Subject: Re: Need Some Advice on Seperating Out Some Data
Posted by rdellsy on Tue, 08 Aug 2006 22:20:13 GMT
View Forum Message <> Reply to Message

I'm a tad confused about what you're suggesting. I'll try and work it
out, but I'm still fairly new to IDL, so if you could give an IDL or
pseudo-code example of what you're trying to explain, I would
appreciate. If that's too much work, I understand, and I'll just try to
puzzle it out on my own.
Thanks,
Rob

JD Smith wrote:
> On Tue, 08 Aug 2006 16:57:28 -0400, Ben Tupper wrote:
>> Hi,
>>
>> Just an end-of-the-day wildcard, but I would bin the data into a 2d
>> histogram (ala JD's HIST_ND or the built-in HIST_2D). Then I would try to
>> find the "saddle" between the data and noise. You'll have to fiddle with
>> the binsize a bit to balance "lumping" and "splitting" - maybe that can be
>> done dynamically. I dunno. But it should be quick.
>>
>> It is an interesting problem that we have face here with flow cytometry -
>> but we work the region manually as you do. I'll be interested to see what
>> your final solution is.
>
> A related concept would be to:
>
> 1. Bin the original data into a 2D image, with HIST_ND, with using
>   REVERSE_INDICES (call this RI#1).
> 2. Threshold this binned image so that it's zero below, and 1 above
>   some threshold value representing the "no data" saddle. This
>   threshold could be zero, but doesn't have to be (e.g. to take care
>   of random noisy points in the distribution). As Ben mentions,
>   you'll have to experiment to pick a good bin size.
> 3. Use LABEL_REGION to find all contiguous blobs of data in the
>   bi-valued, thresholded, binned image.
> 4. Use HISTOGRAM with REVERSE_INDICES (RI#2) on the resulting "label
>   image" to find the extents/centroid/etc. of the data in each "blob"
>   (either roughly via the bin positions present in the blob, or more
>   precisely using RI#2 and RI#1 to locate the original un-binned data
>   which fall in the blob, performing an average over the data).
> 5. Pick the blob which is at the lower-right, and is large enough,
>   etc. The criteria you use here can be quite flexible, assuming the
>   "blobs" always arrive in the same pattern. You might even choose
>   just to exclude certain blobs that have a given shape and relative
>   position, and then take everything else.

> 6. Find the bins which belong to the chosen blob(s), using RI#2, and
>   then locate the data points within these original bins, with RI#1.
> 7. Give yourself a raise.
>
> This is actually a very good exercise to try if you want to know
> everything about HISTOGRAM and REVERSE_INDICES.
>
> JD