

DSLR 카메라의 RAW 파일 포맷 분석

노광현*

*한성대학교 산업시스템공학과

e-mail:khrho@hansung.ac.kr

An Analysis on RAW File Format of DSLR Cameras

Kwang-Hyun Ro*

*Dept of Industrial & System Engineering, Hansung University

요 약

최근 일반인들에게 널리 보급되고 있는 DSLR 디지털 카메라로 촬영되는 정지 이미지는 범용의 JPEG 포맷이나 제조사별 독자적인 RAW 파일 포맷으로 저장매체에 저장된다. 이와 함께 PMP 등의 휴대용 멀티미디어 기기를 활용하여 촬영된 RAW 파일 포맷을 재생 및 확인하기 위한 요구가 증대되고 있지만, RAW 파일 포맷이 디지털 카메라 제조사별로 다르고, 해당 포맷에 대한 구체적인 정보가 공개되지 않는 이유로 대부분의 휴대용 멀티미디어 기기에서 RAW 파일 재생 기능을 지원하지 못하고 있는 실정이다. 본 연구에서는 RAW 파일 포맷을 분석하여 향후 임베디드 멀티미디어 기기에서 모든 RAW 파일 포맷을 지원할 수 있는 라이브러리 구축을 위한 사전 작업을 수행하였다. Canon의 CRW, CR2 포맷과 Nikon의 NEF 포맷을 분석한 결과를 설명한다. 분석 결과를 바탕으로 RAW 파일 포맷에서 JPEG 파일을 성공적으로 추출하였다. MRW, PEF, DCR, DNG 등의 RAW 파일 포맷을 추가로 분석하여 모든 RAW 파일을 재생할 수 있는 임베디드 장비 및 PC용 RAW 파일 재생 라이브러리를 구축할 것이다.

키워드 : RAW file format, DSLR camera, CR2, CRW, NEF

1. 서론

최근 DSLR 디지털 카메라와 휴대용 멀티미디어 기기 사용자의 급증으로 DSLR 카메라 사용자들 사이에서 휴대용 멀티미디어 기기들의 RAW 파일 지원에 대한 요구가 증대되고 있다. 보급형 DSLR 카메라들의 보급 확대와 함께 하이엔드급 디지털 카메라도 점차 RAW 파일을 지원하기 때문이다.

일반 디지털 카메라로 촬영된 이미지는 주로 JPEG 포맷의 파일로 저장되지만, DSLR 카메라는 JPEG 뿐만 아니라 고해상도의 RAW 파일을 지원한다. Raw 파일은 카메라 센서에서 촬영된 사진에 어떠한 처리도 하지 않은 원본 그대로의 상태를 말한다. 기본적으로 RAW 파일은 많은 정보를 포함하므로 JPEG과 비교하여 약 3~4배 정도 파일 사이즈가 크다. 그런데, 현재 DSLR 카메라에서 주로 사용되는 저장매체의 용량이 4~8GB정도이므로 저장매체에 다수의 RAW 파일을 저장하기 어렵다. 또한, 고해상도의 RAW 파일을 DSLR 카메라의 소형 LCD로 재생하여 확인하는 것도 쉽지 않다. 이러한 문제점을 해결할 수 있는 방안으로 제시되고 있는 것이 휴

대용 멀티미디어 기기에서의 RAW 파일 재생 지원이다. DSLR 카메라의 저장매체에 저장되어 있는 RAW 파일을 휴대용 멀티미디어 기기에 신속하게 백업하여 디지털 카메라의 저장매체 가용량을 늘릴 수 있고, 휴대용 멀티미디어 기기의 대형, 고해상도 LCD로 통해 RAW 파일의 밝기, 제조, 초점 등을 신속하고 확실하게 확인할 수 있기 때문이다.

하지만, DSLR 디지털 카메라 제조사와 카메라 기종별로 RAW 파일 포맷이 다르며, 공식적으로 공개되어 있지 않기 때문에 현재 출시된 PMP 등의 휴대용 멀티미디어 기기들은 RAW 파일 재생 기능을 완벽하게 지원하지 못하고 있다. 따라서, 휴대용 멀티미디어 기기 개발자와 관련 업체에서는 다양한 RAW 파일 포맷에 대한 구체적인 분석 및 이 파일을 임베디드 장치에서 재생할 수 있는 라이브러리 지원을 필요로 하고 있다.

본 논문에서는 이와 관련하여 중소기업청의 산학공동기술 개발사업으로 수행 중인 'Raw Image Data Viewing'이 가능한 Embedded Platform 개발 과제의 연구 결과 중 Raw Image 포맷 분석에 대해 설명한다.

II. RAW 파일 포맷

국내외에 출시된 다양한 DSLR 디지털 카메라가 지원하는 RAW 파일 포맷은 일반적으로 제조사별로 다르며, 모델 별로도 차이가 있다. RAW 이미지의 명확한 포맷은 카메라 제조업체의 지적재산권을 이유로 일반에게 공식적으로 공개 되지 않고 있다. <표 1>은 제조사별 RAW 파일 포맷과 해당 RAW 파일을 재생할 수 있는 PC 전용 프로그램이다.

<표 1> DSLR 카메라 제조사별 RAW 파일 포맷

제조사	포맷(확장자)	전용 프로그램
Canon	CRW, CR2	Digital Photo Professional
Nikon	NEF	Nikon Capture
SONY	MRW, SRF	Lightbox SR
Pentax	PEF, DNG	Photo Browser Laboratory
Fuji	RAF	RAW File Converter LE
Minolta	MRW	Dimage Viewer, Master
Olympus	ORF	Olympus Master
Kodak	DCR	Kodak Photodesk
Sigma	X3F	SIGMA Photo Pro
SAMSUNG	DNG	Samsung raw converter

위의 표에 나타난 바와 같이 DSLR 디지털 카메라 제조업체들은 고유한 RAW 파일 포맷을 사용하고 있으므로, 임베디드 장비에서 RAW 파일을 재생 기능을 지원하기 위해서는 모든 RAW 파일 포맷에 대한 완벽한 분석 후 라이브러리 구축이 필요하다.

RAW 파일 포맷이 다양하게 존재하지만 이들은 기본적으로 TIFF(Tag Image File Format) 포맷을 기반으로 하고 있다. TIFF 포맷의 기원은 스캐너에서 생성되는 흑백 이미지 저장 및 컴퓨터 출판 분야에 응용하기 위한 이미지 표준방식으로 1986년 Aldus사에서 발표하였다. 이후 계속 버전이 변경되어 1992년 TIFF 6.0이 발표 되었고, 여러 응용분야에서 폭넓게 사용되고 있다. TIFF 포맷이 RAW 파일 포맷의 기본이 된 점은 알아보기 위해 우선 TIFF의 구조를 살펴 보도록 한다.

TIFF 파일은 Image File Header(IFH), Image File Directory(IFD), bitmap data로 구성된다. TIFF 파일은 여러 개의 이미지를 포함할 수 있으며, 이 경우 한 이미지당 한 개의 IFD와 bitmap data가 필요하다. 해상도가 큰 원본 이미지를 미리 확인하기 위한 용도로 원본 이미지를 축소시킨 thumbnail을 두 번째 이미지로 포함되기도 한다.

IFH는 <표 2>와 같이 3개 필드로 구성되며, TIFF 파일 시작 위치에 8바이트 크기로 존재한다. 첫 번째 필드는 Byte-order Identifier로 TIFF 파일의 데이터가 little-endian(Intel format)과 big-endian(Motolora format) 중 어느 방식으로구성되어 있는지를 알려준다. 이 필드값이 "II" 즉 0x4949이면 little-endian, "MM" 즉

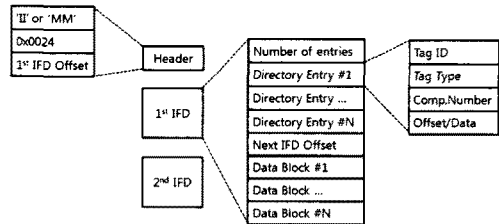
0x4d4d이면 big-endian을 의미한다. 이후의 데이터는 이 필드에서 정해진 byte-ordering scheme에 따라 구성되어 있다. 두 번째 필드는 TIFF 포맷의 버전을 의미하며 항상 42(0x002a)의 값을 가진다. 따라서, TIFF 파일의 처음부터 연속된 4바이트의 값이 0x4949 0x2a00이거나 0x4d4d 0x002a이면 TIFF 파일이 맞다고 판단할 수 있다. 세 번째 필드는 TIFF 파일내의 첫 번째 IFD의 위치를 나타내는 32비트의 offset값이다. 첫 번째 IFD가 IFH 에 이어서 위치하는 경우 이 값은 0x08이 된다.

<표 2> Image File Header of TIFF file

Offset	# of bytes	Description
0x0000	2	Byte-order Identifier ("II" or "MM")
0x0002	2	TIFF version number (0x002a)
0x0004	4	Offset of the first IFD

IFD는 해상도, 크기 등의 이미지 정보뿐만 아니라 카메라의 셔터 스피드, 노출정도, 코멘트, 저작권 관련 공지 등의 카메라와 관련된 사진 정보를 포함할 수 있는 작은 데이터 블록이다.

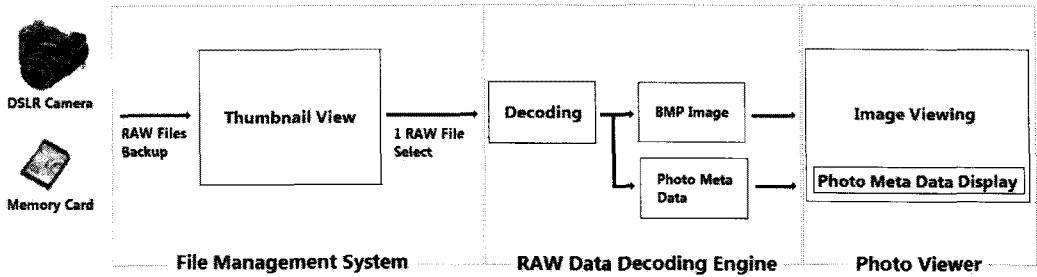
<그림 1>은 Header와 IFD로 구성된 TIFF 파일 포맷 구조이다.



<그림 1> TIFF 포맷 구조

TIFF 기반의 RAW 파일의 생성/재생 과정은 <그림 2>와 같다. RAW 파일은 필름 카메라에서 필름과 같은 정보를 포함하고 있어 촬영된 이미지를 다양하게 활용할 수 있다. 또한, RAW 파일에는 RAW 이미지 데이터와 Thumbnail이 한 개 이상 포함되어 있으며, JPEG 이미지의 개수 및 해상도는 RAW 파일 포맷에 따라 다르다.

카메라 제조사마다 다양한 형식의 RAW 파일 포맷을 만들어 사용하는 상황에서 Adobe사에서 RAW 파일 포맷 표준을 만들고자 DNG(Digital Negative)라는 새로운 RAW 파일 포맷을 만들었다. DNG 파일은 RAW 파일을 통합하여 보관할 수 있는 파일 형식으로 DNG 파일을 해석하기 쉽도록 소스를 오픈하고 있다. 일부 DSLR 카메라에서는 이러한 DNG 포맷을 지원하고 있다[1].



〈그림 2〉 RAW 파일 생성 및 재생 과정

국내 시판 중인 휴대용 멀티미디어기기 중에서 일부 제품만이 RAW 파일 재생 기능을 지원하고 있고, 지원하는 카메라 모델별 RAW 파일 포맷을 명확한 기재하지 않고 있다. 호환되는 카메라 모델이 60%이상으로 가장 높은 PMP 기준은 코윈의 Q5제품이며, 그 뒤를 이어 디지털큐브의 iStation M43, Netforce2가 전부인 상황이다.

III. 대표적인 RAW 파일 포맷 분석

본 장에서는 〈표 1〉의 RAW 파일 포맷 중 Cannon사의 CRW, CR2와 Nikon사의 NEF에 대해 분석한다. CRW와 CR2는 TIFF를 변형하였는데 CR2가 TIFF 포맷에 가깝고, NEF는 TIFF 포맷과 거의 동일하다(2,3,4).

3.1 Cannon CRW File Format

CRW는 Cannon사의 첫 번째 RAW 파일 포맷으로, Cannon D30, D60, 10D, 300D, Power Shot Pro1, G1~G6, S30~S70 등의 모델에서 사용된다.

CRW 파일 포맷은 TIFF 포맷을 기반으로 하며, 1개의 Main IFD(Image File Directory)를 포함하는 구조이다. CRW 파일의 헤더는 〈표 3〉과 같이 TIFF 파일의 헤더를 기본으로 하며 일부 필드가 추가되었다.

〈표 3〉 Header structure of CRW file format

Offset	# of bytes	Description	Value
0x0000	2	Byte order	"II" or "MM"
0x0002	4	Header Length	0x0000001a (26)
0x0006	8	Signature	"HEAPCCDR"
0x0014	4	CRW Version	0x00010002
0x0018	8	Reserved	0

CRW 헤더에 이어 위치하는 루트 디렉토리 블록 구조는 〈표 4〉와 같다. 이 구조는 TIFF의 IFD의 개념을 적용하고 있지만 차이점이 있다.

〈표 4〉 Directory block structure of CRW file format

Offset within block	# of bytes	Value	Description
0	S	-	ValueData

S	2	N	디렉토리 엔트리 수
S+2	N*10	-	CRW 디렉토리 엔트리
S+2+N*10	any	-	Other data
BlockSize-4	4	S	ValueData size

데이터 블록을 분석하기 위해서는 먼저 블록 맨 뒤에 위치한 4바이트를 읽어야 한다. 이 값은 ValueData 크기를 의미하며, 데이터 블록 시작에서부터 ValueData 크기의 위치까지 RAW 이미지 데이터, thumbnail 이미지와 기타 이미지 및 카메라 관련 정보가 저장된다. ValueData에 연속된 2 바이트의 값은 디렉토리내의 엔트리 수를 의미한다. 각 엔트리는 10바이트의 크기를 가지면 다양한 정보를 포함하고 있다.

3.2 Cannon CR2 File Format

CR2는 Cannon RAW version 2를 의미하며, Cannon Digital EOS Camera 350D, 20D, G9, 1D MarkII 등의 모델에서 사용된다.

CR2 파일 포맷은 TIFF 포맷을 기반으로 하며, 4개의 IFD(Image File Directory)를 포함하는 구조이다. CR2 파일의 헤더는 〈표 5〉와 같이 TIFF 파일의 헤더를 기본으로 하며 일부 필드가 추가되었고, CRW와 약간의 차이가 있다.

〈표 5〉 Header structure of CR2 file format

Offset	# of bytes	Description	Value
0x0000	2	Byte order	"II" or "MM"
0x0002	2	TIFF magic word	0x002a
0x0004	4	TIFF offset	0x0000 0010
0x0008	2	CR2 magic word	"CR" or 0x4352
0x000a	1	CR2 major version	0x02
0x000b	1	CR2 minor version	0x00
0x000c	4	RAW IFD offset	RAW IFD offset

IFD 구조는 〈표 6〉과 같고, CRW의 IFD 구조가 변경되었다. 첫 번째 IFD는 촬영된 이미지를 1/4 크기로 축소하여 압축한 JPEG 이미지 정보, EXIF 정보, Maker note 정보를 포함한다. 두 번째 IFD는 160x120 크기의 JPEG 이미지 정보를 포함한다. 세 번째 IFD는 압축되지 않은 작은 이미지 파

일 정보를 포함한다. 네 번째 IFD는 손실없는 JPEG으로 압축된 RAW 데이터를 포함한다. CR2의 IFH의 마지막 필드인 RAW offset 필드값이 네 번째 IFD의 시작 위치를 나타낸다.

〈표 6〉 IFD structure of CR2 file format

Offset	# of bytes	Description
0x0000	2	Number of entries (N)
0x0002	12	entry #0
0x000e	12	entry #1
...
2+12*(N-1)	12	entry #N-1
2+12*N	4	Next IFD offset

3.3 Nikon NEF File Format

NEF(Nikon Electronic Format)은 Nikon사의 RAW 파일 포맷으로 TIFF 포맷과 유사하다. NEF 파일의 헤더는 〈표 7〉과 같이 TIFF 파일 헤더와 동일하다.

〈표 7〉 Header structure of NEF file format

Offset	# of bytes	Description	Value
0x0000	2	Byte order	"II" or "MM"
0x0002	2	TIFF magic word	0x002a
0x0004	4	offset to IFD #0	0x0008

NEF 파일 헤더에 이어 2개의 IFD가 존재한다.

Primary IFD는 헤더 바로 뒤에 위치하며, thumbnail 이미지의 해상도, 압축방법, raster 데이터의 offset, raster 데이터 크기, EXIF 데이터를 포함하는 IFD의 offset 등과 함께 Secondary IFD의 offset 정보를 포함한다.

Secondary IFD는 실제 사진의 raw raster 데이터 관련 정보를 포함한다. 사진의 해상도, 압축방법, raster 데이터의 offset, raster 데이터 크기 등의 정보를 포함한다.

Primary IFD에 EXIF 데이터를 포함하는 SubIFD가 존재하며 offset은 primary IFD에 설정되어 있다. 이 IFD에는 카메라 노출 관련 정보가 포함되어 있고, Maker Note라는 SubIFD가 존재하고, thumbnail 이미지가 존재한다.

따라서, NEF 파일에는 압축되지 않은 raw 데이터 파일, 고해상도의 JPEG 파일, thumbnail용 JPEG 파일, Nikon사의 카메라 모델에서 사용하기 위한 thumbnail용 JPEG 파일이 존재한다.

IV. RAW 파일 포맷 분석 결과

CRW, CR2, NEF 파일 포맷 분석 결과를 확인하기 위해 두 가지 방법을 사용하였다. 첫 번째는 파일값을 16진수로 변환하여 보여주는 WinHex라는 프로그램을 활용하여 CRW, CR2, NEF 실제 이미지를 헤더부터 각 IFD까지 분석하였다. 〈그림 3〉은 WinHex 프로그램을 사용하여 CRW 포맷의 sample.CRW 파일을 16진수값으로 변환한 결과 중 Header 부분만 나타낸 것이다.

```

Offset  0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
0  49 49 1A 00 00 00 48 48 41 90 43 43 42 02 00 01 00 00 00 00
21  00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
42  00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
63  00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
    
```

〈그림 3〉 sample1.CRW의 헤더 정보

두 번째 방법으로 실제 RAW 파일에 포함된 JPEG 파일을 추출할 수 있는 프로그램을 C언어로 개발하였고, JPEG 파일 1개 이상을 성공적으로 추출하였다.

V. 결론

본 연구에서는 DSLR 카메라의 RAW 파일 포맷 중 Cannon의 CRW, CR2 포맷과 Nikon의 NEF 포맷을 분석한 결과를 설명하였다. 이 RAW 포맷들이 TIFF 파일 포맷의 기본적인 구조를 변형하여 구성되었음을 확인하였고, 분석 결과를 바탕으로 CRW, CR2, NEF 파일에서 JPEG 파일을 성공적으로 추출하였다. 향후 MRW, PEF, DCR, DNG 등의 RAW 파일 포맷을 추가로 분석하여 모든 RAW 파일을 재생할 수 있는 임베디드 장비 및 PC용 RAW 파일 재생 라이브러리를 구축할 것이다.

참고문헌

- [1] 김기덕, Camera Raw, 정보문화사, 2008.
- [2] <http://www.pmpinside.com>
- [3] <http://lclevy.free.fr/cr2>
- [4] <http://www.timelesswandering.net/equipment/D100/NEF.html>