Subject: Re: Multi-band sampling strategy
Posted by wita on Wed, 18 Aug 2010 19:46:17 GMT
View Forum Message <> Reply to Message

Dear Jess,

The procedure that you are now using is fetching a complete slice of your image
for each sample location. I do not know how large your image is, but if I take
a typical TM scene (6000x6000 pixels), then many of your 38,000 pixels will
actually be positioned in the same slice. So you are probably fetching the
same slice a couple of times from your image just to take out another sample
location. The more sample locations you have, the worse the speed penalty gets.

Instead, what you could do to speed this up is first sort out which sample
locations fall on which slice and group them together. There is actually a
very efficient way to accomplish this in IDL by using the Histogram() command
in combination with the REVERSE_INDICES. If you have never heard of this then
please have a look at the histogram tutorial on David Fannings site to
understand the trick.

So in pseudo code it could look like this, where I assume that SampleLines,
SampleCols store the line/column numbers of your sample positions:

```
h = Histogram(SampleLines, MIN=0, REVERSE_INDICES=ri)
FOR j=0, N_Elements(h)-1 DO BEGIN
  IF ri[j+1] GT ri[j] THEN BEGIN
     SampleColsinLine = samplecols[ri[ri[j]:ri[j+1]-1]]
     SampledSpectra = Get_Sample_Spectra(fid, j, SampleColsinLine)
  ENDIF
ENDFOR
```

with the function Get_Sample_Spectra() looking like this:

```
FUNCTION get_sample_spectra, fid, line, samplecols
   Slice = ENVI_Get_Slice(FID=fid,line=line, pos=bandsSelected)
   RETURN, Slice[samplecols, *]
END
```

This has the additional advantage that you are progressing through the file
in the order in which it is stored on disk. This automatically takes advantage
of the read-ahead capabilities of operating systems and disk drives.

Creating point ROIs as Maxwell suggest is, in my experience, also becoming
very slow for a large number of samples. The suggestion by Paulo is inherently
limited by the size of your memory. Moreover, it is very inefficient for small
sample sizes.

Hope this helps.

Allard
