Federal Agencies Digitization Guidelines Initiative

NATIONAL DIGITAL INFORMATION INFRASTRUCTURE AND PRESERVATION PROGRAM

FADGI Still Image Working Group Meeting

March 23, 2012

10:00 AM to Noon

1. **Introductions**
2. **Sub-Group Activities**
   a) File Format
   b) Embedded Metadata
3. **Activities**
   a) Compression Analysis Update
   b) CIE Color Accuracy Study Update
   c) Analyzing Photo-Negative Collections to Determine Scanning Resolution
   d) Monitoring Production Scanning with SFR Target and M-Scan/ImCheck
4. **Documents and Wiki**
   a) Discuss Revision Priorities and Timeline
5. **Priorities and New Work**

File Formats

File Format Subgroup assessment of 5 formats considered appropriate.

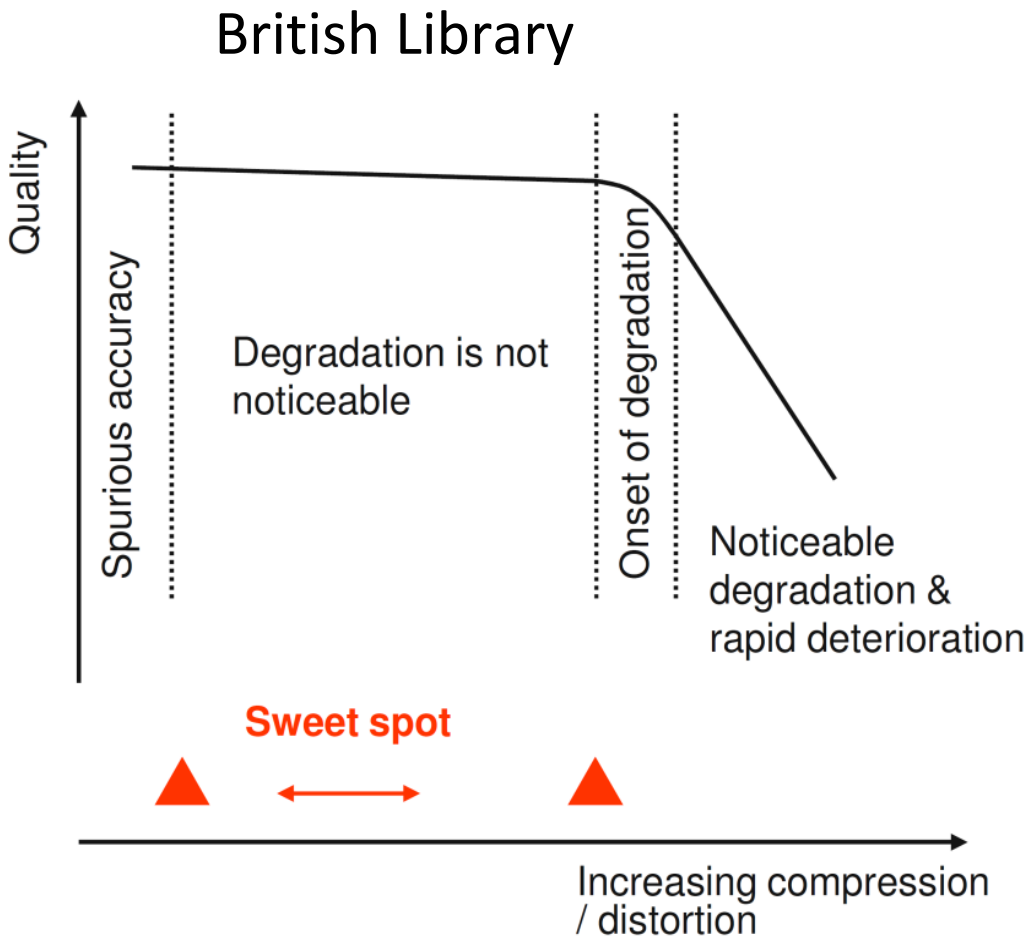
Embedded Metadata

Evaluation of Image Compression and JPEG 2000 Configuration

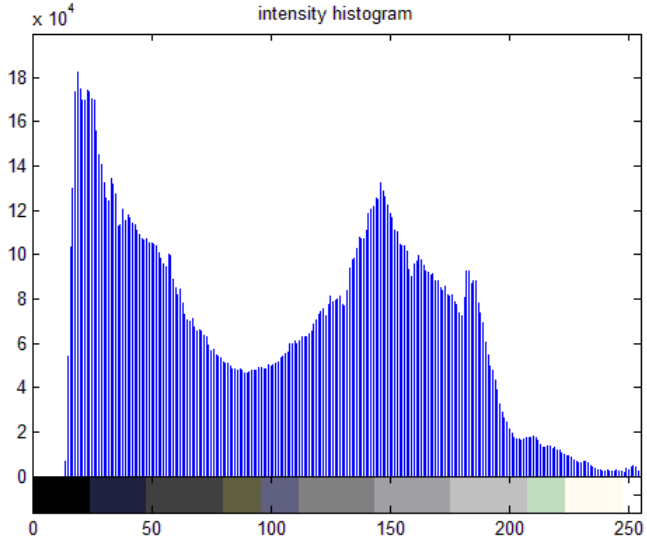
Three approaches to evaluating the effects of compression:

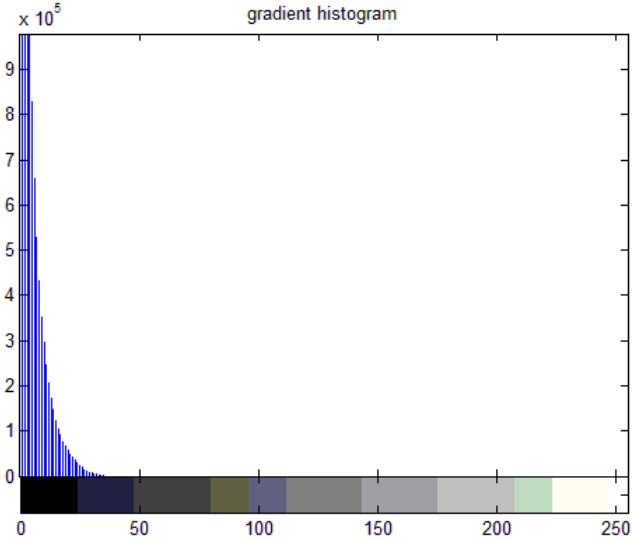
- Visual or subjective evaluation – ranking experiments
- Metric or objective evaluation – Mean Squared Error (MSE), Peak Signal to Noise Ratio (PSNR), Structural Similarity Index (SSIM), etc.
- Task accuracy

Still Image File Formats and Image Compression

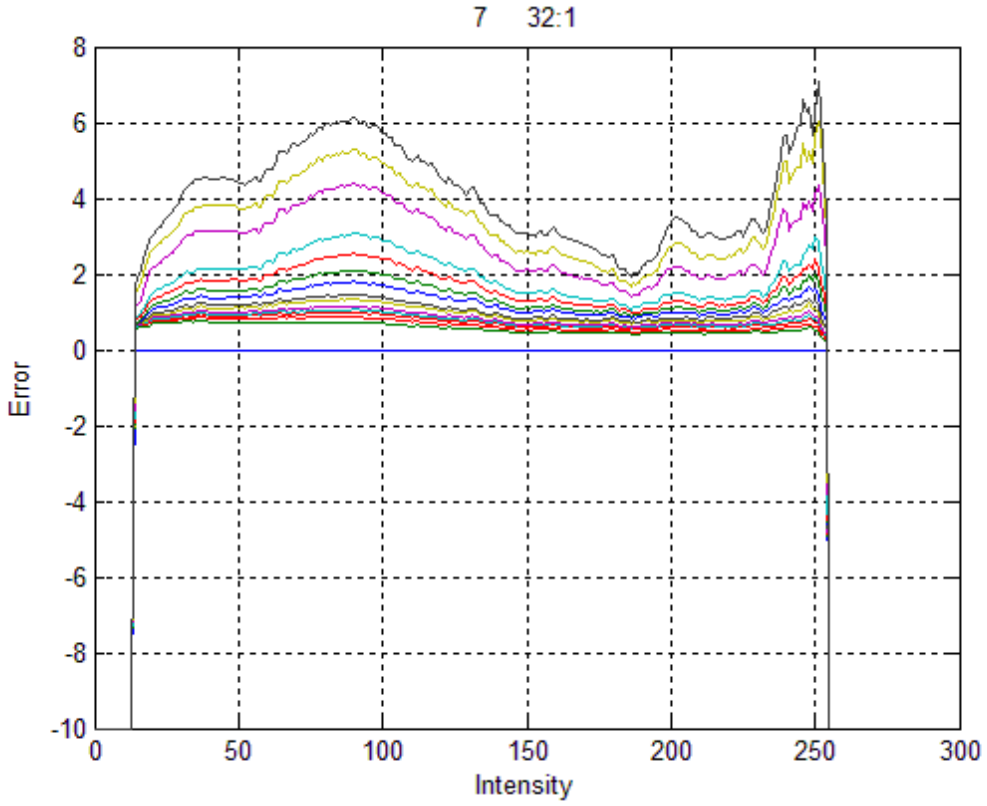


33860
Intensity and gradient histograms

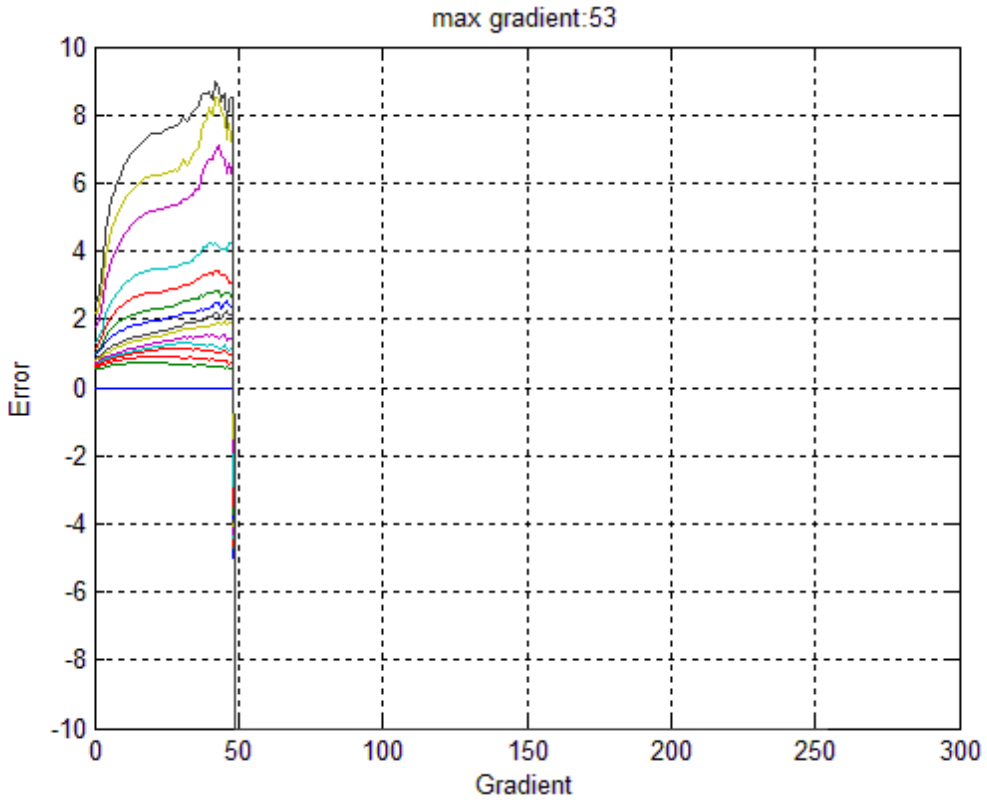




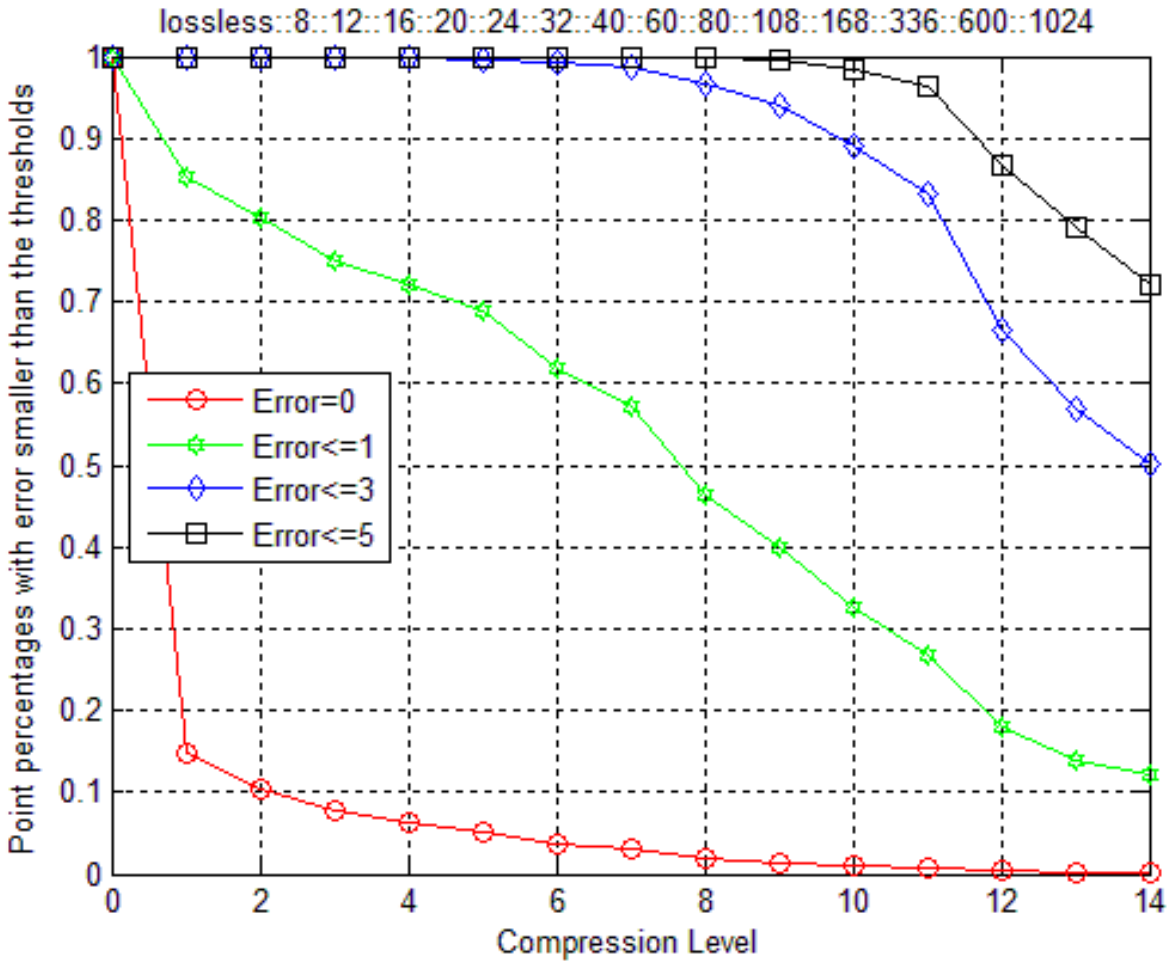
Mean(|Error|) w.r.t. intensity



Mean(|Error|) w.r.t. gradient



Still Image File Formats and Image Compression



Some preliminary conclusions relating to JPEG 2000 compression analysis-

- Error with respect to intensity
  - No obvious relationship between intensity level and error (both mean and variance)
  - The error variance is correlated to the amount of compression
    - Small variance across different intensity levels at low compression levels
    - Large variance at high compression levels

- Error with respect to intensity
  - In most cases, the number of samples at different intensity levels shows an effect on error: a large number of samples result in small error (both mean and variance), especially at high compression levels
  - The error mean curves show same pattern as the variance curves, i.e., large error correlate with large variance.

- Error with respect to gradient magnitude
  - With increased compression levels (e.g., >12:1), error increase rather fast with increased gradient (e.g., from 0 to 15); then error increases slowly with increased gradient.
  - In most cases (e.g., except B05 copyright card), the mean curves show the same pattern as the variance curves, i.e., large error has large variance.
  - At small compression levels (e.g., 8:1 and 12:1), error mean and variance curves are relative “flat”, i.e., small variance.

Still Image File Formats and Image Compression

- Error with respect to gradient magnitude
  - At large gradients with small amount of samples, both mean and variance curves show large variances at large gradients, i.e., may decrease with increased gradient, especially for high compression levels (e.g., >100:1).

Some preliminary conclusions relating to JPEG 2000 compression analysis-

Still Image File Formats and Image Compression

- Even low levels of lossy compression (e.g., 8:1) introduce errors to most pixels (e.g., 60%) as measured by ΔE (a change in color and/or brightness).
- Between 10% and 80% of pixels will have an error of $\Delta E=1$ or only a just noticeable difference.
- The effect differs for different collection types- fewer pixels are effected for color photos compared fine prints and b+w negatives, which have fewer pixels effected than cartoon drawings.

Still Image File Formats and Image Compression

- In practice, a $\Delta E=5$ is considered an acceptable level of difference when comparing standard images. With the exception of cartoons, all collection types evaluated so far have a high tolerance for moderate compression levels (e.g., up to 16:1 in our tests), with more than 90% of pixels having a ΔE of less than 5.
- Lossless compression (about 2:1) has no effect on ΔE. Observers noted differences due to limited ICC color profile support with JP2 files.

Still Image File Formats and Image Compression

- Observers identify compression artifacts:
  - At lower compression levels for grayscale images (average of 24:1) compared to color images (average of 48:1)
  - At lower compression levels for smaller images (total number of pixels) compared to larger images – this varies somewhat depending on collection type
- On average, higher resolution samples (400 ppi) have smaller error than lower resolution samples (300 ppi) at the same compression level (particularly for higher compression).

- More work:
  - Evaluate the effect of sampling efficiency on the perception of compression artifacts
  - Likely more pixels with greater error due to low sampling efficiency compared to the change due to low or moderate lossy compression.

Organizations using or accepting JPEG 2000:

- Biodiversity Heritage Library
- British Library
- National Library of the Czech Republic
- Early European Books - ProQuest
- Google
- Harvard University
- Internet Archive
- Library of Congress
  - Geography and Maps Division
  - National Audio Visual Conservation Center
  - National Digital Newspaper Program
- National Library of the Netherlands
- National Library of Norway
- New York Public Library
- Wellcome Library

Still Image File Formats and Image Compression

Progression Order ('LRCP', 'RLCP', 'RPCL', 'PCRL' or 'CPRL')

Has no effects on the accuracy, i.e., all five different orders produce the same error.

The effect on the time cost is “random”.

Quality Layers (max 20, default 1)

A smaller number of quality layers has better quality than a larger number of layers.

On the other hand, more quality layers provide flexibility in delivering images of different quality to the applications.

Reduction Levels (Decomposition levels, max 8 in MATLAB)

For most images, fewer decomposition levels (e.g. 2) has larger error, but a lower time cost, compared to more levels (e.g. 4 or 8).

The larger number of levels, the more the time cost.

However, with higher numbers of decomposition levels, it becomes possible to separate the noise (artifacts) from the true signals.

Tile Size (min [128 128], default image size)

Small tile size (128x128) results in larger error and time cost.

When the tile size is large enough (e.g. 4096x4096) in our test, the results are the same as the original image without tiles.

More tiles (blocks) limits error propagation in transmission.

Smaller or more tiles may introduce block effect with larger compression error.

Still Image File Formats and Image Compression

JPEG 2000 offers a lot of flexibility in terms of configuration.

Many organizations have determined a JPEG 2000 configuration based on optimizing the end-user experience for web delivery.

Many organizations are using pretty much the same basic JPEG 2000 profile.

Archival Color

- CIE Color Accuracy Study Update

Additional Labs

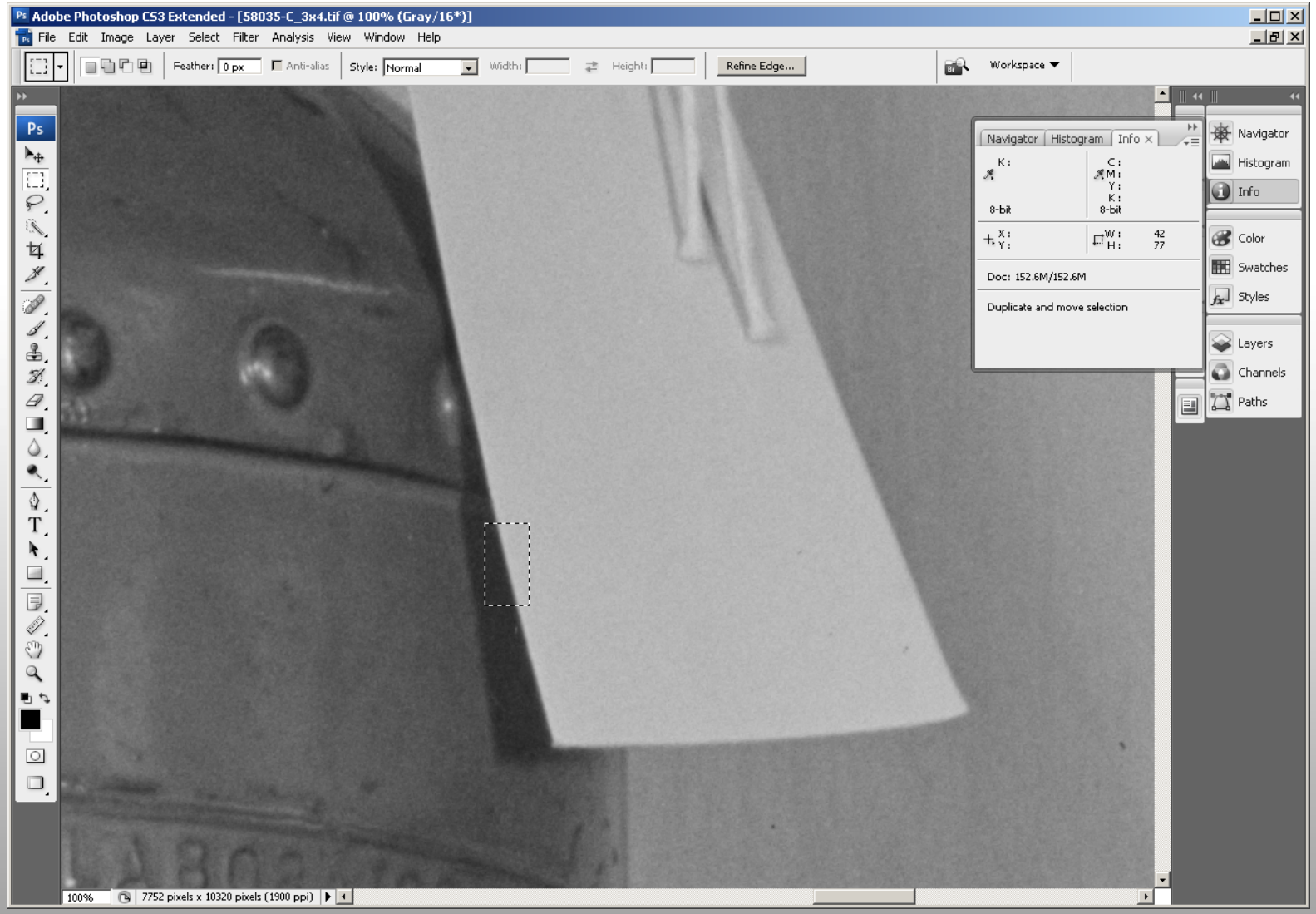
- Second group has imaged samples/targets-
  - Harvard University
  - Stanford University
  - Art Institute of Chicago
  - National Gallery of Art

Next Steps

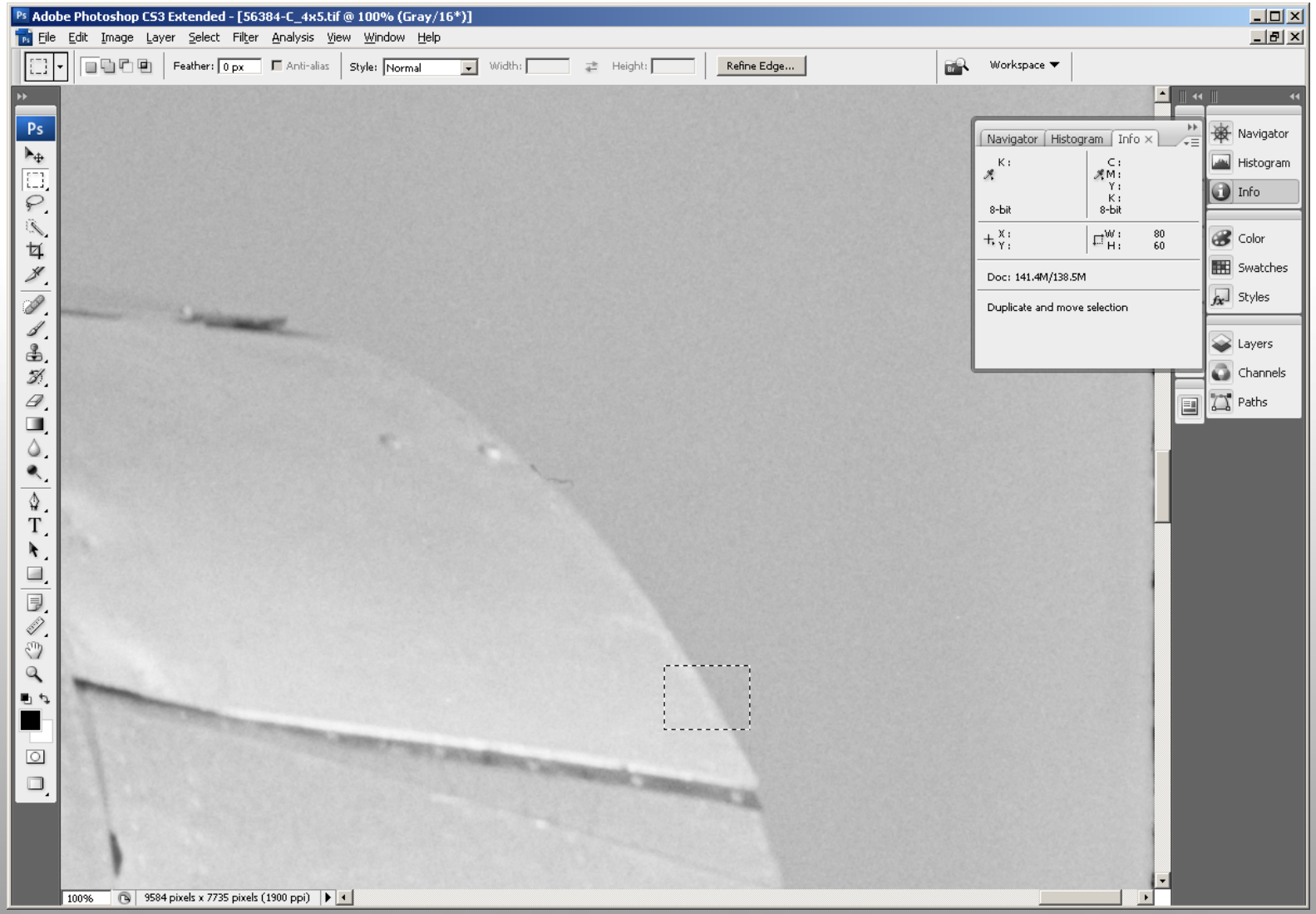
- Analyze data from all 7 North American imaging labs
- 2nd imaging phase by European labs-
  - KB Netherlands
  - van Gogh Museum
  - Rijksmuseum
  - Studio Buitenhof, commercial studio in Netherlands
  - KB Denmark
  - Maybe more
- Update to be presented at IS&T Archiving 2012 in Copenhagen, DK

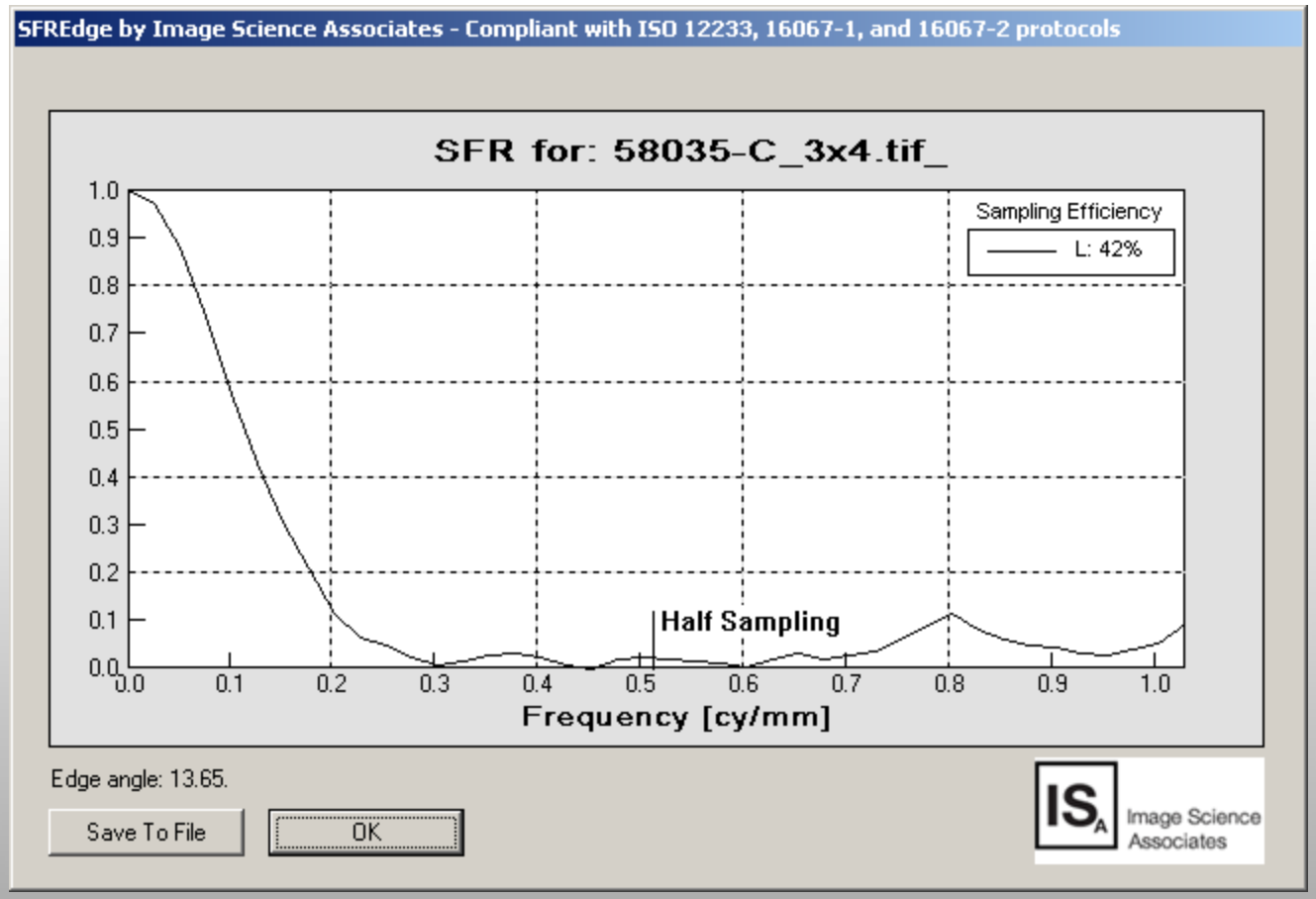
Analyzing Photo-Negative Collections to Determine Scanning Resolution

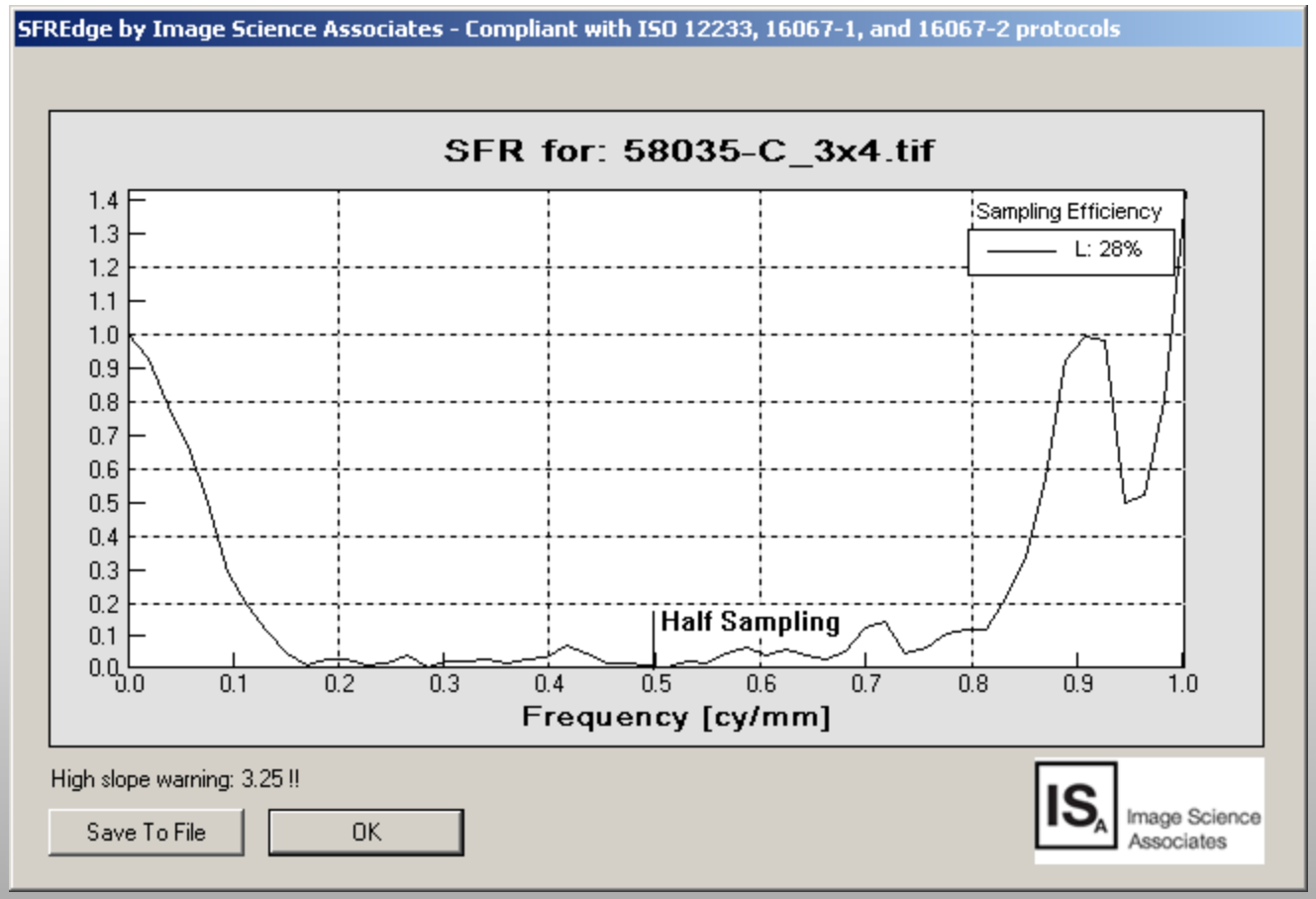
- Sample collection
- Scan at higher than expected resolution – verify with target
- Analyze selected features in image
- Determine appropriate resolution

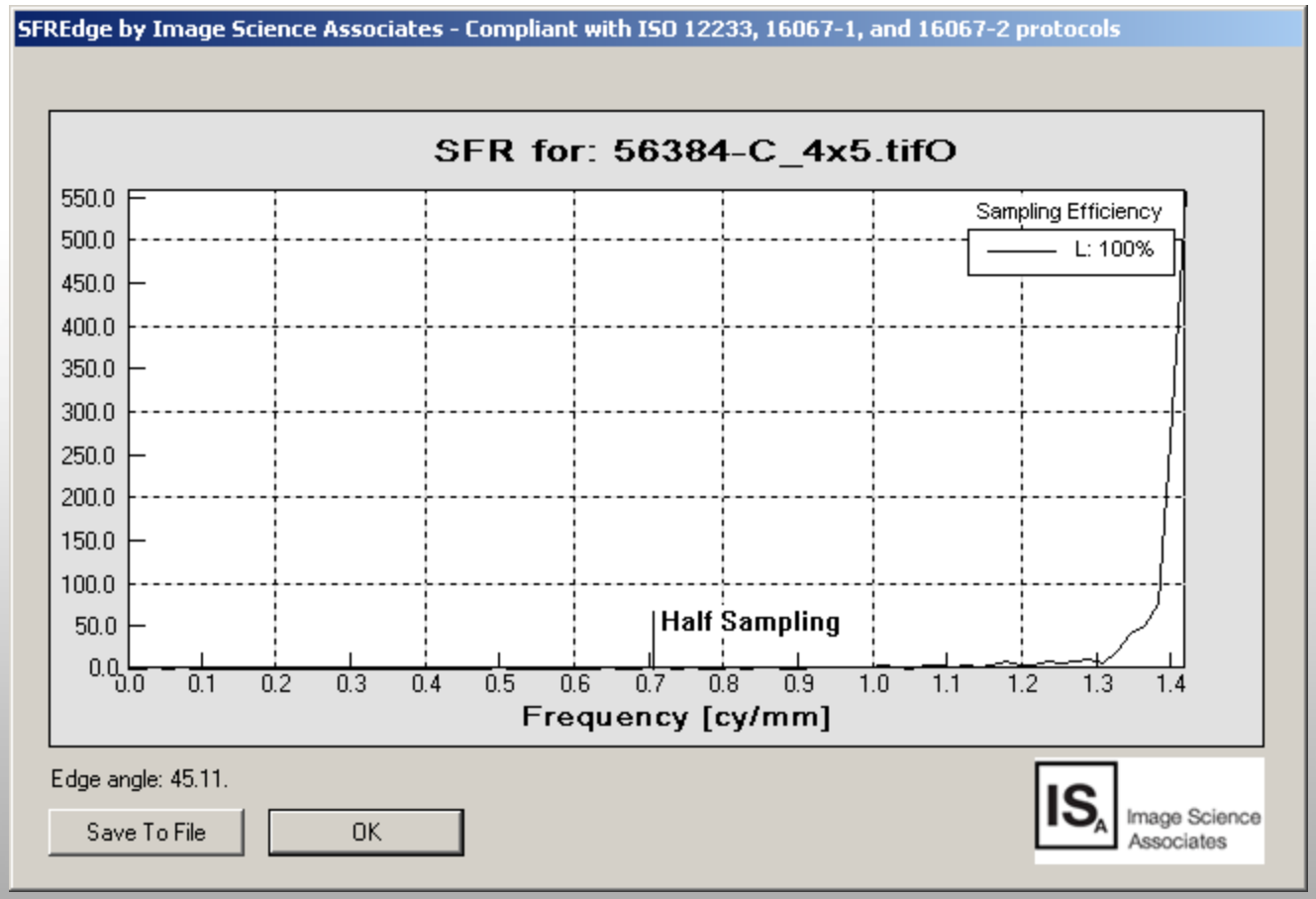

Adobe Photoshop CS3 Extended - [58035-C_3x4.tif @ 100% (Gray/16*)]
File Edit Image Layer Select Filter Analysis View Window Help
Feather: 0 px
Anti-alias
Style: Normal
Width:
Height:
Refine Edge...
Workspace
Navigator
Histogram
Info
K :
8-bit
C :
M :
Y :
K :
8-bit
X :
Y :
W :
H :
42
77
Doc: 152.6M/152.6M
Duplicate and move selection
Navigator
Histogram
Info
Color
Swatches
Styles
Layers
Channels
Paths
100%
7752 pixels x 10320 pixels (1900 ppi)


Adobe Photoshop CS3 Extended - [73305-E_2x2.tif @ 100% (Gray/16*)]
File Edit Image Layer Select Filter Analysis View Window Help
Feather: 0 px
Anti-alias
Style: Normal
Width:
Height:
Refine Edge...
Workspace
Navigator
Histogram
Info
K: 47%
8-bit
C: 49%
M: 40%
Y: 41%
K: 4%
8-bit
X: 3073
Y: 4390
W: 40
H: 64
Doc: 75.9M/74.3M
Click and drag to move layer or move a copy of selection. Shift constrains to 45 degree increments.
Color
Swatches
Styles
Layers
Channels
Paths
100%
6190 pixels x 6428 pixels (1900 ppi)

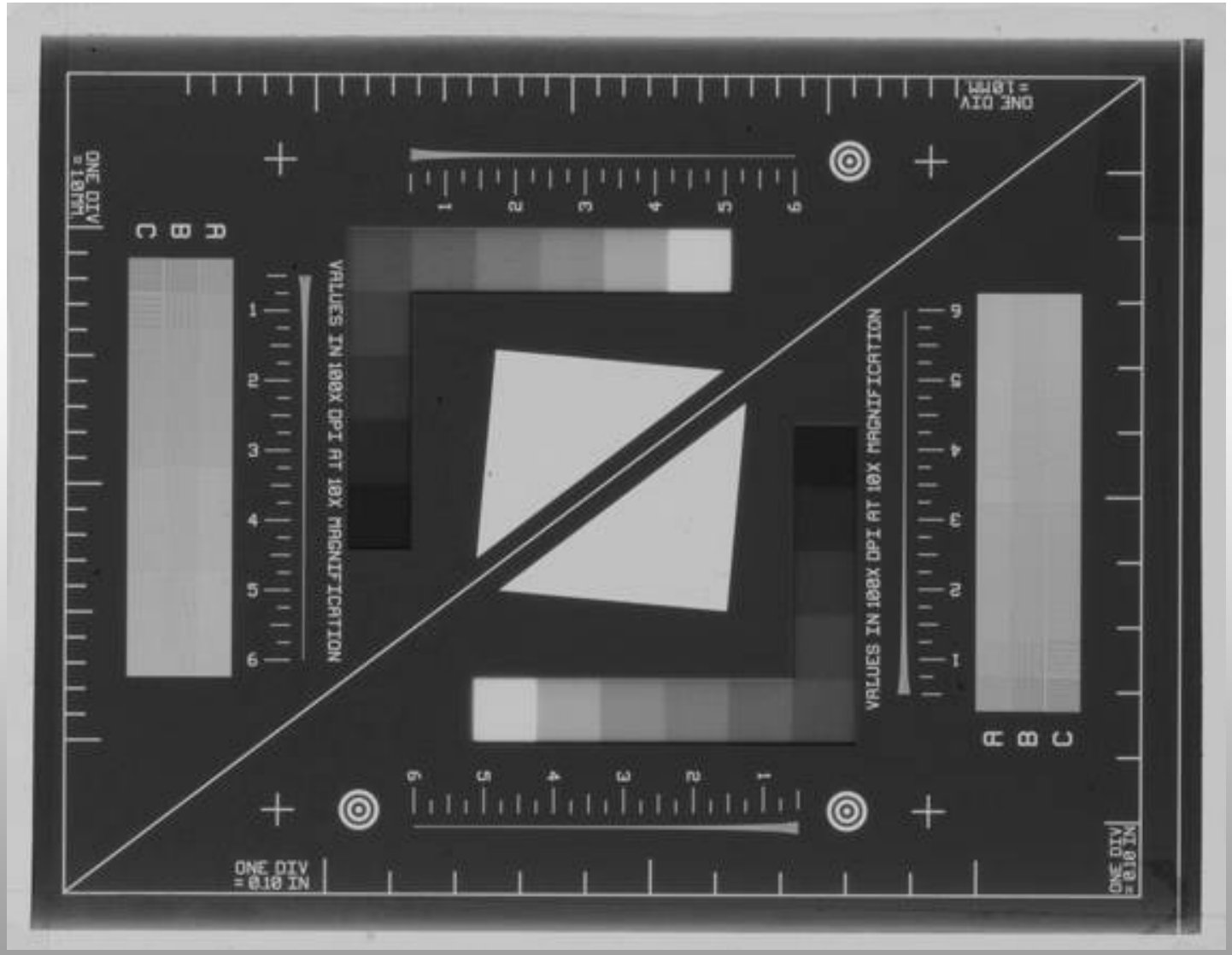

Adobe Photoshop CS3 Extended - [56384-C_4x5.tif @ 100% (Gray/16*)]
File Edit Image Layer Select Filter Analysis View Window Help
Feather: 0 px
Anti-alias
Style: Normal
Width:
Height:
Refine Edge...
Workspace
Navigator
Histogram
Info
K:
8-bit
C:
M:
Y:
K:
8-bit
X:
Y:
W: 80
H: 60
Doc: 141.4M/138.5M
Duplicate and move selection
Navigator
Histogram
Info
Color
Swatches
Styles
Layers
Channels
Paths
100%
9584 pixels x 7735 pixels (1900 ppi)


SFREdge by Image Science Associates - Compliant with ISO 12233, 16067-1, and 16067-2 protocols
SFR for: 58035-C_3x4.tif_
Sampling Efficiency
L: 42%
Half Sampling
Frequency [cy/mm]
Edge angle: 13.65.
Save To File
OK
Image Science Associates


SFREdge by Image Science Associates - Compliant with ISO 12233, 16067-1, and 16067-2 protocols
SFR for: 58035-C_3x4.tif
Sampling Efficiency
L: 28%
Half Sampling
Frequency [cy/mm]
High slope warning: 3.25 !!
Save To File
OK
Image Science Associates


SFREdge by Image Science Associates - Compliant with ISO 12233, 16067-1, and 16067-2 protocols
SFR for: 56384-C_4x5.tifO
Sampling Efficiency
L: 100%
550.0
500.0
450.0
400.0
350.0
300.0
250.0
200.0
150.0
100.0
50.0
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4
Half Sampling
Frequency [cy/mm]
Edge angle: 45.11.
Save To File
OK
IS A
Image Science Associates

Monitoring Production Scanning w/SFR Target and M-Scan/ImCheck Software


ONE DIV
= 10MM
VALUES IN 100X DPI AT 10X MAGNIFICATION
A
B
C
ONE DIV
= 0.10 IN

- Scan target on a daily basis – center, and corners for larger formats
- Plot over time
- Monitor change and variability

WG26 of ISO TC42 –

- New standards group
- Plan to develop standards relating to digitization and related tools, such as targets

Priorities and New Work-

- New work proposal by Mike Horsley, NARA – Consideration of production and quality management metadata – What to keep short-term vs. long-term?

Next Meeting-

Probably May or June.