Subject: help needed to make the program run faster
Posted by gunvicsin11 on Fri, 09 Sep 2016 06:04:39 GMT
View Forum Message <> Reply to Message

Hi all,
I need to read 500 fits files and do analysis for all this,

So im doing like this,

```
file=file_search('*.fts')
nn=n_elements(file)
for ii=0,nn-1 do begin
img=readfits(file(ii),h)
----
---some analysis----

endfor
end
```

in the analysis part also i have some for loops so the program takes so much time to process this job.

So can anybody let me know whether any other faster methods are there to do this.

thanking you

Subject: Re: help needed to make the program run faster
Posted by Markus Schmassmann on Fri, 09 Sep 2016 09:11:57 GMT
View Forum Message <> Reply to Message

On 09/09/2016 08:04 AM, sid wrote:
> I need to read 500 fits files and do analysis for all this,
>
> So im doing like this,
>
> file=file_search('*.fts')
> nn=n_elements(file)
> for ii=0,nn-1 do begin
> img=readfits(file(ii),h)
> ----
> ---some analysis----
>
> endfor
> end
>
> in the analysis part also i have some for loops so the program takes so much time to process this job.

>
> So can anybody let me know whether any other faster methods are there to do this.
Hi Sid,

- use PROFILER and/or TIC & TOC to figure out what part of your code is slow
- remove loops by vectorising
- if all fits-images have the same dimensions and header structures you
can put all into one array and then do analysis on all images at once, e.g.:

```
file=file_search('*.fts')
nn=n_elements(file)
img0=readfits(file(0),h0)
img=fltarr([size(img0,/dim),nn])
img[*,*,0]=temporary(img0)
h=strarr([size(h0,/dim),nn])
h[*,0]=temporary(h0)
for i=1,nn-1 do begin
 img[*,*,i]=readfits(file(i),hi)
 h[*,i]=hi
endfor
```

---some analysis----

- not knowing what analysis you do it is difficult to tell how to speed
it up, but using, WHERE, SORT, UNIQ, HISTOGRAM, VALUE_LOCATE and the
like sometimes makes it a lot faster

Good luck, Markus

[1] http://www.harrisgeospatial.com/docs/PROFILER.html
[2] http://www.idlcoyote.com/code_tips/slowloops.html
[3] http://www.harrisgeospatial.com/docs/WHERE.html
[4] http://www.harrisgeospatial.com/docs/SORT.html
[5] http://www.harrisgeospatial.com/docs/UNIQ.html
[6] http://www.harrisgeospatial.com/docs/HISTOGRAM.html
[7] http://www.harrisgeospatial.com/docs/VALUE_LOCATE.html

Subject: Re: help needed to make the program run faster
Posted by Craig Markwardt on Fri, 09 Sep 2016 14:35:38 GMT
View Forum Message <> Reply to Message

On Friday, September 9, 2016 at 2:04:57 AM UTC-4, sid wrote:
> Hi all,
> I need to read 500 fits files and do analysis for all this,
>
> So im doing like this,

>
> file=file_search('*.fts')
> nn=n_elements(file)
> for ii=0,nn-1 do begin
> img=readfits(file(ii),h)
> ----
> ---some analysis----
>
> endfor
> end
>
> in the analysis part also i have some for loops so the program takes so much time to process this job.
>
> So can anybody let me know whether any other faster methods are there to do this.

This part of the loop should take very little. Try it yourself. Just remove the "analysis" part, and keep the FOR loop and READFITS() parts. You should see it completes pretty quickly. Or if it doesn't, it means you have a lot of data, and you will have to live with it (or get a faster computer and/or hard drive).

This will help you focus your efforts on the analysis parts. There, of course, you will want to vectorize as much as possible.

Craig

Subject: Re: help needed to make the program run faster
Posted by wlandsman on Fri, 09 Sep 2016 14:39:55 GMT
View Forum Message <> Reply to Message

On Friday, September 9, 2016 at 5:12:00 AM UTC-4, Markus Schmassmann wrote:

> for i=1,nn-1 do begin
> img[*,*,i]=readfits(file(i),hi)
> h[*,i]=hi
> endfor
>

One speed tip is to not use the asterisk above and write it as

```
for i=1,nn-1 do begin
 img[0,0,i]=readfits(file[i],hi)
 h[0,i]=hi
endfor
```

http://www.idlcoyote.com/code_tips/asterisk.html

Wayne

Subject: Re: help needed to make the program run faster
Posted by gunvicsin11 on Mon, 12 Sep 2016 03:58:36 GMT
View Forum Message <> Reply to Message

On Friday, September 9, 2016 at 2:42:00 PM UTC+5:30, Markus Schmassmann wrote:
> On 09/09/2016 08:04 AM, sid wrote:
>> I need to read 500 fits files and do analysis for all this,
>>
>> So im doing like this,
>>
>> file=file_search('*.fts')
>> nn=n_elements(file)
>> for ii=0,nn-1 do begin
>> img=readfits(file(ii),h)
>> ----
>> ---some analysis----
>>
>> endfor
>> end
>>
>> in the analysis part also i have some for loops so the program takes so much time to process this job.
>>
>> So can anybody let me know whether any other faster methods are there to do this.
> Hi Sid,
>
> - use PROFILER and/or TIC & TOC to figure out what part of your code is slow
> - remove loops by vectorising
> - if all fits-images have the same dimensions and header structures you
> can put all into one array and then do analysis on all images at once, e.g.:
>
> file=file_search('*.fts')
> nn=n_elements(file)
> img0=readfits(file(0),h0)
> img=fltarr([size(img0,/dim),nn])
> img[*,*,0]=temporary(img0)
> h=strarr([size(h0,/dim),nn])
> h[*,0]=temporary(h0)
> for i=1,nn-1 do begin
> img[*,*,i]=readfits(file(i),hi)
> h[*,i]=hi
> endfor
>
> ---some analysis----

>
> - not knowing what analysis you do it is difficult to tell how to speed
> it up, but using, WHERE, SORT, UNIQ, HISTOGRAM, VALUE_LOCATE and the
> like sometimes makes it a lot faster
>
> Good luck, Markus
>
>
> [1] http://www.harrisgeospatial.com/docs/PROFILER.html
> [2] http://www.idlcoyote.com/code_tips/slowloops.html
> [3] http://www.harrisgeospatial.com/docs/WHERE.html
> [4] http://www.harrisgeospatial.com/docs/SORT.html
> [5] http://www.harrisgeospatial.com/docs/UNIQ.html
> [6] http://www.harrisgeospatial.com/docs/HISTOGRAM.html
> [7] http://www.harrisgeospatial.com/docs/VALUE_LOCATE.html

Thanks for the info,
Actually the main problem im facing is im using where function and
for example if im searching where(image(*,i) gt threshold,count=c)
for some rows counts will be zero,
so in that case im using if statement, that way my program becomes much slow.
Is there any way to get out of this problem.

thanks

Subject: Re: help needed to make the program run faster
Posted by Markus Schmassmann on Mon, 12 Sep 2016 08:52:32 GMT
View Forum Message <> Reply to Message

On 09/12/2016 05:58 AM, sid wrote:
> On Friday, September 9, 2016 at 2:42:00 PM UTC+5:30, Markus Schmassmann wrote:
>> On 09/09/2016 08:04 AM, sid wrote:
>>> I need to read 500 fits files and do analysis for all this,
>>>
>>> So im doing like this,
>>>
>>> file=file_search('*.fts')
>>> nn=n_elements(file)
>>> for ii=0,nn-1 do begin
>>> img=readfits(file(ii),h)
>>> ----
>>> ---some analysis----
>>>
>>> endfor
>>> end
>>>
>>> in the analysis part also i have some for loops so the program takes so much time to process

this job.
>>>
>>> So can anybody let me know whether any other faster methods are there to do this.
>> - use PROFILER and/or TIC & TOC to figure out what part of your code is slow
>> - remove loops by vectorising
>> - if all fits-images have the same dimensions and header structures you
>> can put all into one array and then do analysis on all images at once, e.g.:
>>
>> file=file_search('*.fts')
>> nn=n_elements(file)
>> img0=readfits(file(0),h0)
>> img=fltarr([size(img0,/dim),nn])
>> img[*,*,0]=temporary(img0)
>> h=strarr([size(h0,/dim),nn])
>> h[*,0]=temporary(h0)
>> for i=1,nn-1 do begin
>>   img[*,*,i]=readfits(file(i),hi)
>>   h[*,i]=hi
>> endfor
>>
>> ---some analysis----
>>
>> - not knowing what analysis you do it is difficult to tell how to speed
>> it up, but using, WHERE, SORT, UNIQ, HISTOGRAM, VALUE_LOCATE and the
>> like sometimes makes it a lot faster
>>
>> Good luck, Markus
>>
>>
>> [1] http://www.harrisgeospatial.com/docs/PROFILER.html
>> [2] http://www.idlcoyote.com/code_tips/slowloops.html
>> [3] http://www.harrisgeospatial.com/docs/WHERE.html
>> [4] http://www.harrisgeospatial.com/docs/SORT.html
>> [5] http://www.harrisgeospatial.com/docs/UNIQ.html
>> [6] http://www.harrisgeospatial.com/docs/HISTOGRAM.html
>> [7] http://www.harrisgeospatial.com/docs/VALUE_LOCATE.html
>
> Thanks for the info,
> Actually the main problem im facing is im using where function and
> for example if im searching where(image(*,i) gt threshold,count=c)
> for some rows counts will be zero,
> so in that case im using if statement, that way my program becomes much slow.
> Is there any way to get out of this problem.
depending on the analysis you make, you can use

image[where(image(*,i) gt threshold,/null),i]

which is !null for rows with count 0. If you can get it work without

throwing an error, you should be fine

Subject: Re: help needed to make the program run faster
Posted by Jeremy Bailin on Thu, 22 Sep 2016 21:17:34 GMT
View Forum Message <> Reply to Message

On Sunday, September 11, 2016 at 10:58:46 PM UTC-5, sid wrote:
> On Friday, September 9, 2016 at 2:42:00 PM UTC+5:30, Markus Schmassmann wrote:
>> On 09/09/2016 08:04 AM, sid wrote:
>>> I need to read 500 fits files and do analysis for all this,
>>>
>>> So im doing like this,
>>>
>>> file=file_search('*.fts')
>>> nn=n_elements(file)
>>> for ii=0,nn-1 do begin
>>> img=readfits(file(ii),h)
>>> ----
>>> ---some analysis----
>>>
>>> endfor
>>> end
>>>
>>> in the analysis part also i have some for loops so the program takes so much time to process this job.
>>>
>>> So can anybody let me know whether any other faster methods are there to do this.
>> Hi Sid,
>>
>> - use PROFILER and/or TIC & TOC to figure out what part of your code is slow
>> - remove loops by vectorising
>> - if all fits-images have the same dimensions and header structures you
>> can put all into one array and then do analysis on all images at once, e.g.:
>>
>> file=file_search('*.fts')
>> nn=n_elements(file)
>> img0=readfits(file(0),h0)
>> img=fltarr([size(img0,/dim),nn])
>> img[*,*,0]=temporary(img0)
>> h=strarr([size(h0,/dim),nn])
>> h[*,0]=temporary(h0)
>> for i=1,nn-1 do begin
>> img[*,*,i]=readfits(file(i),hi)
>> h[*,i]=hi
>> endfor
>>
>> ---some analysis----

>>
>> - not knowing what analysis you do it is difficult to tell how to speed
>> it up, but using, WHERE, SORT, UNIQ, HISTOGRAM, VALUE_LOCATE and the
>> like sometimes makes it a lot faster
>>
>> Good luck, Markus
>>
>>
>> [1] http://www.harrisgeospatial.com/docs/PROFILER.html
>> [2] http://www.idlcoyote.com/code_tips/slowloops.html
>> [3] http://www.harrisgeospatial.com/docs/WHERE.html
>> [4] http://www.harrisgeospatial.com/docs/SORT.html
>> [5] http://www.harrisgeospatial.com/docs/UNIQ.html
>> [6] http://www.harrisgeospatial.com/docs/HISTOGRAM.html
>> [7] http://www.harrisgeospatial.com/docs/VALUE_LOCATE.html
>
> Thanks for the info,
> Actually the main problem im facing is im using where function and
> for example if im searching where(image(*,i) gt threshold,count=c)
> for some rows counts will be zero,
> so in that case im using if statement, that way my program becomes much slow.
> Is there any way to get out of this problem.
>
> thanks

Sometimes you can use masks instead of where to speed things up... but again, it depends what exactly you are doing with it. For example, to double every pixel that is greater than the threshold:

positive_mask = image gt threshold
image += image * positive_mask

-Jeremy.