Exiv2 - Bug #977

Tagging RAW images for Canon EOS-1Ds corrupts them

06 Aug 2014 13:19 - Kory Roberts

Status:	Assigned		**Start date:**	06 Aug 2014
Priority:	Normal		**Due date:**	
Assignee:	Robin Mills		**% Done:**	10%
Category:	metadata		**Estimated time:**	10.00 hours
Target version:	0.28			

Description

I first reported to digiKam, but they directed me here since this appears to be the library at issue. I'm sorry for not digging into the relevant command line inputs, but hopefully testing can be done from what's provided here. This issue seems pretty critical since RAW files are completely destroyed and unrecoverable simply by trying to tag them.

.....

I have digital RAW images taken with a Canon EOS-1Ds (mark 1, original). Tagging these images in digiKam corrupts them.

When these files have a .TIF extension, tagging always corrupts them. When these files have a .RAW extension and "If possible write Metadata to RAW files (experimental)" in the digiKam settings is unchecked, sidecar XMP files are created and the originals spared.

When corruption occurs, what should be 9-11 MB files are resaved as 280ish KB files (embedded thumbnail?). No warnings or any other obvious signs hint that this has occurred.

A sample raw image for testing can be downloaded:

https://drive.google.com/file/d/0ByXO8US0zFPicXFYVV9ISDY0Rlk/edit?usp=sharing

History

#1 - 06 Aug 2014 13:25 - Kory Roberts

I might mention that, for some bizarre reason, Canon gave these raw images a .TIF extension which has causes all kinds of issues along the way! Usually renaming to .RAW extension solves minor issues, such as an image program correctly identifying the image as RAW format, rather than a true TIF (and only "seeing" the small embedded thumbnail).

This issue was the final straw for me personally (thank goodness I did have backups!!!), so I'm now in the process of converting my entire library of older 1Ds RAW photos to DNG. Even to do that with digiKam's converter, I'm having to first rename from .TIF to .RAW.

#2 - 06 Aug 2014 16:26 - Robin Mills

- Status changed from New to Assigned

- Assignee set to Robin Mills

- Priority changed from Urgent to Normal

- Target version set to 0.25

I agree that this is a serious bug. The library should never destroy/corrupt an image. Thanks for providing the test image. I will look at this tomorrow.

It sounds as though your work-around is to convert all images to DNG and use that exclusively in your work-flow. Can you confirm that is working correctly? Adobe provide a free downloadable convertor to create DNGs. I believe it can operate on a directory of images, so it's probably easy to import images from your camera and convert them immediately.

Please be aware that even if this is fixed tomorrow in the library, the fix will probably not be available until included in a future version of DigiKam. We hope to release Exiv2 v0.25 in November and it could be early 2015 when available in DigiKam. So, if your workflow based on DNG is working, you will be using that for some time to come and will probably never revert to a pure RAW workflow. Once you live in the world of DNG, you will probably be using other tools are also broken in RAW format. So when you adopt DNG, you'll probably be very happy.

#3 - 06 Aug 2014 17:03 - Kory Roberts

Thanks, Robin. Yes, I'm aware that bug-fixes can take some time to propagate into future releases, but appreciate your explanation. Basically, I'm hoping to save someone else from completely fubaring their 1Ds raw archives! I think Canon really messed up and was not forward-thinking in the way they named (TIF???) and formatted these files.

I've been moving forward today in converting all of these files to DNG, careful to convert before accidentally trying to tag first. I have tested conversion (digiKam converter) and keyword/metadata application (digiKam) with these DNG files. Other than having to first rename from .TIF to .RAW extension, all appears to be working as expected.

I will hold back a few unaltered originals (possibly renamed?) in case these are needed for further testing. I was unsuccessful in locating any others online to download.

Thanks.

#4 - 06 Aug 2014 18:10 - Robin Mills

Kory

I've downloaded you image and the latest Adobe DNG convertor.

I've reproduced the corruption you describe. I think your guess that we have removed the image and only retained the thumbnail is probably correct. Exiv2 operates on the file with either extension .tif or .raw. The .tif extension may have implications for image recognition in DigiKam or other applications, however the Exiv2 library works (equally badly) with either.

I've converted the image to DNG and applied the command from the Exiv2 man page:

```
614 rmills@rmillsmbp:~/Pictures/DNG $ exiv2 -M"set Exif.Photo.UserComment charset=Ascii Comment added by rmill
s" bug.dng
Error: Directory Canon with 25665 entries considered invalid; not read.
Error: Directory Canon with 25665 entries considered invalid; not read.
615 rmills@rmillsmbp:~/Pictures/DNG $ exiv2 -pa bug.dng  | grep -i rmills
Error: Directory Canon with 25665 entries considered invalid; not read.
Exif.Photo.UserComment                    Undefined  31  Comment added by rmills
616 rmills@rmillsmbp:~/Pictures/DNG $ ls -alt *.dng
-rw-r--r--  1 rmills  staff  9311394  7 Aug 02:05 bug.dng
617 rmills@rmillsmbp:~/Pictures/DNG $
```

That seems to be working OK. I will investigate the warnings.

However to return focus to your original report, I will step the code tomorrow in the debugger. It's 2am here in England and time to go to bed. Maybe the simple fix is to report "Unsupported file format. Please convert to DNG.". Doing nothing is less harmful than corrupting files.

#5 - 08 Aug 2014 22:48 - Kory Roberts

Robin Mills wrote:

> Doing nothing is less harmful than corrupting files.

I am in full agreement. I think many--including me!!!--would consider tagging a few keywords to an image a very benign action, and might not take the necessary precautions to safeguard the image against corruption. I lost a whole year's worth of raw images before I caught what was going on. I did have backups though...thank goodness!

#6 - 09 Aug 2014 05:18 - Robin Mills

We agree. Corrupting files is not acceptable.

Good News. I've found a fix which will cause exiv2.exe to report:

```
C:\>\Users\rmills\gnu\exiv2\trunk\msvc2005\bin\x64\DebugDLL\exiv2.exe -M"set Exif.Photo.UserComment charset=As
cii Comment added by rmills" c:\20030803
-021.TIF
Warning: Unsupported date format
Warning: Unsupported time format
Exiv2 exception in modify action for file c:\20030803-021.TIF:
c:\20030803-021.TIF: No image found in file.  Try converting to DNG

C:\>
```

There are three matters to deal with before committing the fix:
1) This disturbs a couple of our standard tests and I want to investigate those before committing a fix.
2) The "Unsupported data/time format" message deserves investigation, although I think that's something in your file and unrelated to this issue.
3) I'd like to discuss my fix with Andreas, our Project Lead Engineer.

When resolved, I would like to add a test to our test harness. Do you have a smaller file than the file you have provided.

Just for the record, here is my current patch:

```
505 rmills@rmills-laptop:~/gnu/exiv2/trunk $ svn diff
Index: src/error.cpp
===================================================================
--- src/error.cpp       (revision 3288)
+++ src/error.cpp       (working copy)
@@ -105,7 +105,8 @@
        { 49, N_("TIFF directory %1 has too many entries") }, // %1=TIFF directory name
        { 50, N_("Multiple TIFF array element tags %1 in one directory") }, // %1=tag number
        { 51, N_("TIFF array element tag %1 has wrong type") }, // %1=tag number
-       { 52, N_("%1 has invalid XMP value type `%2'") } // %1=key, %2=value type
+       { 52, N_("%1 has invalid XMP value type `%2'") }, // %1=key, %2=value type
+       { 53, N_("%1: No image found in file.  Try converting to DNG") }, // %1=path
     };

 }
Index: src/tiffimage.cpp
===================================================================
--- src/tiffimage.cpp   (revision 3288)
+++ src/tiffimage.cpp   (working copy)
@@ -198,6 +198,8 @@

     void TiffImage::writeMetadata()
     {
+            // Issue: #977.  Don't re-write TIFs if there is no image!
+            if ( !pixelHeight_ || !pixelWidth_ ) throw Error(53, io_->path());
 #ifdef DEBUG
        std::cerr << "Writing TIFF file " << io_->path() << "\n";
 #endif
506 rmills@rmills-laptop:~/gnu/exiv2/trunk $
```

#7 - 09 Aug 2014 08:28 - Kory Roberts

Robin Mills wrote:

> Do you have a smaller file than the file you have provided.

Since these are RAW images there are limits to how small, but I've discovered those with a lot of black tend to be smaller. I've uploaded one that is just over 5 MB. It has been renamed from the original file name, now with a .raw extension, but I think the EXIF information might provide the original file name, if that even matters.

https://drive.google.com/file/d/0ByXO8US0zFPiOVdjLXRzRU5PRmc/edit?usp=sharing

#8 - 09 Aug 2014 09:27 - Robin Mills

Thanks for the smaller file.

The video test files are downloaded on demand using svn. We only download the file once of course, not on every test. We had a complaint about our test directory being too fat, and we put it on a diet by storing the video and eps files in a different SVN branch and downloading them on demand. When I add a test for your file, I'll use that mechanism.

I regret that your files were corrupted and I understand the inconvenience. I believe Exiv2 to be solid and reliable code. I didn't write it - however I've contributed to this project for 6 years, and I don't recall anybody saying their files had been corrupted. But, you know how it is with software, we can't know what we don't know and RAW files are a software standards nightmare. Maybe we should tighten up and recommend DNG whenever it's not JPEG/PNG/GIF/BMP - then we move responsibility to the Adobe guys!

Anyway, I hope we get this fixed in the trunk in the next few days and it'll be in v0.25 which we hope to ship in November 2014.

Robin

#9 - 18 Aug 2014 05:40 - Phil Harvey

I encountered this same problem in ExifTool 9 years ago. My solution was to specifically test for the 1D magic bytes at offset 8 in all TIFF images (0xba 0xb0 0xac 0xbb), and if found treat this as a 1D RAW file instead of a TIFF. (ExifTool does not support writing a 1D RAW file, so the response is to issue an error when attempting to write a file like this.)

#10 - 18 Aug 2014 07:23 - Robin Mills

Thanks, Phil.
If that's good enough for ExifTool, it's good enough for Exiv2. I'll incorporate your strategy in our code.

Kory
I still have to investigate the tests which failed with my fix above and resolve that before committing a fix to the trunk.

#11 - 09 May 2015 08:18 - Robin Mills

- Target version changed from 0.25 to 0.26

Deferred to v0.26. Insufficient time to deal with this for v0.25.

#12 - 23 May 2015 08:39 - Robin Mills

- Assignee deleted (Robin Mills)

#13 - 12 Oct 2015 09:02 - Robin Mills

- Assignee set to Robin Mills

#14 - 04 Jan 2016 10:47 - Robin Mills

- Estimated time set to 10.00 h

I haven't looked at this for a long time. However, I'm adding a SWAG to the outstanding tasks for v0.26 to attempt to understand how much work remains.

#15 - 04 Jan 2016 11:37 - Robin Mills

- % Done changed from 0 to 10

#16 - 13 Jun 2016 06:23 - Robin Mills

- Target version changed from 0.26 to 0.28

Deferred to v0.27 to make the schedule for v0.26.