

Exiv2 - Bug #1244

exiv2 without EXV_HAVE_MMAP throws an exception

15 Oct 2016 08:58 - Robin Mills

Status:	Closed	Start date:	15 Oct 2016
Priority:	Normal	Due date:	
Assignee:	Robin Mills	% Done:	100%
Category:	tiff parser	Estimated time:	6.00 hours
Target version:	0.26		
Description http://dev.exiv2.org/boards/3/topics/2756			
Related issues: Related to Exiv2 - Feature #1245: Better I/O implementation when EXV_HAVE_MMA... Assigned 17 Oct 2016			

Associated revisions

Revision 4633 - 15 Oct 2016 09:04 - Robin Mills

#1244 Fix submitted.

Revision 4637 - 17 Oct 2016 16:32 - Robin Mills

#1244. Removing experimental APIs introduced by r4637. I submitted those APIs just to retain the code somewhere. I have no plan to release such as API.

Revision 4638 - 17 Oct 2016 19:24 - Robin Mills

#1244. Correction to r4637. Added bigBlock_(NULL) to Basiclo::Basiclo().

Revision 4639 - 18 Oct 2016 06:35 - Robin Mills

#1244 Fix crwimage.cpp to read into memory (to make CRW work with Remotelo).

History

#1 - 15 Oct 2016 09:39 - Robin Mills

- Status changed from Resolved to Closed

Fix submitted: [r4633](#)

I've successfully run the test suite with EXV_HAVE_MMAP unset in include/exiv2/config.h:

```
#undef EXV_HAVE_MMAP
#undef EXV_HAVE_MUNMAP
//
// That's all Folks!
#endif // _CONFIG_H_
```

Time for test suite (without MMAP):

```
real    1m11.472s
user    0m28.147s
sys     0m36.597s
```

Time for test suite (with MMAP):

```
656 rmills@rmillsmbp:~/gnu/exiv2/trunk $ time make tests >/dev/null
```

```
real    1m4.627s
user    0m27.299s
sys     0m34.565s
657 rmills@rmillsmbp:~/gnu/exiv2/trunk $
```

I'm not surprised that the time is similar. The test suite does not have large files. However the difference 5X when reading a 20mb .NEF

Without MMAP:

```
$ time exiv2 -pa --grep Software DSC_0002.NEF
Exif.Image.Software          Ascii          10 Ver.1.00

real    0m0.068s
user    0m0.007s
sys     0m0.036s
```

With MMAP:

```
$ time exiv2 -pa --grep Software DSC_0002.NEF
Exif.Image.Software          Ascii          10 Ver.1.00

real    0m0.015s
user    0m0.006s
sys     0m0.005s
```

#2 - 15 Oct 2016 10:41 - Robin Mills

We should not read the whole file when EXV_HAVE_MMAP is not in use. There is code in the Remotelo class called a "Block Map". In Remotelo, we wanted to avoid reading the whole file. To achieve that, we allocate a large block of memory which is sufficient to hold the complete file - however it is not populated. We maintain a parallel map with one boolean for every "block" (of 8k or so). When we read, or write, we consult the blockmap and populate the memory block *just in time*. I'm confident that it is straightforward to promote the block map from Remotelo to Basiclo and use this strategy in Filelo. This will make a huge difference to the amount of reading being performed.

When this is done, we should also pay attention to the size of the "big block". There is no need to allocate a block to hold the complete file. We can realloc that block when necessary.

I am confident that we can make a huge improvement to the I/O and Memory demands of class Filelo when EXV_HAVE_MMAP is not set. No changes are required in the TiffXxxx classes as all of this activity will be performed invisibly within Basiclo.

The project to use the "BlockMap" within class Filelo cannot be undertaken for v0.26 as it involves too much work and risk at a very late stage in the project.

#3 - 20 Oct 2016 18:26 - Robin Mills

Discussion with Asdiel Echevarria

We really like your idea and implementation for reading only the metadata blocks while still using File I/O and we are thinking to try to back port it to 0.26 once 0.26 is released. We will of course share it back in the repository in case you guys do a release between 0.26 and 0.27.

My reply:

I've backported the necessary code from v0.26 to v0.25. The changes to make that happen are mostly in src/*image.cpp and src/basiclo.cpp (and their .hpp companions). It's not as trivial as I say because you have to update the build and other consequential magic. There are new files in v0.26 (src/webpimage.cpp, src/ini.cpp). However I've done everything in about two hours. It builds and executes the v0.25 test suite without crashing. The test suite reports various matters which have been fixed in v0.26. The formatted output from the command exiv2 -pS is slightly different in v0.26. For certain this is sufficient to be sent to your test/QE people. <http://clanmills.com/exiv2/exiv2-0.25+.tar.gz> and I attach a patch for v0.25.

It reads TIFFs over the internet very efficiency. I added instrumentation to HttpIo to see the blocks being fetched. 11 blocks of 1024bytes.

```
1052 rmills@rmillsmbp:~/gnu/exiv2 $ ssh secret@clanmills.com ls -alt www/files/Reagan.tiff
-rw-r--r-- 1 secret secret 8628164 Oct 16 10:45 www/files/Reagan.tiff
1053 rmills@rmillsmbp:~/gnu/exiv2/v0.25/build $ bin/Debug/exiv2 -pa --grep Software http://clanmills.com/files
/Reagan.tiff
HttpIo::HttpImpl::getDataByRange: 0,0
HttpIo::HttpImpl::getDataByRange: 8416,8416
HttpIo::HttpImpl::getDataByRange: 8417,8417
HttpIo::HttpImpl::getDataByRange: 8418,8418
HttpIo::HttpImpl::getDataByRange: 8419,8422
HttpIo::HttpImpl::getDataByRange: 8423,8425
Exif.Image.Software          Ascii          29 Adobe Photoshop CS Macintosh
1054 rmills@rmillsmbp:~/gnu/exiv2/v0.25/build $
```

If/When you make the changes for the network drive, I will be very happy to accept a patch. I'll review and test it, then put it on the trunk after v0.26 has shipped. From my point of view, there is no hurry at all with this.

Incidentally, I pulled down all the raw images yesterday from here: <https://www.rawsamples.ch/index.php/en/> Exiv2 reads all 322 without a single stumble when they are on local storage. <https://www.rawsamples.ch/index.php/en/> The project for 2017 to enhance our raw image support and test will investigate that every image can be read efficiently over the internet. I'm planning to recruit a Google Summer of Code student for that project. So it would be good to have your patch by May 2017.

```
529 rmills@rmillsmbp:~/gnu/exiv2/trunk $ time build/bin/Debug/exiv2 -pa -g Software http://clanmills.com/files
/Reagan.tiff
```

