Applied Geography 48 (2014) 31–41

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog

Random forest classification of urban landscape using Landsat archive and ancillary data: Combining seasonal maps with decision level fusion

Aniruddha Ghosh*, Richa Sharma, P.K. Joshi

Department of Natural Resources, TERI University, New Delhi, India

Keywords:
Decision level fusion
DEM
Landsat
Multi-season
Random forest
Urban

ABSTRACT

Mapping landscapes in a rapidly urbanizing region can contribute significantly to quantifying, monitoring and understanding the complex process of urbanization. However, mapping such urban areas is a challenging task due to issues of spatial heterogeneity and dynamic land use practices. In this study we propose an operational mapping algorithm using multi-season Landsat and ancillary data with minimum image pre-processing and limited training samples. The methodology was applied to produce a detailed land use land cover (LULC) map of National Capital Region of India. Seasonal maps (with nine LULC classes) were produced by using Random forest (RF). A second classification involving seasonal maps with decision level fusion based on expert knowledge resulted in an annual composite map with increased number (eleven) of LULC classes. These detailed maps have moderately high (>60%) overall accuracies. The maps generated over different seasons are especially significant in identifying areas with mixed land use practices (like agriculture) occurring over an annual cycle. The annual map as the end product of the decision fusion summarizes the LULC dynamics of the study area with the help of eleven LULC classes. The significance of this work lies not only in generating accurately classified LULC maps, but also in detecting the seasonal dynamics of land use practices in a complex urbanizing landscape. Furthermore, reproducibility of the developed methodology will aid the extension of research for different time periods and with newer sensors in investigating the patterns and dynamics of land use and urban planning activities.

Introduction

Urbanization is a major form of anthropogenic land-use activity influencing the climate (Kalnay & Cai, 2003). In the last hundred years the urban population as per cent to total global population has increased from 10% to more than 50% (Grimm et al., 2008). The current rate of spatial expansion of urban areas is twice as fast as their population (Angel, Parent, Civco, Blei, & Potere, 2011). Urban expansion – whether planned or unplanned – threatens biodiversity (McKinney, 2002), affects quality of life, causes habitat loss (McDonald, Kareiva, & Forman, 2008) and results in loss of above-ground carbon storage (Imhoff et al., 2004). In addition, many of the urban agglomerations are exposed to natural hazards such as flood, drought or earthquake (Cutter, 1996; Sherbinin, Schiller, & Pulsipher, 2007). Existence of these phenomena strongly validates the argument for sustainable cities. One of the foremost requirements for planning, design, management and development activities leading to sustainability are accurate and reliable information of the current spatial distribution of urban components. In addition, this information is also essential for estimating green-house emission and predicting the likely growth of urban areas.

Remotely sensed data from various earth observation satellites can provide accurate and timely geospatial information of urban and peri-urban areas at diverse spectral, spatial and temporal scales (Taubenböck et al., 2012). These dataset are increasingly becoming an attractive alternative to ground based survey and mapping methods due to the advantages of cost and time saving for larger areas. However, use of remote sensing data to map an urban area is often challenging given the spatial complexity and dynamics of the area under study (Schneider, 2012). Often the characteristic scale of occurrence and shape of many urban features lead to the problem

---

* Corresponding author. Tel.: +91 11 26122222; fax: +91 11 26122874.
E-mail addresses: aniru123@gmail.com (A. Ghosh), richa.sharma85@gmail.com (R. Sharma), pkjoshi27@hotmail.com (P.K. Joshi).

0143-6228/$ – see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.apgeog.2014.01.003

of 'mixed pixel' (Small, 2003) which can be addressed by the use of very high resolution (VHR) dataset with spatial resolution less than 5 m from sensors such as IKONOS, QuickBird or World View 2 (Myint, Gober, Brazel, Grossman-Clarke, & Weng, 2011; Small, 2003; Welch, 1982). But, along with low swath, the VHR dataset also suffers from the constraint of limited scene availability (spatially and temporally) which restricts its wide use. Medium-resolution Landsat has a regularly sampled historical archive of 40 years and provides an optimum solution to minimize the trade-off between coverage, temporal and spatial resolution thereby satisfying the requirement for local to global scale studies (Goward, Masek, Williams, Irons, & Thompson, 2001).

Various agencies concerned with the issues of urbanization require accurate and timely geospatial information about the patterns and dynamics of urban components. Given the rapid growth of urban areas in developing countries like China and India, annual monitoring is more essential than decadal monitoring. In areas with high inter- and intra-annual variability of land use practices, it is nearly impossible to use medium-resolution data sources to separate LULC classes at a single point in time (Heller et al., 2012). Thus generating a single annual map is difficult owing to the fact that many of the LULC features are under the process of rapid change such as transitional urban (exposed land which has been cleared for construction purposes) or exhibit different phenological behavior in forms of different stages of crop (seedling, bloom, and harvesting stage), forest types, and leaf-on and leaf-off conditions (O'Hara, King, Cartwright, & King, 2003). In this context, few studies have attempted to use multi-seasonal remote sensing data for monitoring LULC classes in urban areas (Schneider, 2012). Multi-seasonal information has proven beneficial for reducing class confusion for vegetation by exploiting the phenological behavior of different vegetation types in complex landscapes (Langley, Cheshire, & Humes, 2001; Joshi, Roy, Singh, Agrawal, & Yadav, 2006; Rodriguez-Galiano, Chica-Olmo, Abarca-Hernandez, Atkinson, & Jeganathan, 2012). Yuan, Sawaya, Loeffelholz, and Bauer (2005) used spring–summer Landsat imagery to map Twin Cities Metropolitan Area of Minnesota. Punia, Joshi, and Porwal (2011) explored the potential of seasonal (Monsoon–Winter–Summer) IRS-P6 AWIFS data for mapping LULC of Delhi. These studies reported significant increase in overall accuracy regardless of the difference in the study areas.

Majority of previous studies stacked the seasonal data to create a single raster following a radiometric normalization and classified this raster using exhaustive training sets involving all the seasonally/temporally varying class information (for example cropping patterns, transitional lands) (Punia et al., 2011; Schneider, 2012). The difficulty with such mapping process is that it is largely dependent on the classifiers' ability to handle the vast and diverse dataset. At the same time, generation of training database consisting of all LULC classes present in study area throughout the year is a challenging task and requires multiple rigorous field visits and special emphasis on seasonally/temporally varying classes. In the light of these considerations, this study proposes a simple operational methodology to produce annual LULC map based on decision level fusion of the classified multi-seasonal data, which requires relatively lower *a priori* spatial information of the temporally varying classes. We tested this methodology for three different seasons corresponding to the crop cultivation calendar, i.e., *kharif* or monsoon (June–November), *rabi* or winter (October–March) and *zaid* or summer (March–July) for three different annual crop cultivation cycles (1998–99, 2002–03, 2010–11). DEM derived three ancillary input features were also used to improve the classification accuracy. Furthermore we assessed the importance of the input variables and analyzed their seasonal variability across three different years.

Methodology

Study area and class description

Delhi, the capital of India, is the second most populated megacity in the world with nearly 17 million people with an area of 1483 km^2. Population in Delhi has risen from 1.7 million in 1951 to 13 million in 2001, finally crossing the mark of 16.7 million in 2011 (Census of India, 2011), as one of the fastest growing urban areas in history (United Nations, 2012). This unprecedented escalation has attracted a wide community of researchers to map the complex LULC of Delhi (Cole, Wentz, & Christensen, 2005; Mookherjee & Hoerauf, 2004; Punia et al., 2011; Rahman, Kumar, Fazal, & Bhaskaran, 2011; Sokhi, Sharma, & Uttarwar, 1989; Wentz, Nelson, Rahman, Stefanov, & Roy, 2008). In 1991, the urban core of Delhi was expanded to nearby interstate city areas such as Ghaziabad and Noida of Uttar Pradesh and Faridabad and Gurgaon of Haryana forming the National Capital Region (NCR). Following this elevated status, this region has experienced rapid spatial changes in its LULC patterns. These changes – whether conversion of agricultural land to industry and civic facilities, or sparse built-up areas to dense built-up areas – have impacted the residential patterns, agricultural practices and urban forest distribution in the city. Delhi is characterized by highly heterogenous land use practices with impervious surfaces dominating the commercial, industrial as well as residential areas; and patches of forests at the heart of the city, in the north and the south. Peri-urban areas and banks of river Yamuna are predominantly covered with agricultural lands. Such high heterogeneity of land use provides high dynamism of land transformations over both long-term as well as short-term scales. Over short-term, seasonal changes in crop and plant phenology play an important role to discriminate between the vegetation classes. The city experiences five major seasons; Winter (Nov–Feb), Spring (late Feb–Mar), Summer (Apr–Jun), Monsoon (Jul–Sep), and Post-monsoon (Oct–early Nov). Based on these seasons, four cropping practices are followed; *kharif* (Monsoon crop), *rabi* (Winter crop), zaid (Summer crop) and *double crop* (two or more than two cropping is done) (Punia et al., 2011) (Fig. 1).

For selecting the representative LULC of the study area, we did not follow any particular classification scheme. We conducted four extensive field visits to identify the dominant LULC classes during 2010–11. Based on the field visits and expert knowledge, we identified 9 major LULC classes for each season (Table 1). For the other two years (1998–99 and 2002–03), we followed this designed classification schemes. The annual LULC map for each time period has eleven classes (Table 1). The details of generation of the annual map are presented in knowledge based classification of seasonal maps to generate the annual map section.

Collection of training samples

For each season in each period, an inventory of unique training sites was generated. For 2010–11, training sites were collected through field visits using handheld GPS unit. For 2002–03, training sites were selected following on screen digitization through visual interpretation of seasonal Landsat data and Google Earth® imagery. For 1998–99, due to the unavailability of Google Earth® imagery, only seasonal Landsat imagery and expert knowledge derived from other periods were used to select training sites. During the collection of training samples of agriculture land use, only cropped or non-cropped lands were identified for seasonal data classification, which mean no *a priori* knowledge for, *kharif* (monsoon), *rabi* (winter), *zaid* (summer) and double crop practices were present for classifying the seasonal data. To minimize any kind of errors related to mixed pixels, homogenous polygons were selected as training