

Exiv2 - Feature #1245

Better I/O implementation when EXV_HAVE_MMAP is not set

17 Oct 2016 15:33 - Robin Mills

Status: Assigned	Start date: 17 Oct 2016
Priority: Normal	Due date:
Assignee: Robin Mills	% Done: 0%
Category: design	Estimated time: 0.00 hour
Target version: 0.28	
Description See discussion in #1244 .	
Related issues:	
Related to Exiv2 - Bug #1244: exiv2 without EXV_HAVE_MMAP throws an exception	Closed 15 Oct 2016
Related to Exiv2 - Feature #992: Better raw file support and test	Assigned 18 Sep 2014

History

#1 - 20 Oct 2016 18:24 - Robin Mills

- File *v025plus.patch.zip* added

- Status changed from New to Assigned

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/basicio.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-user-name@clanmills.com ls -alt www/files/Reagan.tiff
-rw-r--r-- 1 clanmill clanmill 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
Exif.Image.Software          Ascii          29  Adobe Photoshop CS Macintosh

real    0m1.582s
user    0m0.014s
```

```

sys      0m0.012s
530 rmills@rmillsmbp:~/gnu/exiv2/trunk $ time curl -O http://clanmills.com/files/Reagan.tiff
% Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
           Dload  Upload   Total   Spent    Left   Speed
100 8425k  100 8425k    0     0  787k      0  0:00:10  0:00:10  --:--:-- 1601k

real      0m10.745s
user      0m0.074s
sys       0m0.319s
531 rmills@rmillsmbp:~/gnu/exiv2/trunk $ ls -alt Reagan.tiff
-rw-r--r--+ 1 rmills staff 8628164 Oct 21 12:00 Reagan.tiff
532 rmills@rmillsmbp:~/gnu/exiv2/trunk $

```

#2 - 01 Nov 2016 08:39 - Robin Mills

Here's a discussion with Asdiel that I would like to share with Hanno who is interested in fuzzing in [#1248](#).

I like your idea to use MMAP on local drives. I've not so enthusiastic about class Networklo because I think there is a simpler way! There's an option "useCurl" in the ImageFactory. Perhaps you could expand that from a simple boolean to an enum. Something like:

```

typedef enum { kifNone = 0x0 , kifUseCurl = 0x01, kifUseMMAP = 0x02 , kifUseOtherMagic = 0x04 } ImageFactoryOption;

```

You'll need code in Filelo which I've already added to Remotelo to read the blockMap. However try to avoid cutting'n'pasting code. Promote code in the class hierarchy where possible.

How the get from here to happiness?

As you know, getting a product to ship is a tough time. I've fixed TiffImage/Remotelo by calling TiffImage::printStructure(kpsRecursive) in TiffImage::readMetadata(). I'll have to implement CrwImage::printStructure(). Gosh, our project to review and strengthen RawImage support is scheduled for v0.27. If I pursue every dark hole, I'll be here forever!

My agenda is to finish v0.26 and do not want to be diverted into making those changes in Filelo. However, you can do the work and test it. I'll accept (review and test) your patch and put it on the trunk AFTER v0.26 has shipped.

I didn't write Exiv2, so there are areas of the code that are unfamiliar. I don't remember how the image factory works, however you can make your own AlienFactory from which you may instance Alienlo classes and/or our classes such as Filelo classes.

However, rather than start a parallel universe, I feel it would be better if you build on the excellent foundations of Andreas and Brad. Build on Filelo and share the code for everybody to use! After all, you get this magic and a lot of my time for nothing. Share with the community - that's the open-source way!

Files

v025plus.patch.zip	41 KB	20 Oct 2016	Robin Mills
--------------------	-------	-------------	-------------