

JPEG2000 and Google Books

Jeff Breidenbach



Google's mission is to organize the world's information and make it universally accessible and useful.



Mass digitization

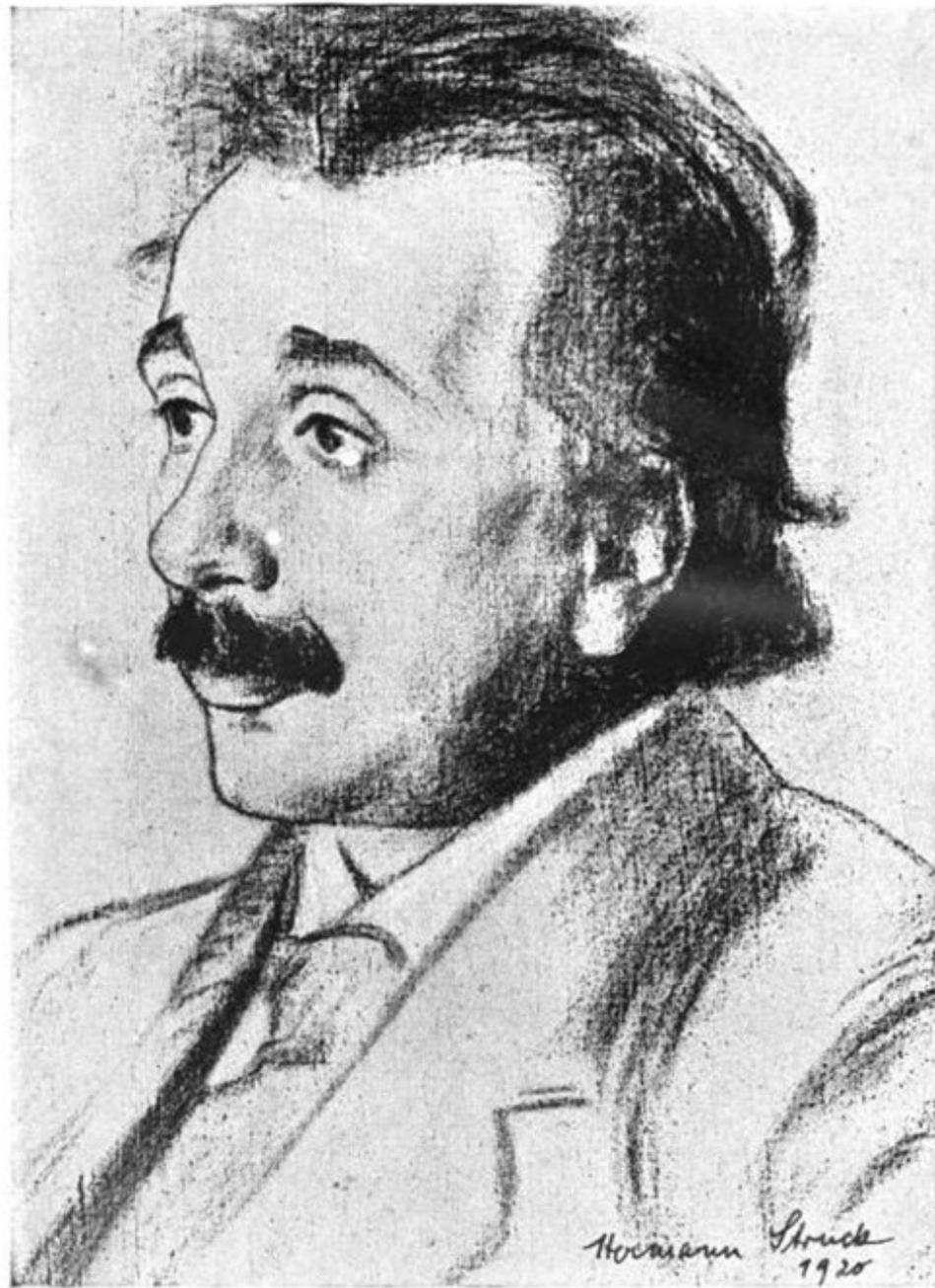
- broad coverage
- iteratively improve quality (reprocess, rescan)
- more than XXM books out of $XXXM$ since 2004
- XXX nominal pages per book
- billions of images, petabytes of data

JPEG2000

- pre-processed images
- processed illustrations and color images
- library return format
- illustrations inside PDF files



EPHESUS.



A. Einstein

Dem aller durchleuch-
tigsten/ Großmechtigsten Fürsten vnd
Herrn/ Herrn Ferdinando/ Römischen/ zu
Ungern vnd Böhem 1c. König/ Ertz-
herzog zu Oesterreich 1c. Meinem
aller Gnedigsten Herrn.



Aler Durchleuch-
tigster/ Großmechtig-
ster Römischer König
1c. Aller Gnedigster
HERR/ Wiewol vil
hochgelerte züvor/ mit
höchstem fleiß nicht on-
nuz vnnnd frucht der
glaubigen/ von der hai-
ligen Mes geschriben/
hab ich doch den güt-

hertzigen zü trost/ dise meine schrifft vnd Predig
von dem Ampt der heyligen Mes/ auß Christen-
lichem eyffer auch wöllen mittailen / Dieweil die
verachtung der Mes bey etlichen vngehorsamen
nicht abnimbt/ sonder täglich wechsť/ so auch die
ainfeltigen Catholischen mit irriger lehř / als ob
die Mes ein verdamlicher grewol were/ zü berede
vnd zü verwirren ganz geschwind vñ abgericht.

* ij Disen

Dem aller durchleuch-
tigsten/ Großmechtigsten Fürsten vnd
Herrn/ Herrn Ferdinando/ Römischen/ zu
Ungern vnd Böhem 1c. König/ Ertz-
herzog zu Oesterreich 1c. Meinem
aller Gnedigsten Herrn.

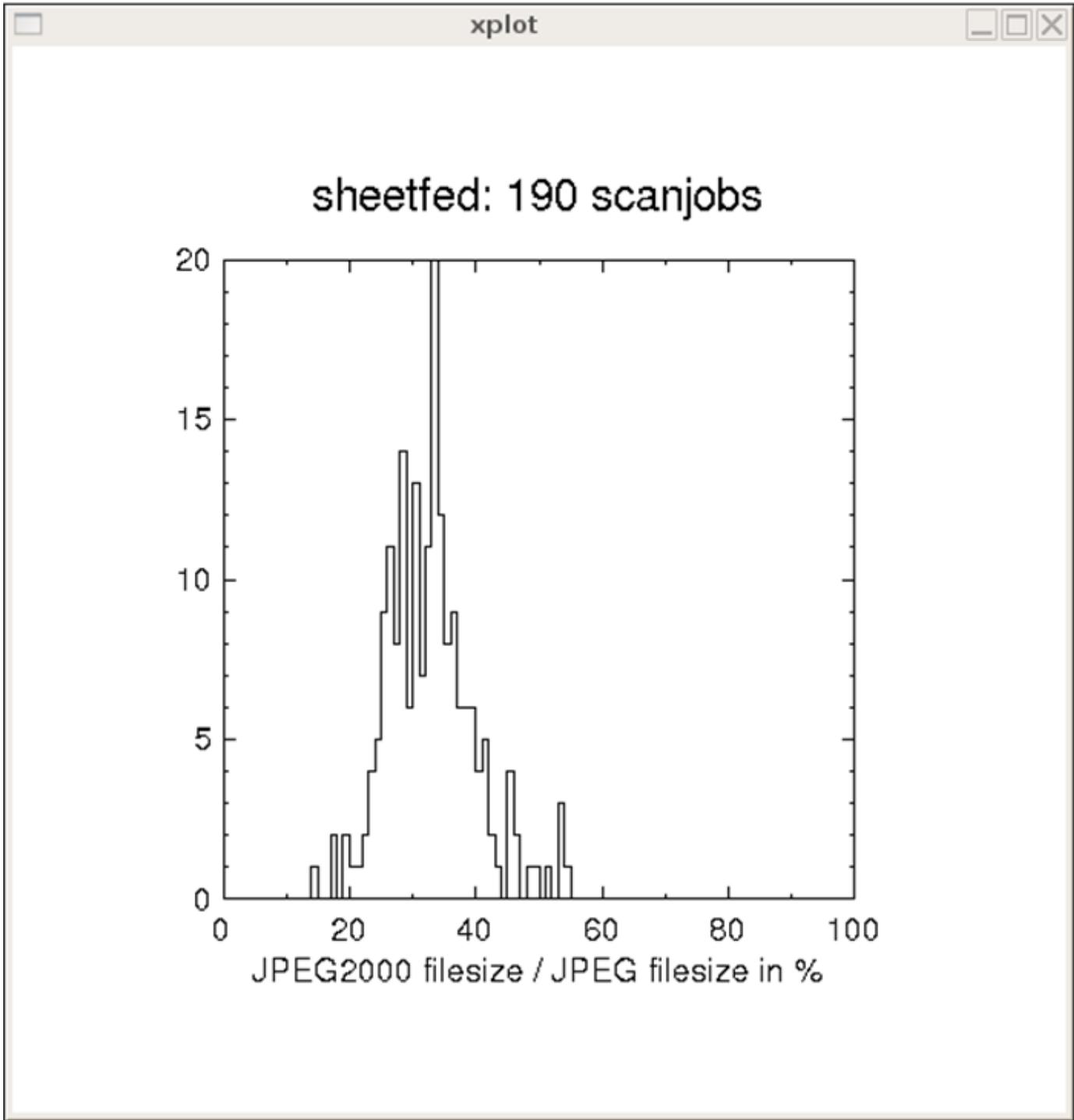


Aler Durchleuch-
tigster/ Großmechtig-
ster Römischer König
1c. Aller Gnedigster
HERR/ Wiewol vil
hochgelerte züvor/ mit
höchstem fleiß nicht on-
nuz vnnnd frucht der
glaubigen/ von der hai-
ligen Mes geschriben/
hab ich doch den güt-

hertzigen zü trost/ dise meine schrifft vnd Predig
von dem Ampt der heyligen Mes/ auß Christen-
lichem eyffer auch wöllen mittailen / Dieweil die
verachtung der Mes bey etlichen vngehorsamen
nicht abnimbt/ sonder täglich wechsť/ so auch die
ainfeltigen Catholischen mit irriger lehř / als ob
die Mes ein verdamlicher grewol were/ zü berede
vnd zü verwirren ganz geschwind vñ abgericht.

* ij Disen







**JPEG
Quality**

JPEG

JPEG2000

100

no. 180

no. 180

93

no. 180

no. 180

85

no. 180

no. 180

75

no. 180

no. 180

Jhove (Rel. 1.4, 2009-07-30)

Date: 2011-05-03 20:06:36 PDT

RepresentationInformation: 00000001.jp2

ReportingModule: JPEG2000-hul, Rel. 1.3 (2007-01-08)

LastModified: 2006-09-22 11:01:00 PDT

Size: 231249

Format: JPEG 2000

Status: Well-Formed and valid

SignatureMatches:

JPEG2000-hul

MIMEtype: image/jp2

Profile: JP2

JPEG2000Metadata:

Brand: jp2

MinorVersion: 0

Compatibility: jp2

ColorspaceUnknown: true

ColorSpecs:

ColorSpec:

Method: Enumerated Colorspace

Precedence: 0

Approx: 0

EnumCS: sRGB

UUIDs:

UUIDBox:

UUID: -66, 122, -49, [...]

Data: 60, 63, 120, [...]

Codestreams:

Codestream:

ImageAndTileSize:

Capabilities: 0

XSize: 1165

YSize: 2037

XOSize: 0

YOSize: 0

XSize: 1165

YSize: 2037

XTOSize: 0

YTOSize: 0

CSize: 3

SSize: 7, 1, 1

XRSize: 7, 1, 1

YRSize: 7, 1, 1

CodingStyleDefault:

CodingStyle: 0

ProgressionOrder: 0

NumberOfLayers: 1

MultipleComponentTransformation: 1

NumberDecompositionLevels: 5

CodeBlockWidth: 4

CodeBlockHeight: 4

CodeBlockStyle: 0

Transformation: 0

QuantizationDefault:

QuantizationStyle: 34

StepValue: 30494, 30442, 30442, 30396, 28416, 28416, 28386, 26444, 26444, 26468, 20483, 20483, 20549, 22482, 22482, 22369

NisImageMetadata:

MIMEtype: image/jp2

ByteOrder: big-endian

CompressionScheme: JPEG 2000

ImageWidth: 1165

ImageLength: 2037

BitsPerSample: 8, 8, 8

SamplesPerPixel: 3

Tiles:

Tile:

TilePart:

Index: 0

Length: 229827

Jhove



```
<?xpacket begin=" id='W5M0MpCehiHzreSzNTczkc9d'?>
<x:xmpmeta xmlns:x='adobe:ns:meta/'>
<rdf:RDF xmlns:rdf='http://www.w3.org/1999/02/22-rdf-syntax-ns#'>
<rdf:Description rdf:about="" xmlns:tiff='http://ns.adobe.com/tiff/1.0/'>
<tiff:ImageWidth>1165</tiff:ImageWidth>
<tiff:ImageLength>2037</tiff:ImageLength>
<tiff:BitsPerSample>
<rdf:Seq>
<rdf:li>8</rdf:li>
<rdf:li>8</rdf:li>
<rdf:li>8</rdf:li>
</rdf:Seq>
</tiff:BitsPerSample>
<tiff:Compression>34712</tiff:Compression>
<tiff:PhotometricInterpretation>2</tiff:PhotometricInterpretation>
<tiff:Orientation>1</tiff:Orientation>
<tiff:SamplesPerPixel>3</tiff:SamplesPerPixel>
<tiff:XResolution>300/1</tiff:XResolution>
<tiff:YResolution>300/1</tiff:YResolution>
<tiff:ResolutionUnit>2</tiff:ResolutionUnit>
<tiff:DateTime>2004-04-27T00:00:00+08:00</tiff:DateTime>
<tiff:Artist>Google, Inc.</tiff:Artist>
<tiff:Make>MDP</tiff:Make>
<tiff:Model>Photostation v1</tiff:Model>
<tiff:Software>Photostation v1 scanning software</tiff:Software>
</rdf:Description>
<rdf:Description rdf:about="" xmlns:dc='http://purl.org/dc/elements/1.1/'>
<dc:source>jp2k/0345430573/00000001.jp2</dc:source>
</rdf:Description>
</rdf:RDF>
</x:xmpmeta>
<?xpacket end='w'?>
```

Embedded XMP

The Joys of Slope Rate Distortion

// Lower number == less distortion == higher fidelity

const int [kJp2kLosslessQuality](#) = 0;

// decent

const int [kJp2kOceanDefaultNDRawQuality](#) = 50980;

// same space as JPEG-75

const int [kJp2kOceanDefaultNDCleanQuality](#) = 51180;

// removes JPEG artifacts

const int [kJp2kOceanDefaultSFRawQuality](#) = 51315;

// GEFGW

const int [kJp2kOceanDefaultSFCleanQuality](#) = 51350;

// acceptable

const int [kJp2kOceanGRINImagePageQuality](#) = 51492;

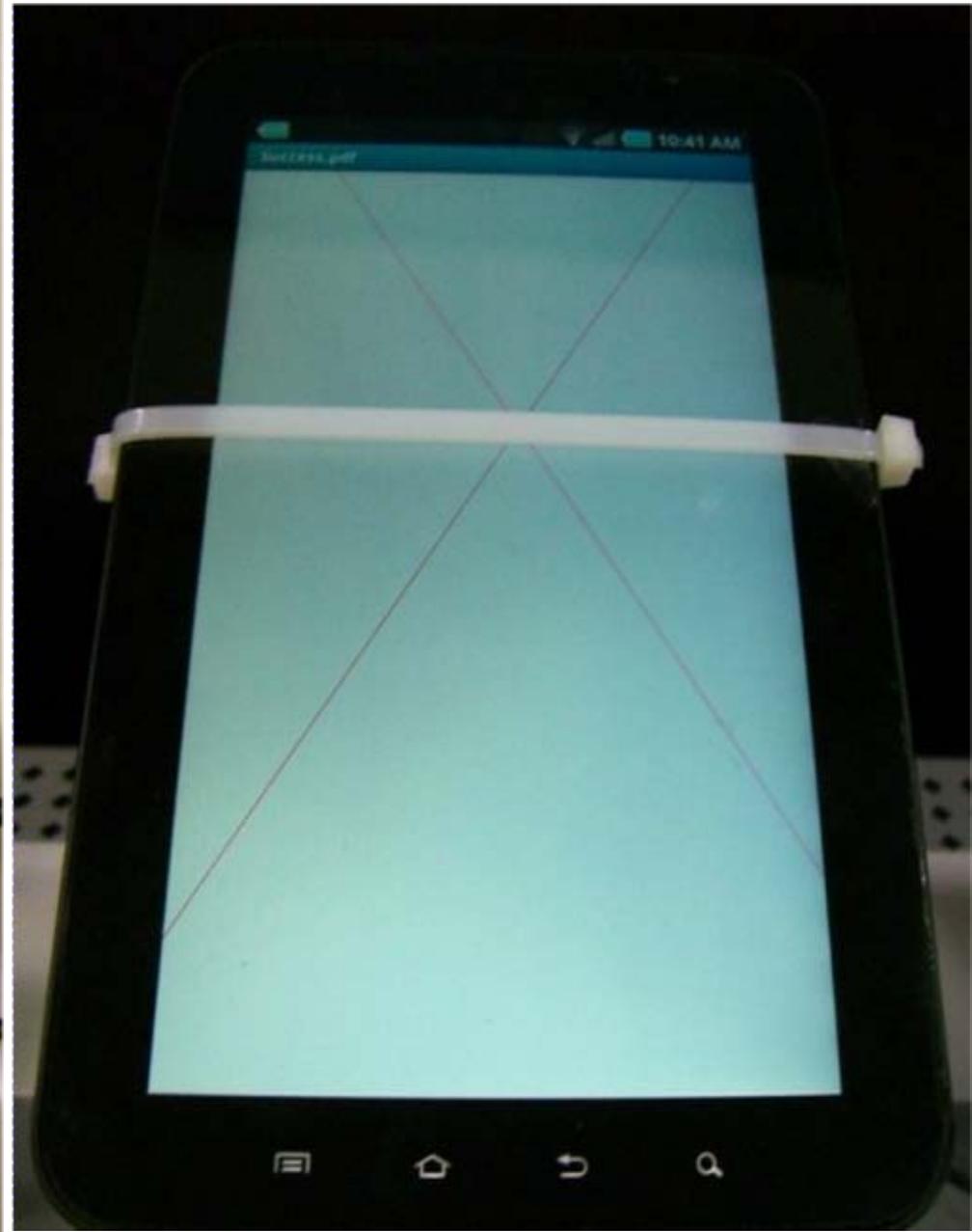
// marginal

const int [kJp2kOceanGRINTextPageQuality](#) = 52004;



JP2K/PDF Compatibility

[Credit: Mike Cane](#)



Thank you / questions

Backup Slides

•WebP

mainly for publishing on the Web
very efficient coding (e.g: segmentation) esp.
at low bitrate. Comparable to h264.
block-based: decoding footprint is very light
(memory scales with the width)
fast decoding: 2x-3x slower than jpeg (9x for
JP2K with Kakadu)
encoding still slow, being worked on
royalty-free
evolving quickly with container features