Algorithm to select Optimal Spectral Bands for Hyperspectral Index of Feature Extraction

P. Murugan¹²*, R. Sivakumar¹, R. Pandiyan² and M. Annadurai²

¹Department of Civil Engineering, SRM University, Kattankulathur - 603203, Tamil Nadu, India; palanimuruga64@gmail.com, sivakumar.r@ktr.srmuniv.ac.in, pandiyan@isac.gov.in, madurai@isac.gov.in
²ISRO Satellite Centre, Bangalore - 560017, Karnataka, India; muru@isac.gov.in

Abstract

Background/Objectives: Water body management and food-grains supply will be challenging tasks. Selecting spectral bands for accurate area estimation of water body and selected vegetation (crop) is the objective of this study. Methods/Statistical Analysis: Indices such as Normalized Difference Water Index (NDWI), Normalized Difference Vegetation Index (NDVI) are used for extracting water and vegetation respectively from other features. Generally bands providing high index values for the target are utilized for extraction. In this study the bands giving high index value for selected target and low index for surrounding are selected. Bands selected in this method provide better extraction and accurate area estimation. Findings: The NDWI and NDVI are based on multispectral data and have less number of combinations for band sets selection. Though the proposed method is derived from NDWI, it utilizes the Hyperspectral data that has narrow and hundreds of bands. Another advantage of this method is it utilizes the index value of target and surrounding features. It selects suitable band set by iteration method and provides accurate extraction of the targets and area estimation. The performance of the bands selected in this method was tested with coastal and inland water bodies. Area estimated with this method matches with NDVI and MNDWI values. Applications/Improvements: This method selects suitable bands to estimate area of water body and vegetation. This estimation will be useful for water body management and food production forecasting.

Keywords: Algorithm, Hyperspectral, Index, Vegetation, Water Body

1. Introduction

Water bodies and vegetation enclosures are important resources of Earth. Their quantity and quality are to be monitored for better utilization and preservation. The quality of groundwater and surface water sources are affected by the human activities such as urbanization¹, improper disposal of chemical and dyeing industrial effluents² and excess usage of fertilizers. The storage capacity of water bodies are affected by sedimentation, encroachment by increasing population etc. Flood area extraction is very important activity for search and rescue during disaster. It is now a well known fact that the satellite remote sensing is an economical and efficient way of monitoring the resources at regular time intervals. Feature identification and extraction is one of the major activities in remote sensing data analysis. Various methods are used to identify and extract required features from multispectral satellite data³⁴ and SAR data⁵. Among them the index methods are simple and effective.

The popular way to extract and estimate the water surface area from multispectral data is Normalized Difference Water Index (NDWI)⁶. The NDWI was modified as Modified Normal Difference Water Index (MNDWI)² using MIR band instead of NIR band in another study⁴ to extract water bodies with the background of buildings. Water moisture content in the vegetation was estimated with NDWI using RED and MWIR bands⁷. For clarity and to avoid the confusion between Gao’s NDWI and McFeeters’ NDWI, Gao’s NDWI is called as Normalized
Difference Moisture Index (NDMI)\textsuperscript{14}. These indices are used for extraction of water bodies\textsuperscript{12,13} and spatial-temporal change studies\textsuperscript{14}. The threshold is adjusted to get better extraction results\textsuperscript{13}.

Similarly the spread area of forest and cultivation is estimated using Normalized Difference Vegetation Index (NDVI)\textsuperscript{16}. NDVI is used in many studies\textsuperscript{15,16} to estimate the vegetation area. Various indices like Renormalized Difference Vegetation Index (RDVI), Weighted Difference Vegetation Index (WDVI)\textsuperscript{15}, Soil Adjusted Vegetation Index (SAVI)\textsuperscript{17} have also been evolved to monitor the vegetation area.

The Normalized Difference Snow Index (NDSI)\textsuperscript{21} was developed by and extensively used in many studies\textsuperscript{22–26}. Other familiar indices are Normalized Difference Pond Index (NDPI)\textsuperscript{27}, Normalized difference Built-up Index (NDBI)\textsuperscript{28}, Optimized Soil Adjusted Vegetation Index (OSAVI)\textsuperscript{29}.

Some more indices are also developed to extract built up area of urban, derive temperature of specified location, discriminate snow etc. using multispectral images. Table 1 provides some indices and spectral bands used for feature extraction.

Table 1. Feature extraction spectral indices

<table>
<thead>
<tr>
<th>Index</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normalized Difference Water Index</td>
<td>NDWI = (Green-NIR)/(Green+NIR)</td>
</tr>
<tr>
<td>Modified NDWI</td>
<td>MNDWI = (R_{green}-R_{MIR})/(R_{green}+R_{MIR})</td>
</tr>
<tr>
<td>Normalized Difference Water Index for vegetation</td>
<td>NDWI = (R_{NIR}-R_{MIR})/(R_{NIR}+R_{MIR})</td>
</tr>
<tr>
<td>Normalized Difference Vegetation Index</td>
<td>NDVI = (R_{NIR}-R_{RED})/(R_{NIR}+R_{RED})</td>
</tr>
<tr>
<td>Soil Adjusted Vegetation Index</td>
<td>SAVI = (1+L)(R_{670}-R_{870})/(R_{670}+R_{870}+L) L = 0.5</td>
</tr>
<tr>
<td>Optimized Soil Adjusted Vegetation Index</td>
<td>OSAVI = (1+0.16)(R_{670}-R_{870})/(R_{670}+R_{870}+0.16)</td>
</tr>
<tr>
<td>Normalised difference Snow Index</td>
<td>NDSI = (Green- SWIR)/(Green + SWIR)</td>
</tr>
<tr>
<td>Normalized difference Built-up Index</td>
<td>NDBI = (Band5-Band4)/(Band5+Band4)</td>
</tr>
<tr>
<td>Normalized difference Pond Index</td>
<td>NDPI = (R_{NIR}-R_{green})/(R_{NIR}+R_{green})</td>
</tr>
</tbody>
</table>

The multispectral data gives spectral information of the target in few spectral bands. Selecting relatively better band combination for the index is easy. As Hyperspectral imagery has hundreds of narrow spectral bands, selecting most suitable band combination for a particular feature surrounded by other features is a difficult task. The band set used in the above indices are selected based on high index value obtained for the interested feature only. In this paper, a supervised band selection method for index is presented. In this method band sets are selected based on the index values of the interested target feature and surrounding features. The performance of the extraction index will be better if the index gives higher values for interested feature and lower values for surrounding features. An algorithm to select suitable bands for an index is presented in this paper.

The proposed algorithm can be used for satellite or Ariel remote sensing data. When we use algorithm on satellite data, the spectral bands affected by the atmospheric absorption are to be removed from study.

2. Methods and Materials

2.1 Algorithm

The proposed algorithm considers the spectral reflectance of target feature reflectance along with surrounding features reflectance also in selecting the spectral bands for an index. The difference between each surrounding feature index value and target feature index value are calculated and combined suitably. The spectral band set giving highest combined difference value is selected for index application.

The algorithm to select suitable bands for Index is given below:

<table>
<thead>
<tr>
<th>Step 1:</th>
<th>Select a target feature which is to be extracted from the image. (It can be a Water body, Vegetation etc).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 2:</td>
<td>Obtain the DN values of selected target in all spectral bands.</td>
</tr>
<tr>
<td>Step 3:</td>
<td>Convert them to radiance values.</td>
</tr>
</tbody>
</table>
### Step 4:
Calculate Index (NDVI, NDWI, NDSI etc.) values for target (T) with all possible band combinations using Equation (1).

\[ T(n,m) = \frac{B_{n} - B_{m}}{B_{n} + B_{m}} \]  

Where \( B_{n} \) and \( B_{m} \) are radiance values of selected bands to compute the index.

\[ \sum_{n=0}^{N} B_{n} \sum_{m=0}^{N} B_{m} \Rightarrow I_{nm} \]

N number of bands gives \( N^2 \) number of combinations. Where \( B_{n} \) is the Band \( n \), \( B_{m} \) is Band \( m \) and \( I_{nm} \) is the Index value calculated with band number n and m.

### Step 5:
Select surrounding feature (Reference Feature).

### Step 6:
Obtain DN values.

### Step 7:
Convert them to radiance values.

### Step 8:
Similar to step 4, calculate index for surrounding feature with all possible band combinations using Equation (2).

\[ S_{x}(n,m) = \frac{(B_{n} - B_{m})}{(B_{n} + B_{m})} \]

Where \( x \) is the surrounding feature number.

### Step 9:
Find the difference between target index and surrounding index of corresponding set of spectral bands.  
The difference Index value (D) for selected feature is obtained using Equation (3).

\[ D_{x}(n,m) = T(n,m) - S_{x}(n,m) \]

Where \( x \) is the surrounding feature number.

### Step 10:
To add another surrounding feature, go to step 4 else go to step 11.

### Step 11:
Use suitable method (Arithmetic mean, geometric mean etc) to obtain cumulative index values from difference index values for each band set.  
Cumulative index for each band set can be calculated by averaging the index values as given in Equation (4) or by finding geometric mean of index values as given in Equation (5).

\[ C(n,m) = \left( \frac{D_{1}(n,m) + D_{2}(n,m) + \ldots + D_{x}(n,m)}{x} \right) \]

i.e \( C(n,m) = \left( \frac{\sum_{i=1}^{x} D_{i}(n,m)}{x} \right) \)

\[ C(n,m) = \left( \prod_{i=1}^{x} D_{i}(n,m) \right)^{\frac{1}{x}} \]

Weighted average methods also can be used based on the abundance of each surrounding feature.

\[ C(n,m) = \left( \frac{A_{1}*D_{1}(n,m)+A_{2}*D_{2}(n,m)+\ldots+A_{x}*D_{x}(n,m)}{weightage\ factor} \right) \]

\[ C(n,m) = \frac{A_{1}*D_{1}(n,m)*A_{2}*D_{2}(n,m)*\ldots*A_{x}*D_{x}(n,m)}{weightage\ factor} \]

Where \( A_{1}, A_{2}, \ldots, A_{x} \) are the weightage values for related surrounding features.

### Step 12:
Find the highest cumulative index value.

### Step 13:
Find the spectral channels (bands) corresponding to the highest cumulative index value.
Step 14: Use the selected band set in the index equation to extract the feature.

Bands selected in this method provide better contrast between the target feature and surrounding feature.

2.2 Data set for Algorithm Verification

The algorithm verification was done with Hyperion data. Hyperion is a medium spatial resolution Hyperspectral imager flown on the NASA’s new millennium mission EO-1. Hyperion has 242 spectral bands ranging from 355–2577 nm. Due to low Signal to Noise Ratio (SNR) 198 bands are only processed.

Some salient features of Hyperion data are given below:

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial resolution</td>
<td>30 m</td>
</tr>
<tr>
<td>Number of pixels</td>
<td>256</td>
</tr>
<tr>
<td>Swath</td>
<td>7.5 km</td>
</tr>
<tr>
<td>No. of bands</td>
<td>242</td>
</tr>
<tr>
<td>No. of calibrated bands</td>
<td>198</td>
</tr>
<tr>
<td>Spectral coverage</td>
<td>400 nm to 2500 nm</td>
</tr>
<tr>
<td>Band width</td>
<td>~10 nm</td>
</tr>
<tr>
<td>Image coverage</td>
<td>7.5 x 100 km (max)</td>
</tr>
</tbody>
</table>

The wide spectral coverage of Hyperion is achieved by two systems namely Visible and Near Infrared (VNIR, 400-1000 nm) and Short Wavelength Infrared (SWIR 900–2500 nm). The VNIR data is scaled by 40 and SWIR data is scaled by 80. There is a spectral overlap between the VNIR and SWIR bands. Radiometrically and geometrically corrected data is provided by United State Geological Survey (USGS) through the website ‘http://earthexplorer.usgs.gov/’. The ‘Level 1Gst’ data obtained from USGS is already radiometrically corrected, resampled for geometric correction, registered to geographic map projection and ortho-corrected using Digital Elevation Models (DEM). This study was carried out with the spectral bands of 8 to 57 from VNIR and 79 to 224 of SWIR. This selection avoids non-calibrated bands and duplicate calibrated bands. This selection gives a continuous spectrum of 435 nm to 2365 nm. The downloaded data was subjected to atmospheric correction and noise reduction processes. Data preprocessing was done with the commercially available software and the algorithm was incorporated and tested with indigenously developed software.

2.3 Test Sites

The algorithm was tested with Hyperion data of two different areas to cover both Case-1 and Case-2 water bodies. The first test area is the upper part of Pulicat Lake which is between 13° 26’ and 13° 43’ N latitude and 80 ° 03’ and 80 ° 18’ E longitudes. It is located 60 Km North of Chennai and covers areas of Andhra Pradesh and Tamil Nadu states. The Lake is a well known productive eco-

![Figure 1. First test site (Part of Pulicat lake) as seen by Hyperion.](image1)

![Figure 2. Second test site (Borana kanive) as seen by Hyperion.](image2)
systems and second biggest brackish water lake in India, formed out of backwaters of the Bay of Bengal, covering an area of 600 sq.km\textsuperscript{31}. The average depth of the lake is about 1.5 m and the depth varies between 0.5 to 6.0 m. The depth increases in rainy season. The inlet to this lagoon is by three seasonal rivers namely Arani, Kalangi and Swarnamukhi. This test data covers sea water of Bay of Bengal, turbid water of Pulicat Lake, mangrove vegetation and different types of sand. The second test area, Borana Kanive dam is at 13.35.48 N Latitude and 76.37.21 E Longitude. It is an inland clean water body surrounded with barren land, hillock and vegetation. It is located in Huliyar village near Sira town of Karnataka state, India. The water source for this dam is seasonal rain water from hillocks. Selected sites are shown in Figures 1 and 2.

The DN numbers of this image were converted into radiance and reflectance as per the guidelines given in Hyperion Data book. The reflectance values were used for analysis. Typical radiance and reflectance values obtained from Hyperion are shown in Figure 3 and Figure 4 respectively.

Hence, the index values are shown in image form. The index values vary from -1 to +1. The values are biased by adding 1 to avoid negative values and linearly stretched to the range of 0 to 255 for display purpose. Bands B\textsubscript{n} are in x axis and bands B\textsubscript{m} are in Y-axis. Band numbers increase from left to right and bottom to top. When n = m the index values will be 0. The brighter points show the index near +1 and dark area shows the index near -1. Though algorithm for the band selection is applicable for all indices which are given in Table 1, here it was tested with NDWI and NDVI only.

Using the Hyperion data of Pulicat Lake and Borana Kaneve, the algorithm was verified. The NDWI index values with various possible combinations of spectral bands for the water body were calculated using Pulicat Lake and the values are shown in Figure 5.

As number of bands selected for study is high, representing the results in mesh or grid form is not possible. Hence, the index values are shown in image form. The index values vary from -1 to +1. The values are biased by adding 1 to avoid negative values and linearly stretched to the range of 0 to 255 for display purpose. Bands B\textsubscript{n} are in x axis and bands B\textsubscript{m} are in Y-axis. Band numbers increase from left to right and bottom to top. When n = m the index values will be 0. The brighter points show the index near +1 and dark area shows the index near -1. Though algorithm for the band selection is applicable for all indices which are given in Table 1, here it was tested with NDWI and NDVI only.

Using the Hyperion data of Pulicat Lake and Borana Kaneve, the algorithm was verified. The NDWI index values with various possible combinations of spectral bands for the water body were calculated using Pulicat Lake and the values are shown in Figure 5.

Similarly the NDWI index values for the surrounding feature is calculated and presented in Figure 6.

The difference between target feature and the surrounding feature (reference) index values are calculated and displayed in Figure 7.

The Figure 7 shows difference of index values of target and surrounding (one feature). Similarly difference index values for other surrounding features were also calculated. The difference values were suitably combined to generate cumulative index values. These cumulative index values provide the difference between the target index and surrounding index.
The Figure 8 shows the cumulative values obtained by averaging of difference indices obtained from various reference features surrounding the target. The brighter points give the better set of spectral band for extraction of selected target. In the above exercise, the water body is chosen as target and vegetation and sand as surrounding features. The cumulative index values obtained using multiplication method is presented in Figure 9. The cumulative value can be obtained using geometric mean method also. In that method, if one difference index value is zero, the cumulative index for that band combination will become zero.

Figure 6. NDWI of different band sets for a reference (surrounding) feature.

Figure 7. Difference between target index values and reference index values.

Figure 8. Cumulative index values obtained by arithmetic mean method.

Figure 9. Cumulative index values obtained by geometric mean method.

Figure 10. Cumulative index values for vegetation target.
Similar concept was followed to generate the vegetation index and values are provided in the Figure 10.

2.4 Verification of Results
The results were verified with two different data sets for NDWI and NDVI. The bands corresponding to highest cumulative index values were used for feature extraction and the results are presented in Figures 11 and 12.

Though SWIR bands of Hyperion data is affected by noise and images are with stripes, the algorithm is working satisfactorily.

The water body extraction results using the second test site data without threshold adjustment is shown in Figure 13. The Figure 14 shows the result with suitable threshold which eliminated gray objects.
The cumulative difference index image shows the regions of suitable spectral bands for a particular feature extraction. The band sets corresponding to dark areas are not suitable to extract the feature as they have almost equal index for target and background. The spectral bands corresponding to the brighter points will give better extraction results.

2.5 Noisy Channels

Though the proposed method is efficient in selecting suitable band sets in ideal condition, in practical, bands with low signal to noise ratio may not give expected result. The radiance levels of some bands are almost zero because of atmospheric water molecule absorption, low reflectance and imaging system noise. Considering these channels in index calculation may lead to extraction with noise and strips as shown in Figure 12. Better feature extraction can be achieved by eliminating these low SNR channels from consideration. A minimum cut off signal value can be set to eliminate the channels which give low signal. These eliminated channels will not be considered for band set selection. After elimination the index values are presented in Figure 15.

![Figure 15. Bands eliminated based on low signal levels.](image)

3. Results and Discussions

3.1 Results with Water Index

The band sets from brighter areas of Figure 16 were used, extraction was carried out and the results are shown in Figures 17 to 20.

![Figure 17. Water body extracted with B1 = 498 nm and B2 = 854 nm (A).](image)
3.2 Results with Vegetation Index

Similarly noise corrected index values were calculated for vegetation target and shown in Figure 21.

Figure 21. Index values for vegetation extraction corrected by noise matrix.

It can be noted that interchanging the $b_w$ and $b_m$ of water index (Figure 16) provides the vegetation index (Figure 21).

The vegetation area extraction also exercised with the algorithm. The results with different spectral combinations selected from the Figure 21 are shown in Figures 22, 23 and 24.

Figure 22. Vegetation area extraction with $B_1 = 854$ nm and $B_2 = 498$ nm (A).
Algorithm to select Optimal Spectral Bands for Hyperspectral Index of Feature Extraction

3.3 Road Extraction

The suitable bands were selected for road extraction and the results are shown in the Figure 25. In the figure the road near Bangalore and road and railway of the north Chennai are shown. Some other features like waterbodies, buildings are seen after extraction, but they are much away from target feature.

3.4 Comparison with NDWI and MNDWI

The results obtained with optimized bands water index were compared with the NDWI and MNDWI methods. For NDWI and MNDWI calculations, Hyperion bands corresponding to the middle value of Landsat spectral bands were selected i.e Hyperion Band No. 21, 48 and 150 were used in place of Landsat Band 2, Band 4 and Band 5. The images generated with these three indices are shown in Figure 26. It can be seen that the optimized bands water index provides better contrast to the selected features with surrounding features.

In the image at locations 1 and 2 the extraction of water by NDWI is not clear, while other two methods have extracted clearly. As the MNDWI has higher background noise, the water and other features are not separated well. The water index with optimized bands gives better discrimination between water and other features. The result with second test site is given in Figure 27.

The area of extracted water bodies using these three methods at different sub scenes of these data sets are presented in terms of number of pixels in Table 3. Here it is to be noted that the pixel extraction is associated with threshold adjustment, a subjective activity. As the contrast between target and surrounding better in Band Optimized Water Index (BOWI), adjusting the threshold to extract the area is simple and faster. Threshold setting is time consuming in the case of MNDWI.

Figure 23. Vegetation area extraction with $B_1 = 1598$ nm and $B_2 = 498$ nm (D).

In the case of vegetation the effect of noise is less as the signal is higher than that of water body.

Figure 24. Vegetation area extraction with $B_1 = 1608$ nm and $B_2 = 813$ nm (E).

Figure 25. Roads and railways extracted using the algorithm.
Table 3. Area of water surface of sub scenes (in number of pixels).

<table>
<thead>
<tr>
<th>Sub Scene No</th>
<th>NDWI Bn = 559 nm Bm = 835 nm</th>
<th>WI with selected Bands (Varies Scene to Scene)</th>
<th>MNDWI Bn = 559 nm Bm = 1598 nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>33445</td>
<td>35505</td>
<td>37275</td>
</tr>
<tr>
<td>2</td>
<td>5532</td>
<td>5562</td>
<td>5637</td>
</tr>
<tr>
<td>3</td>
<td>1152</td>
<td>1171</td>
<td>1202</td>
</tr>
<tr>
<td>4</td>
<td>12930</td>
<td>12853</td>
<td>13453</td>
</tr>
<tr>
<td>5</td>
<td>8345</td>
<td>8389</td>
<td>8482</td>
</tr>
</tbody>
</table>

4. Conclusion

The method presented in this paper provides a better set of bands for an accurate and easy feature extraction using Hyperspectral data. It was tested for water and vegetation extraction. The results show a clear extraction of the targets. While area extracted by NDVI and BOWI are closely matching, MNDWI values are higher. As the SWIR bands of Hyperion data are with noise and low SNR, they are not giving expected output; in spite of that the algorithm works well and the extraction is found to be better.

5. Acknowledgement

Authors thank Management of Indian Space Research Organization and SRM University for their continuous support and encouragement during this study.

6. References

Algorithm to select Optimal Spectral Bands for Hyperspectral Index of Feature Extraction
