A Hierarchical Approach of Hybrid Image Classification for Land use and Land cover Mapping


Received: August 22, 2017 | Revised: October 6, 2017 | Accepted: November 27, 2017
DOI: 10.5937/22-16620

Abstract

Remote sensing data analysis can provide thematic maps describing land-use and land-cover (LULC) in a short period. Using proper image classification method in an area, is important to overcome the possible limitations of satellite imageries for producing land-use and land-cover maps. In the present study, a hierarchical hybrid image classification method was used to produce LULC maps using Landsat Thematic mapper TM for the year of 1998 and operational land imager OLI for the year of 2016. Images were classified using the proposed hybrid image classification method, vegetation cover crown percentage map from normalized difference vegetation index, Fisher supervised classification and object-based image classification methods. Accuracy assessment results showed that the hybrid classification method produced maps with total accuracy up to 84 percent with kappa statistic value 0.81. Results of this study showed that the proposed classification method worked better with OLI sensor than with TM. Although OLI has a higher radiometric resolution than TM, the produced LULC map using TM is almost accurate like OLI, which is because of LULC definitions and image classification methods used.

Keywords: Error matrix, Hierarchical classification, Iran, land use/cover, Plasjan sub-basin

Introduction

Satellite data are often used to prepare land-use and land-cover maps. (Chrysoulakis et al., 2010; Lakshmi et al., 2015). Selection of proper land-use classification method is crucial in many inventories especially in watersheds’ uplands, which are usually water sources for wetlands (Anderson, 1976; Purkis et al., 2006; Mie et al., 2015; Tian et al., 2015). When satellite images data are used to produce LULC map, it is often very difficult to identify spectrally unique land-use/cover classes because of similar spectral responses arising from different features (Roy et al., 2014; Knudby et al., 2014; Estoque & Murayama, 2015; Lakshmi et al., 2015). Several methods can be employed to produce LULC by employing remote sensing data (Purkis & Klemas, 2011; Lakshmi et al., 2015; Al-doski et al., 2013). However, it should be noted that in case land surface objects have a similar reflectance or a small area, most of them could not provide high accurate maps (Gao & Xu, 2016). Using low radiometric resolution imageries, land classification can be a serious challenge because of spectral mixing of different surface elements and landscape complexity (Julien et al., 2011; Stenzel et al., 2016).

* Natural Resource Faculty, University of Zabol, Iran, Vahid_rahdary@yahoo.com; soffianian@cc.iut.ac.ir; Spmanafi@yahoo.com
B Griffith Centre for Coastal Management, Griffith University, QLD, Australia, R.mosadeghi@griffith.edu.au
C Soil & Geomorphology in Iranian Soil and Water Research Institute, Isfahan, Iran, hghaiumi@gmail.com
* Corresponding author: Alireza Soffianian, e-mail: soffianian@cc.iut.ac.ir
In such cases, application of hybrid classification approach in hierarchical way will produce better land-use/land-cover maps (Lakshmi et al., 2015). In this method, land-use/land-cover maps are produced by combining different methods like unsupervised, supervised, object-based methods and different indices produced from satellite images (Anderson, 1976; Homer et al., 2004; Di Gregorio, 2005; Disperati et al., 2015; Misra & Balaji, 2015; Lakshmi et al., 2015).

Several studies compared the accuracies obtained by different image classification methods and applied imageries with different radiometric resolutions for producing accurate LULC maps (Benfield et al., 2007; Alves et al., 2012; Vieira et al., 2012; Al-doski et al., 2013; Knudby et al., 2014; Estoque & Murayama, 2015; Mei et al., 2015; Disperati et al., 2015; Poursanidis et al., 2015; Mirsa & Balaji, 2015; Knudby et al., 2015).

The objective of this study is to develop a hybrid classification method to prepare accurate land-use/cover maps even when imageries with lower radiometric resolutions are used.

Methods and data

The study area
The study area was Pelasjan sub-basin including the western part of the Gavkhooni watershed located in central Iran and covering approximately 412,999 hectares. The Zayandehrood is the major river in Gavkhooni watershed to which Pelasjan sub-basin gives the highest portion of water. The Gavkhooni wetland is located in the eastern part of Gavkhooni watershed and is the terminal basin of the Zayandehrood River. Pelasjan sub-basin average temperature is 8-13 C° with 400-1250 mm precipitation. Agriculture activities and animal husbandry are the main activities of people living in these areas. Figure 1 shows the location of the Zayanderood River Basin and Pelasjan sub-basin in the western part of the Gavkhooni wetland in Iran.

Dataset
The Operational Land Imager (OLI) and Thematic Mapper (TM) sensors were launched with Landsat satellites and are useful in natural resources studies. OLI sensor measures in the visible, near infrared, and short wave infrared portions of the spectrum in 11 bands. TM sensors with 7 bands are in visible wavelengths and in infrared. The temporal resolution of both TM and OLI are 16 days. Considering the fact that the study area was located between two Landsat paths, 164 and 165, two images were downloaded from the USGS website. Because the highest vegetation cover was in June and August, satellite images were downloaded for susceptible months. Table 1 shows satellite data selected for this study.
A Hierarchical Approach of Hybrid Image Classification for Land use and Land cover Mapping

In addition, aerial images, digital elevation model 1:25,000 topographic maps were used to best understand the study area’s situation.

Field studies
Field studies were conducted to collect training areas for each LULC class to be used in the image classification. Positions of the lands under agricultural areas were determined with GPS. To check the status of the vegetation cover crown percentage (VCCP), 270 plots with 7.3 meters were measured.

In this study, samples from each LULC were collected by paying attention to imageries spatial resolution (30m); and they were collected in a homogenous area of LULC. In this order, we take samples in homogenous areas, which at least are more than 30-meters distant from margins. Therefore, by avoiding marginal land-use/land-cover reflectance, we achieve almost pure reflectance samples for each LULC.

Because there were not enough data for 1998 image, by using topographic maps and aerials and by comparing TM and OLI image false color composites (FCC), NDVI images values in TM and OLI and field studies, VCCP in each recorded plot was predicted.

LULC classification
Based on the available data and field studies, 7 LULC classes were defined for the study area (Table 2).

Table 1. Satellite sensor specification

<table>
<thead>
<tr>
<th>Date</th>
<th>Satellite</th>
<th>Sensor</th>
<th>Sensor ID</th>
<th>Pixel size</th>
</tr>
</thead>
<tbody>
<tr>
<td>13-AUG-1998</td>
<td>Landsat</td>
<td>TM</td>
<td>Path/row 165/37</td>
<td>30</td>
</tr>
<tr>
<td>06-AUG-1998</td>
<td>Landsat</td>
<td>TM</td>
<td>Path/row 164/37</td>
<td>30</td>
</tr>
<tr>
<td>10-JUN-1998</td>
<td>Landsat</td>
<td>TM</td>
<td>Path/row 165/37</td>
<td>30</td>
</tr>
<tr>
<td>03-JUN-1998</td>
<td>Landsat</td>
<td>TM</td>
<td>Path/row 164/37</td>
<td>30</td>
</tr>
<tr>
<td>03-Jan-1998</td>
<td>Landsat</td>
<td>TM</td>
<td>Path/row 164/37</td>
<td>30</td>
</tr>
<tr>
<td>09-JUN-15</td>
<td>Landsat</td>
<td>OLI</td>
<td>Path/row 165/37</td>
<td>30</td>
</tr>
<tr>
<td>02-JUN-15</td>
<td>Landsat</td>
<td>OLI</td>
<td>Path/row 164/37</td>
<td>30</td>
</tr>
<tr>
<td>12-AUG-15</td>
<td>Landsat</td>
<td>OLI</td>
<td>Path/row 165/37</td>
<td>30</td>
</tr>
<tr>
<td>5-AUG-15</td>
<td>Landsat</td>
<td>OLI</td>
<td>Path/row 165/37</td>
<td>30</td>
</tr>
</tbody>
</table>

Satellite image Pre-processing
Earth atmosphere is a mixed of gases, liquid and solid particles, most of these are optically active causing absorption, diffusion and scattering. Signals which measured at the satellite is the emergent radiation from the earth surface atmosphere system in the sensor observation direction. The radiance measured at sensor is known as Top of Atmosphere (TOA) radiance. Atmospheric corrections aim to convert the TOA radiance of the objects into the near earth reflectance (Lakshmi et al., 2015). Atmospheric correction was done using Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubus (FLAASH) algorithm. FLAASH was developed to provide accurate, physics-based derivation of atmospheric properties in Envi 5.1. FLAASH includes correction for the adjacency effect, cirrus and opaque cloud classification and adjustable spectral polishing for artifact suppression (Jia et al, 2014; Lakshmi et al., 2015).

Satellite image processing – Hierarchical image classification
First, for image processing, the conceptual model of the three-level earth’s surface matrix that was shown in Figure 2 was applied on both TM and OLI data.

Table 2. LULC classification (Feranec, et al., 2007; Feranec, et al., 2010)

<table>
<thead>
<tr>
<th>Short description</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agricultural area which mainly are drainage with other</td>
<td>Drainage agriculture</td>
</tr>
<tr>
<td>rain water</td>
<td></td>
</tr>
<tr>
<td>Agriculture areas that irrigate only with rain water</td>
<td>Rain-fed agriculture</td>
</tr>
<tr>
<td>Natural vegetation upper than 50%</td>
<td>Dense range land</td>
</tr>
<tr>
<td>Vegetation cover crown less than 50%</td>
<td>Sparse range land</td>
</tr>
<tr>
<td>Trees with rangelands</td>
<td>Forest</td>
</tr>
<tr>
<td>Housing developments</td>
<td>Residential areas</td>
</tr>
<tr>
<td>water bodies including; dam, natural and manmade lakes</td>
<td>Water bodies</td>
</tr>
</tbody>
</table>
First step
At this step, the lower cover crown percentage of rain-fed agriculture was considered as the threshold of 50% cover crown for separating from sparse rangeland and other LULC. To map VCCP, NDVI index was used as follows (Equation 1) (Mukherjee, 2004; Oldeland et al., 2010; Peña & Brenning, 2015):

\[ NDVI = \frac{NIRb - REDb}{NIRb + REDb} \]

Simple linear regression was done between samples taken as the dependent variables and their NDVI values for each image as independent variables. Using prepared VCCP models, the VCCP maps prepared two classes of dense vegetation and sparse vegetation. Field control and comparison with FCC image showed that there were some mixings between the cultivation area (especially rain fed agriculture) and dense rangeland. In this step, because of drainage, agriculture areas separate more correctly.

Second step
For initial separation of the rain-fed agricultural area and dense rangeland during field studies, and by overlaying first step vegetation map on slope percentage image, it become clear that 30 percent slope was the threshold between rain-fed area and dense rangeland. On the other hand, there was no rain-fed cultivation over 30 percent slope in mountain areas. By applying 30 percent slope threshold in the first step, dense vegetation in more than 30 percent, which were mostly dense rangelands and forests, were separated from dense vegetation in less than 30 percent slope that were mostly drainage and rain-fed agricultural areas (Figure 2).

Third level
On the third level of the hierarchical model (Figure 2), four categories: drainage, rain-fed agriculture, forest and dense rangeland were considered as the sub-classes for the dense vegetation (>50% vegetation coverage). On the other hand, three categories were determined as the sub-classes for the low-density vegetation (<50%) including residential areas, sparse rangeland, and land under water. Because agriculture lands have geomatics shapes, by paying attention to their reflectance and their shape, they were classified as rain-fed and drainage agriculture using object-based image classification method, and were separated from satellite data. Other LULCs were classified with Fisher supervised image classification. Residential area maps for TM image were produced using TM image for January 1998 when the land was totally covered with snow, only residential areas did not have snow cover; and residential areas were separated by applying Fisher image classification method.

Finally, all the individual layers were combined to produce LULCs maps.

Maps accuracy assessment
For accuracy assessment, samples were collected in field studies and were used for TM images. Some areas were selected as samples by considering field studies results and FCC images comparison. The overall accuracy and Kappa coefficient, commission error and omission were also determined.
Results and discussion

Satellite image classification

In supervised classification methods, especially in Fisher classification method, to produce accurate maps, it is important to take samples that are really pure samples of each land-use reflectance (Al-doski et al., 2013). Therefore, samples for each land-use must be prepared in areas far from margins of a land-use/land-cover. In this study, by proposing a sampling in heterogeneous area of each land-use/land-cover and by taking at least 30 meters distance from marginal land-use/land-cover (pixels in border of two land uses/land covers), we managed to achieve pure samples that were really samples of a land use/land-cover (LULC).

Image classifications results have shown that LULC classes with similar reflectance values in different bands have more errors. Moreover, small patches of isolated land covers can increase the classification errors because of impacts of the reflections from the adjacent pixels. In their studies, Luna, Cesar (2003); Yuan et al. (2005); Kamusoko and Aniya (2007); Lakshmi et al. (2015); Estoque and Murayama (2015), mentioned that similarity between LULCs increase errors in image classification. Kamusoko and Aniya (2007) explained that the accuracy of the classification depends on the degree of differentiation among the spectral reflections of LULC classes. Figures 3a and 3b show graphs signatures over used bands for TM and OLI simultaneously either as a spectral response pattern or mean reflectance (b1...n= band number).

As shown in figures 3a and 3b, especially the dense rangeland and rain-fed agriculture follow almost similar reflectance patterns in all bands of the imagery. In these images, rain-fed, dense rangeland and forest almost have the same reflectance trend. Residential areas have high reflectance in all bands, and water reflectance is the lowest after near infrared band.

In this study, by paying attention to LULC similarity and complexity, hierarchical scheme of LULC was designed for satellite image classification. Disperati et al. (2015), for satellite image classification, designed 3 and 4 levels for land classes and mentioned they produced land maps in each level; and at the end, they combined all results together to achieve the final LULC map.

In this study, for VCCP, models were prepared using NDVI index that is a common and useful vegetation index for surviving different kinds of plants (Mukherjee, 2004; Oldeland et al., 2010; Jovanović et al., 2015). Formulae 2 and 3 have shown the vegetation cover crown percentage model.

Equivalent 2:

\[ Y = 179.3X + 24.89 \]

\[ R^2 = 0.89 \text{ P-value < 0.01 for the year 2016} \]
Equivalent 3:
\[ Y = 240.95x + 16.13 \]

\[ R^2 = 0.80 \text{ P-value < 0.0 for the year 1998} \]

Where \( Y \) is vegetation cover crown percentage, and \( X \) is values in NDVI index.

The range of NDVI values are -1 to 1, the lower values show lower VCCP and the upper values are related to areas with more VCCP. In this study, initial image classifications showed that it was not possible to separate agricultural area especially rain-fed agriculture from dense rangeland. In addition, NDVI classification on the basis of produced modes could not separate dense rangeland from agriculture areas in the first level. The field sampleings and overlaying FCC images on slope map indicated that dense rangelands are normally located on slopes greater than 30 percent in mountain areas, while drainage and rain-fed agriculture were located on slopes less than 30 percent slope. Thus, by overlaying the slope layer on the satellite images in GIS area using multiple method, dense rangelands which were mostly located on mountain areas were separated.

Finally, LULC maps were proposed using the conceptual model, and images were classified with hybrid method in 7 layers for years 1998 and 2016. Yuan, et al. (2005); Kamusoko and Aniya (2007) used hybrid image classification and explained that this method is applicable in land with complex reflectance. Figures 4a and b and 5a and b, respectively show LULC maps of the area in the second and third stages. Table 4 shows each class area in hectares.

Residential areas that were small patches were distributed across the study area, and therefore their reflectance was influenced by the neighboring pixels (Malmir et al., 2015). Fisher classification method was able to separate the residential areas in both sensors. In some cases, residential areas and low-density rangeland were classified as one class in TM sensor images. Fisher supervised classification method can separate LULC with high accuracy when training sites were collected accurately (Al-doski et al., 2013). In this study, samples from each LULC were collected by paying attention to OLI spatial resolution; and they were collected in homogenous areas. Table 4 shows there is no significant change in drainage agriculture area and increase in rain-fed agriculture. From 1998 to 2016, water area in The Zayandehrood dam had a 1671-hectare decrease. On the basis of table 2, during this time, dense rangelands and forests decreased, too.

![Figure 4a. Dense and sparse vegetation (level 2)](image)

![Figure 4b. LULC map (level 3)](image)

<table>
<thead>
<tr>
<th>LULC class/ Area (ha)</th>
<th>Drainage agriculture</th>
<th>Rain-fed agriculture</th>
<th>Dense range land</th>
<th>Sparse range land</th>
<th>Forest</th>
<th>Residential area</th>
<th>Water body</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class area 1998</td>
<td>58665</td>
<td>18285</td>
<td>58135</td>
<td>257929</td>
<td>12028</td>
<td>3467</td>
<td>4490</td>
</tr>
<tr>
<td>Class area 2016</td>
<td>58029</td>
<td>25685</td>
<td>48854</td>
<td>262406</td>
<td>11102</td>
<td>4269</td>
<td>2654</td>
</tr>
</tbody>
</table>
A Hierarchical Approach of Hybrid Image Classification for Land use and Land cover Mapping

In this study, for classification accuracy assessment using field study samples and produced maps, error matrices produced for hybrid method results, and kappa coefficient, overall accuracy, precision of producer and user, commission and omission errors were calculated and shown in Tables 5 and 6 (Lunetta & Lyon, 2004; Benfield et al., 2007; Al-doski et al., 2015, Lakshmi et al., 2015).

Error matrices, tables 5 and 6, show that most errors related to misclassified areas are related to rain-fed and drainage agriculture areas and dense range

Table 5. LULC error matrix for map1998 (Pixel)

<table>
<thead>
<tr>
<th>Classes</th>
<th>Drainage agriculture</th>
<th>Rain-fed agriculture</th>
<th>Dense range land</th>
<th>Sparse range land</th>
<th>Forest</th>
<th>Residential area</th>
<th>Water body</th>
<th>Total</th>
<th>ErrorC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drainage agriculture</td>
<td>21695</td>
<td>448</td>
<td>1825</td>
<td>21</td>
<td>857</td>
<td>164</td>
<td>0</td>
<td>25037</td>
<td>0.1335</td>
</tr>
<tr>
<td>Rain-fed agriculture</td>
<td>1850</td>
<td>3017</td>
<td>1850</td>
<td>511</td>
<td>627</td>
<td>13</td>
<td>0</td>
<td>6069</td>
<td>0.5017</td>
</tr>
<tr>
<td>Dense range land</td>
<td>166</td>
<td>539</td>
<td>24564</td>
<td>2114</td>
<td>1947</td>
<td>1</td>
<td>0</td>
<td>29331</td>
<td>0.1625</td>
</tr>
<tr>
<td>Sparse range land</td>
<td>408</td>
<td>209</td>
<td>1772</td>
<td>42715</td>
<td>3721</td>
<td>550</td>
<td>0</td>
<td>49375</td>
<td>0.1351</td>
</tr>
<tr>
<td>Forest</td>
<td>0</td>
<td>0</td>
<td>1660</td>
<td>454</td>
<td>6973</td>
<td>0</td>
<td>0</td>
<td>7887</td>
<td>0.1159</td>
</tr>
<tr>
<td>Residential area</td>
<td>63</td>
<td>0</td>
<td>0</td>
<td>425</td>
<td>0</td>
<td>3636</td>
<td>0</td>
<td>3944</td>
<td>0.1021</td>
</tr>
<tr>
<td>Water body</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>14552</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>24119</td>
<td>4213</td>
<td>31671</td>
<td>46240</td>
<td>14125</td>
<td>4681</td>
<td>14552</td>
<td>136195</td>
<td>-</td>
</tr>
<tr>
<td>ErrorO</td>
<td>0.1028</td>
<td>0.2839</td>
<td>0.2244</td>
<td>0.0762</td>
<td>0.5063</td>
<td>0.2232</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

ErrorO = Errors of Omission (expressed as proportions)
ErrorC = Errors of Commission (expressed as proportions)
90% Confidence Interval = +/- 0.000815 (0.046447 – 0.048076)
95% Confidence Interval = +/- 0.000971 (0.046291 – 0.048232)
99% Confidence Interval = +/- 0.001278 (0.045984 – 0.048539)
Kappa: 0.81
Overall accuracy: 0.84
Lands together. Figures 3.a and b show that these misclassifications are related to similarity between drainage and rain-fed agriculture in dense rangeland areas. TM error matrix (table 5) has shown that most misclassifications are for relating rain-fed agriculture to dense rangelands and forests.

Table 6 indicates that in the map prepared by OLI image, rain-fed agriculture, drainage agriculture, dense and sparse rangelands were separated correctly. In prepared maps, some drainage agricultural pixels are wrongly related to residential areas because of small green spaces in residential areas. On the other hand, because forests included trees and rangeland together, in this class, there is a high misclassification between different vegetation classes, and therefore it has high commission and omission error in two imageries.

Using TM and OLI images, land-use/land-cover was extracted due to different reflectance behaviors of water compared to other phenomena (Figure 3a and b). Tables 5 and 6 show produced map using TM sensor has overall accuracy 84% and with kappa coefficient 0.81, and produced map using OLI sensor has overall accuracy 91% and with kappa coefficient 0.87, which is more than TM. This difference was predictable because of OLI characteristics like more radiometric and spectral resolution.

### Conclusion

LULC spectral profiles have shown LULCs digital numbers were more separated in OLI with 16-bit than TM data, so this is the reason for the less accuracy in TM map with 8-bit radiometric resolution (Figures 3a and 3b). Applying proposed hybrid method inland hierarchy concept could produce almost the same accurate maps for two imageries data. In this study, Fisher classification (Al-doski et al., 2013), object-based classification (Blaschke, 2010; Vieira et al., 2012; Phinn et al., 2012), and NDVI vegetation index (Peña & Brenning, 2015; Oldeland et al., 2010) were used in designed hybrid classification method.

Error matrices have shown more accurate image classification results in the map provided by OLI sensor than TM, especially in mapping different vegetation types and separating land surfaces such as residential areas.

Considering the similarity of some LULC reflectance in this study, in the hierarchy concept of land matrix, hybrid method can produce acceptable LULC maps. Thus, providing detailed maps of LULC that have small areas and similar reflectance will be possible through appropriate methods for each defined land level using different imageries.
References


