Volume Editor

Hans Weghorn
Faculty of Mechatronics
BA-University of Cooperative Education, Stuttgart
Germany

Proceedings of the 4th Annual Meeting on
Information Technology and Computer Science – ITCS, Volume 2008
Stuttgart, Germany, February 2008
Hans Weghorn (Ed.)

Printed in Stuttgart, Germany

ISSN 1614-2519

The RAW Image File Format Problem - Applications of Digital SLR Cameras in Astronomy and Science

Thilo Bauer

bluewater multimedia concepts e.K., Kölnstraße 191, D-53757 St. Augustin, Germany
tbauer@astro.bwmc.net

Abstract. This work describes the application of digital single lens reflex (SLR) cameras in astronomy science. A definition of the term raw file format is given. As with formats like FITS or TIFF, the CR2 file format is explained beeing a disclosed, and well documented raw file format. Results from raw image file decomposition with different software tools, and a comparison of imaging with a digital SLR camera and a professional grade CCD detector are presented. The analysis of noise and image histograms will help to recover problems with raw file format decomposition tools and their adjustment. Without doubt, digital SLR cameras are suited for special scientific applications. Further implications from this work are a detailed analysis of noise characteristics to achieve high quality imaging at very low light levels from raw files taken with digital SLR cameras.

1 Introduction

Scientific imaging at low light levels require imaging sensors with a high linearity and low noise characteristics. The most common applications within the optical astronomy are stellar photometry, the measurement of light intensities, and astrometry, the measurement of positions of stars, planets, quasi stellar objects or complex star formations. Similar requirements of measuring intensities and positions at low light level will be found in optical microscopy, biology, pathology, physics and other sciences. Popular imaging sensors are divided into three sensor technologies: monochrome CCD sensors, CMOS cameras, partially using color filter arrays to distiguish different wavelengths, and a technology often referred as the Foveon technology [1]. The latter two technologies form the basic sensors of modern digital SLR color cameras. The Canon EOS digital camera models are based on proprietary CMOS sensors. A Bayer color filter array in front of the silicon chip separates different colors into red, green and blue pixel values (RGB).

A raw file format herewith shall be defined as containing a two dimensional set of digital numbers obtained from a given imaging sensor, obtained lossless, without any preprocessing, and having a well defined value range and bit representation as yielded by the digitization unit. Scientific analysis furthermore requires linear response or at least a detector response which allows conclusions about the true nature of the original signal.

The FITS file format is the de-facto standard within astronomy science across all physical wavelenghts [3]. FITS is provided and supported by the International Astronomical Union (IAU) and its FITS Working Group (IAU-FWG) as the international control authority [4,5]. One of the early adopted image and graphics standard file formats for general use is the Tagged Image File Format (TIFF) created by Adobe [6]. Similar to the FITS format, TIFF supports different pixel bit counts: 1 to 8, and 16 bit of linear dynamic range with binary, grey scale or false color images (color tables representing different gray values), and true color RGB content. At least three compression methods are provided with this specification. Manufacturers of commercial, digital SLR cameras tend to provide different approaches for compact and reliable file formats to the photographer, which presently leads to a zoo of file formats with every manufacturer. Proprietary raw file formats often are argued as beeing undisclosed, thus having harmful disadvantages with processing and long-term archiving [7]. Work is done for replacements, like the Adobe DNG specification [8]. Undocumented in the sense of having a format with details undisclosed to the public seems especially true for cameras like the Canon digital EOS cameras.

2 Imaging with Astronomical Telescopes

NGC 7635 is a catalogue name of a H-II region of interstellar gas, often referred as the *Bubble Nebula* [9], and invisible for the naked eye. The nebula was choosen as a test candidate of a very faint light source. The author took images with a Canon EOS 400D camera mounted on a Vixen VC200L Cassegrain 20cm telescope on Aug 4, 2007. Conditions were a few days after full bright moon taken from the authors garden and with typical light pollution from the moon and city light illumination. To achieve a higher dynamic range 30 single exposures were recorded at a temperature of +14°C. The camera provides an angular resolution of about 1.3 arcsec/pix at 3888x2588 pixel in the primary focus of the telescope. Having 12 bit dynamic range, and not to clip the intensities of bright stars, the camera was set to 30 seconds of exposure time at a gain representing 800 ASA. Internal pre-processing, noise reduction, color and gamma correction were switched off. The images were processed with modifications of a stacking method developed by the author [10,11]: automatic detection of field stars above 10 noise sigma, detection of the stellar positions, selection of images without position movement of the stars (due to turbulent atmosphere and guiding errors during exposure), and shift & add the selected individual images recorded. The resulting image yielded a recentered composite of 27 selected frames. As a comparison with the equipment explained above, a single image of NGC 7635 taken by the author with the 1m Cassegrain telescope at the Hoher List Observatory is presented. The image was recorded on Nov 3, 1993. Exposure time was 600 seconds with an interference filter blocking background light except the red hydrogen emission from the nebula. A scientific CCD imaging system from Astromed Inc., UK at 416x576 pixel resolution was used within this period providing a similar angular resolution per pixel with the focal reducer of the telescope. With cooling down to about -70°C, using slow scan mode, the camera achieved 16 bit dynamic range. Unfortunately this CCD imaging device was destroyed in the meanwhile due to an irreparable hardware defect.

3 First Results

A novel image processing software written in Java with native C libraries for fast image processing is presented with this work. Within the tool-chain the author used these tools for image decomposition and further image processing: the original software provided from the camera manufacturer[12], and DCRAW, Revision: 1.397, provided by Dave Coffin [13].

A quick reengineering of the Canon raw files showed, that the metadata extracted from the raw files represent a well formed TIFF structure with EXIF extension [14]. The TIFF compression tag identifies a JPEG compression method provided with the TIFF standard [6]. As JPEG is generally beleived as an image format with lossy compression, the original JPEG specification also defines lossless compression methods [15,16]. Unfortunately, not every image processing software supports all possible TIFF variants and derivatives, like EXIF. Moreover, and to make confusion perfect, Canon created a different file extension for the EXIF raw image files: CR2, officially named „Canon Raw, Version 2“, which is a TIFF derivate.

A code review of DCRAW with its many options confirms the use of several preprocessing techniques used for image decomposition. Compared to conventional photography, standard astronomical interpretation of colors, like stellar color photometry, relies on different rules and generally will avoid non-linearities, color and noise correction. The developer of DCRAW advises to use certain options only for obtaining linear and unprocessed image data [17]. Option „*-D*“ produces linear grayscale output representing the isolated R, G (2x) and B pixel values to be rearranged by the application software. The options „*-h -k 0 -r 1.0 1.0 1.0 1.0 -o 0 -4 -T*“ provides linear 16 bit TIFF output no additional processing at half of the pixel resolution of the sensor and yielding virtual RGB pixels. This is confirmed by the image analysis. Histograms show uniformly distributed values as it would be expected from distributing 12 bit values into 16 bit dynamic resolution from a multiplication by 16. Decomposition with the latter options will result in a color shift of about +/- 0.5 pixel in R and B colors at an angle of -45° due to the geometry of the Bayer matrix which is arranged as a square RGGB pixel pattern.

With the original Canon software the raw files have also been temporarily decomposed and converted into TIFF (16 bit, RGB). No options exist to control image decomposition and conversion into more appropriate formats like standard TIFF. From reading these TIFF files, unequal distributed digital numbers, noticable asymmetric, non-gaussian, and clipped histogram profiles have been found. The noise of decomposed dark frames show broadened histograms with about 5-10 times the noise at FWHM compared to the results of DCRAW from the same raw files. This indicates a complex, non-linear but not very successful pre-processing stage like non-linear gamma scaling and intensity clipping at the lower and bright end of the image intensities. This is a very surprising result, as even a conventional photographers would expect noise properties beeing at least as small as the camera will it provide raw and digital. Especially, if the user would expect a further image optimization stage from such a software tool.

4 Noise Properties, Sensitivity and the Effects of Correlated Noise Terms

Noise was estimated from histograms of single dark exposures. From a sample of the frames about 13 DU (digital units) at FWHM for the EOS 400D (30 seconds exposure, 12 bit, 4096 total units), and about 30 DU for the Astromed camera (600 seconds exposure, 16 bit, 65536 total units) were found respectively. As a result of a relatively larger read-out noise at less dynamic range, sky illumination (photon noise), and reduced camera sensitivity in the red light (~20-25% relative quantum efficiency at hydrogen emission), the faint red nebula degraded in contrast with the EOS 400D compared to the result with the scientific camera. The pixel dimension of the EOS camera is smaller compared to the scientific camera with in gaps between different color pixels, and reduced full-well capacity. However, both images show stars of nearly the same limiting stellar magnitude (faintest stars detected) with both camera and telescope setups. The limiting stellar magnitude is not a result of the aperture diameter of the telescopes, but related to the pixel dimension, illumination intensity, optical spread of point light sources, signal-to-noise ratio, and finally the focal length of the optics [18]. The ratio between the focal length and pixel dimension was nearly identical within this seemingly unequal comparison of two astronomical telescopes having very different aperture diameter.

Different noise models exist within astronomy depending on detectors and algorithms. However, a standard ISO procedure is defined to determine noise characteristics of digital SLR cameras and comparable imaging detectors [19,20]. First results from the astronomical observations indicate, that there remain additional and correlated noise terms close to and below read-out and fixed pattern noise level. Asymmetric (multiple gaussian) noise profiles from histograms of the EOS 400D are obviously caused by line bias variations within individual frames (temporal noise). As a preliminary result of this work, these terms can be corrected resulting in a slightly better S/N. Conversion of analog image intensities into digital numbers may introduce effects like deferred charge, hot sensor areas, bad pixels, periodic terms and so on [21]. Regularily distributed peaks at powers of two (e.g. distances of 2, 4, 8,...) indicate non-linear bit conversion errors from the histograms. This effect was observed from the files of the scientific camera, but invisible with raw files of the EOS 400D. Vertical, non-temporal bias variations remain after correction of line bias variations. Assuming constant pixel dimension and response is not applicable to real imaging detectors. The local light intensity response may vary over the whole pixel area within subpixel range. This is true for both devices, the digital SLR camera and the scientific camera. A Canon EOS 40D (received 26 Jan., 2008) digital SLR camera offers 14 bit of dynamic range from a new signal processing unit. With the same settings (t=30 sec, 800 ASA), dark exposures were taken at room temperature (T=22°C). Histograms again yield asymmetric profiles from dark frames having temporal line bias variation. However, a smaller total noise distribution of 27 DU of noise at FWHM was found. Thus, gain in S/N with this camera is about a factor of two compared to the EOS 400D model. Set to 400 ASA, the camera presents 17 DU noise at FWHM (T=1°C). Correction of line bias again yields a slightly better S/N. Gaussian distributed noise profiles from the compensated raw dark frames now tend to a S/N similar to that of the scientific camera. A detailed analysis of noise properties and a research for compensation of correlated noise terms is ongoing.

4 Critical Evaluation and Remarks

Concerns about disadvantages of camera raw file formats appear unsubstantiated under certain conditions, like those described within this paper. Some digital SLR cameras provide long-exposure imaging representing unprocessed sensor characteristics with disclosed raw file formats. This may imply some advantages like simplified use of cameras without the need for dedicated image recording hardware with some astronomical applications. Scientists would prefer high-end camera models operating at 14 bit or more dynamic range. A noticable remark is a surprisingly, more noisy result and additional non-linearities obtained from image decomposition with original software from camera manufacturers. This, however, is untolerable even for conventional photographers expecting high quality imaging at low light levels. Thus, the term raw image file format in no way will automatically be a guarantee, or a synonym for high quality imaging under any circumstance, as it possibly would be expected by the user. Instead, results of image decomposition may depend on the software tools and adjustments. Scientists are strongly advised to consider their hardware requirements, tools and techniques for the decomposition of the raw image files and before applying further image processing and analysis. Implications from this work are detailed measurements of noise properties, and algorithmic development to achieve high accuracy with astronomical observations. The algorithm used within this work to compensate temporal line bias noise seems robust, but cumbersome.

References

1. H. T. Hytti, Characterization of digital image noise properties based on RAW data, SPIE 2005
2. Joint Picture Expert Group, http://www.jpeg.org/ (Referred Jan, 7 2008)
3. Wells, D. C., Greisen, E. W., & Harten, R. H., FITS - a Flexible Image Transport System, A&A. Supp. Ser. V. 44, P. 363, 1981
4. http://fits.gsfc.nasa.gov/iaufwg/ (Referred Jan, 7th, 2008)
5. http://fits.gsfc.nasa.gov/fits_intro.html, Last revised: 12 September 2002 (Referred Jan 7, 2008)
6. Adobe Systems Inc., TIFF Revision 6.0, final, June 3, 1992
7. Website of the Open Raw Group, http://www.openraw.org/ (Referred Jan, 7 2008)
8. Adobe Systems Inc., Digital Negative (DNG) Specification, San Jose, Version 1.1.0.0, February 2005
9. Moore, B.D. et al.: 2002, Astron. J. 124, 3313
10. Bauer, T., Astron. Ges. Abstract Series, Vol. 9, p.201, 1993
11. Bauer, T., Mustererkennung in der Astronomie, SuW Sterne und Weltraum, Vol. 31, 12, p.774-778, 1992

12. Website of Canon Inc., http://www.canon.com, (referred, Jan 7, 2008)
13. Dave Coffin, http://cybercom.net/~dcoffin/dcraw/ (referred, Jan 7, 2008)
14. Exchangeable image file format for digital still cameras: Exif Version 2.2, Japan Electronic and Information Technology Association, JEITA CP-3451, April 2002
15. CCITT Recommendation T.81, Iinformation Technology – Digital Compression and Coding of Continous-Tone Still Image – Requirements And Guidelines, International Telecommunication Union (ITU), 09/92
16. E. Hamilton, JPEG File Interchange Format Version 1.02, September 1, C-Cube Microsystems, 1992
17. Dave Coffin, Jan 4, 2007, private email communication
18. E. H. Geyer, Bemerkungen zur Nachweisgrenze astronomischer Objekte bei erdgebundenen Beobachtungen, DFG Rundgespräch und Workshop, Bad Honnef Nov. 1984, (manuskript recieved by private communication)
19. ISO 14524: Photography – Electronic still-picture cameras – Methods for measuring opto-electronic conversion functions (OECFs), 1999
20. ISO 15739: Photography – Electronic still-picture imaging – Noise measurements, 2003
21. Steve B. Howell, Handbook of CCD Astronomy, Cambridge University Press, 2006, ISBN-13 978-0-521-61762-8