

Archivage numérique du film et de la vidéo

Digitale Archivierung von Film und Video

Reto Kromer • AV Preservation by reto.ch

MAS/CAS ALIS
Universität de Berne, 12 novembre 2021

1

Table des matières

Inhalt

- son et image numériques
digitaler Ton und digitales Bild
- structure et format de fichiers
Dateiaufbau und Dateiformate
- sauvegarde et maintenance des données
Datensicherung und Datenpflege

2

Recommandations Memoriav

Memoriav-Empfehlungen

Agathe Jarczyk, Reto Kromer, Yves Niederhäuser
& David Pfluger:

- **L'archivage numérique des films et vidéos: fondements et orientations**
→ memoriav.ch/fr/dafv/
- **Digitale Archivierung von Film und Video: Grundlagen und Orientierung**
→ memoriav.ch/de/dafv/

3

Son et image numériques

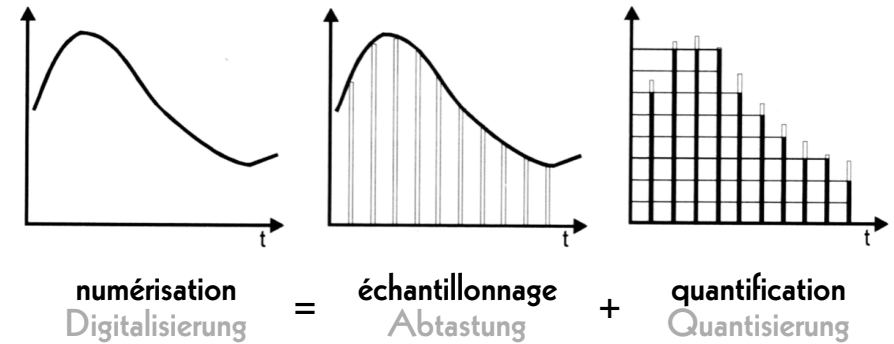
Digitaler Ton und digitales Bild

4

Un peu d'histoire Etwas Geschichte

- **technique télévisuelle**
Fernsehtechnik
- **enregistrements magnétiques**
Magnetaufzeichnung
- **vidéo**
Video

5



6

Son numérique Digitaler Ton

- **échantillonnage**
Abtastung
- **quantification**
Quantisierung

7

Image numérique

- **définition**
- **profondeur de couleurs**
- **linéaire, exponentielle, logarithmique**
- **espace colorimétrique**
- **sous-échantillonnage et compression**
- **illuminant**

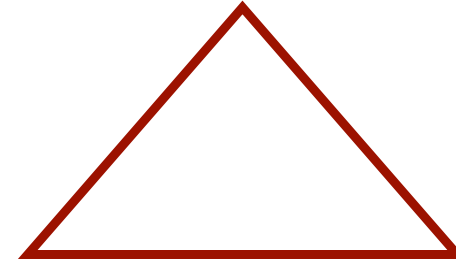
8

Digitales Bild

- Auflösung
- Quantisierungsauflösung
- linear, Potenzfunktion, logarithmisch
- Farbraum
- Farbunterabtastung und Kompression
- Normlicht

9

qualité de l'image Bildqualität



temps d'encodage
Kodierzeit

taille du fichier
Dateigrösse

10

Compression

- non compressé
- compressé sans perte
- compressé avec perte
- sous-échantillonnage
- compressé à la création

11

Kompression

- nicht komprimiert
- verlustfrei komprimiert
- verlustbehaftet komprimiert
- Farbunterabtastung
- komprimiert generiert

12

Non compressé

- + données plus faciles à traiter
- + logiciels plus rapides à exécuter
- fichiers plus lourds
- fichiers plus lents à écrire, lire et transmettre

Exemples: TIFF, DPX, DNG, OpenEXR

13

Nicht komprimiert

- + Daten sind leichter zu bearbeiten
- + Software läuft schneller
- grössere Dateien
- langsames Schreiben, Übermitteln und Lesen der Dateien

Beispiele: TIFF, DPX, DNG, OpenEXR

14

Compressé sans perte

- + fichiers plus légers
- + plus rapides à lire, écrire, transmettre
- données plus complexes à traiter
- logiciels plus longs à exécuter

Exemples: JPEG 2000, FFV1

15

Verlustfrei komprimiert

- + kleinere Dateien
- + schnelleres Schreiben, Übermitteln und Lesen der Dateien
- Daten sind komplexer zu bearbeiten
- Software läuft langsamer

Beispiele: JPEG 2000, FFV1

16

Compressé avec perte

- optimisé pour l'acquisition et/ou pour la postproduction
- optimisé pour la diffusion

Exemples (mezzanine): ProRes 422, ProRes 4444; DNxHD, DNxHR

Exemples (diffusion): H.264 (AVC), H.265 (HEVC), H.266 (VVC); AV1

17

Verlustbehaftet komprimiert

- optimiert für Aufnahme und/oder Postproduktion
- optimiert für Zugang und Distribution

Beispiele (Mezzanine): ProRes 422, ProRes 4444; DNxHD, DNxHR

Beispiele (Zugang): H.264 (AVC), H.265 (HEVC), H.266 (VVC); AV1

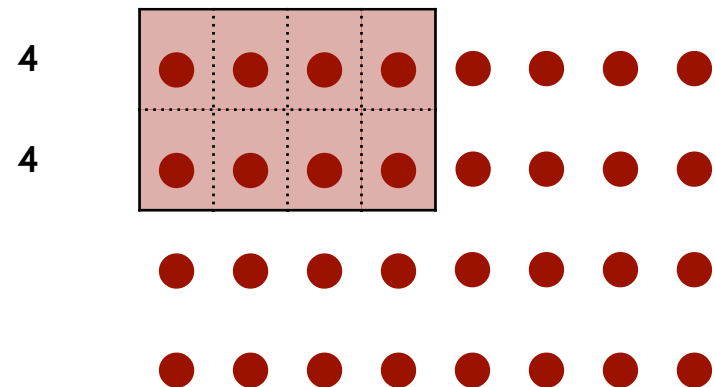
18

Sous-échantillonnage Farbunterabtastung

- 4:4:4
- 4:2:2
- 4:2:0 / 4:1:1

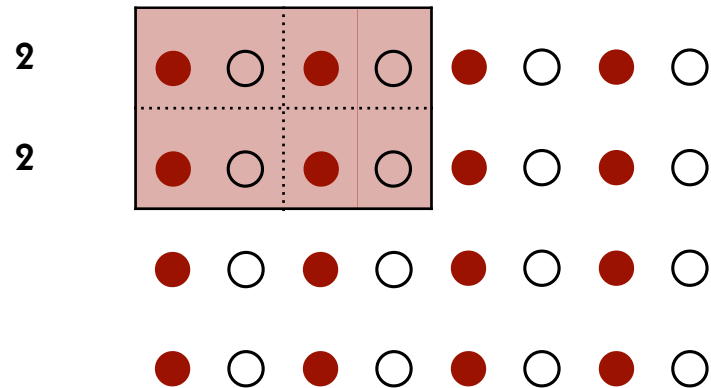
19

4:4:4



20

4:2:2



21

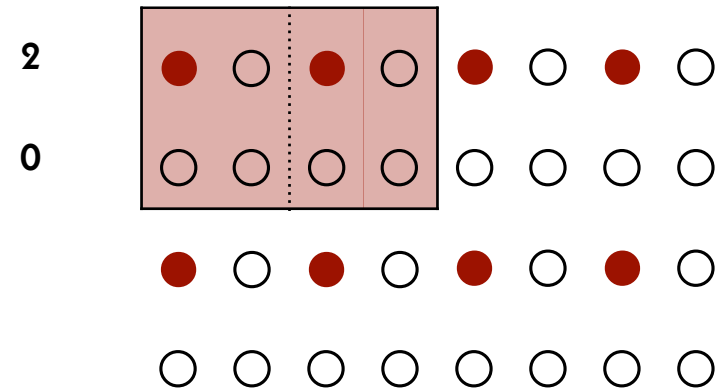
Compressé à la création Komprimiert generiert

- optimisé tant pour l'acquisition que pour la postproduction
sowohl für Aufnahme als auch für Postproduktion optimiert

Exemples: CineForm RAW, ProRes RAW,
Blackmagic RAW

23

4:2:0

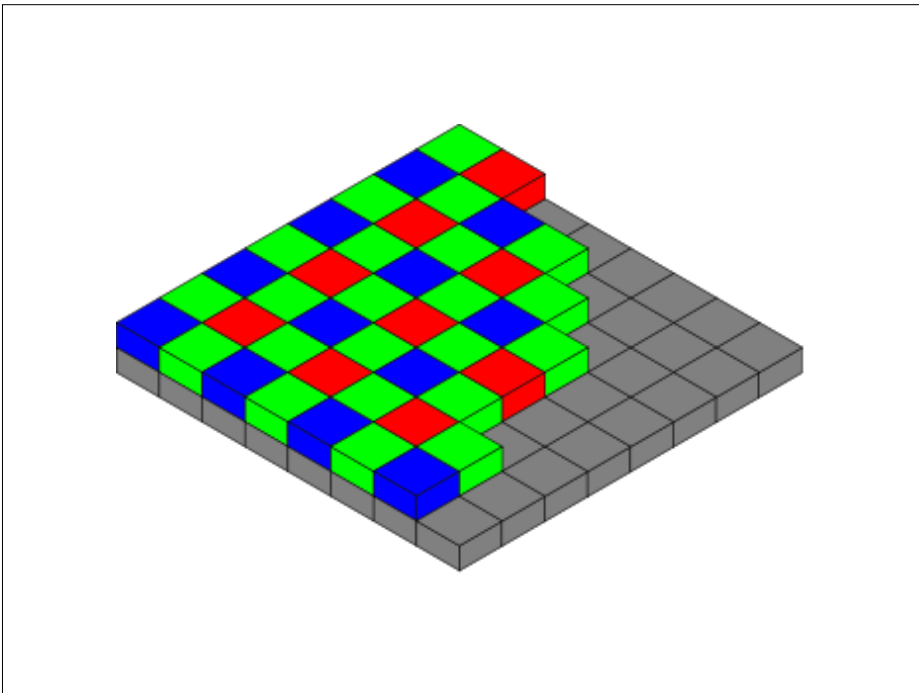


22

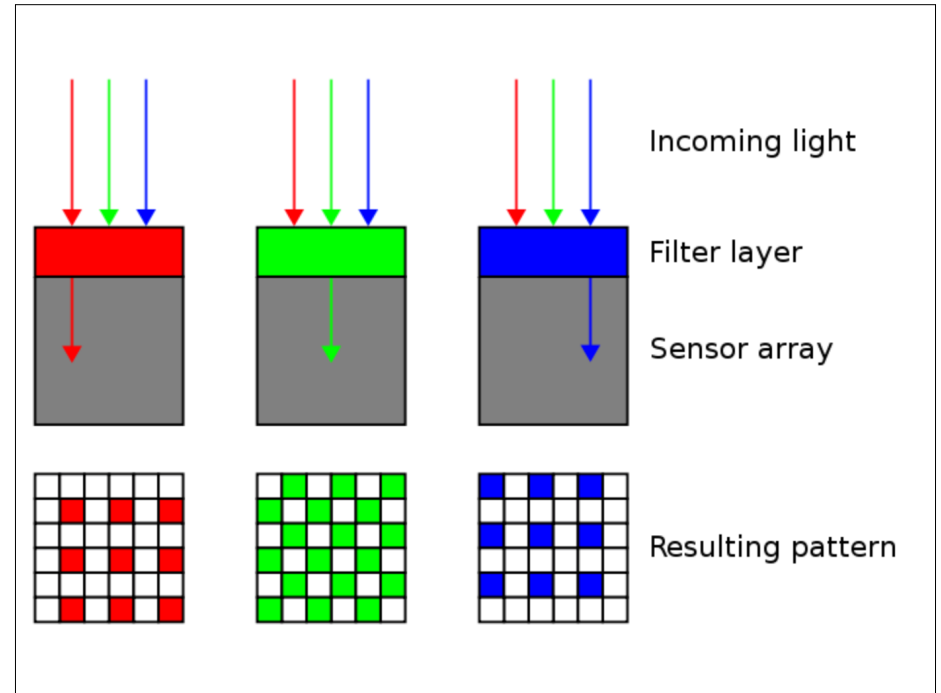
Des vérités qui dérangent Unbequeme Wahrheiten

- les capteurs sont daltoniens
Sensoren sind farbenblind
- les capteurs Bayer ne produisent pas une image RGB complète
Bayer-Sensoren erzeugen kein vollständiges RGB-Bild

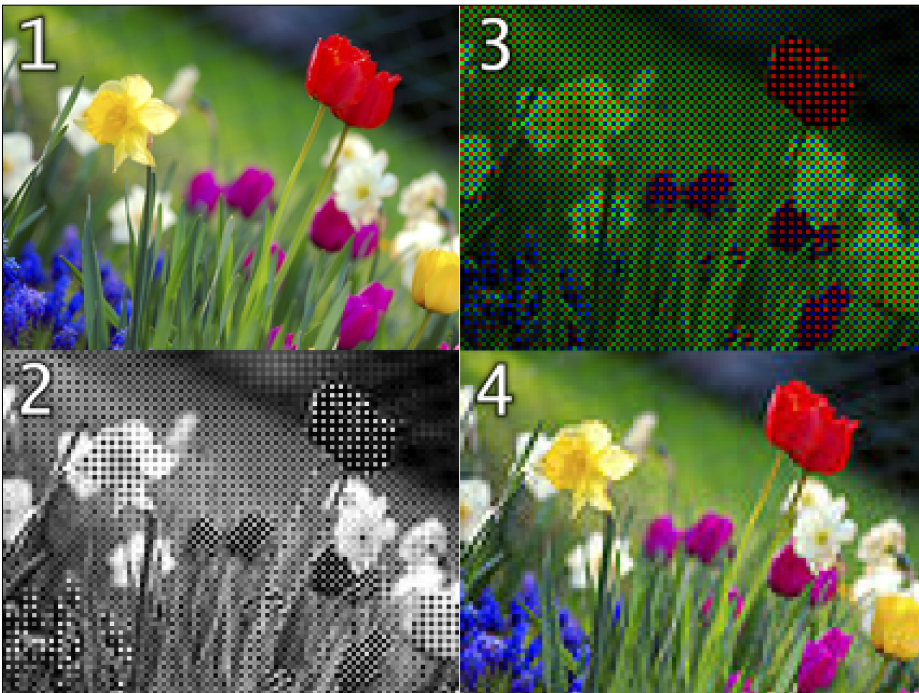
24



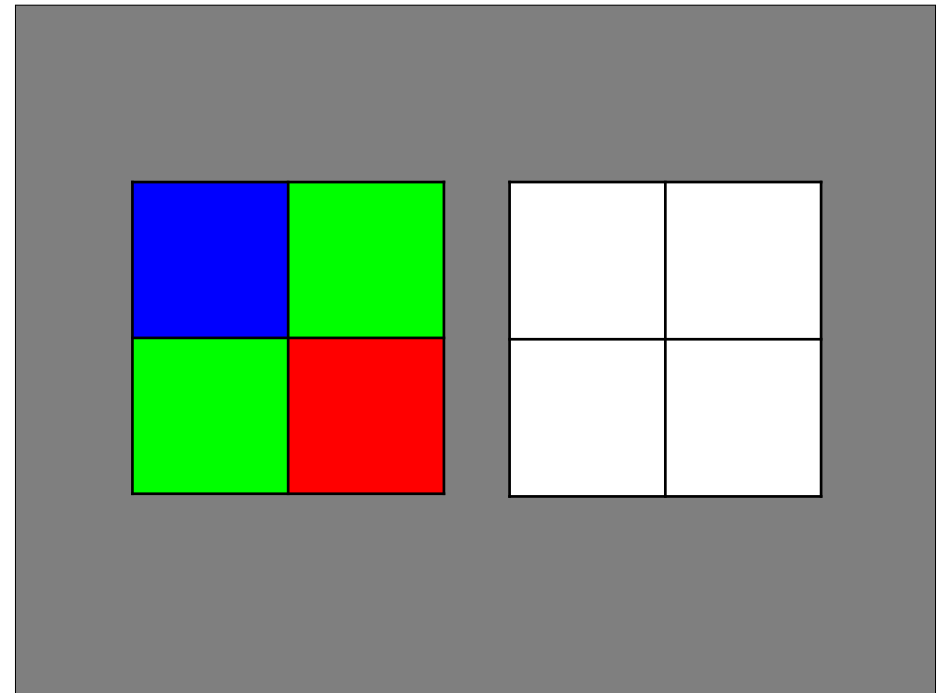
25



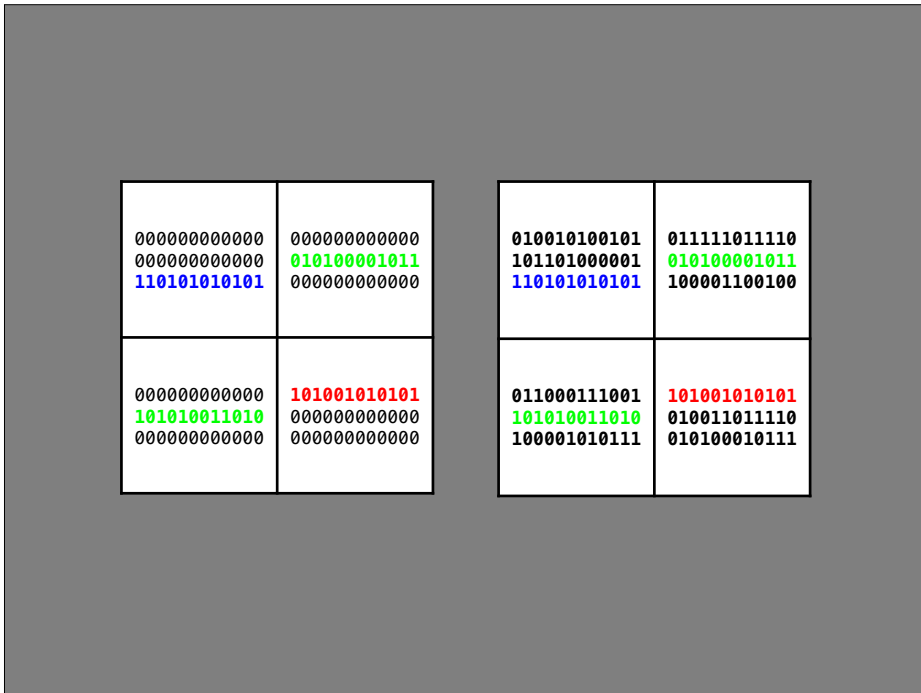
26



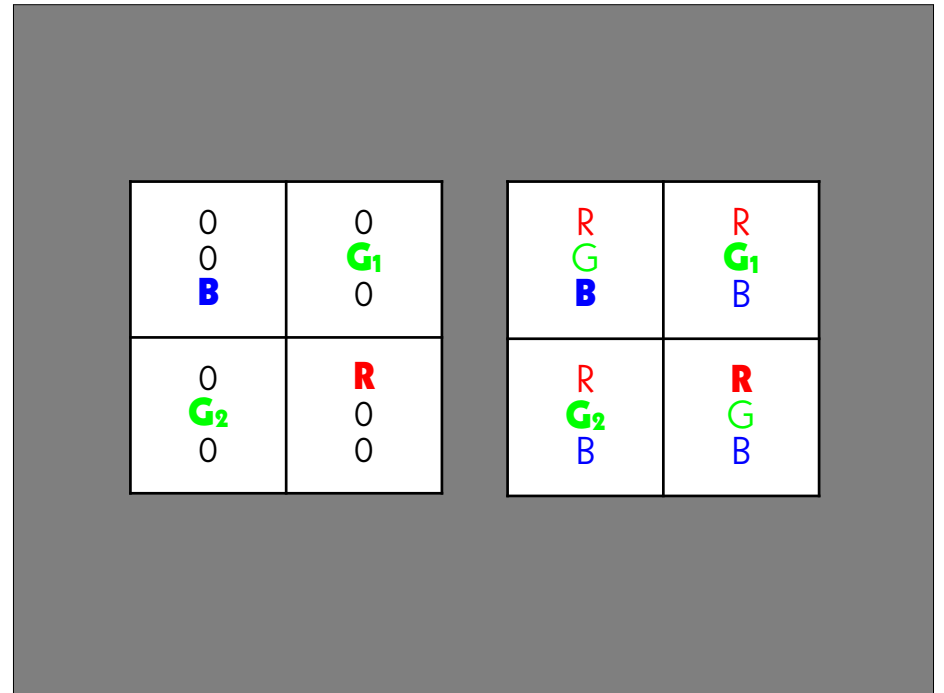
27



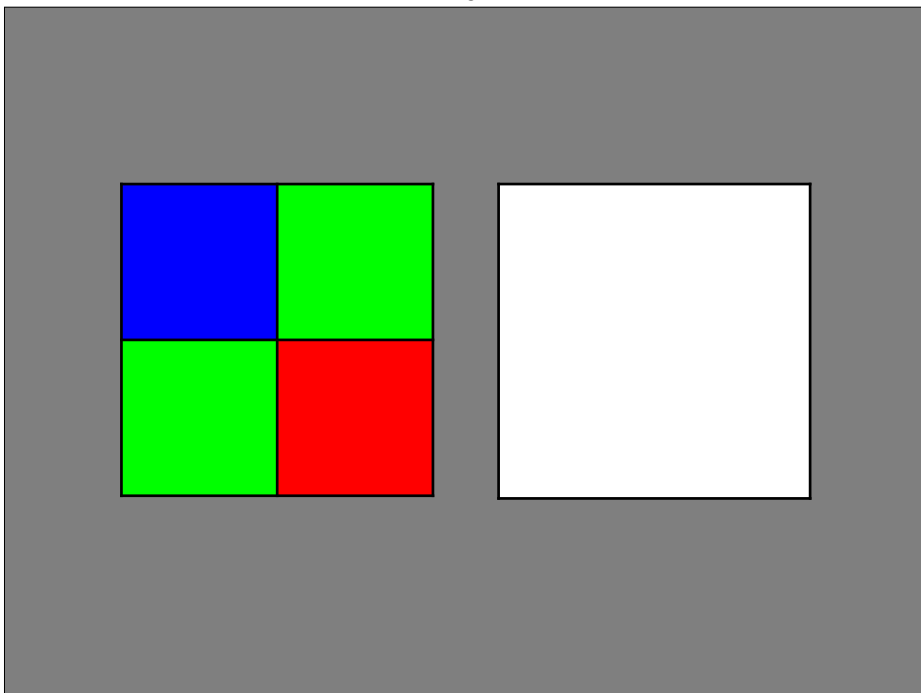
28



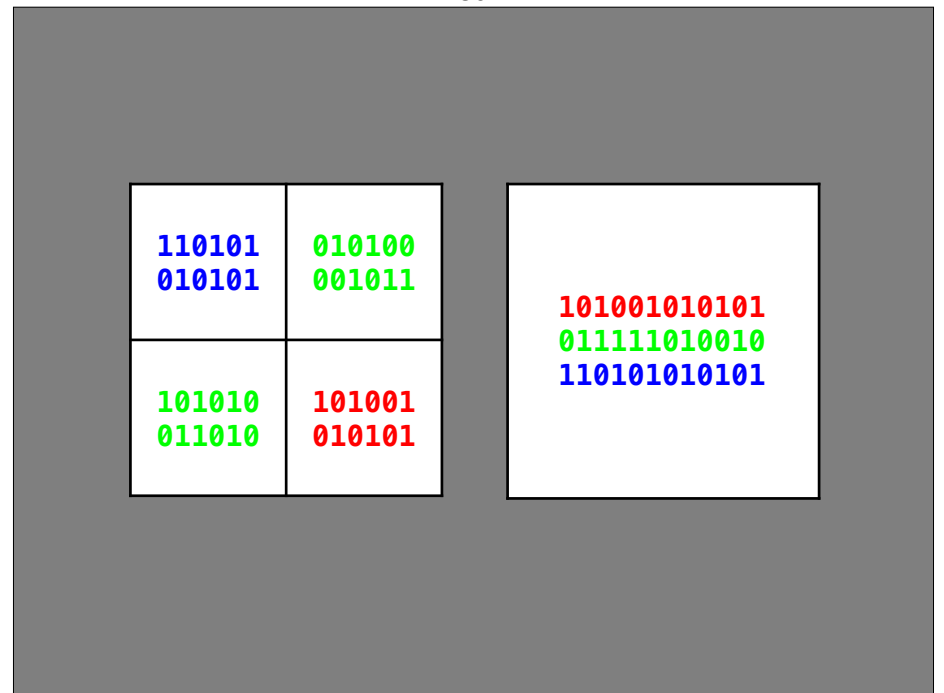
29



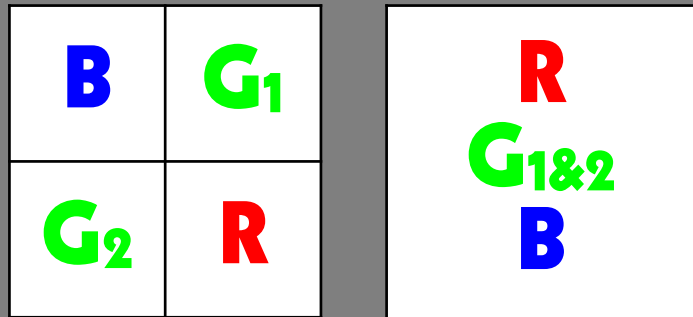
30



31



32



33

Zwei Möglichkeiten, Bayer-Daten zu verwenden

digitales Aufblasen auf RGB

- die generierten Daten werden verdreifacht
- die Datei hat die volle Sensorauflösung
- nur die Hälfte der Daten ist real

digitale Reduktion auf RGB

- drei Viertel der generierten Daten sind gespeichert
- die Datei hat die halbe Sensorauflösung
- die gesamten Daten sind real

35

Deux façons d'utiliser les données Bayer

gonflage numérique au RGB

- 3 fois la quantité de données générées
- le fichier a la pleine résolution du capteur
- seulement $\frac{1}{3}$ des données sont réelles

réduction numérique au RGB

- $\frac{3}{4}$ de la quantité de données générées
- le fichier a $\frac{1}{2}$ de la résolution du capteur
- toutes les données sont réelles

34

Structure et format de fichiers

Dateiaufbau und Dateiformate

36

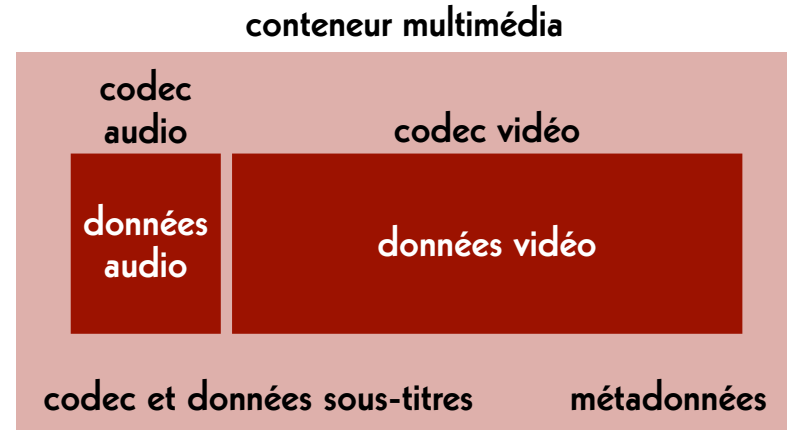
```

0111010100101010100010110101011110
0100110101010101010100001011101010
0111010100101010100010110101011110
0001110101010101010100001011101010
0110101010010101010001011101010111
0010101010101010100001011101010000
0111010100101010100010110101011110
0101010101010101000010111010100110
1001011101010010101010001011010101
1110010101010101010000101110101010
0111010100101010100010110101011110
010101010101010100110101010000001
00101000101010101010010101010101

```

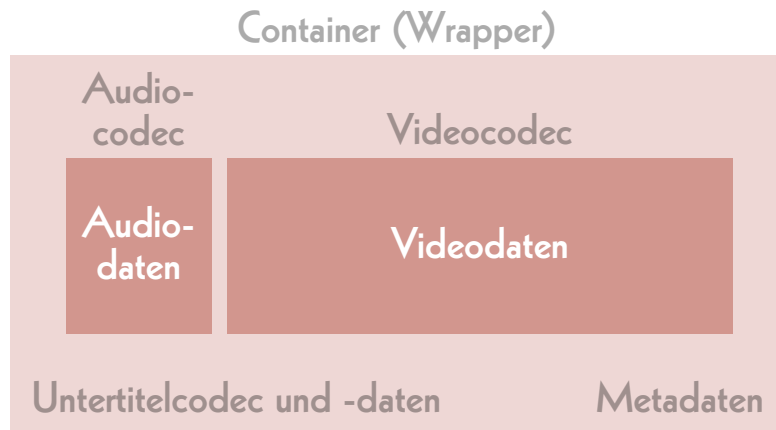
37

Structure des fichiers



38

Dateiaufbau



39

Formats d'archivage (actuel)

Archivformate (heute)

film

- dossier, TIFF, 2K ou 4K, RGB, 16 bit
- MXF, DPX, 2K ou 4K, R'G'B', 10 bit

vidéo

- AVI, «raw», HD, Y'CbCr 4:2:2, 10 bit
- Matroska, FFV1, HD, Y'CbCr 4:2:2, 10 bit

audio

- WAVE, 96 kHz, 24 bit
- BWF, 96 kHz, 24 bit

40

Formats mezzanine (actuel) Mezzanine (heute)

vidéo

- ProRes 4444, 2K
- DNxHR, 2K
- ProRes 422 HQ, HD
- DNxHD 175x, HD

audio

- WAVE, 48 kHz, 24 bit
- BWF, 48 kHz, 24 bit

41

Formats de diffusion (actuel) Zugangsformate (heute)

MP4

vidéo

- H.264, SD, Y'CbCr 4:2:0, 8 bit, lossy
- H.264, «HD», Y'CbCr 4:2:0, 8 bit, lossy

audio

- AAC, 44,1 kHz, 16 bit
- AAC, 48 kHz, 16 bit

42

Formats d'archive et mezzanine Archivformate und Mezzanine

film

- Matroska, FFV1, 4K ou 2K, R'G'B', 12 bit

vidéo

- Matroska, FFV1, «HD», Y'CbCr 4:2:2, 10 bit

audio

- Matroska, FLAC, 192 kHz, 32 bit

43

Formats de diffusion Zugangsformate

MP4

vidéo

- H.265, «HD», Y'CbCr 4:2:0, 8 bit, lossy
- H.266, «HD», Y'CbCr 4:2:0, 8 bit, lossy
- AV1, «HD», Y'CbCr 4:2:0, 8 bit, lossy

audio

- AAC, 96 kHz, 16 bit

44

Un seul format?

Ein einziges Format?

Reto Kromer: **Matroska and FFV1: One File Format for Film and Video Archiving?**

In: «Journal of Film Preservation», n° 96 (avril 2017), FIAF, Bruxelles, Belgique, p. 41–45

→ retokromer.ch/publications/JFP_96.html

45

conteneur:

- dossier
- TAR
- ZIP
- MXF
- Matroska
- AXF

codec:

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

47

Avantages et inconvénients

Vor- und Nachteile

46

	avantages	désavantages
TIFF DPX OpenEXR	traitement plus simple	fichiers plus lourds
JPEG 2000 FFV1	fichiers plus légers	traitement plus complexe

48

	Vorteile	Nachteile
TIFF DPX OpenEXR	Daten leichter zu bearbeiten	grössere Dateien
JPEG 2000 FFV1	kleinere Dateien	Daten komplexer zu bearbeiten

49

Conteneur MXF (.mxf)

MXF-Container (.mxf)

codec vidéo

- DPX
- JPEG 2000
- DNxHD, DNxHR
- ProRes 422, ProRes 4444

50

SMPTE RDD 48:2018

SMPTE REGISTERED DISCLOSURE DOCUMENT



MXF Archive and Preservation Format Registered Disclosure Document

Page 1 of 113

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.

All other inquiries in respect of this document, including inquiries as to intellectual property requirements that may be attached to use of the disclosed technology, should be addressed to the proponent identified below.

Proponent Contact Information:
 Kate Murray
 Library of Congress
 101 Independence Ave, S.E.
 Washington, DC 20540-1300

Email: kmur@loc.gov

51

MXF / DPX

MXF

→ SMPTE RDD 48:2018

DPX

→ SMPTE ST 268M:2015

52

MXF / JPEG 2000

MXF

→ SMPTE RDD 48:2018

JPEG 2000

→ ISO/IEC 15444-1:2019

→ etc.

53

MXF / DN_x

MXF

→ SMPTE RDD 48:2018

DN_xHD, DN_xHR

→ non publié

nicht veröffentlicht

54

MXF / ProRes

MXF

→ SMPTE RDD 48:2018

ProRes 422, ProRes 4444

→ SMPTE RDD 36:2015

55

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.

All other inquiries in respect of this document, including inquiries as to intellectual property requirements that may be attached to use of the disclosed technology, should be addressed to the proponent identified below.

Proponent contact information:

ProRes Program Office
Apple Inc.
1 Infinite Loop, MS: 77-2YAK
Cupertino, CA 95014
USA

Email: ProRes@apple.com

56

Conteneur Matroska (.mkv)

Matroska-Container (.mkv)

codec vidéo

- FFV1
- ProRes 422, ProRes 4444

57

Matroska/FFV1

Matroska (.mkv)

→ IETF Internet Draft

FFV1

→ IETF RFC 9043

58

Stream: Internet Engineering Task Force (IETF)
RFC: [9043](#)
Category: Informational
Published: August 2021
ISSN: 2070-1721
Authors: M. Niedermayer D. Rice J. Martinez

RFC 9043

FFV1 Video Coding Format Versions 0, 1, and 3

Abstract

This document defines FFV1, a lossless, intra-frame video encoding format. FFV1 is designed to efficiently compress video data in a variety of pixel formats. Compared to uncompressed video, FFV1 offers storage compression, frame fixity, and self-description, which makes FFV1 useful as a preservation or intermediate video format.

Status of This Memo

This document is not an Internet Standards Track specification; it is published for informational purposes.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Not all documents approved by the IESG are candidates for any level of Internet Standard; see Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at <https://www.rfc-editor.org/info/rfc9043>.

59

Matroska / ProRes

Matroska (.mkv)

→ IETF Internet Draft

ProRes 422, ProRes 4444

→ SMPTE RDD 36:2015

60

Format de fichier OpenEXR (.exr) OpenEXR-Dateiformat (.exr)

OpenEXR

- licence BSD à 3 clauses
3-Klausel-BSD-Lizenz
- non standardisé
nicht normiert

61

Sauvegarde et maintenance des données Datensicherung und -pflege

62

LTO

- LTO = Linear Tape-Open
- réponse de l'informatique aux banques et assurances
- génération LTO-1 introduite en 2000
- actuellement LTO-9
- actuellement le consortium LTO se compose de Hewlett Packard Enterprise, IBM et Quantum

63

LTO

- LTO = Linear Tape-Open
- Antwort der IT an den Banken- und Versicherungsektor
- LTO-1 kam 2000 auf den Markt
- LTO-9 ist die neuste Generation
- zur Zeit bilden Hewlett Packard Enterprise, IBM und Quantum das LTO Consortium

64

LTO-8

- rétrocompatibilité de lecture d'une seule génération
- le format M8 sont des cartouches LTO-7 formatées comme LTO-8
- le format M8 ne peut être utilisé que sur les lecteurs LTO-8

65

LTO-8

- nur LTO-7 kann gelesen werden, nicht aber LTO-6
- im Format M8 sind LTO-7-Kassetten, die als LTO-8 formatiert sind
- M8 kann ausschliesslich von LTO-8-Geräten gelesen werden

66

LTO-9

- appareils fabriqués uniquement par IBM
- cartouches fabriquées uniquement par Fujifilm et Sony
- rétrocompatibilité de lecture pour le LTO-8 standard (L8), mais pas le M8

67

LTO-9

- Geräte werden ausschliesslich von IBM hergestellt
- Kassetten werden ausschliesslich von Fujifilm und Sony hergestellt
- LTO-9-Geräte können reguläre LTO-8-Kassetten (L8) lesen, nicht aber M8

68

Capacité de stockage Speicherkapazität

génération Generation	non compressé nicht komprimiert
LTO-5	1,5 TB
LTO-6	2,5 TB
LTO-7	6 TB
LTO-8	12 TB
LTO-9	18 TB

69



70

Formattage

TAR

- de LTO-1 à LTO-4 il n'y avait que ce format
- il est toujours possible aujourd'hui

LTFS

- possible (et recommandé) dès LTO-5

71

Die Formatierung

TAR

- von LTO-1 bis LTO-4 gab es grundsätzlich nur diese Möglichkeit
- diese Möglichkeit ist nach wie vor möglich

LTFS

- gibt es ab LTO-5

72

TAR

- TAR = TApe Archive
- TAR seul
 - taille des blocs
 - nombre d'archives par cartouche
 - archives nécessitant plus qu'une cartouche
- TAR avec une structure des données propriétaire (p.ex. BRU, Retrospect)

73

TAR

- TAR = TApe Archive
- reines TAR
 - Blockgrösse
 - Anzahl Archive auf der Kassette
 - Archive über mehrere Kassetten
- TAR mit einer proprietären Datenstruktur (z. B. mittels der Software «BRU» oder «Retrospect»)

74

LTFS

- plusieurs versions
- nombreuses implémentations, mais...
- ... «ltfs» et «mklts» sont des commandes communes
- compression sans perte (réglage par défaut) ou pas de compression
- avec cryptage des données ou sans cryptage (réglage par défaut)

75

LTFS

- LTFS = Linear Tape File System
- mehrere Versionen (2.5.1 ist die letzte)
- Implementierungen teils Herstellerabhängig...
- ... aber «ltfs» und «mklts» sind gemeinsame Befehle
- verlustfreie oder keine Kompression
- unverschlüsselte oder verschlüsselte Daten

76

Assurer la pérennité

- conditions de stockage des cassettes
- trois copies ...
- ... géographiquement séparées
- vérification de l'intégrité des données
- migration des données
- disponibilité des appareils LTO

77

Langzeit

- Lagerung der Kassetten
- drei Kopien ...
- ... in drei geografisch getrennten Orte
- Prüfung der Datenintegrität
- Datenmigration
- Verfügbarkeit von LTO-Lesegeräte

78

Prévoir la migration

- convention de nommage
- utiliser des code-barres
- utiliser des sommes de contrôle
- ajouter la liste de tous les fichiers de la cartouche
- inclure les métadonnées techniques
- ajouter le code pour récupérer les fichiers

79

Die nächste Migration planen

- Dateinamen
- Strichcode
- Prüfsumme
- das volle Verzeichnis auf die Kasette schreiben
- technische Metadaten
- Code zum Wiederherstellen der Dateien

80

Sommes de contrôle Prüfsummen

cryptographique kryptografische

- MD5
- SHA-1
- SHA-256
- SHA-512

non cryptograph. nicht kryptogr.

- CRC-32
- xxHash 32
- xxHash 64
- xxHash 128

81

lire → modifier → écrire
lesen → ändern → schreiben

script pour modifier:

Skripte zum Verändern von:

- conteneur
- codec
- conteneur et codec
- métadonnées
- nommage des fichiers

82

Le bon côté de la migration Die gute Seite der Migration

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

→ retokromer.ch/publications/IASA_49.html

83

Open Source

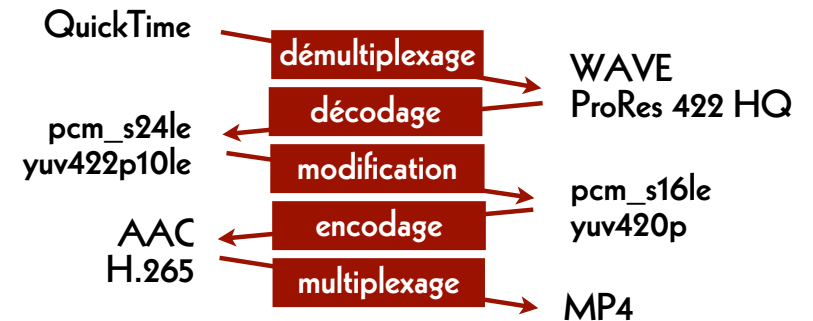
84

Aides à l'identification Hilfsmittel zur Identifizierung

- Brian R. Pritchard: «Identifying 35 mm Films», «Identifying 16 mm Films» & «Identifying Motion Picture Film Sound Tracks»
→ www.brianpritchard.com/IMPF.htm
- «Videotape Identification and Assessment Guide»
→ www.arts.texas.gov/wp-content/uploads/2012/04/video.pdf

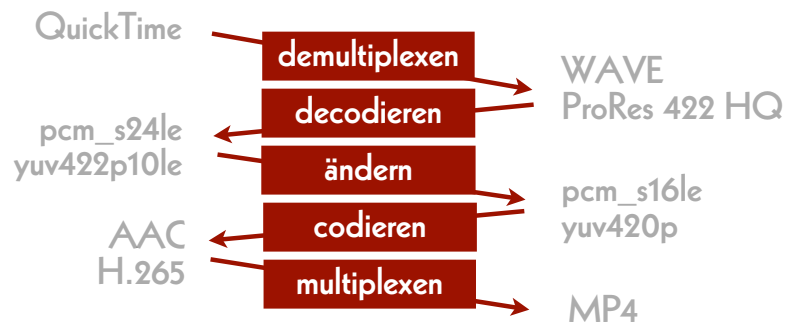
85

Exemple: transformation d'un fichier audiovisuel



86

Beispiel: Umwandlung von Bild und Ton



87

Changements de format Formatumwandlungen

ffmpeg (CLI)

→ ffmpeg.org

Cookbook for Archivists

→ avpres.net/FFmpeg/

ffmpegprovisr

→ amiaopensource.github.io/ffmpegprovisr/

88

Une interface graphique pour FFmpeg Ein GUI für FFmpeg

FFCommand Engine (GUI)

→ github.com/ColorlabMD/FFCommand_Engine

version 0.6 for macOS via Homebrew:

```
brew tap avpres/formulae
```

```
brew install --HEAD ffcommand-engine
```

89

Extraction des métadonnées Metadaten extrahieren

MediaInfo (GUI, CLI)

→ mediaarea.net/MediaInfo

ffprobe (CLI)

→ ffmpeg.org

90

Lecteurs multimédia Mediaplayer

VLC (GUI)

→ www.videolan.org/vlc/

mpv (CLI)

→ mpv.io

ffplay (CLI)

→ ffmpeg.org

91

L'anneau manquant Die Brücke zwischen den zwei Welten

RAWcooked (CLI)

→ mediaarea.net/RAWcooked

92

RAWcooked

- encodage dans Matroska (.mkv) en utilisant le codec vidéo FFV1 et le codec audio FLAC
- toutes les métadonnées conservées
- décodage avec réversibilité bit à bit
- possibilité d'intégrer des fichiers «sidecar» (p.ex. MD5, LUT, XML)
- compatibilité avec lecteurs multimedia

93

RAWcooked

- Kodierung in Matroska (.mkv) mit FFV1-Video codec und FLAC-Audiocodec
- alle Metadaten bleiben erhalten
- vollständige, bitgenaue Reversibilität
- Möglichkeit zum Einbetten von Sidecar-Dateien (z. B. MD5, LUT, XML)
- Kompatibilität mit Mediaplayern

94

Contrôle de qualité Qualitätskontrolle

QCTools (GUI), **qcli** (CLI)

→ bavc.org/programs/preservation/preservation-tools/

QCTools Documentation

→ bavc.github.io/qctools/

AV Artifact Atlas

→ www.avartifactatlas.com

95

Extraction du son optique Extrahieren des optischen Tons

AEO-Light (GUI)

→ usc-imi.github.io/aeo-light/

AEO-Light Manual

→ github.com/usc-imi/aeo-light/releases/download/v2.2-beta/AEO-Light-2.2-beta-Manual.pdf

96

Encodeur (et lecteur) DCP DCP-Encoder (und -Player)

DCP-o-matic (GUI)

→ dcpomatic.com

DCP-o-matic users' manual

→ dcpomatic.com/manual/html/

Create DCPs: DCP-o-matic

→ avpres.net/training/DCP-o-matic.html

97

AMIA Open Source

vrecord (CLI)

→ github.com/amiaopensource/vrecord

audioqc

→ github.com/amiaopensource/audioqc

... etc.

98

MediaArea

DVRRescue

→ mediaarea.net/DVRRescue

MediaConch

→ mediaarea.net/MediaConch

... etc.

99

AV Preservation by reto.ch

Bash Scripts for Audio-Visual Preservation

→ avpres.net/Bash_AVpres/

... etc.

100

FADGI

Federal Agencies Digital Guidelines Initiative (USA)

→ digitizationguidelines.gov

101

CECO KOST

Centre de coordination pour l'archivage à long terme de documents électroniques (Suisse)

Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen (Schweiz)

→ kost-ceco.ch

102

Nestor

Eine kleine Enzyklopädie der digitalen Langzeitarchivierung (Deutschland)

→ nestor.sub.uni-goettingen.de/handbuch/

103

Questions d'éthique Ethische Fragen

104

1^{er} principe 1. Grundsatz

- Toute action entreprise doit augmenter la possibilité qu'une œuvre ou un document audiovisuel reste disponible dans son intégrité aussi à l'avenir.

Die Wahrscheinlichkeit, dass in ein Werk in seiner Integrität weiter erhalten bleibt, ist grösser.

105

3^e principe 3. Grundsatz

- Chaque acte effectué doit être soigneusement documenté.

Jeder Bearbeitungsschritt wird sorgfältig dokumentiert.

107

2^e principe 2. Grundsatz

- Toute action entreprise doit maintenir ouvertes toutes les possibilités existant auparavant, afin que nos successeurs puissent améliorer ou refaire le travail, en utilisant le matériel d'origine dans les mêmes conditions que nous.

Alle Möglichkeiten der Bearbeitung, die vor einem Eingriff gegeben waren, bleiben auch nach dem Eingriff weiter bestehen.

106

AV Preservation by reto.ch

zone industrielle Le Trési 3
1028 Préverenges
Switzerland

Web: reto.ch
Twitter: @retoch
Email: info@reto.ch



108