Searching in High Dimensional Space

Jeffrey Bowles and David Gillis
Naval Research Laboratory
Washington, DC 20375
Jeffrey.Bowles@nrl.navy.mil
Overview

- Hyperspectral Imaging and statement of the problem
- The approach developed at the NRL
  - Initially for use in anomaly detection work
- The complicated coastal ocean
- Retrieval of environmental parameters using a large Lookup Table (LUT)
Hyperspectral Imaging for Environmental Characterization

- A hyperspectral imager records the spectrum of the reflected light from each pixel in the scene.
- The spectral information can be exploited to retrieve detailed information about the scene.
- Coastal hyperspectral data products:
  - over water:
    - Bathymetry
    - Bottom Type
    - Chlorophyll Concentration
    - Colored Dissolved Organic Matter (CDOM)
    - Total Suspended Sediment (TSS)
    - Total Optical Attenuation Coefficient $K_a(\lambda)$
    - Optical Absorption Coefficient $A(\lambda)$
    - Optical Backscatter Coefficient $B_b(\lambda)$
    - Horizontal Visibility
  - over land:
    - Vegetation Type Maps
    - Soil Type Maps
    - Beach Trafficability …

The method of analysis and the imager performance requirements depend on the scene and the desired information.
What is the Problem

• Hyperspectral imager produce data at fast rates
  – Our CASI -1500 is about 30 GB/hr
  – Other systems have rates that are much higher

• There are often times when the data cubes are too large for the desired algorithm to process on a reasonable time scale

• There are applications that can require the searching large spectral libraries for the best match to a single spectrum
“Prescreener” Algorithm

- Replace the full hyperspectral image with a representative subset
  - One possible approach is to find a subset, \( S = \{s_1, s_2, s_3 \ldots \} \), called exemplars such that for all image pixels, \( I_i \),

\[
\cos^{-1} \left( \frac{I_i \cdot S_j}{\|I_i\| \cdot \|S_j\|} \right) \leq \varepsilon
\]

for at least one \( s \), where \( \varepsilon \) is an error criterion – typically 1 to 2 degrees
Algorithm

• Each exemplar has a “hypersphere” that it represents
  • Most image pixels have many exemplars that satisfy the above inequality (best match?)
  • Often, we will keep track of which image pixels matched which exemplars (codebook)
  • Here we are explicitly working with spectral shape and magnitude does not come into play
  • There is interesting behavior in how the exemplars and image spectra interact with changing angle
Building Up Exemplars

• We build the exemplars up one by one from the spectra in the image
  \{\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5, \ldots\}

• Each image pixel is compared with the current exemplar list to determine if it is already represented

• When the list is small, an exhaustive check is possible, is there an s such that
  \[ \frac{I_i \cdot S_s}{\|I_i\| \cdot \|S_s\|} \leq \varepsilon \]
  – However, the list is not sufficiently small for very long
Can We Limit the Possible Matches?

- An exhaustive search is not practical
- We can limit what needs to be checked by using the concept of a reference vector, $R_1$.

- An image spectra, $I_1$, can only be within $\varepsilon$ of $S_1$ if

$$S_1 R_1 - \delta \leq I_1 \cdot R_1 \leq S_1 R_1 + \delta$$

$$\delta = \sqrt{2(1 - \cos(\varepsilon))}$$
Algorithm

\[ \sigma_{\text{min}}^i = \langle X_i, R \rangle - \sqrt{2(1 - \theta_T)} \]

\[ \sigma_{\text{max}}^i = \langle X_i, R \rangle + \sqrt{2(1 - \theta_T)} \]

Sorted list of projected exemplars
Fine Tuning

- Can use more than one reference vector
  - The method is a trade off between dot products and logical decisions
  - Or physically, the more that is known about the spectra the better
- Have found that if reference vectors are the mean centered PCA directions the method can be very fast
  - Physically, one can consider this a method to only compare similar spectra – a water spectrum won’t be compared to a vegetation spectrum
Extend to multiple dimensions
Other aspects

• There is a big difference between doing this in radiance space or doing in digital numbers, or counts
  – In counts, DN have same weight regardless of wavelength
• For some applications, it is important to keep track of with which exemplar the image spectra matched – something we called the codebook
• There is interesting behavior in the results as the error angle is varied
  – As the error angle gets smaller, the number of different exemplars that can represent an image spectrum gets larger
  – The number of exemplars needed to describe a fixed number of image spectra increases strongly as error angle decreases
• One good aspect of this is that we can add spectra to the library and maintain efficiency without changing reference vectors.
  – Can support calculate as you go
Applications

• Initially this was used to limit spectra needed in calculations required to determine endmembers

• However, there are many applications
  – Speed up physical modeling approaches
    • Process only a limited set of spectra that were chosen in a representative manner
  – Searching large libraries
Exemplars to Speed Up Processing

• In this scene only water spectra were considered
• Yellow dots indicate the location of exemplars in the image
• Process exemplars only and then fill in results
• Could you get sufficiently accurate results processing only the yellow dots (3%) in the picture below?
Bio-optical Coastal Oceanography

- Monitoring the coastal ocean areas of the world is needed for many applications.
- It is often desired to determine bathymetry, bottom type information and water constituents.
- Water constituents include phytoplankton, Colored Dissolved Organic Matter (CDOM) and suspended solids.
- The coastal ocean is a very complicated place.
- However, forward radiative transfer modeling of the propagation light through the water column works well.
NRL Tafkaa Atmospheric Removal Algorithm

• Atmosphere and water surface reflection algorithm designed for maritime use
  - Uses atmospheric information in the hyperspectral data itself
    • Pixel-by-pixel -- does not assume horizontal homogeneity
  - Uses look-up table approach to find atmospheric parameters used to calculate correction.
Ecolight Radiative Transfer Model

• Ecolight is part of a code developed by Curt Mobley of Sequoia Scientific
  – Give the code all needed parameters and it will provide Remote Sensing Reflectance

• To create large library use that code to calculate $R_{RS}$ for millions or combinations of parameters
  – Depth, bottom reflectance, phytoplankton, CDOM, suspended sediments, and other details
HICO Instrument

- HICO was built at the NRL using commercial off the shelf (COTS) parts
- It is a VNIR (350-1050 nm measured) spectrometer with ~90 m GSD, and very good SNR in the blue
Launch To The ISS

Launched from Tanegashima Space Center, Japan, September 10, 2009, on Japanese HTV

HTV payload module carrying HREP docked to Space Station September 17

HREP on Japanese Remote Manipulator arm

HREP docked to Japanese Exposed Facility September 24

HICO viewing slot

Photographs courtesy NASA
# LUT Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Values</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water Depth</td>
<td>0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0, 11.0, 12.0, 13.0, 14.0, 15.0, 16.0, 17.0, 18.0, 19.0, 20.0, 21.0, 22.0, 23.0, 24.0, 25.0, 26.0, 27.0, 28.0, 29.0, 30.0 meter</td>
<td></td>
</tr>
<tr>
<td>Chlorophyll A</td>
<td>0 to 1 by 0.05; 1 to 20 by 1</td>
<td>mg/l</td>
</tr>
<tr>
<td>CDOM Extinction</td>
<td>0 to 1 by 0.1; 1.4, 1.8</td>
<td>1/meter</td>
</tr>
<tr>
<td>Bottom Type</td>
<td>CleanSeagrass (EL), RedAlgae (EL), GreenAlgae (EL), CoralSand (EL), BrownAlgae (EL), 18%Gray, Cladophora (FL), Dictyota (FL), DisturbedSand (FL), FilmySand (FL), GreenAlgae (FL), ImpactedTurf (FL), RedAlgae (FL), Sand (FL), TurfAlgae (FL), Gray18% (50)_RedAlgae (50), Gray18%(50)_TurfAlgae(50), Thalassia, BrownMud (MB), Cymodocea (MB), Ovalis (MB), Spinulosa (MB), Ulva (MB), WhiteSand (MB), Zostera (MB), TurfAlgae (LSI), CoralMontastria (LSI), CoralDichococenia (LSI), BiosandandGrass (LSI), OoidSand (LSI), DarkSediment (LSI), Macrophyte (LSI), Seagrass (LSI), DarkSediment_SeaGrass Mixtures, DarkSediment_TurfAlgae Mixtures, OoidSand_SeaGrass Mixtures, OoidSand_TurfAlgae Mixtures, Sand (WA), Cobble (WA), Ulva (WA)</td>
<td>Spectra</td>
</tr>
</tbody>
</table>
Mechanics of LUT Search

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Spectrum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chl 1, CDOM 1, BT 1, Depth 1</td>
<td>Spectrum</td>
</tr>
<tr>
<td>Chl 2, CDOM 1, BT 1, Depth 1</td>
<td>Spectrum</td>
</tr>
<tr>
<td>Chl 3, CDOM 1, BT 1, Depth 1</td>
<td>Spectrum</td>
</tr>
<tr>
<td>Chl 4, CDOM 1, BT 1, Depth 1</td>
<td>Spectrum</td>
</tr>
<tr>
<td>Chl 5, CDOM 1, BT 1, Depth 1</td>
<td>Spectrum</td>
</tr>
<tr>
<td>Chl 1, CDOM 2, BT 1, Depth 1</td>
<td>Spectrum</td>
</tr>
<tr>
<td>Chl 2, CDOM 2, BT 1, Depth 1</td>
<td>Spectrum</td>
</tr>
<tr>
<td>Chl 3, CDOM 2, BT 1, Depth 1</td>
<td>Spectrum</td>
</tr>
<tr>
<td>Chl 4, CDOM 2, BT 1, Depth 1</td>
<td>Spectrum</td>
</tr>
<tr>
<td>Chl 5, CDOM 2, BT 1, Depth 1</td>
<td>Spectrum</td>
</tr>
<tr>
<td>Chl 1, CDOM 3, BT 1, Depth 1</td>
<td>Spectrum</td>
</tr>
<tr>
<td>Chl 2, CDOM 3, BT 1, Depth 1</td>
<td>Spectrum</td>
</tr>
</tbody>
</table>

Matched Spectra

Determine Best Matched Spectra

Build up map of parameters
HICO Key Largo, FL
Key Largo, FL Rrs Comparison

In addition to shape (represent by angle) we optimize Euclidean distance
Key Largo, FL

CWST Bottom Type Class Map retrieved using the following flat file parameters:

- No. of Spectra = 1,924,146
- Depth = 0-30m + Optically Deep
- TSS = 0
- CDOM = 0-1 m-1
- Pigment = ChlA 0-5 mg/m^3
- Bottom Types = Coral Sand, Clean Seagrass, Brown Algae, Green Algae, Red Algae, and 18% Gray, Macrophyte, Coral Dichocoenia, Dark Sediment, Turf Algae, Ooid Sand, Biosand & Grass, Seagrass, Coral Montastria
Key Largo, FL

CWST Depth Map retrieved using the following flat file parameters:

No. of Spectra = 1,924,146
Depth = 0-30m + Optically Deep
TSS = 0
CDOM = 0-1 m-1
Pigment = ChlA 0-5 mg/m^3
Bottom Types = Coral Sand, Clean Seagrass, Brown Algae, Green Algae, Red Algae, and 18% Gray, Macrophyte, Coral Dichocoenia, Dark Sediment, Turf Algae, Ooid Sand, Biosand & Grass, Seagrass, Coral Montastria

Note: White is no match
Summary

• Finding the nearest neighbor in spectral space has a number of important applications
  – Many physical model based algorithms can be much faster if exemplars can be substituted for image spectra
  – Library searching such as shown here
• We believe that there are a number of approaches related to ours that can speed up the process significantly
• Using this search method to search large libraries for environmental work is manageable
Quick Math Question

- I don’t know if this is already understood
- The approach above depends on checking ranges
  - To see if I*R is within the necessary range to match with an exemplar
  - That comparison is time consuming
  - By “packing” the answers into a long integer the ranges can be check all at once
Packing for Faster Range Checking

• This search approach can benefit from fast range checking/ordering of exemplars
• The range is $S_1R_i +/\ - \sqrt{2*(1-\cos(\varepsilon))}$ – want to know if an exemplar is in this range for each reference vector
• Take the four values and multiply by $2^7$

<table>
<thead>
<tr>
<th>High</th>
<th>0 _ _ _ _ _ _ 0 _ _ _ _ _ _ 0 _ _ _ _ _ _ 0 _ _ _ _ _ _ 0 _ _ _ _ _ _</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_j^R$</td>
<td>0 _ _ _ _ _ _ 0 _ _ _ _ _ _ 0 _ _ _ _ _ _ 0 _ _ _ _ _ _ 0 _ _ _ _ _ _</td>
</tr>
<tr>
<td>Low</td>
<td>0 _ _ _ _ _ _ 0 _ _ _ _ _ _ 0 _ _ _ _ _ _ 0 _ _ _ _ _ _ 0 _ _ _ _ _ _</td>
</tr>
<tr>
<td>Mask</td>
<td>1 _ _ _ _ _ _ 1 _ _ _ _ _ _ 1 _ _ _ _ _ _ 1 _ _ _ _ _ _ 1 _ _ _ _ _ _</td>
</tr>
</tbody>
</table>

• Then if (((High – Image) or (Image-Low)) & Mask) is not zero one of the range tests failed
  – This is faster than doing 8 compares
• Some DSP have special single clock commands that facilitate this type of calculation