area of the settlement clusters

$$\text{Population of a settlement polygon} = \frac{\textit{Total population of the MAU} \times \textit{Area of the Polygon}}{\textit{Total area of the Settlement of that block}} \quad \textbf{Equation 2}$$

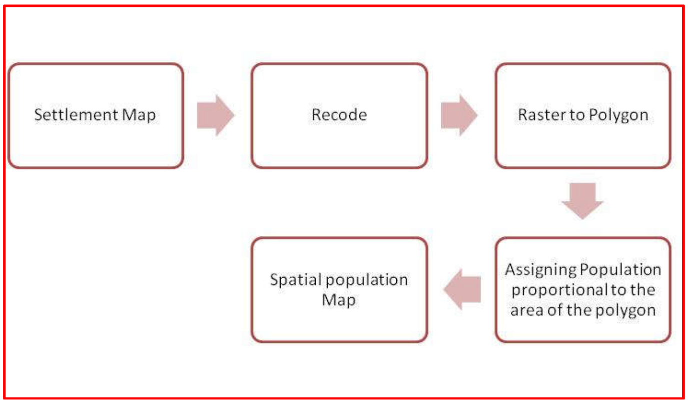

Figure 4 Integration of Population data with settlement Map

RESULT AND DISCUSSION

Probabilistic classification approach for feature extraction using Bayesian network classifier combining with unsupervised classification have produced much more accurate result with less misclassification compared to simple maximum likelihood classification (MLC) technique (Figure 3). Although there are some misclassifications especially within rural settlements, and built up area other than settlement yet current method has been also able to reduce the salt and pepper effect to a sufficient extent. Integration of population data proved to be more convenient than the thematic population map (Figure 6) as we can get the spatial distribution and the number of population simultaneously in the same map. Threshold should be selected in such a way that non-settlement pixels can be included but no settlement pixel can be excluded as we can exclude the non-settlement during unsupervised classification in the next step. Total area under settlement cover in Faridpur Durgapur CD Block found to be 47.5 square km. Majority of the population found urban population and has been agglomerated at Durgapur Municipal Corporation along the Damodar River and mainly around 3 types of Functioning areas i.e. a) Industrial complex b) commercial complex and c) National highway 2 and Railway. Scattered village population can be found in the Eastern and Northern part of the Block (Figure 5).

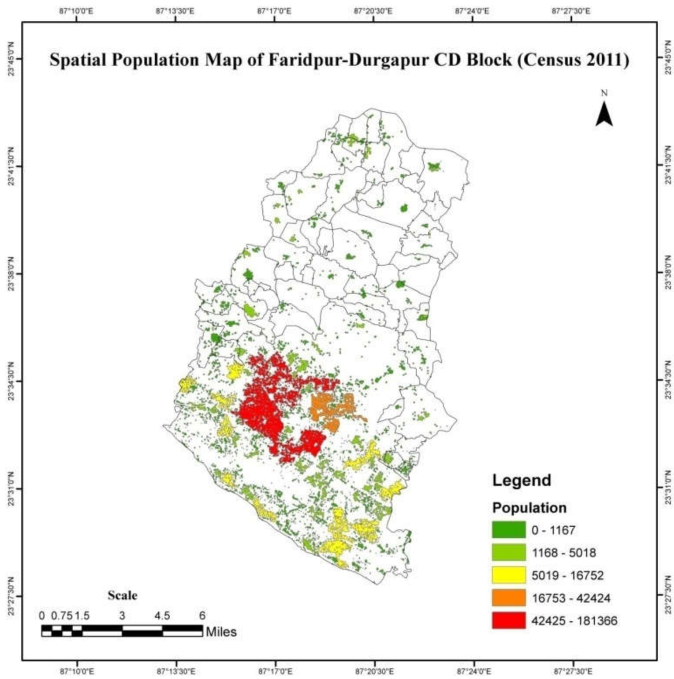

Figure 5 Spatial population map of Faridpur-Durgapur CD block based of 2011 census

Accuracy Assessment: 100 GCPs have been selected and assessed with Google Earth Map to check whether the GCPs fall under settlement class or not. Result of current method has came very accurate and satisfactory with Producer Accuracy 83.33%, User's accuracy 100.00 %, Overall Classification Accuracy = 97.83% and Overall Kappa Statistics = **0.8969** on the other hand result of conventional maximum likelihood classification came out with lesser accuracy and misclassification all over the study area with Producer Accuracy 66.67 %, User's accuracy 66.67 %, Overall Classification Accuracy = 80 % and Overall Kappa Statistics = 0.7290.

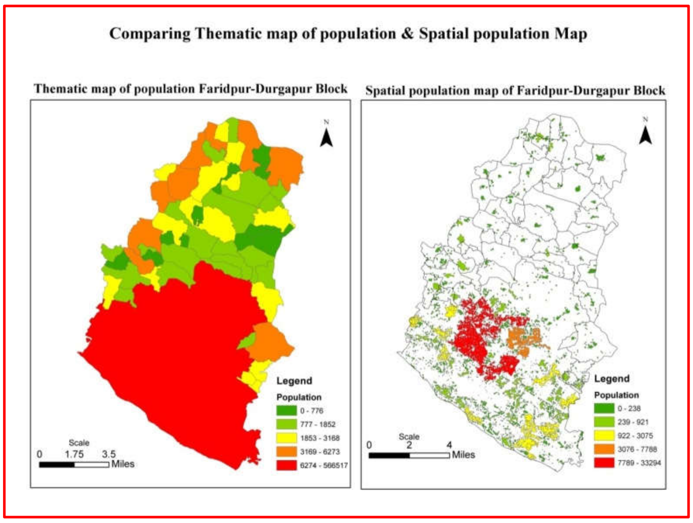

Figure 6 Comparing conventional thematic amp with Spatial Population Map

CONCLUSION

This study was not conducted for the comparison of performance of Maximum Likelihood and the Bayesian network classifier, but has attempted to develop an alternative methodology that produce lesser misclassification and salt pepper effect than the traditional classification system. Probabilistic classification and feature extraction using threshold combining with unsupervised approach reduces the miss classification and salt and pepper affect over the image even with moderate spatial resolution. It is extremely difficult to discriminate between settlement and other built up and a single technique might not be useful for this purpose. Hence integration of several techniques is required to handle such problems. In that case thermal bands can be integrated with conventional technique in future as thermal bands are proved to be useful for built up area study. Micro level (village/ town) study has increased the accuracy of Integrating of population data with the spatial map. Combining unsupervised classification helps to exclude other land use to increases the accuracy.

Acknowledgement

This work is a small subset of the Ph.D. work supported by University Grants Commission at the Department of Geography, Jadavpur University.

Reference

Aryaguna, P.A. and Danoedoro, P., 2016, November. Comparison Effectiveness of Pixel Based Classification and Object Based Classification Using High Resolution Image In Floristic Composition Mapping (Study Case: Gunung Tidar Magelang City). In *IOP Conference Series: Earth and Environmental Science* (Vol. 47, No. 1, p. 012042). IOP Publishing.

Campbell, J.B. and Wynne, R.H., 2011. *Introduction to remote sensing*. Guilford Press.

Chepkochei, L.C., 2011, November. Object-oriented image classification of individual trees using Erdas Imagine objective: case study of Wanjohi area, Lake Naivasha Basin, Kenya. In *Proceedings of the Kenya Geothermal Conference, Nairobi, Kenya* (Vol. 2123).

De Campos, L.M., Cano, A., Castellano, J.G. and Moral, S., 2011, November. Bayesian networks classifiers for gene-expression data. In *Intelligent Systems Design and Applications (ISDA), 2011 11th International Conference on*(pp. 1200-1206). IEEE.

Farrell, S. and Ludwig, C.J., 2008. Bayesian and maximum likelihood estimation of hierarchical response time models. *Psychonomic bulletin & review*, *15*(6), pp.1209-1217

Herold, M., 2009. Some recommendations for global efforts in urban monitoring and assessments from remote sensing. *Global Mapping of Human Settlement*, pp.11-23.

Hu, T., Yang, J., Li, X. and Gong, P., 2016. Mapping urban land use by using landsat images and open social data. *Remote Sensing*, *8*(2), p.151.

Jaynes, E.T., 1974. Probability theory with applications in science and engineering. *Washington University*

Lillesand, T., Kiefer, R.W. and Chipman, J., 2014. *Remote sensing and image interpretation*. John Wiley & Sons.

Lu, D. and Weng, Q., 2007. A survey of image classification methods and techniques for improving classification performance. *International journal of Remote sensing*, *28*(5), pp.823-870.

Myint, S.W., Gober, P., Brazel, A., Grossman-Clarke, S. and Weng, Q., 2011. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote sensing of environment*, *115*(5), pp.1145-1161.

Otukei, J.R. and Blaschke, T., 2010. Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *International Journal of Applied Earth Observation and Geoinformation*, *12*, pp.S27-S31.

Pesaresi, M., Ehrlich, D., Gamba, P. and Herold, M., 2009. A methodology to quantify built-up structures from optical VHR imagery. *Global mapping of human settlement: experience, datasets and prospects, Taylor and Francis, New York*, pp.27-59.

Shengule, V.V., 2015. STUDY AND DEVELOPMENT OF RESTORATION TECHNIQUE FOR REMOTE SENSING IMAGES.

Shlien, S. and Smith, A., 1975. A rapid method to generate spectral theme classification of Landsat imagery. *Remote Sensing of Environment*, *4*, pp.67-77.

Small, C., 2009. The color of cities: an overview of urban spectral diversity. *Global mapping of human settlements*, pp.59-106.

Sun, J., Yang, J., Zhang, C., Yun, W. and Qu, J., 2013. Automatic remotely sensed image classification in a grid environment based on the maximum likelihood method. *Mathematical and Computer Modelling*, *58*(3-4), pp.573-581.

Weng, Q., 2014. *Global urban monitoring and assessment through earth observation*. Crc Press.

Xu, H., 2007. Extraction of urban built-up land features from Landsat imagery using a thematicoriented index combination technique. *Photogrammetric Engineering & Remote Sensing*, *73*(12), pp.1381-1391.

Zha, Y., Gao, J. and Ni, S., 2003. Use of normalized difference built-up index in automatically mapping urban areas from

TM imagery. *International Journal of Remote Sensing*, *24*(3), pp.583-594.

Zhang, H. and Lin, H., 2015, March. Feature selection for urban impervious surfaces estimation using optical and SAR images. In *Urban Remote Sensing Event (JURSE), 2015 Joint* (pp. 1-4). IEEE.

How to cite this article:

Suman Chatterjee and Kaniska Sarkar.2018, Human Settlement Extraction And Population Mapping Using Multispectral Remote Sensing And Census Data At Faridpur Durgapur Community Development Block. *Int J Recent Sci Res.* 9(4), pp. 26359-26364. DOI: http://dx.doi.org/10.24327/ijrsr.2018.0904.2044

****