

Digital Archives and Data Maintenance

Reto Kromer • AV Preservation by reto.ch

Summer School Digital Archives
Filmuniversität Babelsberg Konrad Wolf
6–10 September 2021

1

Global Emergencies

- sanitation emergency
- climate emergency

3

A Biographic Note

1983–1985

1986–1997

1998–2003

- Cinémathèque suisse

2004–2023

- AV Preservation by reto.ch

2

Sanitation Emergency

- physical distancing
- hand hygiene
- surgical masks
- stay home when feeling ill
- get vaccinated when it's your turn

4

Climate Emergency

- embrace low-energy ...
- ... and low-technology solutions
- use less resources ...
- ... and recycle whenever possible

→ slow down

5

Digital Archives

Table of Contents

- Digital Archives
- Data Maintenance
- A Report from the Field

6

Preservation Workflow

- prepare the source elements
- digitise the source elements
- digitally restore the files
- generate derive files
- conserve the source elements
- conserve the files
- document each single step

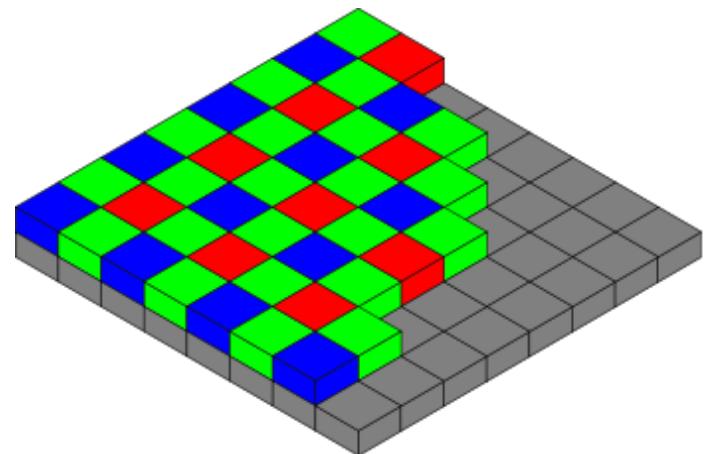
7

8

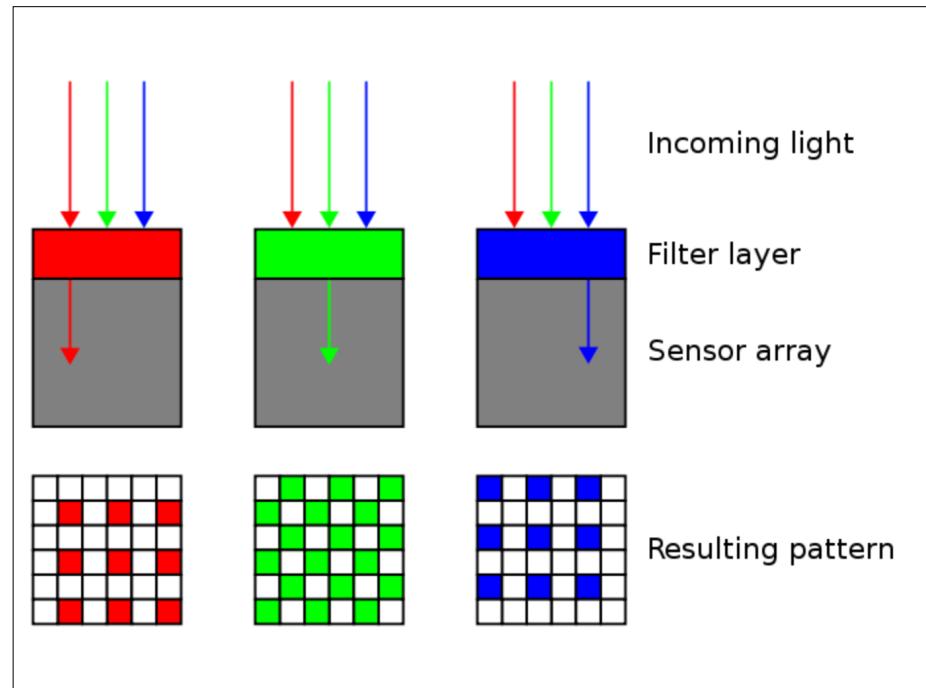
Uncomfortable Truths

- sensors are colour blind
- Bayer sensors do not generate full RGB

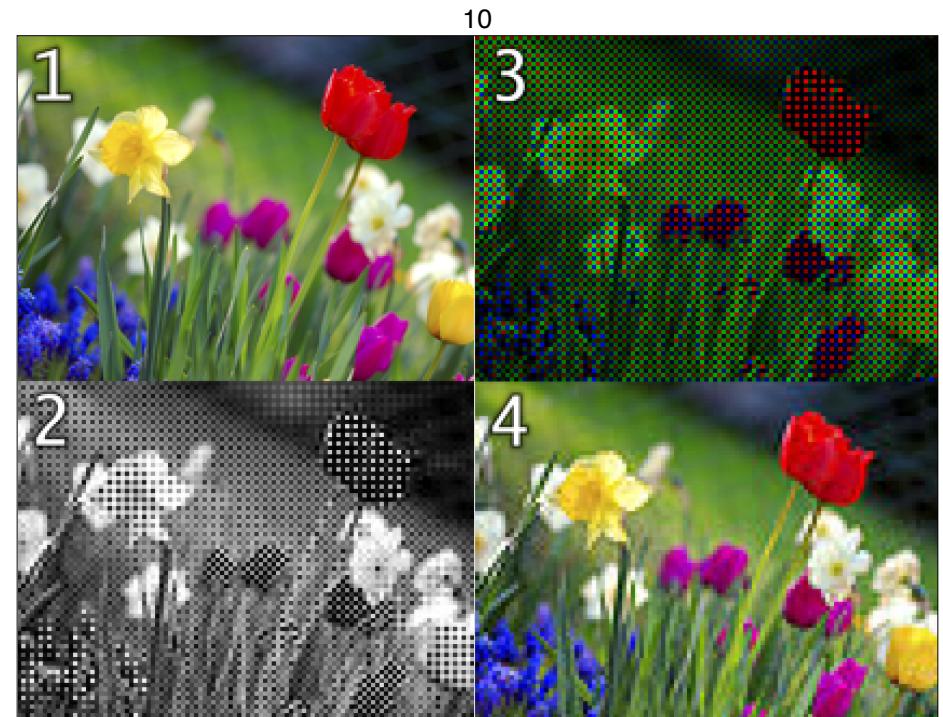
Bayer



9



11



12

```

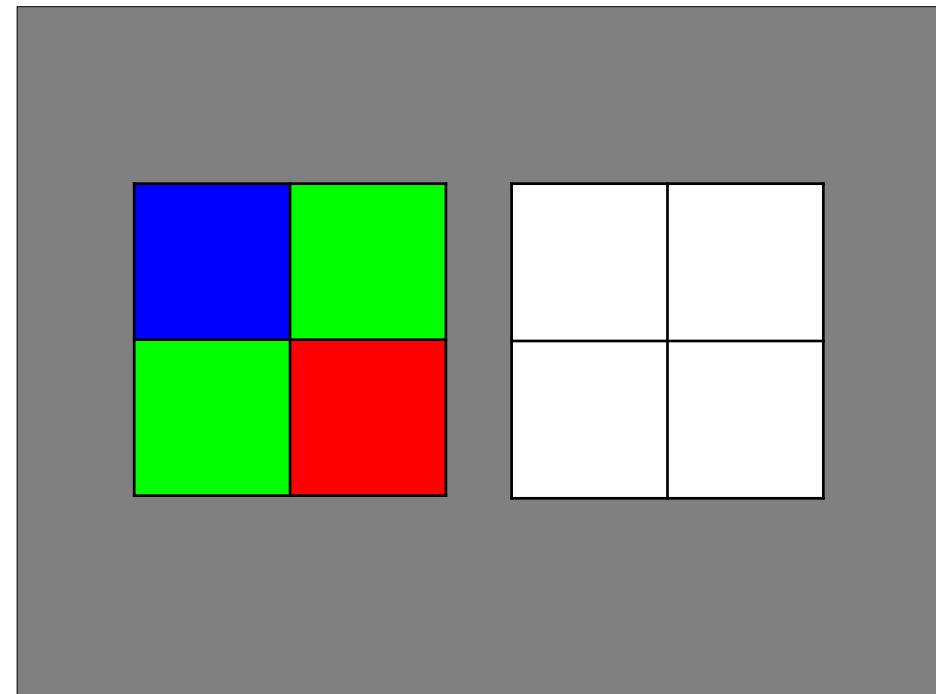
0111010100101010100010110101011110
0100110101010101010100001011101010
0111010100101010100010110101011110
0001110101010101010100001011101010
0110101010010101010001011010101111
0010101010101010000101110101010000
0111010100101010100010110101011110
01010101010101000010111010100110
1001011101010010101010001011010101
11100101010101010000101110101010
01110101001010100010110101011110
01010101010101001101010100000001
0010100010101010100101010101010101

```

13

000000000000 000000000000 110101010101	000000000000 000000000000 01010001011
000000000000 101010011010 000000000000	101001010101 000000000000 000000000000

15

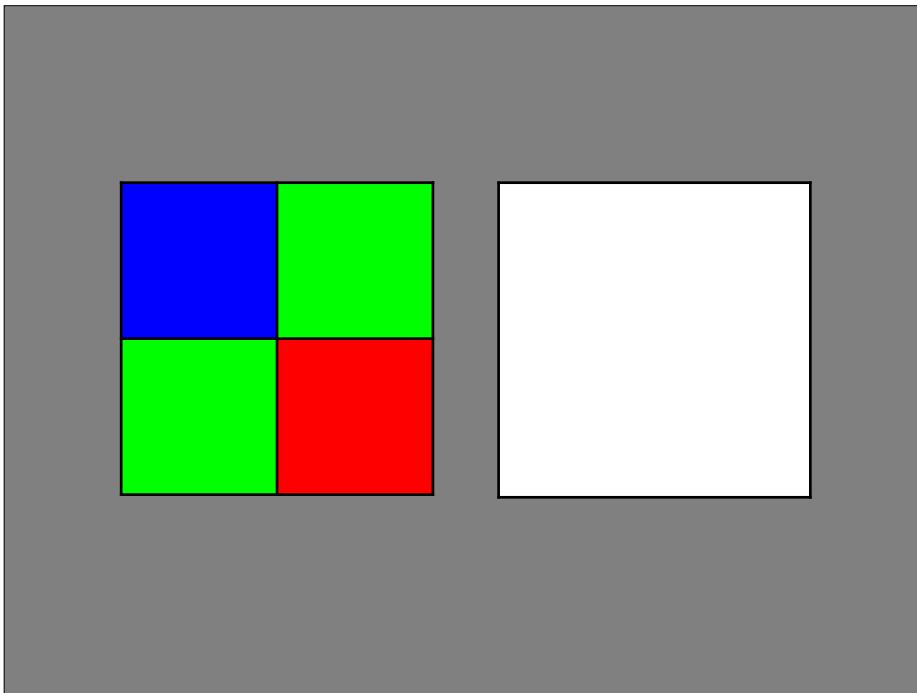


14

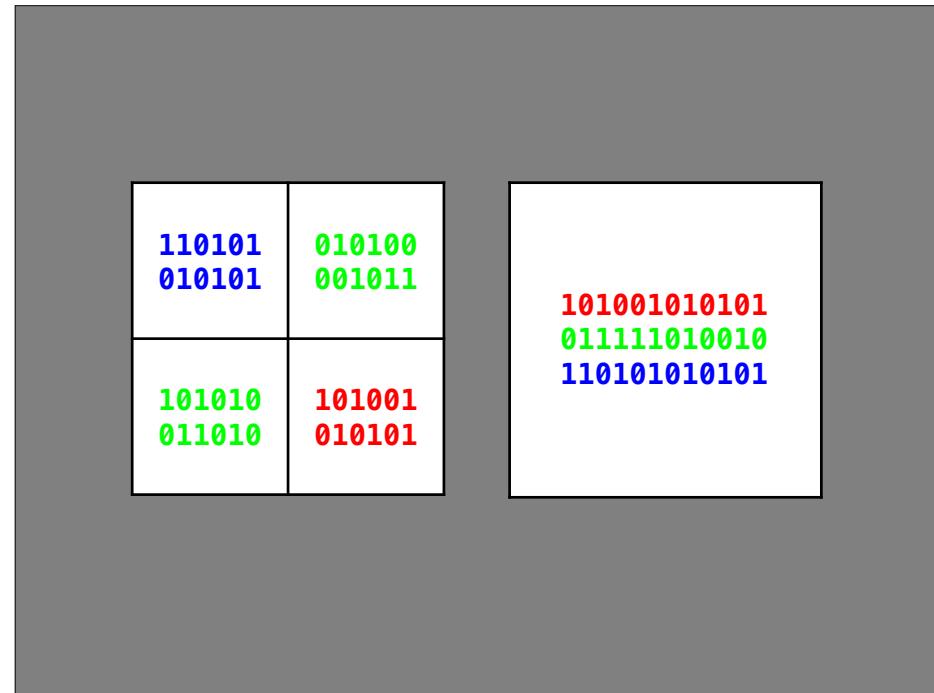
0 0 B	0 G1 0
0 G2 0	R 0 0

R G B	R G B
----------------------------------	----------------------------------

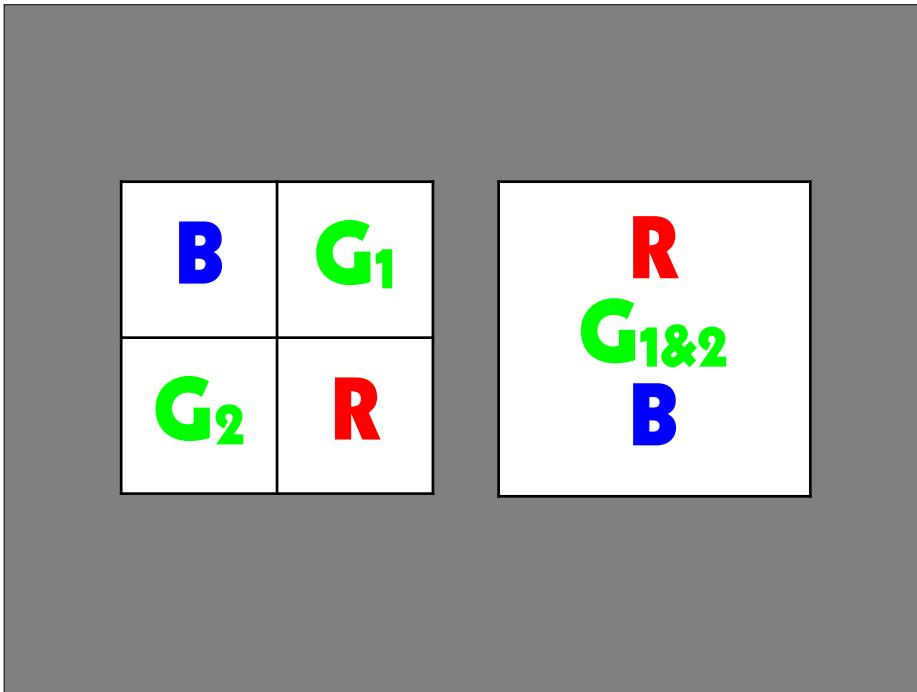
16



17



18



19

Two ways to use Bayer data

digital blow-up to RGB

- 3 times the amount of the generated data
- the file has the full sensor resolution
- only $\frac{1}{3}$ of the data are real

digital reduction to RGB

- $\frac{3}{4}$ the amount of the generated data
- the file has $\frac{1}{2}$ of the sensor resolution
- all data are real

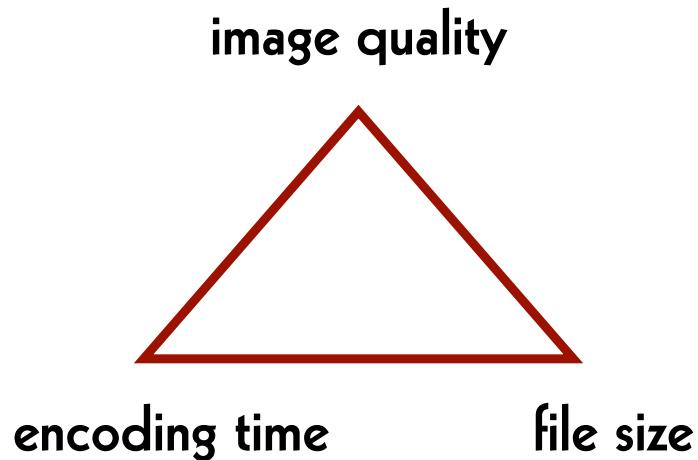
20

File Formats

→ The archive must be able to handle the file formats it holds.

- open source
- simple to use and well documented
- widely used by the community

21



23

AV container:

- folder
- TAR
- ZIP
- MXF
- Matroska

video codec:

- DPX
- TIFF
- JPEG 2000
- FFV1
- OpenEXR
- CineForm RAW
- ProRes RAW
- Blackmagic RAW

22

Compression

- uncompressed
- lossless compression
- lossy compression
- chroma subsampling
- born compressed

24

Uncompressed

- + data simpler to process
- + software runs faster
- bigger files
- slower writing, transmission and reading

Examples: DPX, TIFF, OpenEXR

25

	avantages	disadvantages
TIFF DPX OpenEXR	data easier to process	bigger files
JPEG 2000 FFV1	smaller files	data complexer to process

27

Lossless Compression

- + smaller files
- + faster writing, transmission and reading
- data processing complexer
- software runs slower

Examples: JPEG 2000, FFV1

26

Data Storage

- hard disk drives (HDD)
- solid state drives (SSD)
- magnetic tapes (LTO with LTFS and/or TAR)

28

LTO

- Linear Tape-Open
- answer from the IT industry to the finance and insurance sectors
- in 2000 LTO-1
- currently LTO-8
- currently Hewlett Packard Enterprise, IBM, and Quantum form the LTO Consortium

29

TAR

- standard TAR
 - bloc size
 - number of archives per cartridge
 - archives needing more than one cartridge
- TAR with a proprietary data encoding (e.g. BRU, Retrospect)

31

Formatting

TAR

- only possibility from LTO-1 to LTO-4
- still possible today

LTFS

- possible (and recommended) since LTO-5

30

LTFS

- different versions
- almost one implementation per vendor, but ...
 - ... "ltfs" and "mkltfs" common commands
- lossless compression (default) or uncompressed data
- unencrypted (default) or encrypted data

32

Drive

- internal or external unit
- library

33

Software

- proprietary
- open source
- graphical user interface (GUI)
- command-line interface (CLI)

35

Storage of Tapes

- shelf
- tape library
- fireproof cabinet

34

Plan the Next Migration

- file naming
- barcodes
- checksums
- write the full index onto the cartridge
- technical metadata
- code to retrieve the files

36

Longterm

- storage of the cartridges
- three copies ...
- ... in geographically distant locations
- data integrity check
- data migration
- availability of LTO desks

37

Data Migrations

2014

- our internal archive from LTO-4 to LTO-6
(5.7 PB)

2014–2021

- many migrations for clients

2021

- our internal archive from LTO-6 to LTO-8
(25.2 PB)

39

Data Maintenance

38

Reading

Reto Kromer: **On the Bright Side of Data Migrations**, in «IASA Journal», n. 49 (December 2018), IASA, p. 18–22

→ https://retokromer.ch/publications/IASA_49.html

40

read | script | write

script to modify

- container
- codec
- both container and codec
- metadata
- filename

41

Update the Container

→ read file from source LTO

→ demultiplex file

- ProRes 422, 10 bit [yuv422p10le]
- ProRes 4444, 10 bit [yuv444p10le or yuva444p10le] or 12 bit [yuv444p12le]

→ multiplex file

→ write file to destination LTO

43

#1: ProRes-born Content

from:

- ProRes stored in a QuickTime (.mov) container

to:

- ProRes stored in a Matroska (.mkv) container

42

SMPTE RDD 36:2015

SMPTE REGISTERED DISCLOSURE DOCUMENT

Apple ProRes Bitstream Syntax and Decoding Process



Page 1 of 39 pages

The attached document is a Registered Disclosure Document prepared by the sponsor identified below. It has been examined by the appropriate SMPTE Technology Committee and is believed to contain adequate information to satisfy the objectives defined in the Scope, and to be technically consistent.

This document is NOT a Standard, Recommended Practice or Engineering Guideline, and does NOT imply a finding or representation of the Society.

Every attempt has been made to ensure that the information contained in this document is accurate. Errors in this document should be reported to the proponent identified below, with a copy to eng@smpte.org.

44

#2: Video

from:

- AVI / 8-bit and 10-bit uncompressed
- MOV / 8-bit and 10-bit uncompressed
- MP4 / 8-bit and 10-bit uncompressed

to:

- Matroska / FFV1

45

Container and Codec

→ read file from source LTO

→ demultiplex file

→ decode file

- Y'CbCr, 4:2:2, 8 bit, «raw» [uyvy422]

→ encode file

→ multiplex file

→ write file to destination LTO

46

Container and Codec

→ read file from source LTO

→ demultiplex file

→ decode file

- Y'CbCr, 4:2:2, 10 bit, «raw» [yuv422p10le]

→ encode file

→ multiplex file

→ write file to destination LTO

47

#3: Filename

from:

- Title_YUV422.mkv

to:

- Title_YCbCr422_9d5084b5b0a08d5022b3
9e0e75241d12.mkv

48

Catalogue

49

Kameranegativ

- Arbeitskopie (AK)
- Dup-Positiv
- Dup-Negativ

- Nullkopie (0-Kopie)
- Korrekturkopie (K-Kopie)
- Serienkopie (S-Kopie)

51

Reading

Reto Kromer: **Le catalogage des films**, in
«CinémAction», n. 97 (4^{ème} trimestre 2000), p.
121–125

→ <https://retokromer.ch/publications/CA97.html>

50

Originalton (O-Ton)

- Geräusche
- Musik
- IT-Band
- Kommentar, Sprache
- Mix
- Ton-Negativ

- Nullkopie (0-Kopie)
- Korrekturkopie (K-Kopie)
- Serienkopie (S-Kopie)

52

Always remember:
To do nothing
is never an option!

53

AV Preservation by reto.ch

zone industrielle Le Trési 3
1028 Préverenges
Switzerland

Web: reto.ch
Twitter: [@retoch](https://twitter.com/retoch)
Email: info@reto.ch



54