RAWcooked

Data is never RAW.
Data is always cooked in a way or another.

Jérôme Martinez

CNA Heritage Event, November 2018

MediaArea

Open source software company focused on digital media analysis. We work (different levels of involvement) on:

- MediaInfo
  Convenient unified display of the most relevant technical and tag data for video and audio files
- MediaConch
  Implementation checker, policy checker, & reporter
- QCTools
  Helps users analyze and understand their digitized video files through use of audiovisual analytics and filtering
- BWF MetaEdit, AVI MetaEdit, MOV MetaEdit
  Embedding, validating, and exporting of metadata
- DV Analyzer
  Checking presence of technical errors in DV captures

Raw A/V files

Huge size
(4K+ is coming, can be 100 MB/frame, several TB per hour)

1 file per video frame (thousands of files in a directory)

Not playable as is by several players (VLC...)

So many DPX or TIFF format flavors
(interoperability issues)

FFV1

Lossless video compression format

Open source, patent free

Adopted by several archives

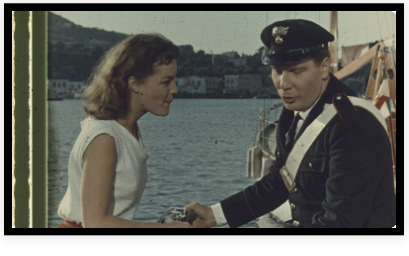
Being standardized (IETF)

Frames are divided by slices, with checksums

Compression

Example with 1 second at 24 fps 10-bit HD film on a 6-core (12-thread) Skylake-X CPU:

- 24 DPX files (or in ZIP/TAR uncompressed): 189 MB
- 1 compressed ZIP file: 175 MB in 10 seconds
- 1 compressed LZMA2 file: 154 MB in 30 seconds
- 1 FFV1/MKV Intra 16-slice file: 105 MB in 1.5 seconds



Tests on SD 10-bit content

- Sponsored by VIAA
- On several TB of MXF/JP2k/PCM content
- Different sources and kinds of material
- E5-2698V3 (16 cores+HT, 2.3-3.6 GHz)

Tests on SD 10-bit content

- This is an average, results varies depending on the content of files
- From 0.4x to 2.4x (average 0.7x) real time/thread (encoding/decoding)
- From 0.7x to 16x (average 3.5x) the speed of JP2k (FFmpeg)
- Better compression ratio by 2% compared to JP2k (9% with additional FLAC audio compression)
- https://viaa.be/files/attachments/.669/VIAA_Preservation_R

Advantages

- Open source and free (FFmpeg)
- Package is 1.5-3x smaller than DPX/TIFF
- Cheksum by "Cluster" (usually 1 second) at container level
- Cheksum by "Slice" (you choose how many per frame) at video level
- Files are natively playable by lot of players (VLC, FFplay...)

Disadvantages

- You lose some metadata
  (DPX/TIFF header: scan software, some colorimetry info, film type, DPX time code, shutter angle, gamma...)
- Complicated command line
- But...

RAWcooked

- Easy: just a short command line "rawcooked YourDirectoryName"
- Store DPX/TIFF headers/footers in a specific Matroska attachment
- Store other sidecar files as Matroksa attachments
- Output is a single Matroska/FFV1/FLAC file
- Encoding is reversible (bit-by-bit to original files) "rawcooked YourMatroskaFileName.mkv"

Usages

- Storage
  Save HDD/LTO space (less €...)
- Transport
  Encode, transport, decode; you save bandwidth (€... and transfer speed) without changing something else in your workflow (same files after revert to DPX/TIFF)

Remaining drawbacks

- More CPU intensive than uncompressed ZIP
  (future project: use GPU and/or SSE/AVX)
- You lose a complete slice content if there is a single corrupted bit
  (Mitigated with several slices per frame, MediaConch can fix a bitflip error and could be extended to a byte error)

Development

Reto.ch commited to main sponsorship end of 2017

Other sponsors showed interest begin of 2018

First Beta releases first half of 2018

Sponsors (and some others) tested with their files during the past 6 months

First stable release in Octobr 2018

Supported input formats

- DPX/Raw: 8/10/12/16 bit, RGB/RGBA
- TIFF/Raw: 16 bit, RGB
- WAV/PCM: 16/24 bit, 1/2/6 channel, 44/48/96 kHz
- AIFF/PCM: 16/24 bit, 1/2/6 channel, 44/48/96 kHz
- Based on files from our sponsors
- More formats or format flavors on request

Our first sponsors

- AV Preservation by reto.ch (main sponsor)
- National Audiovisual Centre Luxembourg (CNA)
- National Library of Norway
- Irish Film Institute (IFI)
- Northwest University Library
- National Library of Wales

Financial sustainability

- Open source code provided without lock to sponsors
- Deliveries on our website are with a lock
- DPX 8/10 bit RGB & WAV 2ch 48kHz flavors are usable by default
- We provide a key for other format flavors and features (temporary key possible)
- 1000 € for first flavor/feature
  + 500 € per additional flavor/feature
- To be compared with storage cost saving
  (storage cost divided by 2)

Potential improvements

- Speed improvement (GPU/SSE/AVX)?
- Error correction code?
- Support of reels?
- CFA/Bayer/RGGB support?
- Graphical interface?
- More input formats?

Stay in touch

MediaArea: https://mediaarea.net, @MediaArea_net

RAWcooked: https://MediaArea.net/RAWcooked

Jérôme Martinez: jerome@mediaarea.net

Slides: https://mediaarea.net/Events

License (except images): CC BY