

Chancen und Risiken des Einsatzes verlustbehafteter Bildkompression in der digitalen Archivierung

Dr. Kai Naumann, Landesarchiv BW
Dr. Christoph Schmidt, Landesarchiv NRW

AK AUdS
Basel, 28.2.2017

Computerzeitschrift c't, Okt. 1997

Qualitätsverlust durch wiederholtes Speichern

PRAXIS & TIPPS | TIPPS & TRICKS

📅 29.10.1997

Ich arbeite momentan viel mit Bilddateien. Weil ich meine Festplatte entlasten möchte, habe ich folgende Frage: Tritt beim mehrmaligen Speichern eines JPG-Bildes ein Qualitätsverlust auf?

Ja! Das JPEG-Format eignet sich nur bedingt zur Zwischenspeicherung, da sich die Ergebnisse mit jedem Speichervorgang verschlechtern. Will man ein Motiv in mehreren Arbeitsschritten bearbeiten, so empfiehlt es sich, ein Bitmap-Format zu verwenden, das ohne Informationsverlust komprimiert. Hier haben sich vor allem das LZW-komprimierte TIFF- und das PNG-Format durchgesetzt (Siehe auch c't 10/97, S. 292). Erst nach dem letzten Bearbeitungsschritt empfiehlt es sich, auf JPEG umzusteigen. ([pen](#))

Computerzeitschrift c't, Nov. 2016



Überlebenskünstler

Gut genug: Seit einem Vierteljahrhundert behauptet sich das JPEG-Format

Vor 25 Jahren erschien libjpeg – eine Open-Source-Bibliothek zum Lesen und Schreiben des JPEG-Bildformats. Seitdem hat der Foto-Standard einige Patentstreitigkeiten und vermeintliche JPEG-Killer überlebt.

Ziel, ein Standard-Dateiformat für die Speicherung von Fotos zu entwickeln. Sie gab nur die Komprimierungsmethode vor, nicht aber, wie die Bilddaten abgelegt werden. 1991 entwickelte Eric Hamilton mit dem Grafikformat JFIF eine solche Methode zum Speichern von JPEG-Dateien.

ergänzt, die zum Standardisierungszeitpunkt noch patentbehafet war.

Bis heute wehrt sich das in seiner Basisausführung etablierte Format zäh dagegen, vom Thron geschubst zu werden: Während GIF aufgrund von Patentstreitigkeiten kurzerhand durch PNG ersetzt wurde, mussten Firmen mit angeblichem JPEG-Patent, die Lizenzgebühren eintreiben wollten, sogar die Aufhebung ihrer Patente befürchten. Die vom JPEG-Komitee selbst zum Nachfolger erkorene JPEG2000-Norm konnte weder Digital-kamera- noch Browser-Hersteller für sich gewinnen. JPEG2000 basiert auf einer neueren Kompressionsmethode, der diskreten Wavelet-Transformation (DWT), die weniger auffällige Artefakte erzeugen sollte als JPEG.

Der vermeintliche Vorteil war allerdings eher theoretischer Natur, weil die gefürchteten JPEG-Blockartefakte nur bei sehr hohen Kompressionsraten auftreten. In der Praxis wird aber überwiegend hohe bis mittlere Qualität benötigt – und hier wirken JPEG-komprimierte Fotos tendenziell sogar schärfer als JPEG2000-komprimierte. JPEG2000 scheiterte außerdem am deutlich höheren Rechenaufwand und einer unklaren Patentsituation. Weitere JPEG-Killer-Anwärter waren WebP von Google und HD Photo von Microsoft.

Ziel und Gliederung des Vortrags

- Ziel: Beförderung einer ergebnisoffenen Diskussion!
- Gliederung
 - Was ist eigentlich ein „Verlust“?
 - Verluste im Labor - Vier Szenarien
 - Exkurs: Verluste durch Bit-Rot
 - Gute Gründe für JPEG
 - Fazit

Was ist eigentlich ein „Verlust“? (1)

- Verlust i.S. Qualitätsverlust ist zunächst nur eine Veränderung
- Wir müssen definieren, welcher Verlust unerwünscht und welcher Verlust erlaubt ist

TIFF

=>

JPEG



Was ist eigentlich ein „Verlust“? (2)

- Verlust i.S. Qualitätsverlust ist zunächst nur eine Veränderung
- Wir müssen definieren, welcher Verlust unerwünscht und welcher Verlust erlaubt ist

TIFF

=>

TIFF (Graustufen)



Was ist eigentlich ein „Verlust“? (3)

- Es gibt keinen absoluten Maßstab dafür, welche Verluste akzeptabel sind.
- Verluste sollten die signifikanten Eigenschaften des Erhaltungsobjekts nicht beeinträchtigen.
- Verluste durch verlustbehaftete Bildkomprimierung lassen sich experimentell feststellen.

Verluste im Labor – Szenario 1

- Szenario 1: Foto aus einer PDF/A-Datei
- Solche PDF/A-Dateien sind sehr verbreitet. Die meisten Kopierer erstellen JPEG in PDF/A.
- Verluste könnten im Archiv bei fortgesetzten verlustbehafteten Formatmigrationen entstehen.
- Doch wie groß werden die Verluste?
- Zwei Experimente sollen Aufschluss geben.



Nadtmann/Schmidt, AOUS Basel, 26.2.2017



Verluste im Labor – Experiment 1a



Differenz zu Ausgangsbild



Veränderung:
0,06 % der Pixel

Differenz zu Ausgangsbild



Veränderung:
1,13 % der Pixel



Verluste im Labor – Experiment 1b



Differenz zu Ausgangsbild



Veränderung:
0,009 % der Pixel

Differenz zu Ausgangsbild



Veränderung:
0,01 % der Pixel



Verluste im Labor Szenario 1



20, 60, 200
Jahre später



- Sind 0,01 oder 1,13 %
der Pixel ein
hinnehmbarer Verlust?



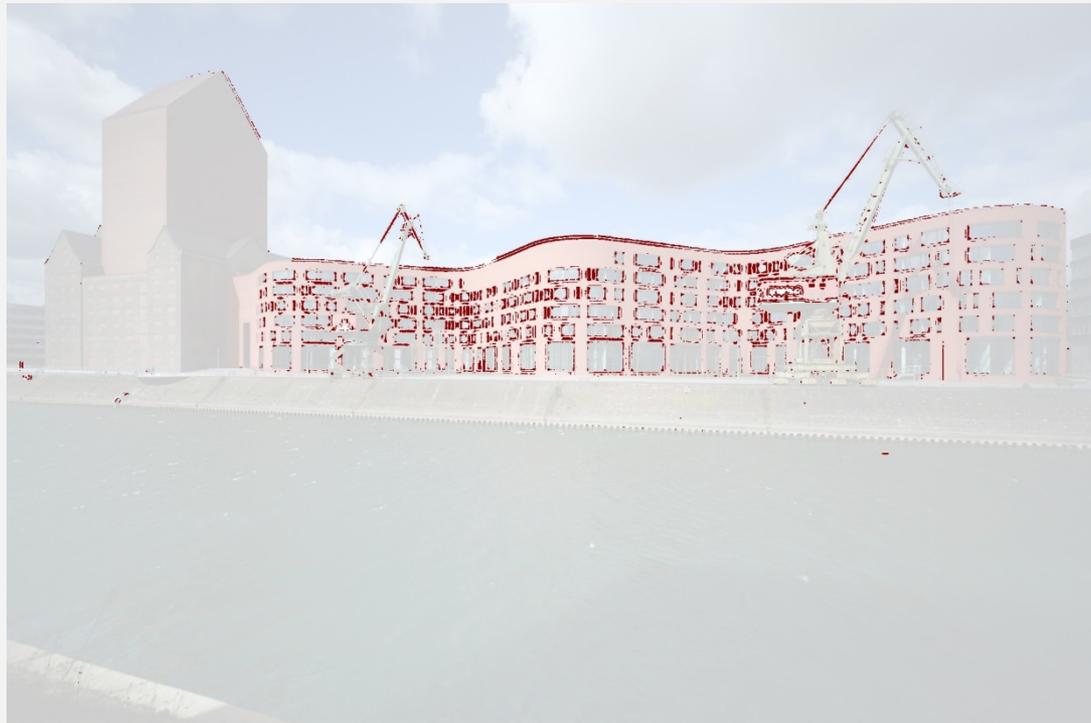
Verluste im Labor – Szenario 2

- Szenario 2: Bildveränderungen durch mehrfaches Speichern



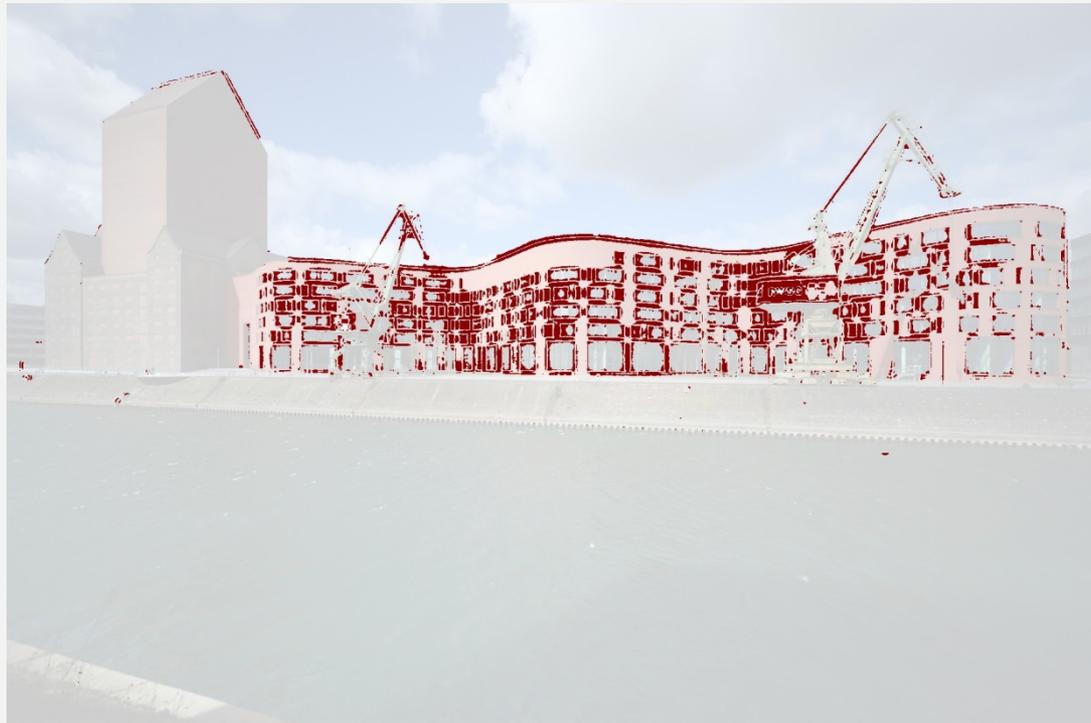
Original

Verluste im Labor – Szenario 2



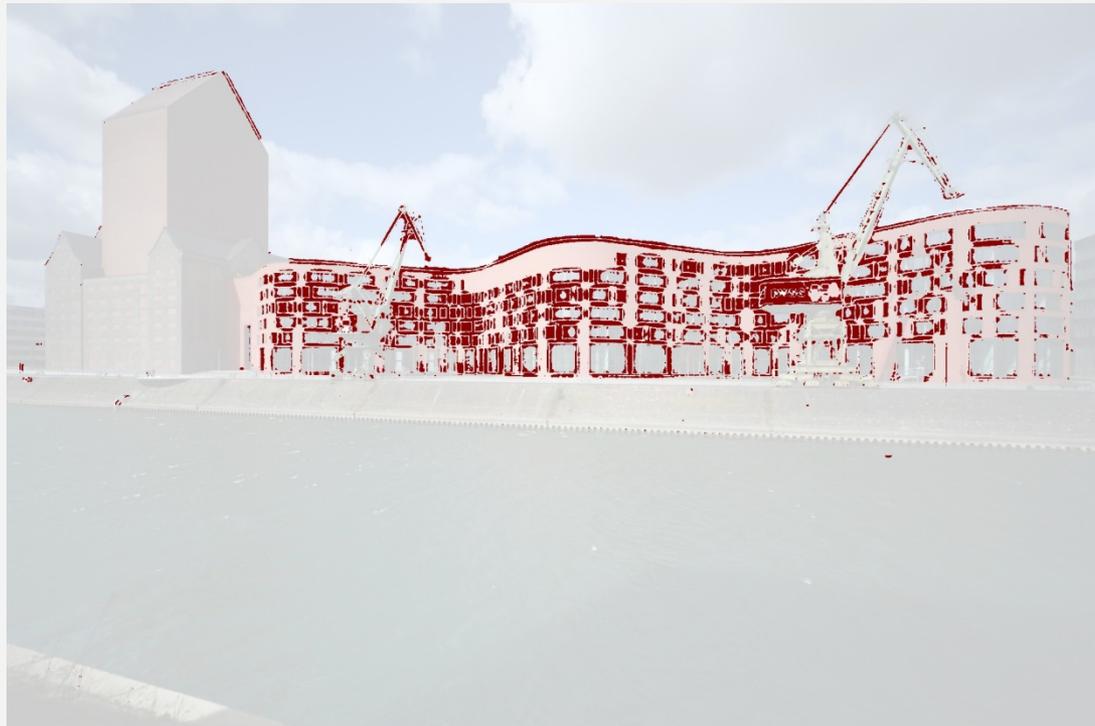
10x neu kodiert: 1,419% Pixelveränderungen

Verluste im Labor – Szenario 2



100x neu kodiert: 3,865% Pixelveränderungen

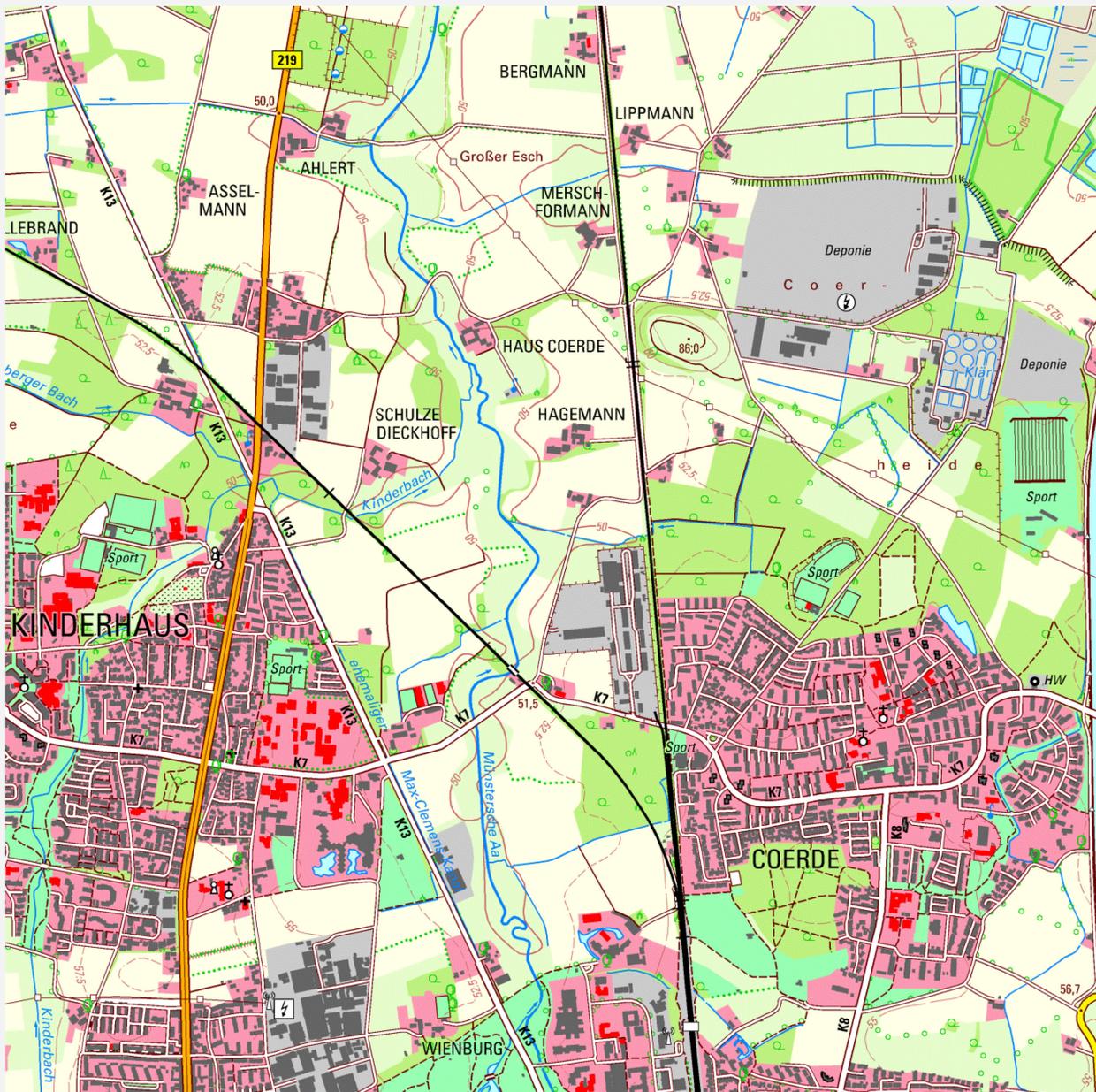
Verluste im Labor – Szenario 2



1000x neu kodiert: 3,873% Pixelveränderungen

Verluste im Labor – Szenario 3

- Szenario 3: Konvertierung einer flächigen TIFF-Grafik nach JPG 100%

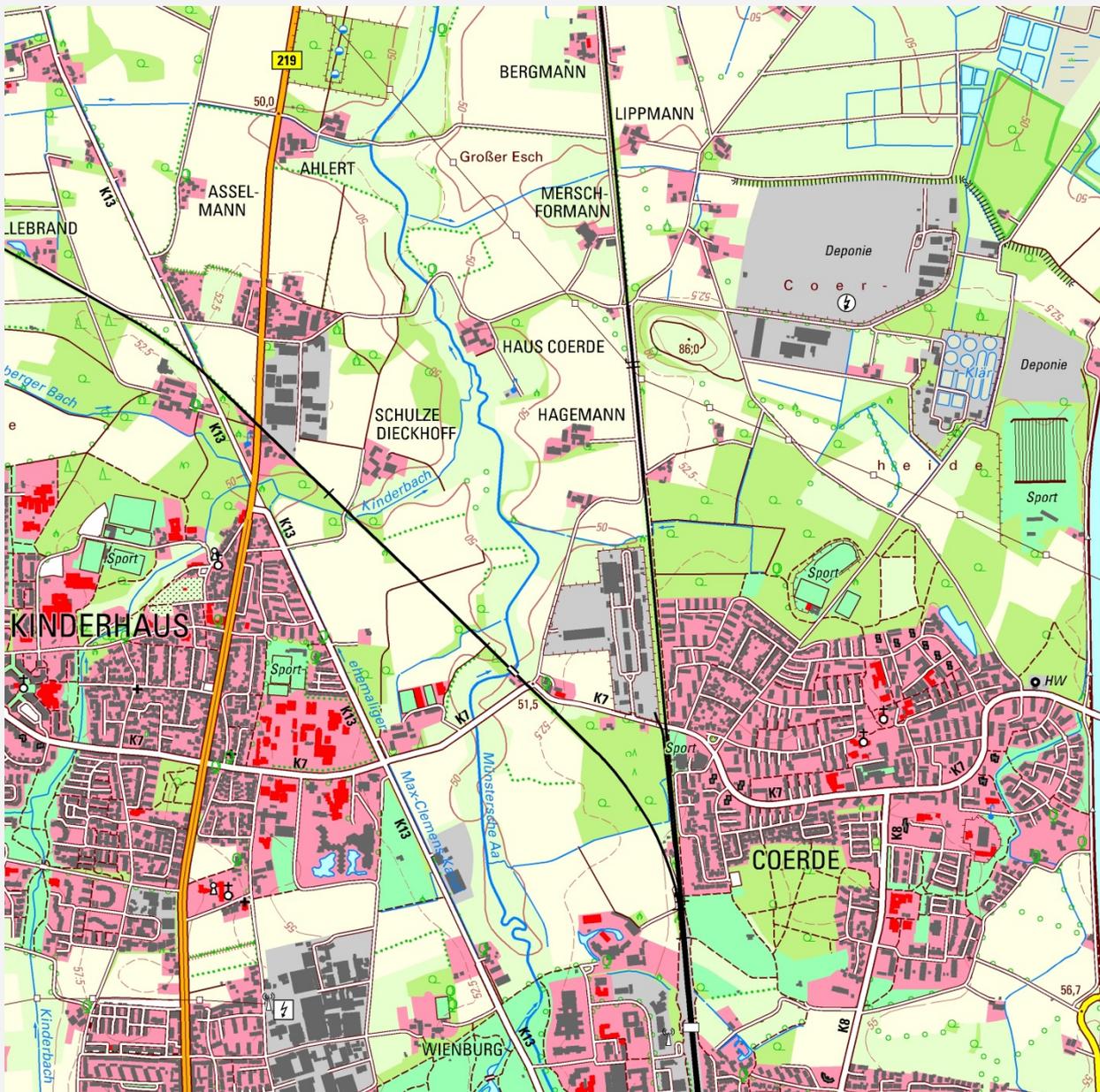


Original:

- Größe: 3200x3200
- Farbtiefe 8 Bit
- Nur 18 Farben (von 256 möglichen) verwendet
- Größe: 1,1 MB
- RLE-komprimiert

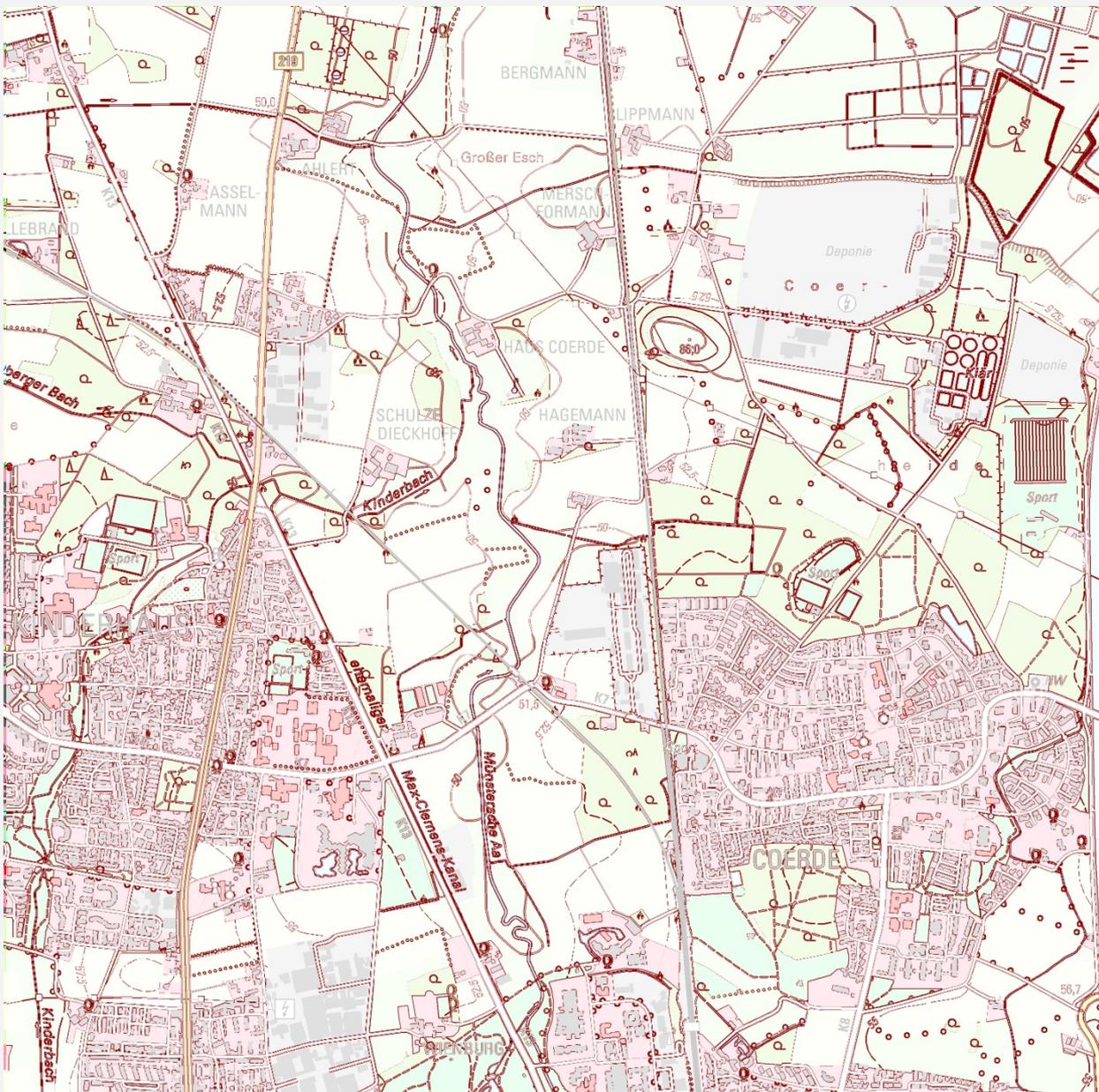
In amtlichen Kartenwerken haben Farben eine signifikante Bedeutung!





JPG 100%:

- Größe: 3200x3200
- **Farbtiefe 24 Bit**
- **36504 Farben real verwendet**
- **Größe: 5,7 MB**
- **Komprimiert**



JPG 100%:

- Größe: 3200x3200
- **Farbtiefe 24Bit**
- **36504 Farben (real)**
- **Größe: 5,7 MB**
- **Komprimiert**
- **9,282%**
abweichende Pixel



Differenzbild

Verluste im Labor – Szenario 4 (1)

- Szenario 4:
Reprowerkstatt
– Gescannt wird ein
Kostümentwurf
(Folio), hier nur
obere Hälfte sichtbar



, 28.2.2017



Ausschnitt TIFF-Datei



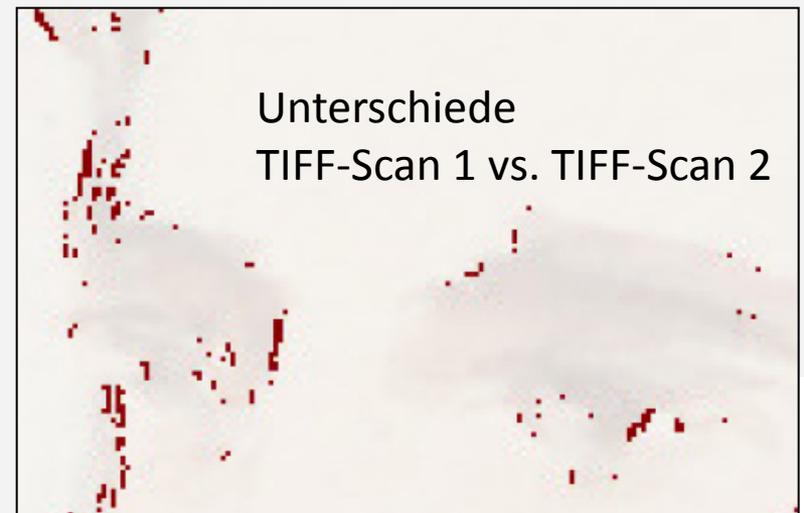
Ausschnitt JPEG-Datei (aus TIFF)

Verluste im Labor – Szenario 4 (2)

- Szenario 4:
Reprowerkstatt
 - Ein neuer Scan des Archivals erzeugt ähnlich viele Unterschiede wie eine TIFF-JPEG-Umwandlung!
 - Sind diese Unterschiede signifikant?



Differenzbilder

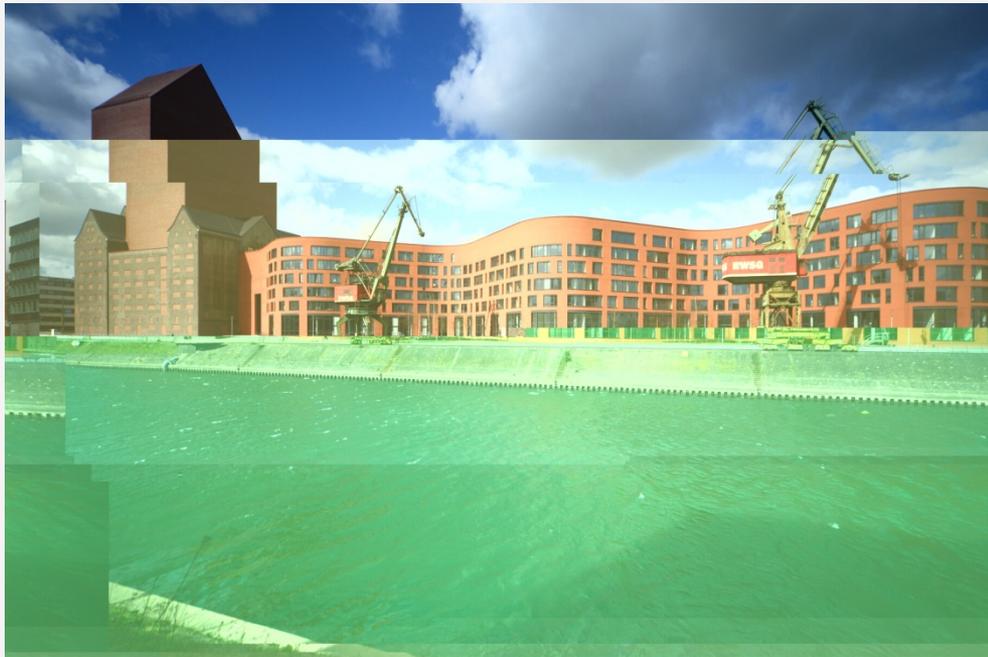


Verluste im Labor – Fazit

- Der Einsatz verlustbehafteter Kompressionsverfahren führt zu nachweisbaren Veränderungen am Bild.
- Der Grad der Veränderungen ist abhängig von:
 - Häufigkeit der Neukodierungen
 - Bildstruktur
 - evtl. Bildbearbeitungen (Weichzeichnung, Scharfzeichnung)
 - Verwendeter Software
- Verlustbehaftete Kompression eignet sich nicht für alle Bildtypen.
- Verlustbehaftete Kompressionsverfahren sollten erst am Ende eines Bearbeitungsprozesses stehen.
 - Gibt es Ausnahmen im quasi unendlichen Bestandserhaltungsprozess?

Exkurs: Verluste durch Bit-Rot (1)

- Komprimierte Dateiformate sind anfälliger gegen Bit-Rot



JPG-Datei, Bit-Rot von 0,0002%, 79,74% Pixelfehler

Exkurs: Verluste durch Bit-Rot (2)

- Komprimierte Dateiformate sind anfälliger gegen Bit-Rot



TIFF-Datei (unkompr.), Bit-Rot von 0,0002%, <1% Pixelfehler

Exkurs: Verluste durch Bit-Rot (3)

Testaufbau: Zufälliger Bit-Rot (einzelne Bits oder Bitfolgen) bei je 50 identischen Bilddateien

Bit-Rot in %	Verlustquote Pixel TIFF (unkompr.)	Verlustquote Pixel JPG
0,01%	55,76%	99,86
0,001%	10,46%	97,43%
0,0002%	2,81%	92,79%

Exkurs: Verluste durch Bit-Rot (4)

Ist Bit-Rot-Prävention ein Argument bei der Auswahl von Formaten?

Ein Archiv für genuin digitale Unterlagen muss Bit-Rot verhindern können, um vertrauenswürdig zu sein.

„Banning compression from archives in the hope of minimizing the damage from bit rot is a foolish preservation strategy.“

(Gary McGath)

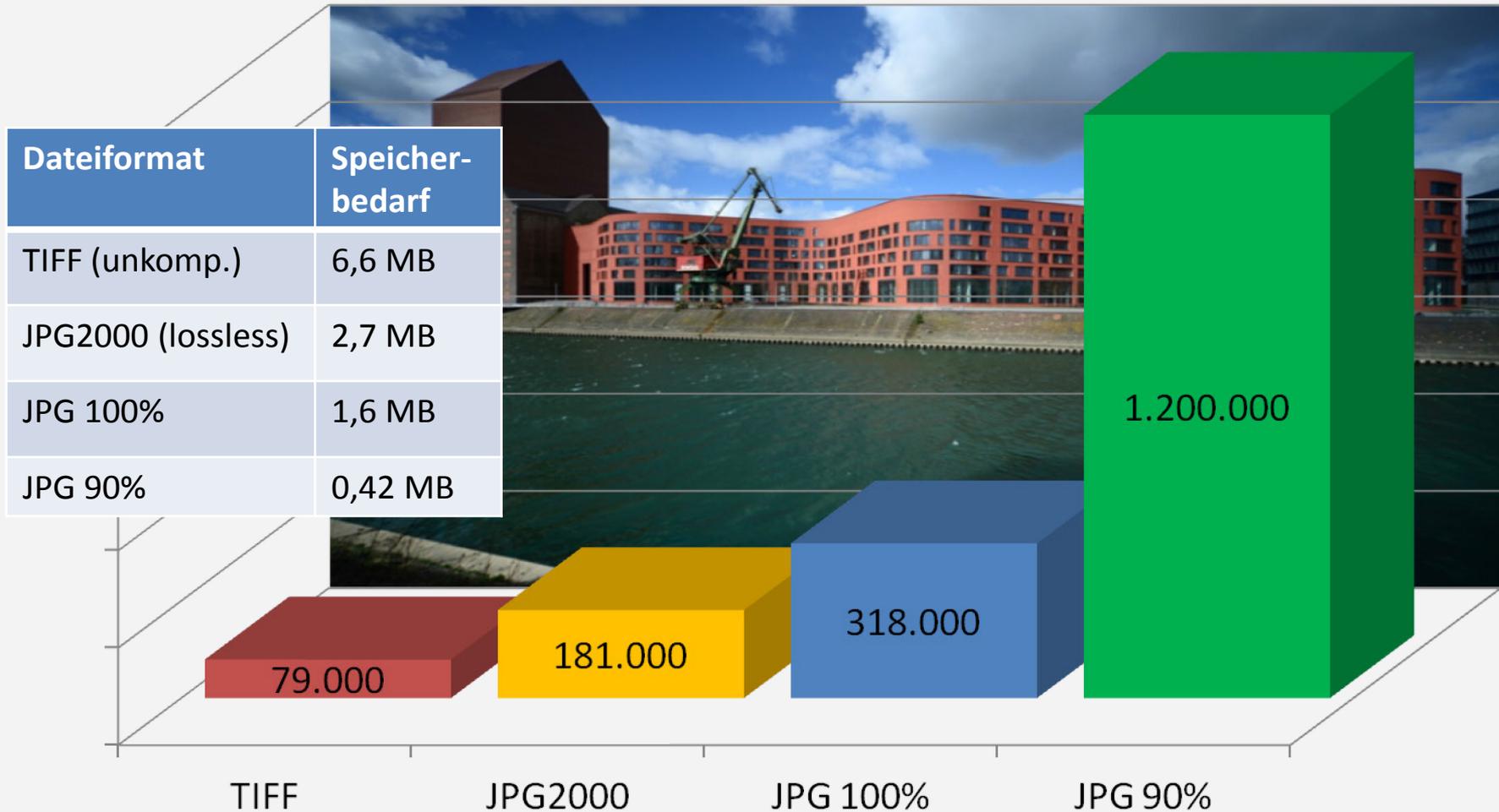
Gute Gründe für JPEG (1)

- JPEG ist längst im Archiv angekommen
- JPEG ist offen standardisiert
- JPEG ist bei Verbrauchern enorm verbreitet
 - Verfügbarkeit verschiedener, freier Viewer und Konverter & Werkzeuge zur Qualitätssicherung ist gewährleistet (anders als bei JPEG2000!)
- Aufgabe der Formatmigration wird breite Teile der IT-Welt betreffen

Gute Gründe für JPEG (2)

- Effiziente Speicherformate sparen Ressourcen bei Ingest, Storage, Preservation Planning und Access
- *Rechenbeispiel DIPS.kommunal:*
 - Kosten pro Mandant pro Jahr: € 20.000,-
 - Inkl. 500 GB Speicher
 - € 3,12 pro zusätzliches GB pro Jahr

Wie oft ließe sich das Beispielbild im Rahmen der Grundgebühr von DIPS.kommunal speichern?



Gute Gründe für JPEG (3)

- Verlustbehaftete Dateiformate können Verluste durch „kalten Kassation“ reduzieren!

Fazit (1)

1. **Verlustbehaftete Komprimierung gehört bereits zum Arbeitsalltag in den Archiven.**

- Wir sollten verantwortungsbewusst, aber entspannt damit umgehen.

Fazit (2)

2. Die Entscheidung für oder gegen verlustbehaftete Komprimierung ist nicht absolut zu treffen. Sie hängt vielmehr ab:

- Von den angenommenen signifikanten Eigenschaften des Archivals.
- Einer strategisch ausgewogenen Kosten-Nutzen-Abwägung.

Fazit (3)

3. Die Diskussion über verlustbehaftete Komprimierung kann versachlicht werden.

- Das Verlustrisiko lauert nicht in der Grundidee der verlustbehafteten Komprimierung, sondern in falschen Parametern.
- Falsche Parameter können im Entstehungsprozess erkannt und korrigiert werden.
 - Beispiel: wenn ein Bild von Format A in Format B migriert wird, und sich mehr als 0,001 Prozent der Pixel verändern, sind Parameter oder Algorithmus ungeeignet.

Fazit (4)

**Man sollte wissen,
was man tut.
Aber sollte man
das nicht immer?**

Danke fürs Zuhören!

Fon: ++49 7141 64854 6331
Dr. Kai Naumann

007

Landesarchiv Baden-Württemberg
Staatsarchiv Ludwigsburg
Arsenalplatz 3
D-71638 Ludwigsburg

Mail:
kai.naumann@la-bw.de

Fon: ☀☀ 49 251 620 650 13
Dr. Christoph Schmidt

0815

Landesarchiv Nordrhein-Westfalen
An den Speichern 11
D-48157 Münster

Mail:
christoph.schmidt@lan.nw.de

Ein paar Links & Literatur

Formatempfehlungen:

- [DFG-Praxisregeln „Digitalisierung“](#)
- [KOST-Katalog archivischer Dateiformate](#)

Speicherkosten:

- [Wirtschaftliche Digitalisierung in Archiven. Empfehlungen der Konferenz der Leiterinnen und Leiter der Archivverwaltungen des Bundes und der Länder](#)

Bildkonversion, Validierung von Konversionen:

- Massenhafte Konversion mit ImageMagick <https://www.imagemagick.org/>
- Pixelweiser Ähnlichkeitsvergleich mit KOST-Simy <http://kost-ceco.ch/cms/index.php?id=329,632,0,0,1,0>
- Veronika Krauß, Arefeh Bahrami: Ist das Bild noch das Bild? Authentizität digitaler Objekte unter Formatttransformationen in Kooperation mit dem Thüringischen Hauptstaatsarchiv, Präsentation AUdS 2016 Potsdam, http://www.staatsarchiv.sg.ch/home/auds/20/jcr_content/Par/downloadlist_1/DownloadListPar/download_0.ocFile/krauss_auds16_authentizitaet_dig_obj.pdf

Bit-Rot:

- Volker Heydegger, Just One Bit in a Million: On the Effects of Data Corruption in Files, in: Maristella Agosti (u.a.) (Hg.), Research and Advanced Technology for Digital Libraries. ECDL 2009, Berlin / Heidelberg 2009, S. 315-326 www.planets-project.eu/docs/papers/Heydegger_JustOneBit_ECDL2009.pdf
- [Gray McGath, Bit-rot tolerance doesn't work \(Blog-Beitrag\)](#)