Federal Agencies Digitization Guidelines Initiative

NATIONAL DIGITAL INFORMATION INFRASTRUCTURE AND PRESERVATION PROGRAM

FADGI Still Image Working Group Meeting

November 8, 2011

9:30 AM to 12:30 PM

1. **Introductions**
2. **Sub-Group Activities**
    a) File Format
    b) Content Categories and Objectives
    c) Archival Color
        1) CIE Color Study Update
    d) Embedded Metadata
    e) Technical Metadata
3. **Activities**
    a) JPEG 2000
    b) DICE
        1) Upgrade and New Targets
4. **Documents**
    a) Discuss Revision Priorities and Timeline
    b) Additional Documents
5. **Wiki**
6. **Priorities and New Work**
7. **Meeting Notes and Advisory Board**

File Formats

Assumptions:

- Considering RGB image capture or "similar grayscale" images, not multi-spectral data
- File formats for reformatted collection materials, not focusing on born digital materials
- Files compliant with ISO specifications for the formats, not what tools produce or can do (e.g., Adobe)
- Anticipate recommending preferred formats for content categories, and recommended profiles or elements that should be in those profiles

For still or raster images, considering:

- TIFF
- JPEG 2000
- JPEG
- PNG
- PDF and PDF/A

Sustainability Factors

1. Disclosure
2. Adoption
3. Transparency
4. Self-documentation
    a) Embedded metadata capabilities
    b) Level of work necessary to embed metadata
5. External dependencies
6. Impact of patents
7. Technical protection mechanisms

Cost Factors

1. Implementation cost
2. Storage cost
3. Network cost
4. Ongoing cost of production (e.g., scanner speed/file transformation and compression)
5. Cost of providing access
6. Cost of preservation processing

System Implementation Factors (Full Lifecycle)

- Level of difficulty/complexity
- Availability of tools
- Compatibility in existing enterprise environment (e.g., OCR-ability)
- Master file size
- Ease and accuracy for validation
- Evaluating and monitoring of Quality

Quality Factors (Pass/Fail????)

1.Clarity

- •Support for bit depths, encodings, multi-channel, multi-layered, multi page files
- •Support for different compression
- •Up to 48 bits?
- •Up to XX pixels?

2.Support Color Maintenance (could have a grading scale for some of these factors)

- •How will does the format support the maintenance of color, including:
  - osRGB, ProPhotoRGB, eciRGB v2, Adobe RGB 1998

3.Maximum File Size up to 4GB?

Revisiting File Formats for Digitization

Content Categories and Objectives

Archival Color

1)CIE Color Study Update

Additional Labs

- Results from first group of labs presented in May at IS&T Archiving 2011-
  - Library of Congress
  - Metropolitan Museum of Art
  - NARA
- A second group has imaged samples/targets-
  - Harvard University
  - Stanford University
  - Art Institute of Chicago

Next Steps

- Re-measure samples and targets
- Analyze data from all 6 North American imaging labs
- 2nd imaging phase by European labs-
  - KB Netherlands
  - Commercial labs in Netherlands
  - KB Denmark
- Submitting abstract to present update at IS&T Archiving 2011 in Copenhagen, DK

Embedded Metadata

Consideration of-

Smithsonian Institutions'

Basic Guidelines for Minimal Descriptive Embedded Metadata in Digital Images

Developed by the Embedded Metadata Working Group (EMDaWG), April 2010

Technical Metadata

JPEG 2000 and Evaluation of Image Compression

Rob Buckley is assisting. Preliminary report-

- Visual or subjective evaluation – ranking experiments
- Metric or objective evaluation – MSE, PSNR, SSIM, etc.
- Task accuracy

British Library

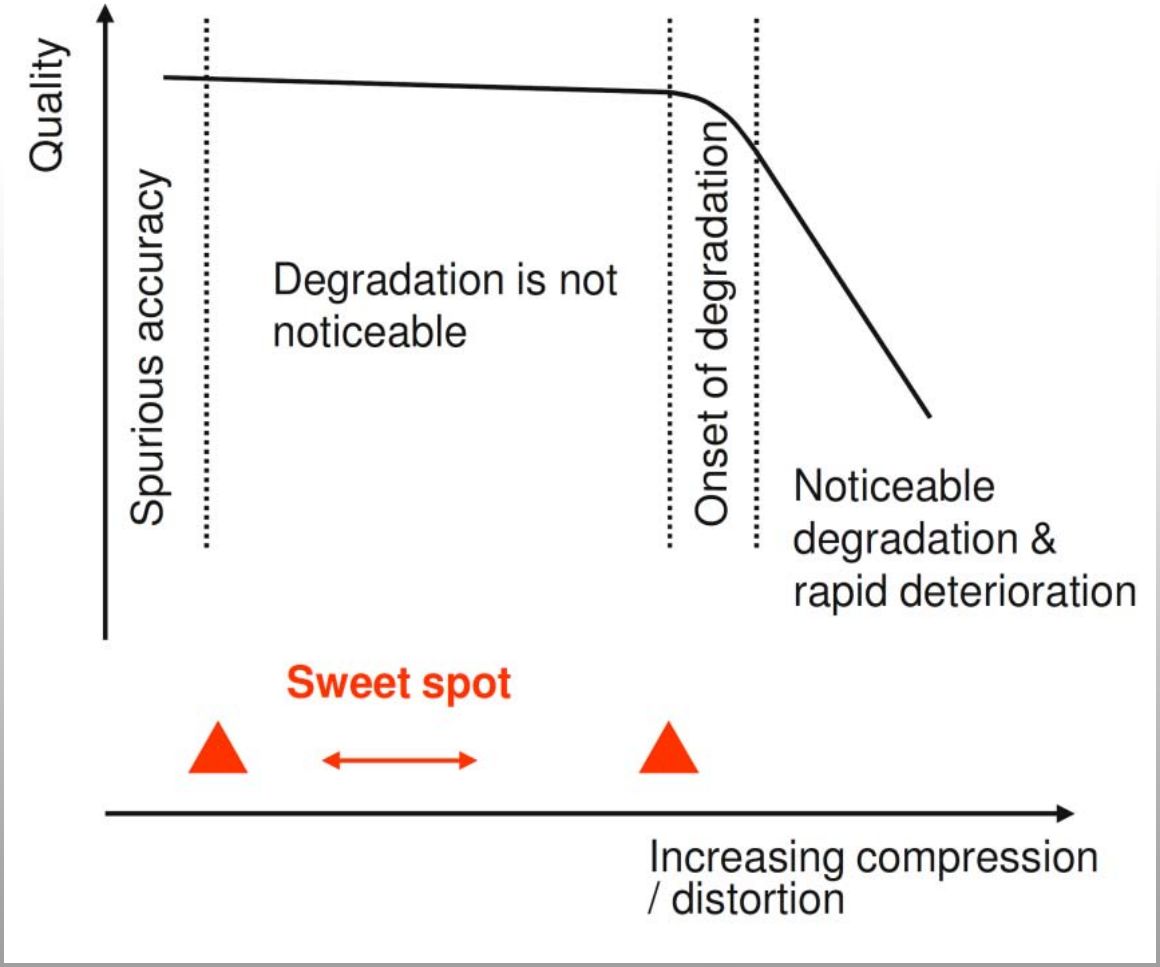


Figure 1: Compression performance plot


Characters Recognized
400
350
300
250
200
150
100
50
0
1.0
10.0
100.0
1000.0
10000.0
Series1
Series2


Character Confidence
80
70
60
50
40
30
20
10
0
1.0
10.0
100.0
1000.0
10000.0
Series1
Series2

LIBRARY OF CONGRESS

Effects of Image Compression on Image Quality

Dr. Lei He and Steven Puglia

Office of Strategic Initiatives

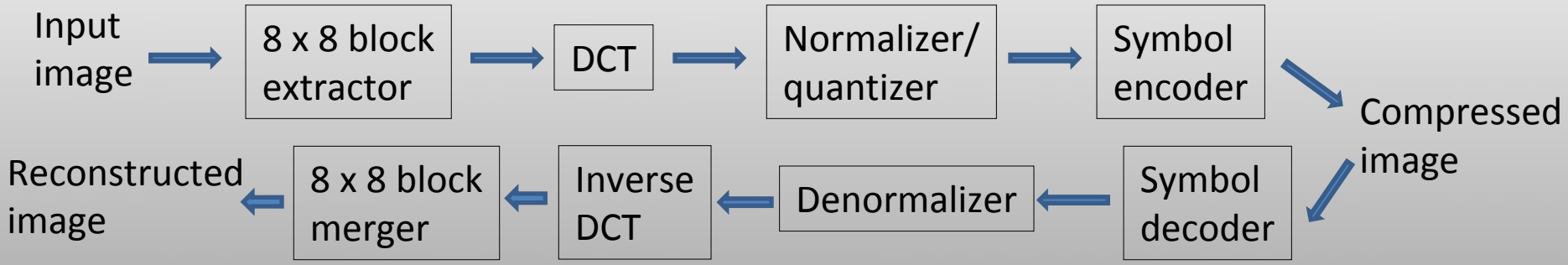
Library of Congress

Goal

- To determine if JPEG 2000 is an appropriate format for digital preservation, i.e., to replace the commonly used TIFF format for master copies.
  - **Why?** JPEG 2000 saves more storage space with larger compression rate.
  - **How?** Identify the appropriate compression rates that is acceptable by human visual perception with unnoticeable artifacts.

Introduction

- JPEG compression
  - Discrete cosine transform (DCT) based compression
  - Block effect (8x8 blocks)



Introduction

- JPEG 2000 compression
  - Wavelet transform-based compression
  - Transform the whole image
  - Better quality and larger compression rate than Jpeg
  - Edge artifacts (ringing)

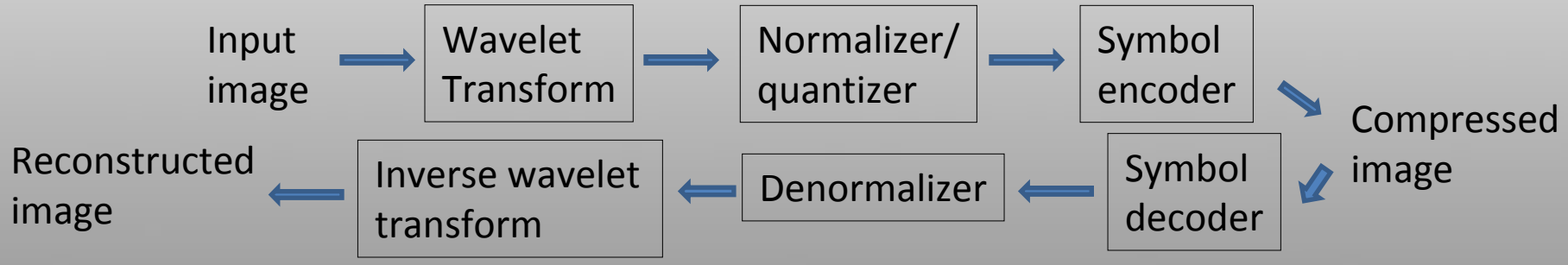
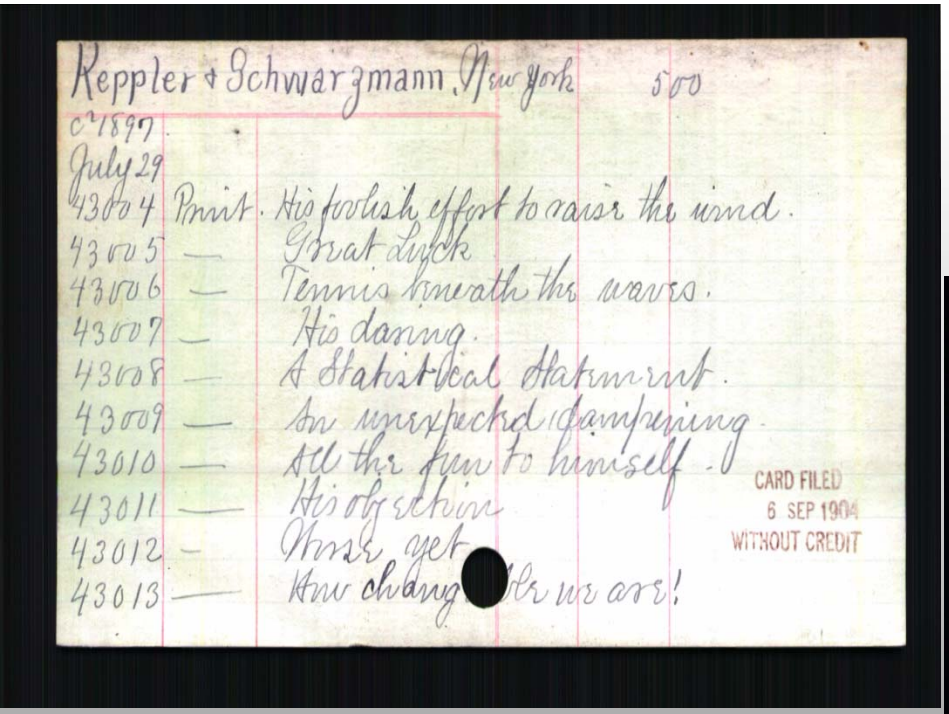


Image Samples

- Copyright cards



#1

#2

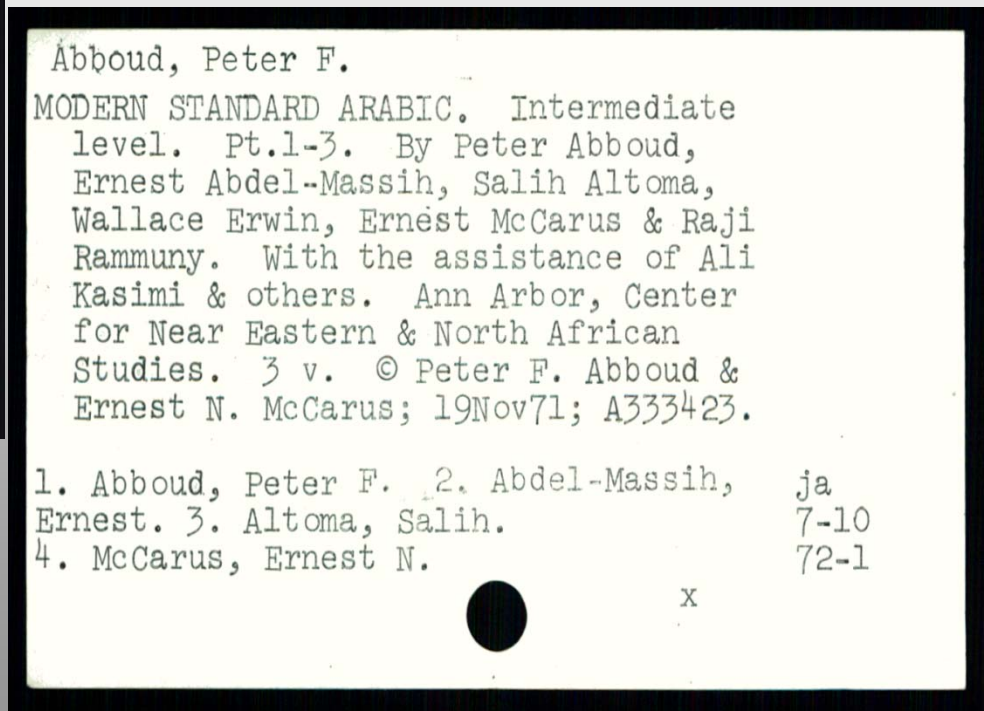


Image Samples

- Photographs: grayscale and color

Image Samples

- Fine prints

Compression Effects Analysis

- Qualitative analysis
  - Ranking
  - Compression threshold identification
- Quantitative analysis
  - Single value measurements
  - Difference (error) distribution analysis

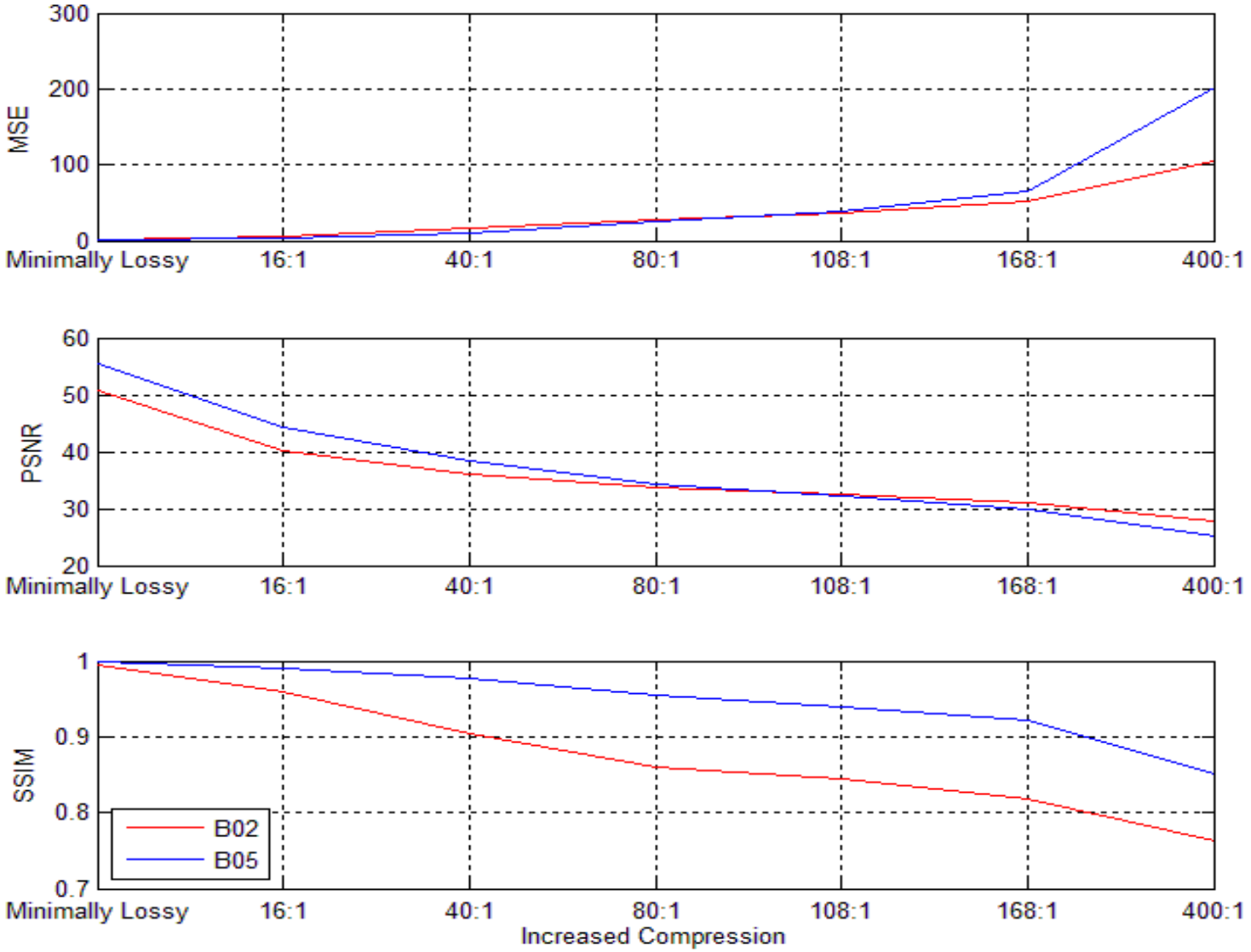
Qualitative Analysis

- Ranking:
  Lossless or minimally lossy, 16:1, 40:1, 80:1, 108:1, 168:1, 336:1 or 400:1
- Compression threshold identification:
  Lossless, 8:1, 12:1, 16:1, 20:1, 24:1, 32:1, 40:1, 60:1, 80:1, 108:1, 168:1, 336:1, 600:1, 1024:1

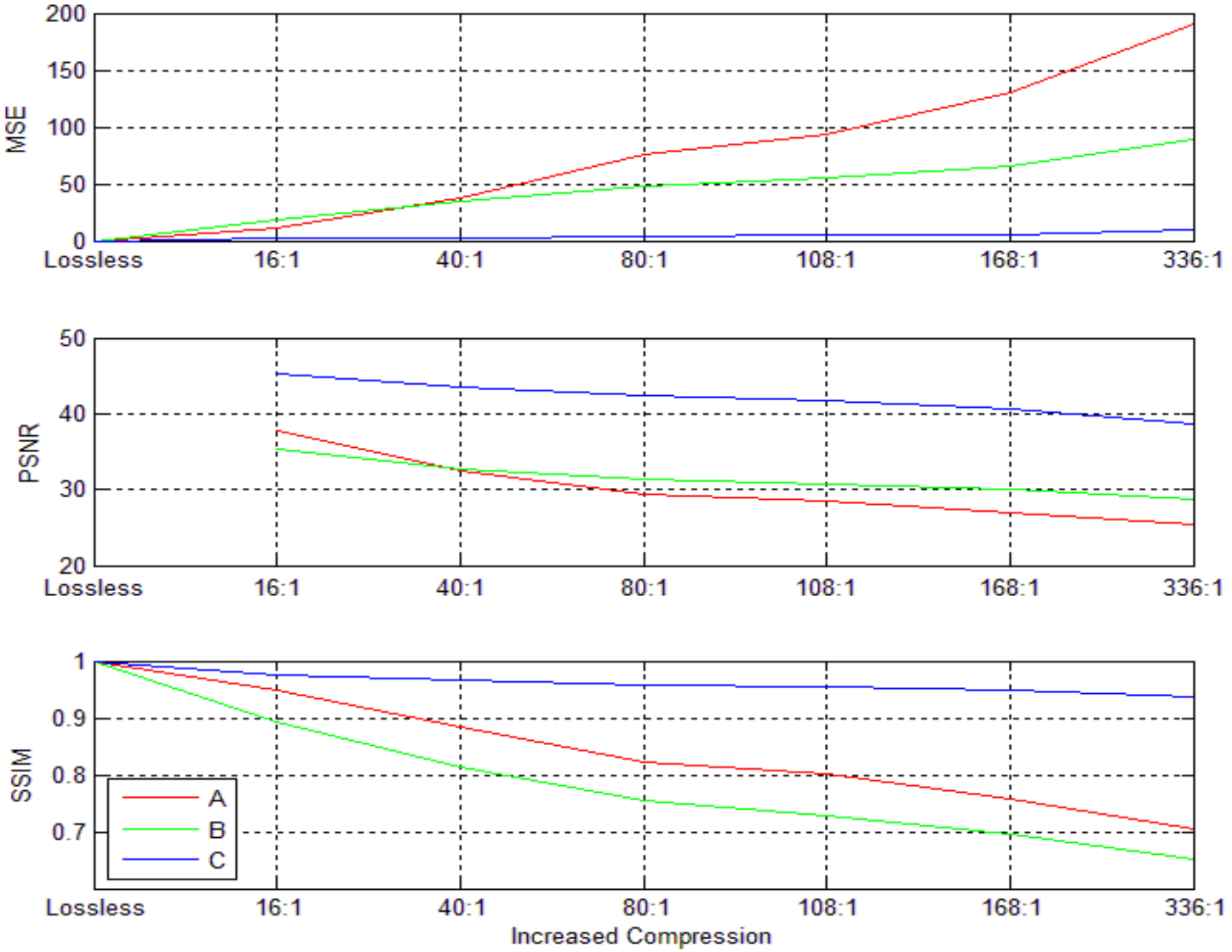
Quantitative Analysis

- Single value error measurements:
  - Peak signal noise ratio (PSNR)
  - Mean squared error (MSE)
  - Structural similarity index (SSIM)

Copyright cards



Photographs & fine prints

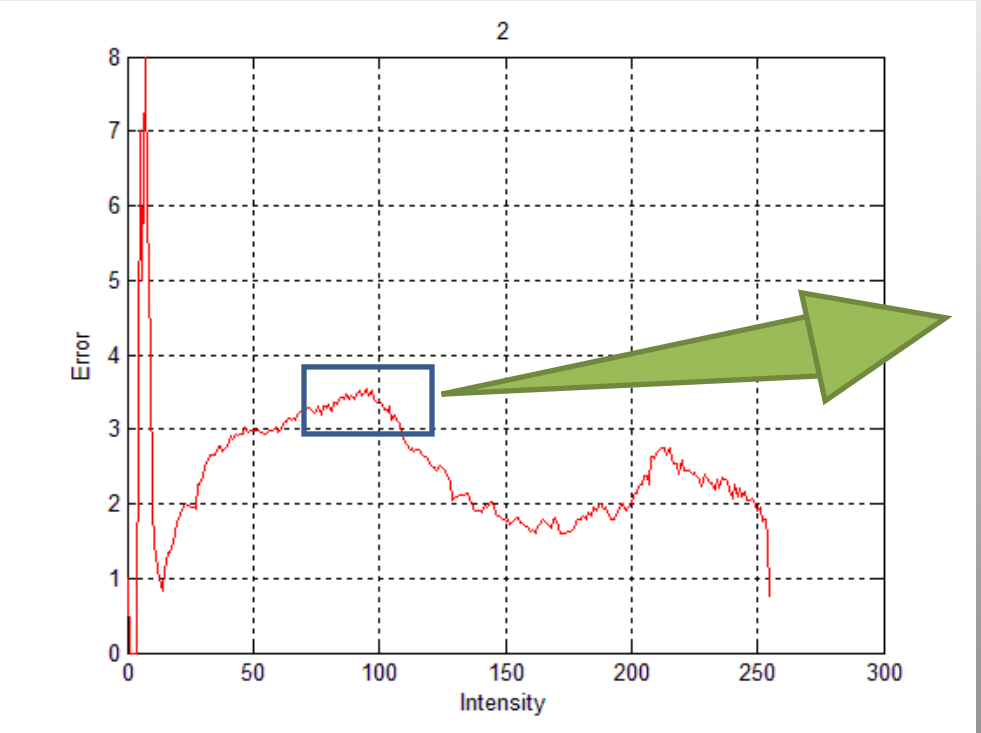


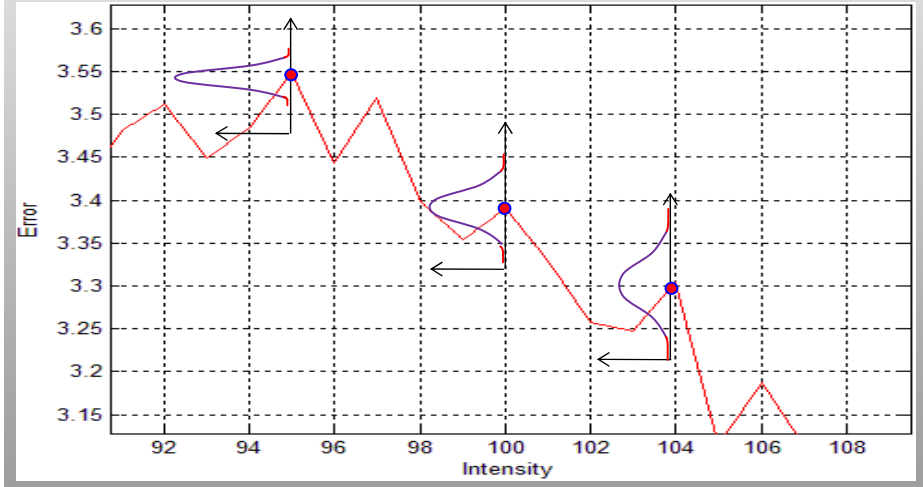
Problem

A single value cannot tell us comprehensive information about the difference between the original image and the compressed ones.

Quantitative Analysis

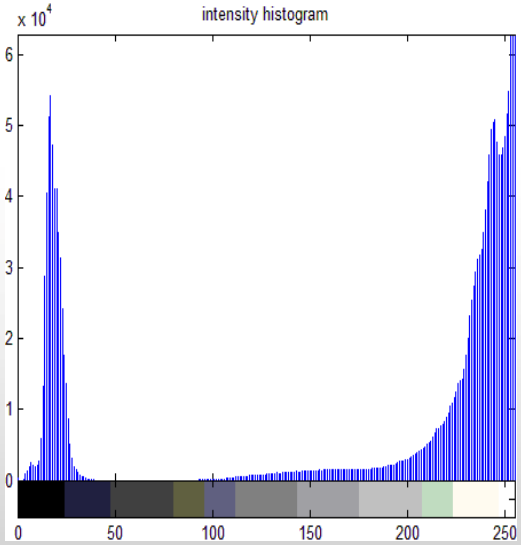
- Difference (error) distribution fitting —— 1D nonlinear Gaussian model: the error at different intensity levels or gradient magnitudes has Gaussian distributions with different mean values and variances.

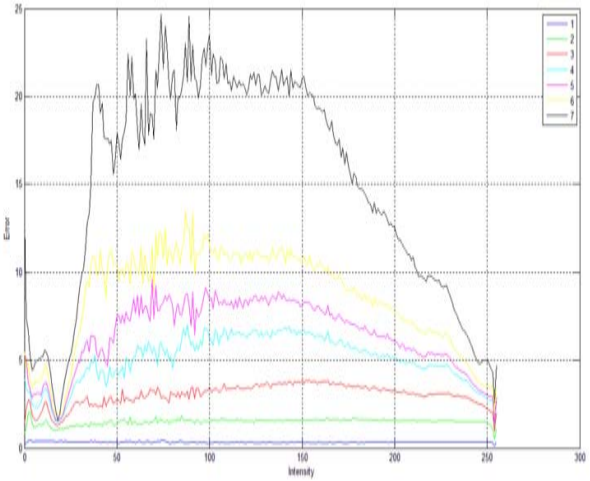


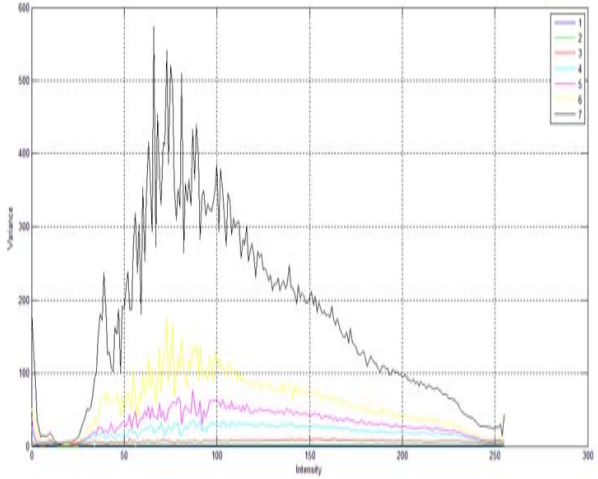


Example 1: Error analysis in the HSI, La*b*, and YCbCr spaces

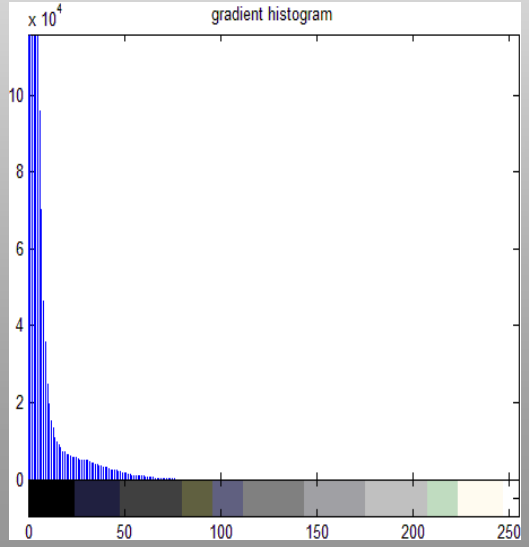
Copyright card sample #1: I channel [0~255]

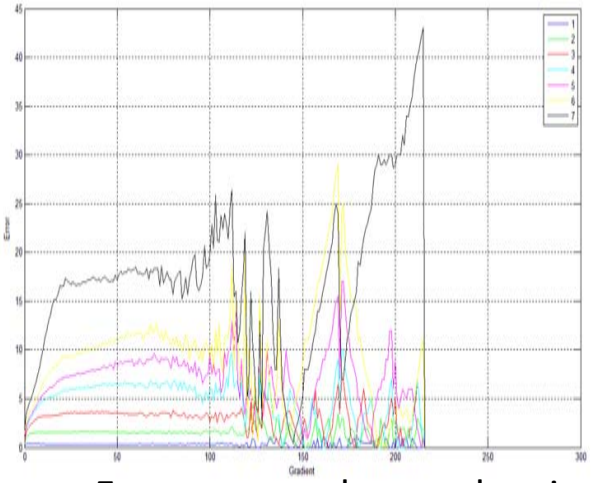


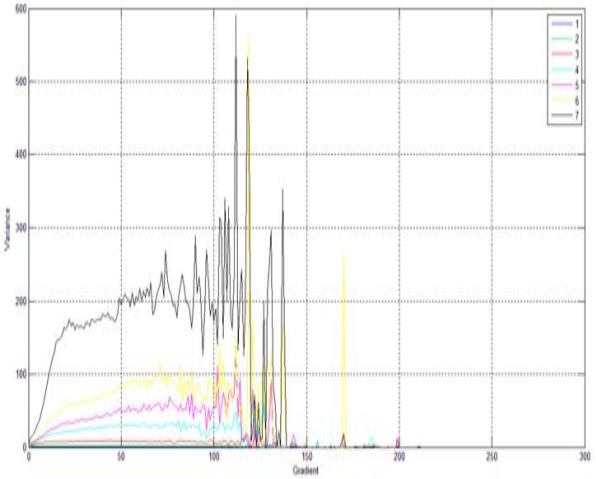




Error mean values and variances w.r.t. intensity levels



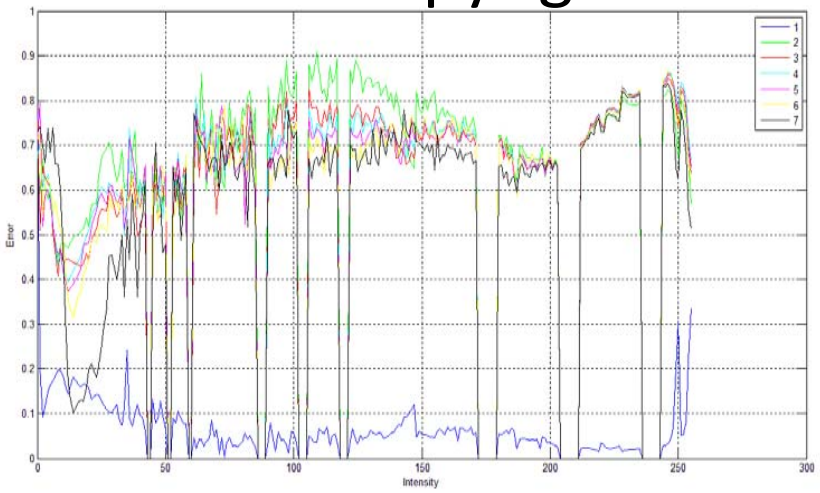


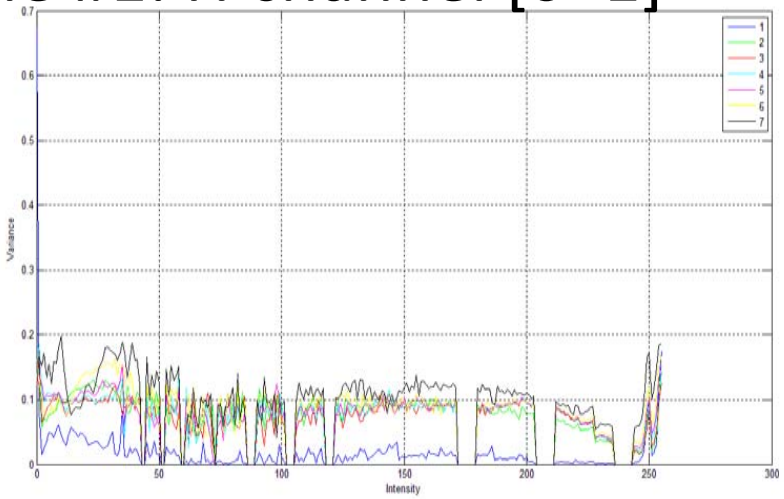


Error mean values and variances w.r.t. intensity gradients

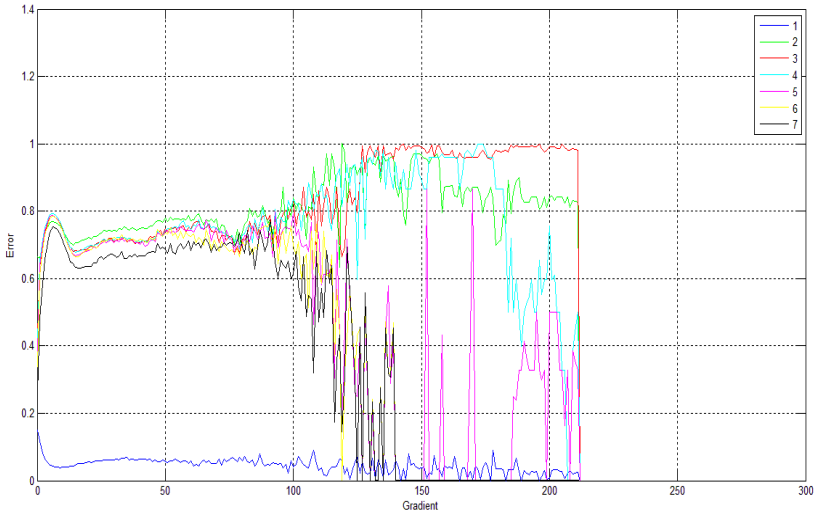
Example 1: Error analysis in the HSI, La*b*, and YCbCr spaces

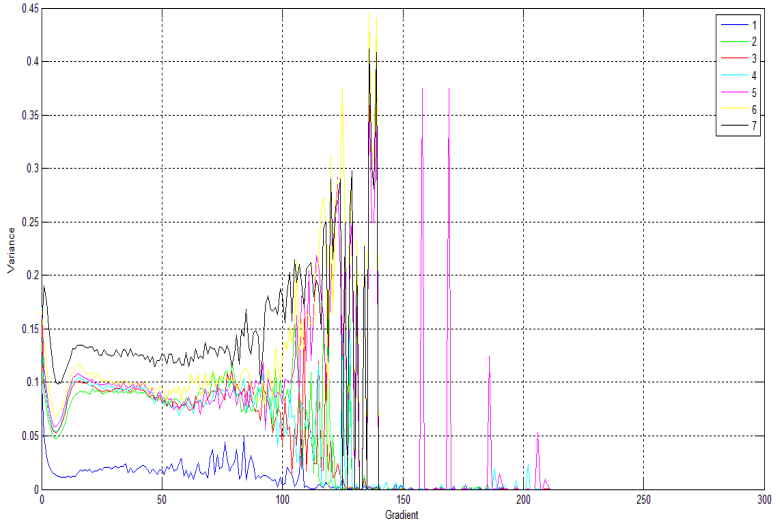
Copyright card sample #1: H channel [0~1]





Error (absolute sine difference) mean values and variances w.r.t. intensity levels

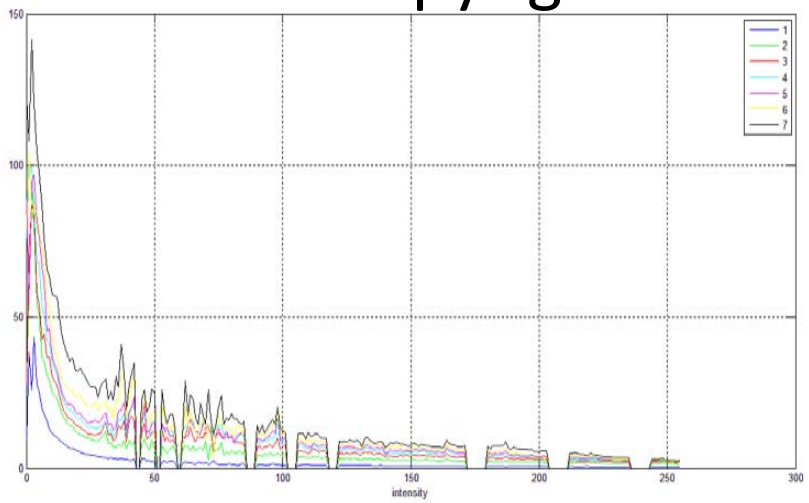


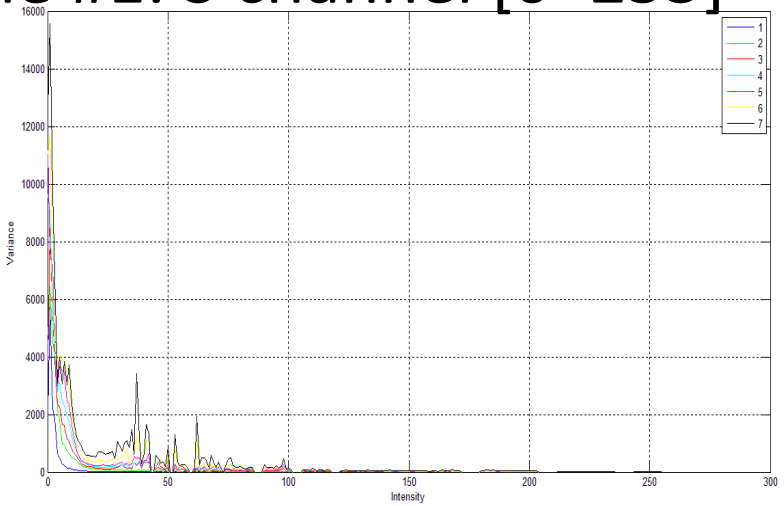


Error (absolute sine difference) mean values and variances w.r.t. intensity gradients

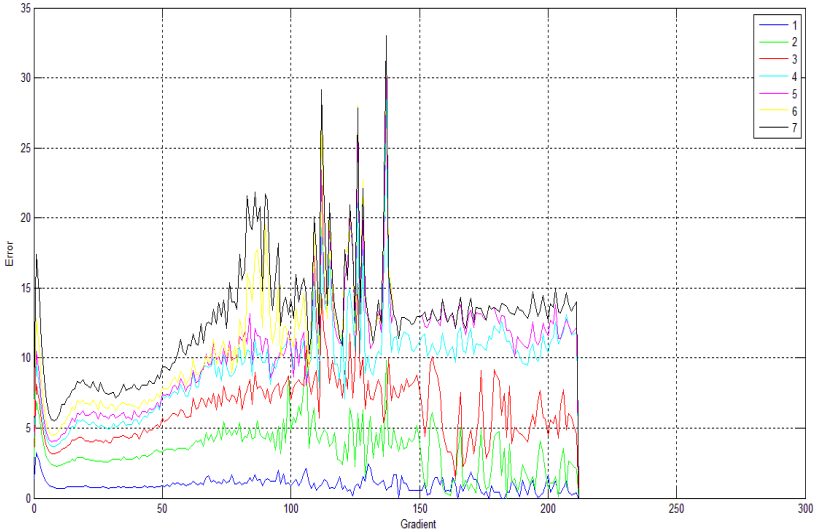
Example 1: Error analysis in the HSI, La*b*, and YCbCr spaces

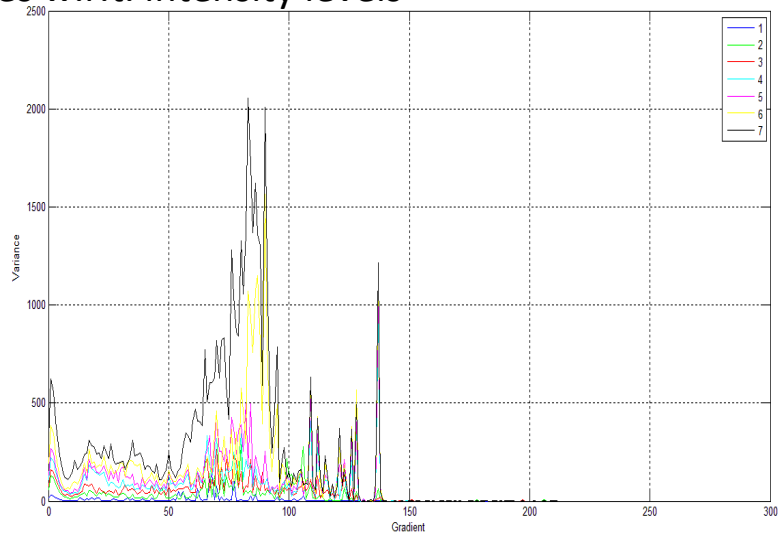
Copyright card sample #1: S channel [0~255]





Error mean values and variances w.r.t. intensity levels

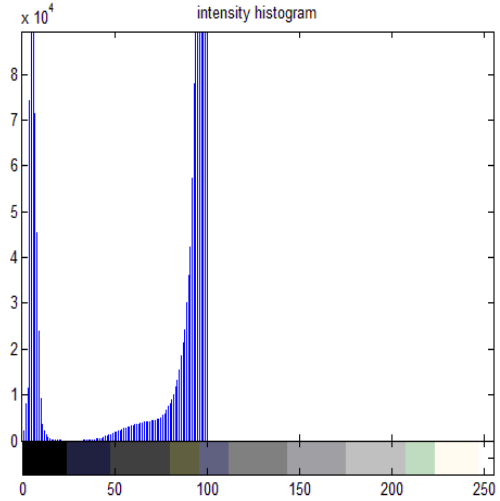


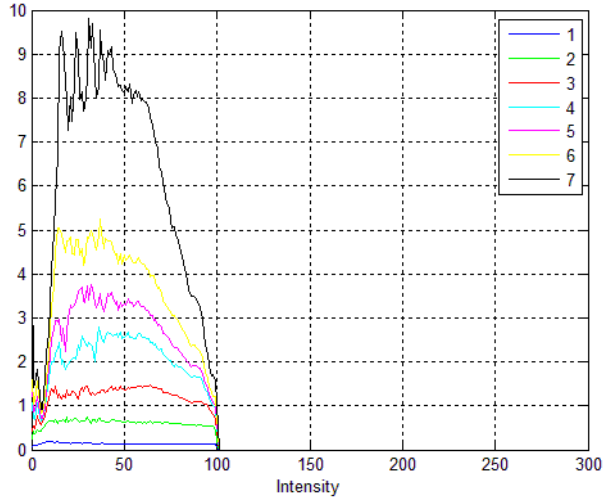


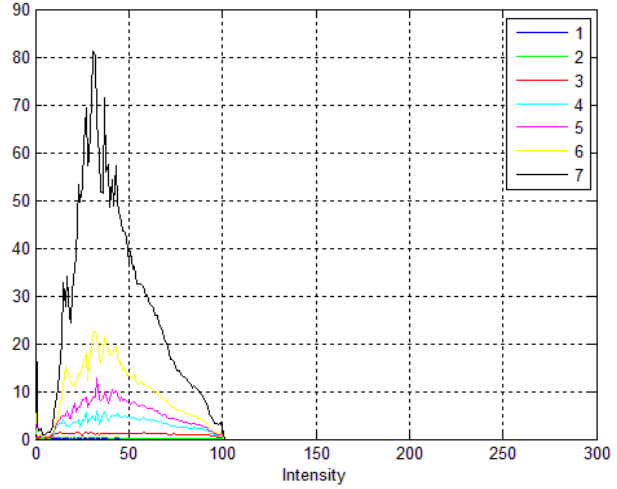
Error mean values and variances w.r.t. intensity gradients

Example 1: Error analysis in the HSI, La*b*, and YCbCr spaces

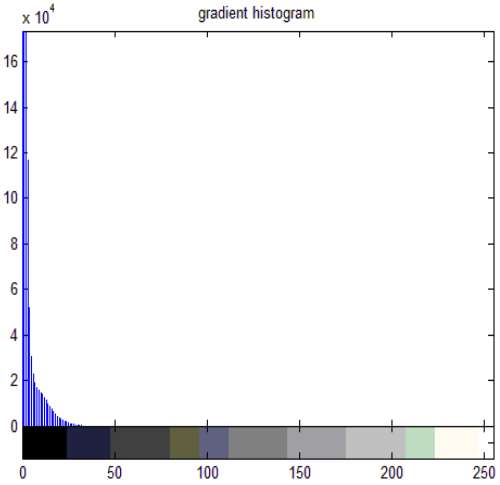
Copyright card sample #1: L channel [0~100]

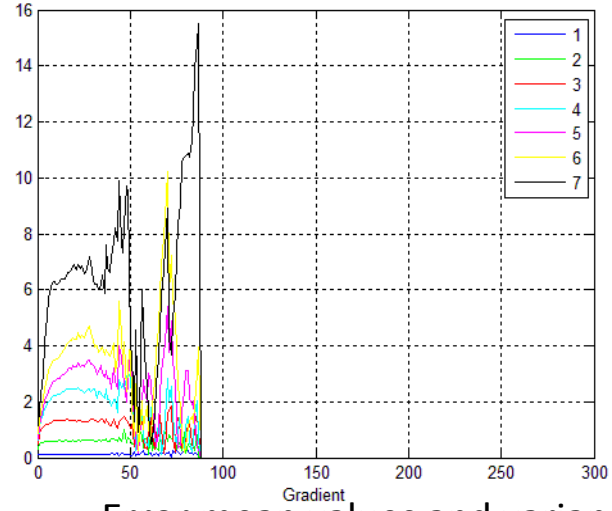


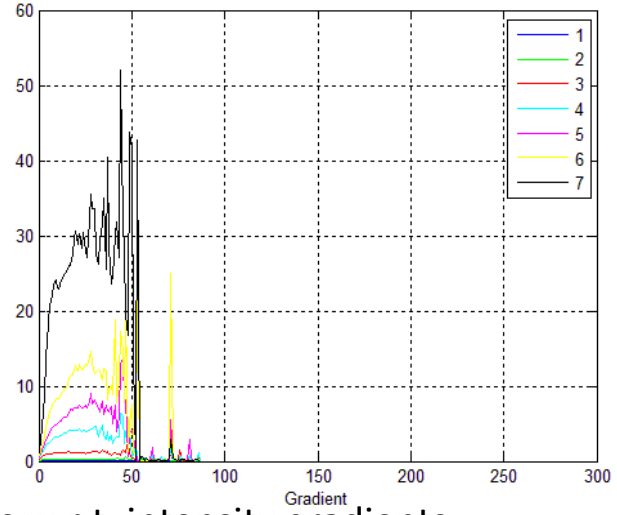




Error mean values and variances w.r.t. intensity levels



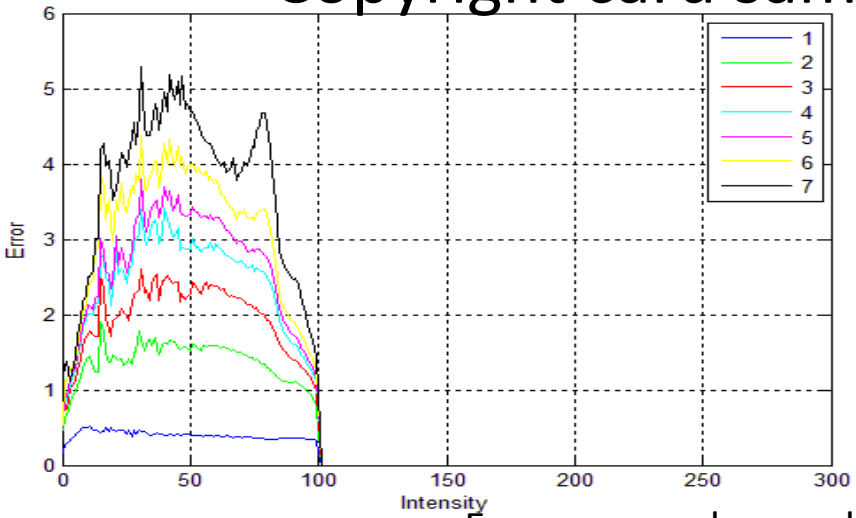


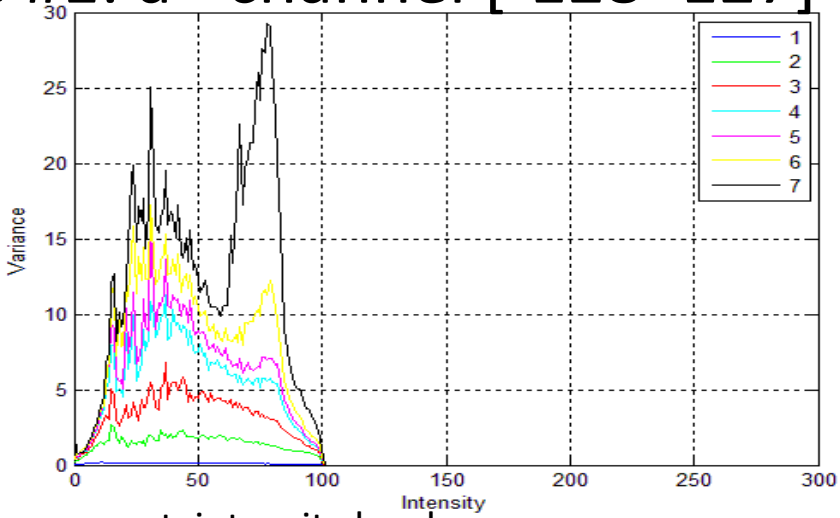


Error mean values and variances w.r.t. intensity gradients

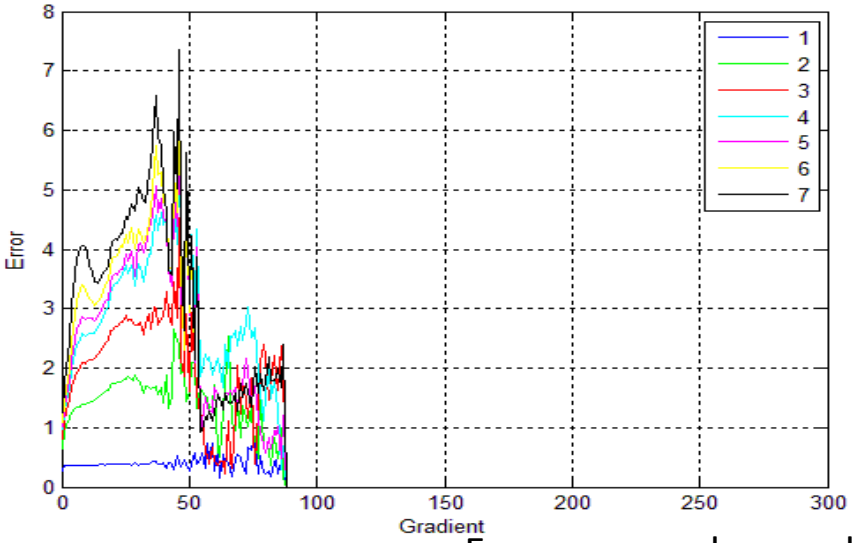
Example 1: Error analysis in the HSI, La*b*, and YCbCr spaces

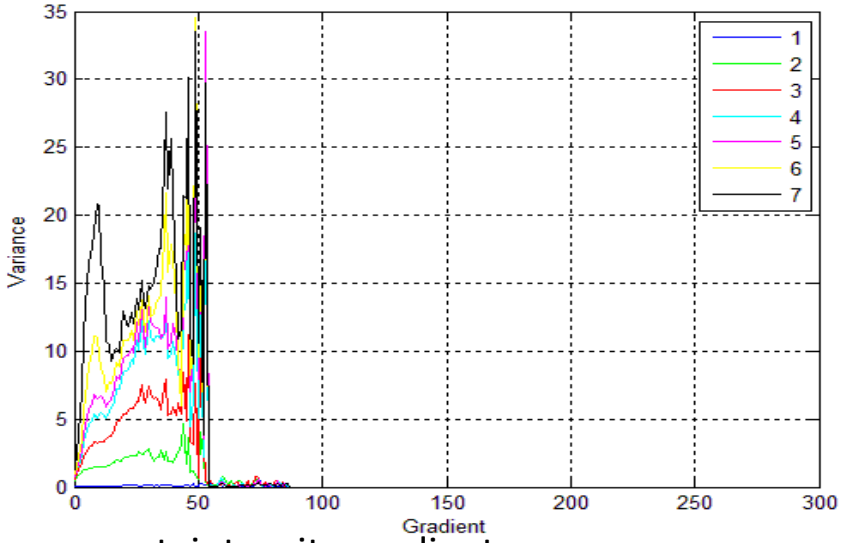
Copyright card sample #1: a* channel [-128~127]





Error mean values and variances w.r.t. intensity levels

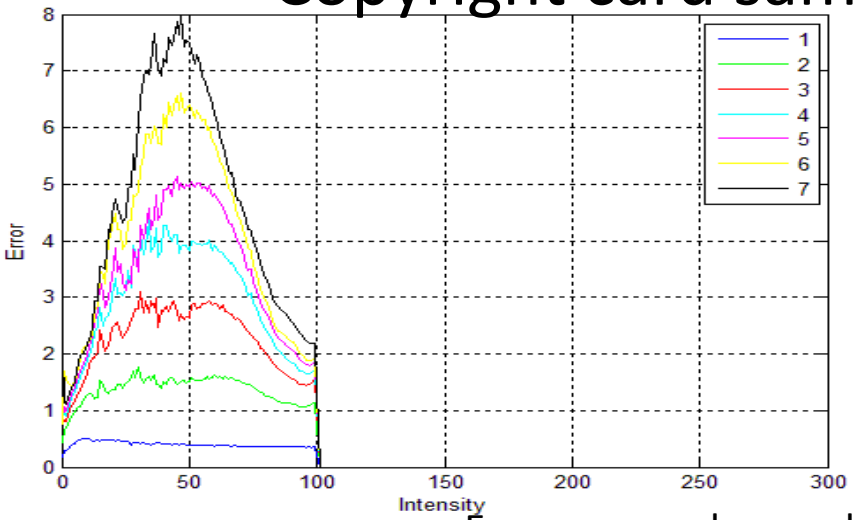


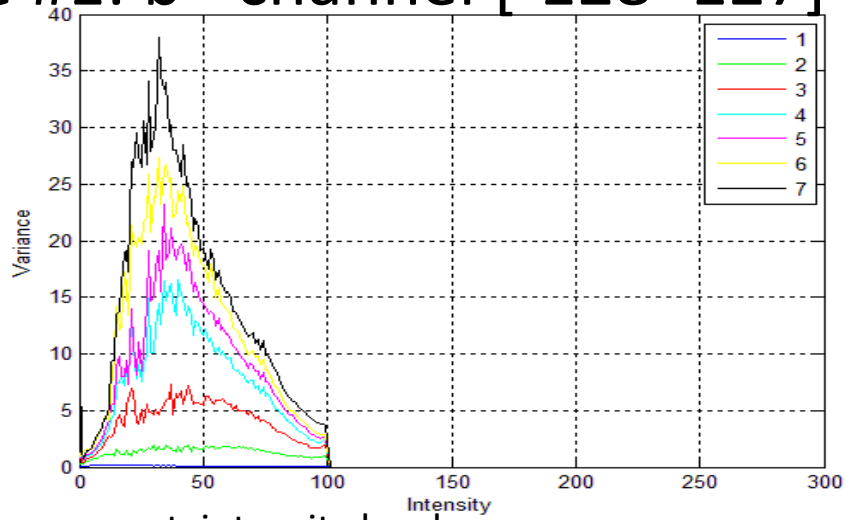


Error mean values and variances w.r.t. intensity gradients

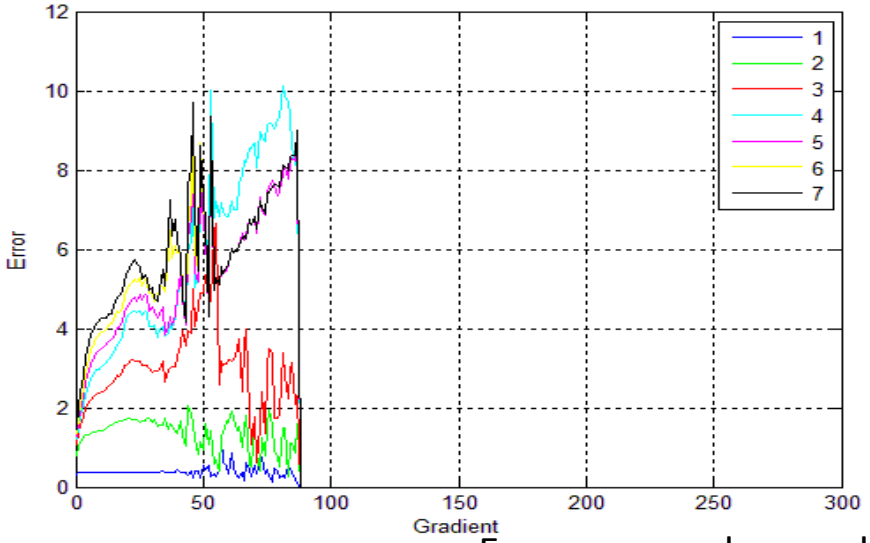
Example 1: Error analysis in the HSI, La*b*, and YCbCr spaces

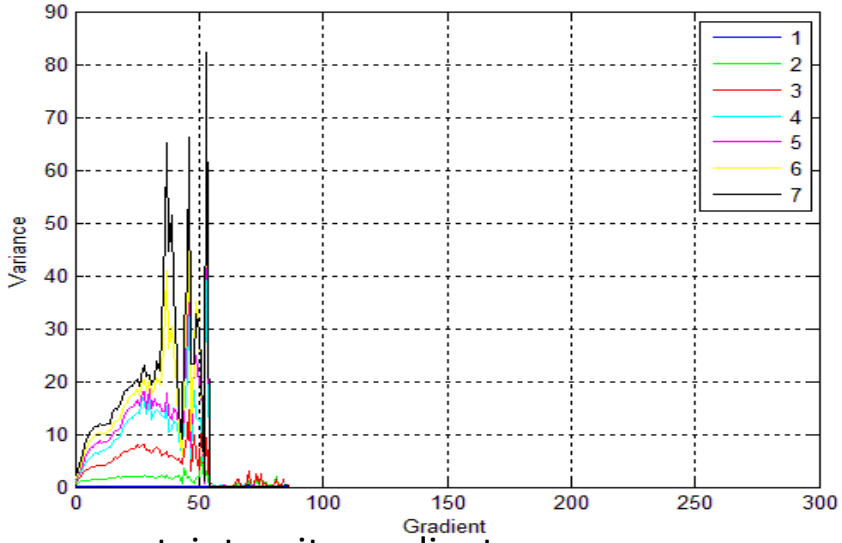
Copyright card sample #1: b* channel [-128~127]





Error mean values and variances w.r.t. intensity levels

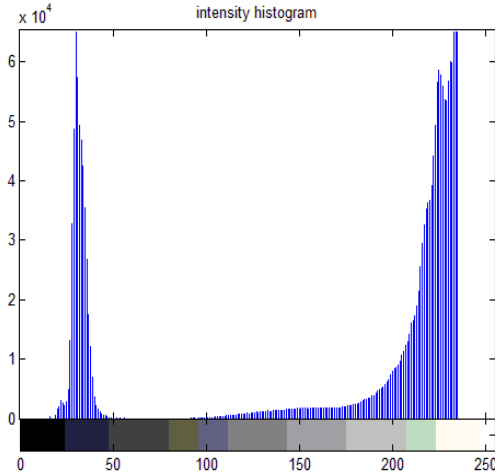


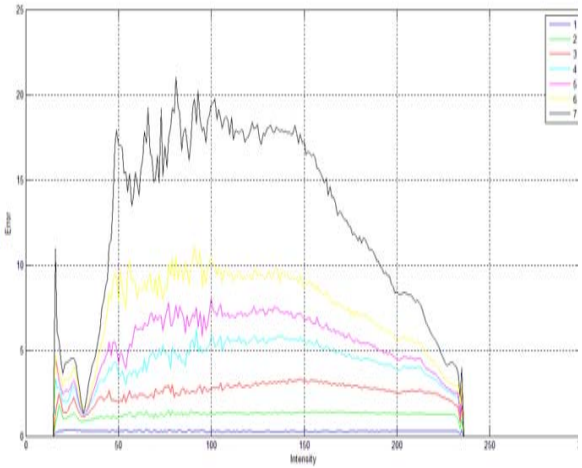


Error mean values and variances w.r.t. intensity gradients

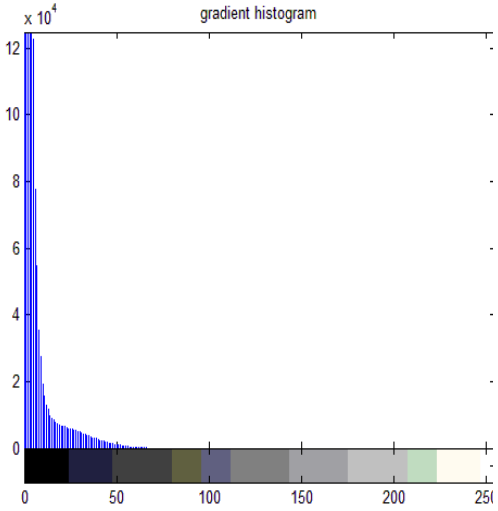
Example 1: Error analysis in the HSI, La*b*, and YCbCr spaces

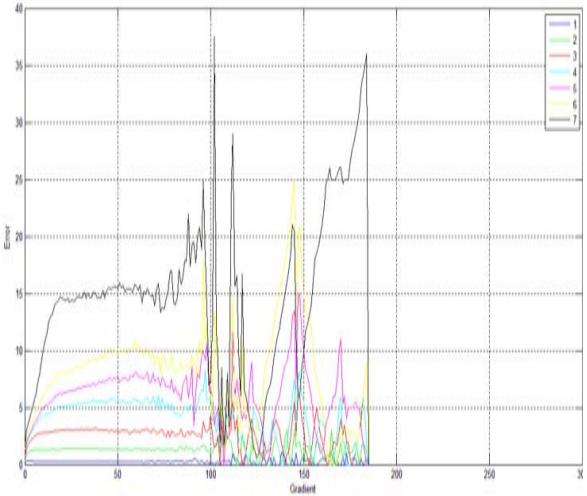
Copyright card sample #1: Y channel [16~235]





Error mean values and variances w.r.t. intensity levels

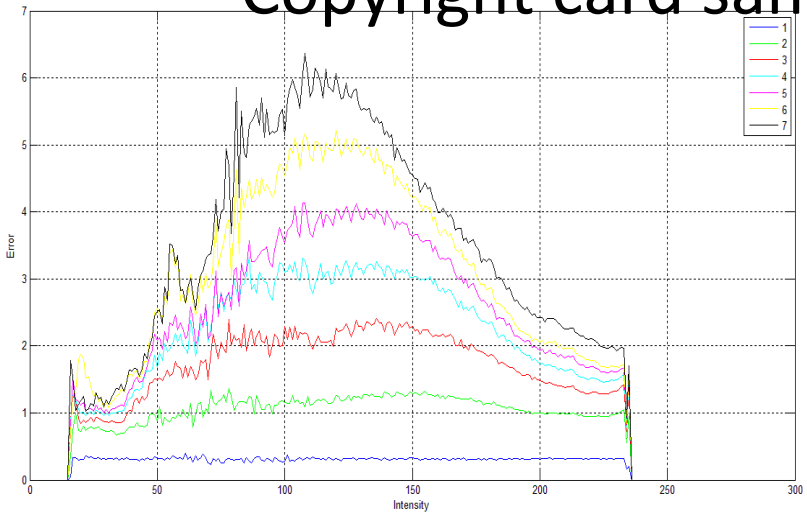


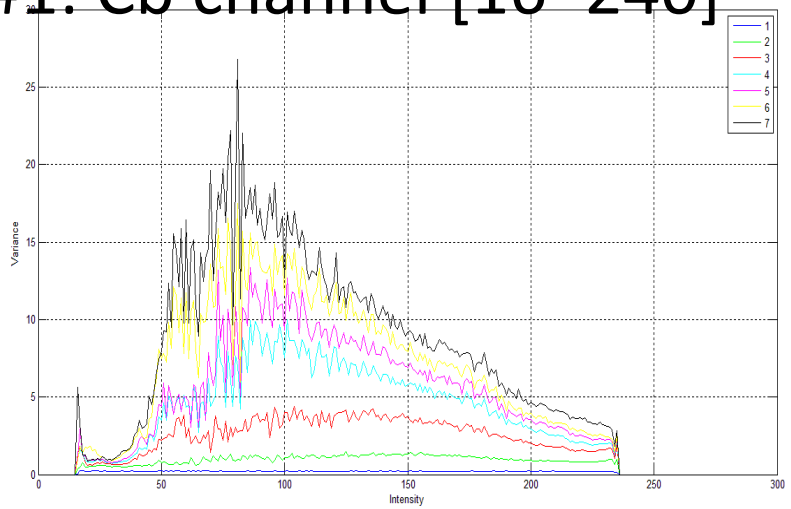


Error mean values and variances w.r.t. intensity gradients

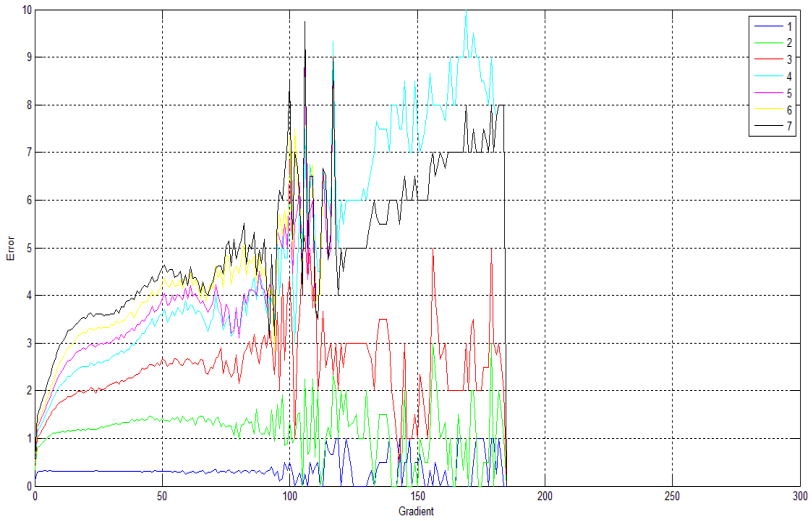
Example 1: Error analysis in the HSI, La*b*, and YCbCr spaces

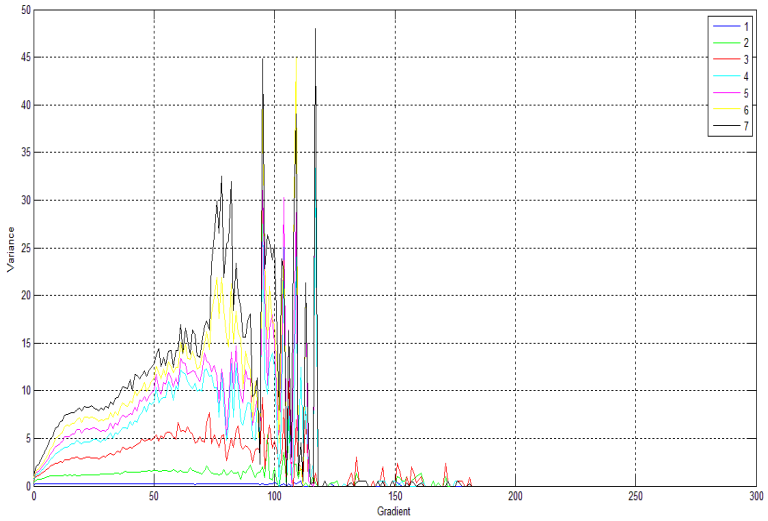
Copyright card sample #1: Cb channel [16~240]





Error mean values and variances w.r.t. intensity levels

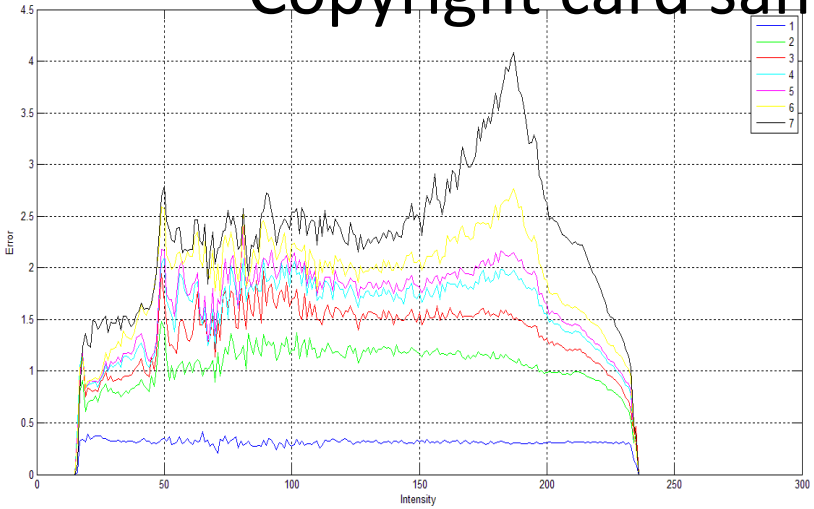


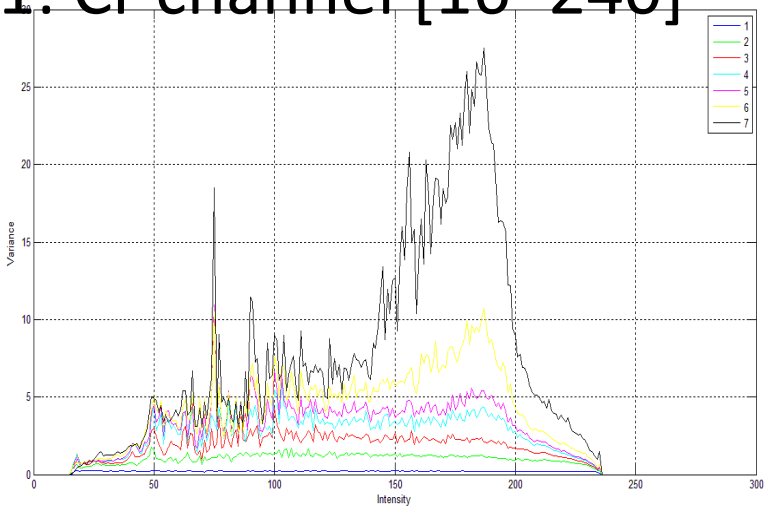


Error mean values and variances w.r.t. intensity gradients

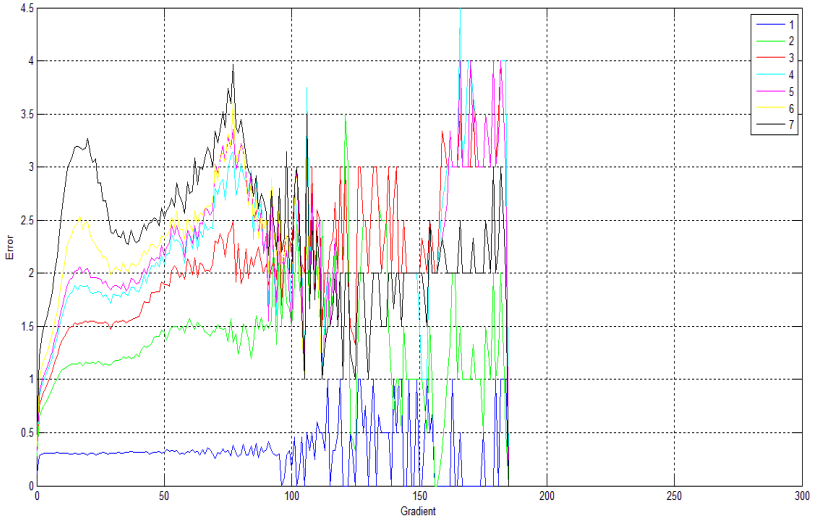
Example 1: Error analysis in the HSI, La*b*, and YCbCr spaces

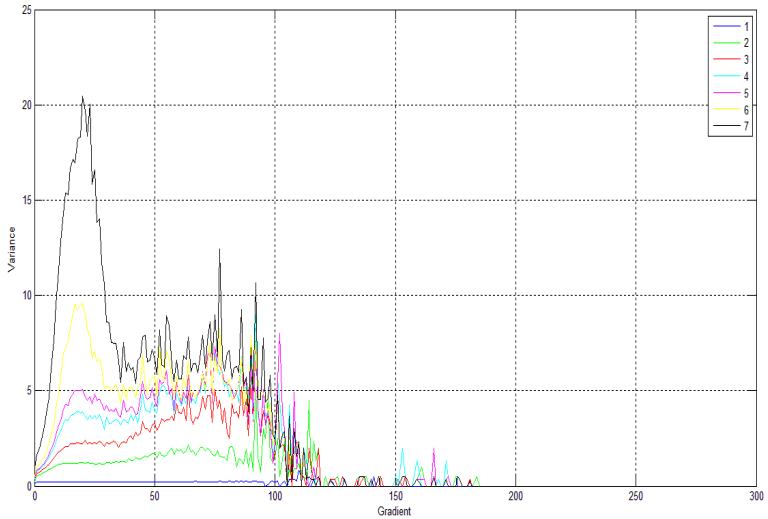
Copyright card sample #1: Cr channel [16~240]





Error mean values and variances w.r.t. intensity levels

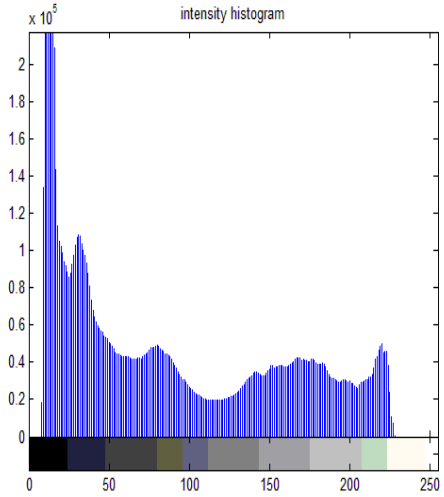


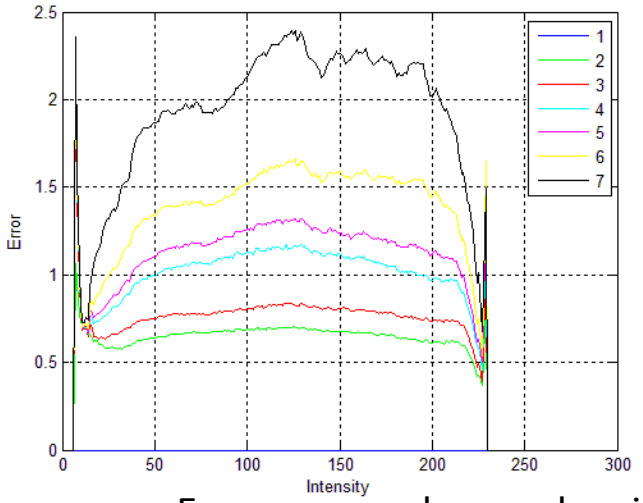


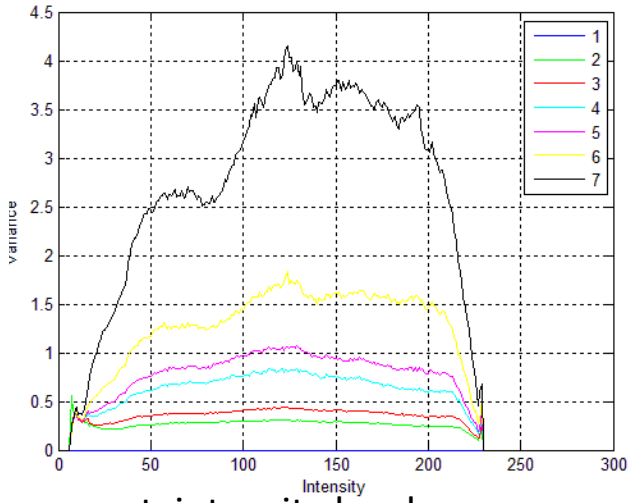
Error mean values and variances w.r.t. intensity gradients

Example 1: Error analysis in the HSI, La*b*, YCbCr spaces

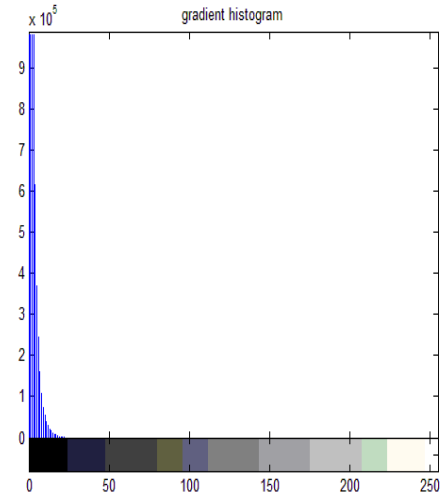
Color Photograph sample: I channel [0~255]

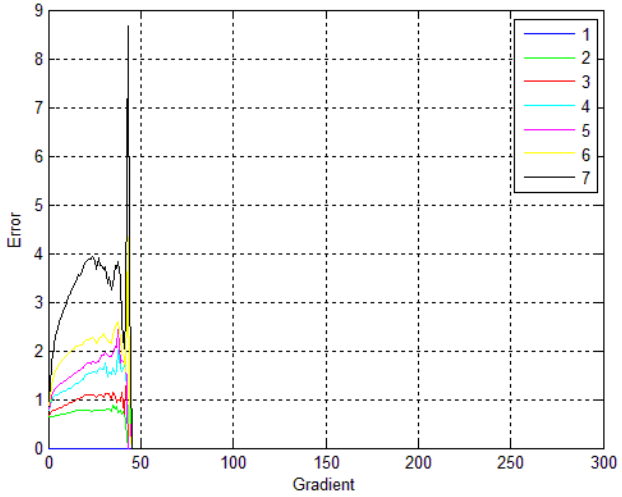


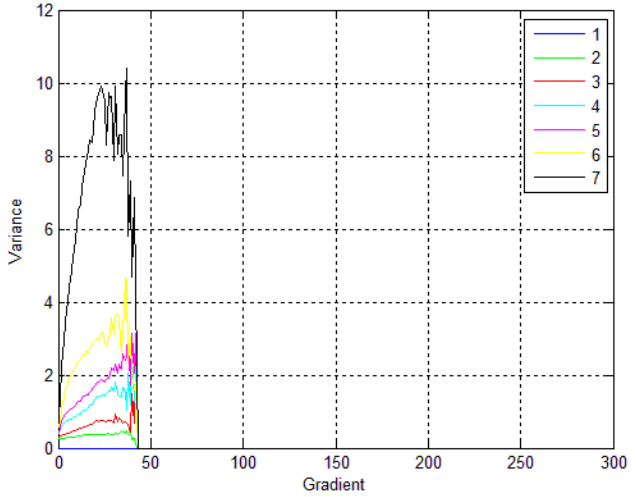




Error mean values and variances w.r.t. intensity levels



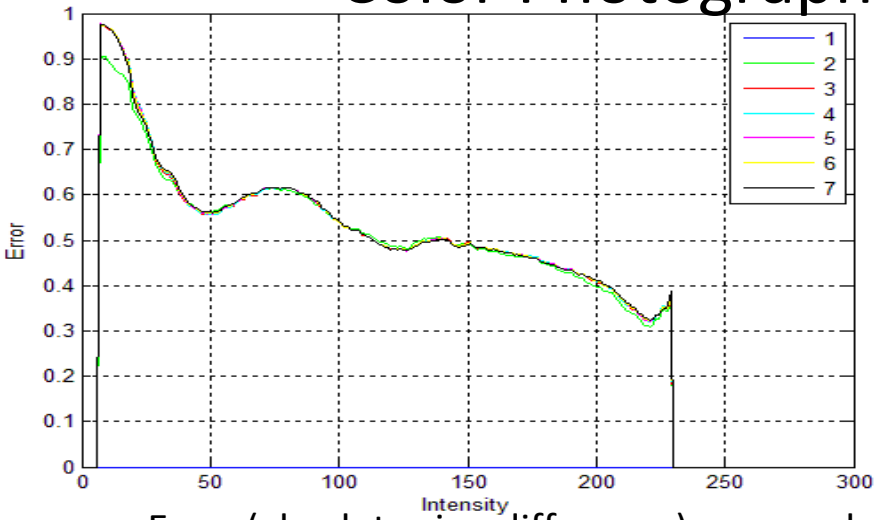


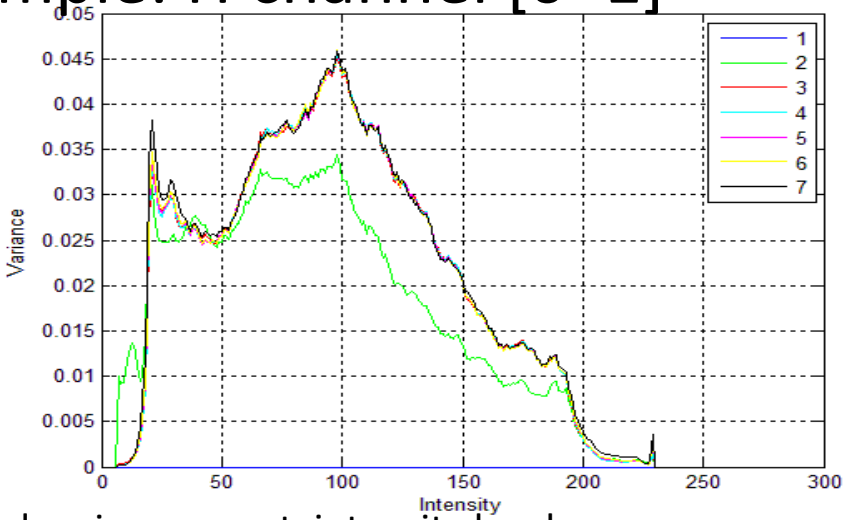


Error mean values and variances w.r.t. intensity gradients

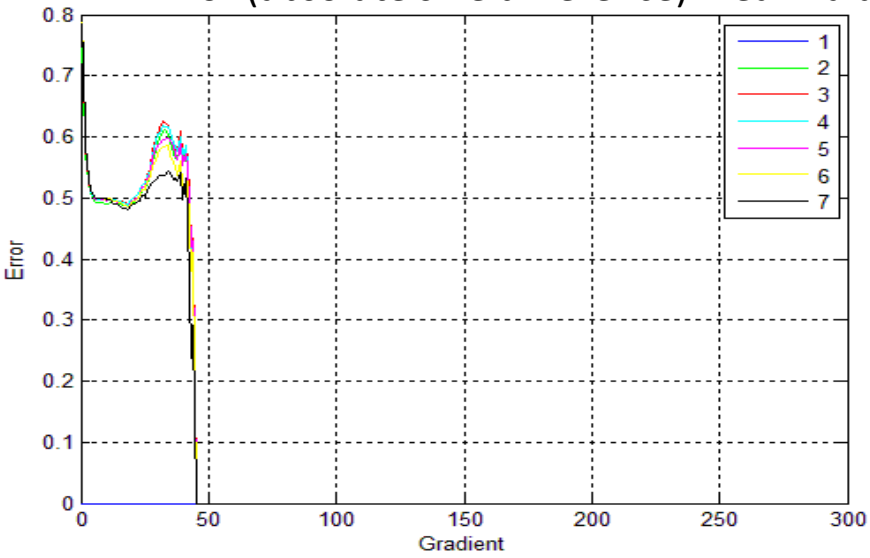
Example 1: Error analysis in the HSI, La*b*, YCbCr spaces

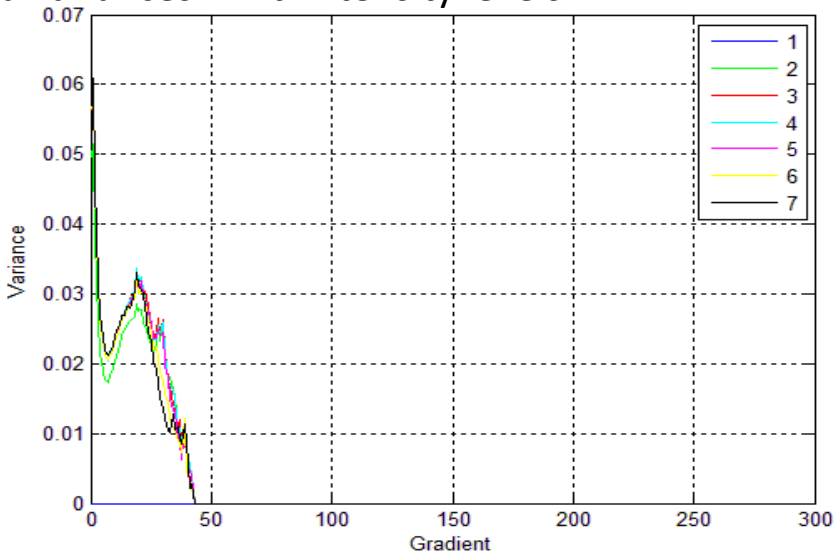
Color Photograph sample: H channel [0~1]





Error (absolute sine difference) mean values and variances w.r.t. intensity levels

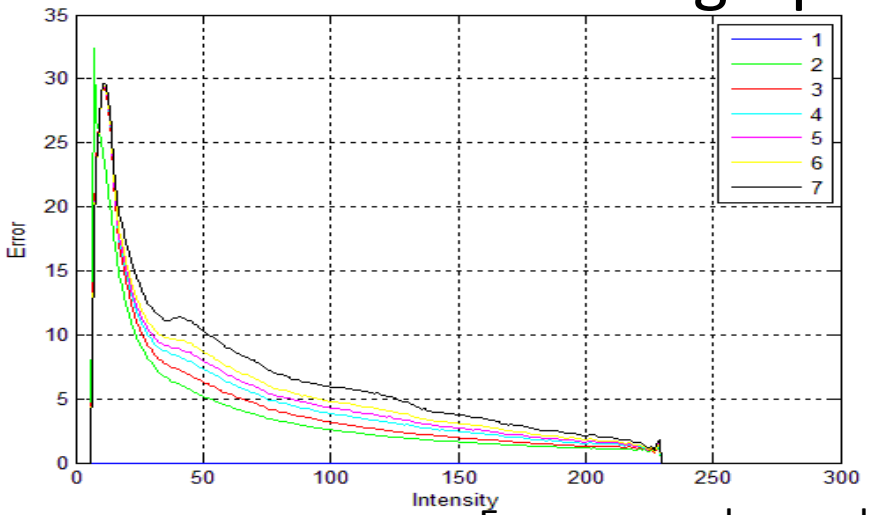


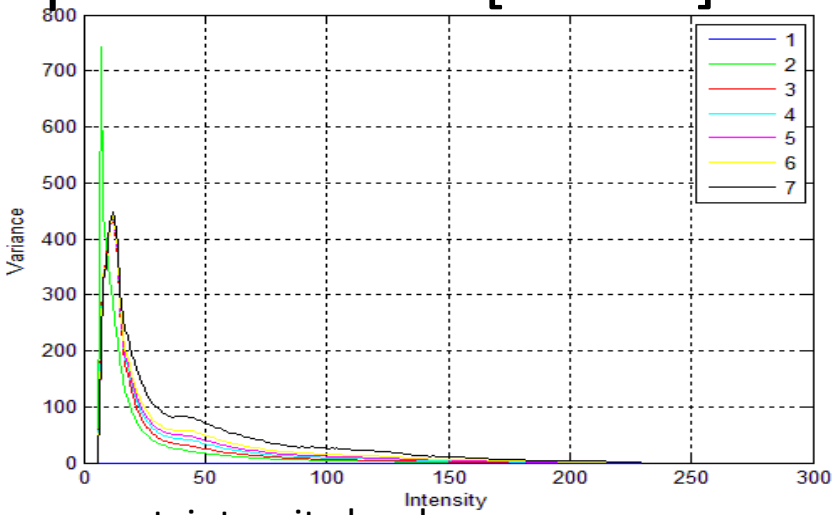


Error (absolute sine difference) mean values and variances w.r.t. intensity gradients

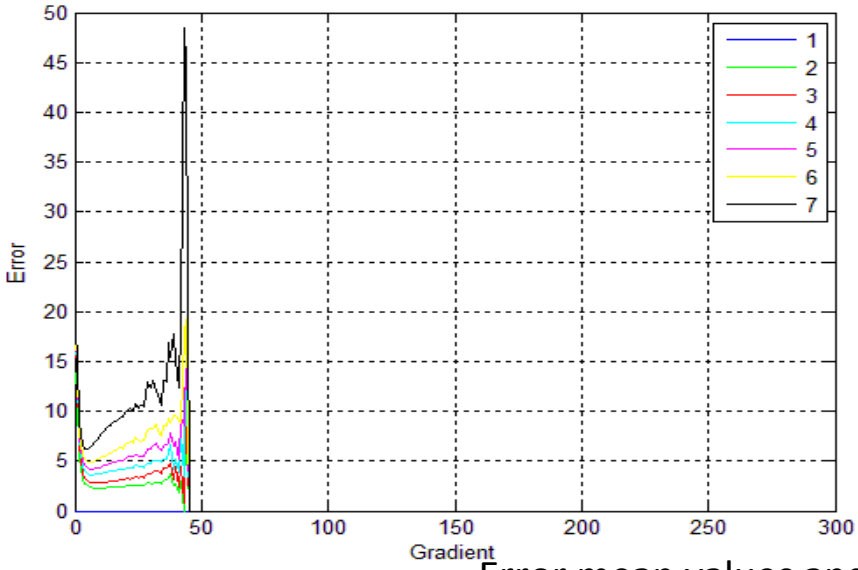
Example 1: Error analysis in the HSI, La*b*, YCbCr spaces

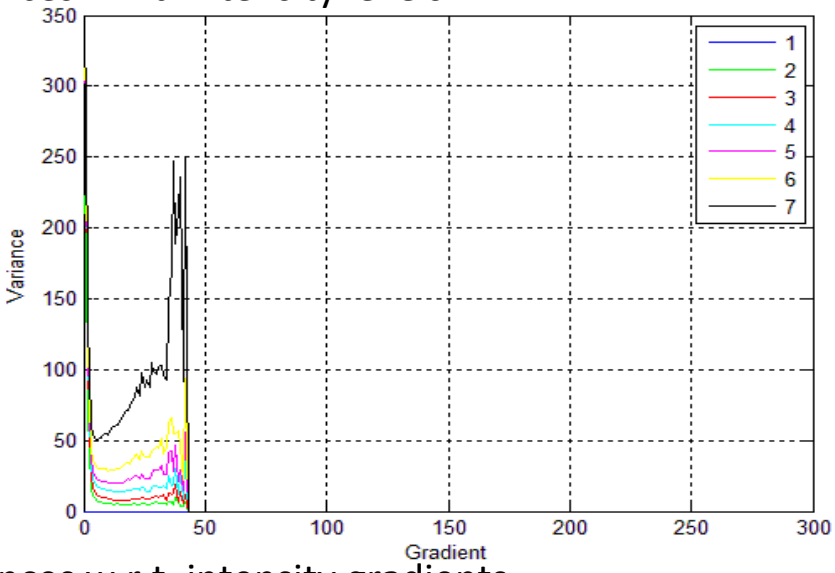
Color Photograph sample: S channel [0~255]





Error mean values and variances w.r.t. intensity levels

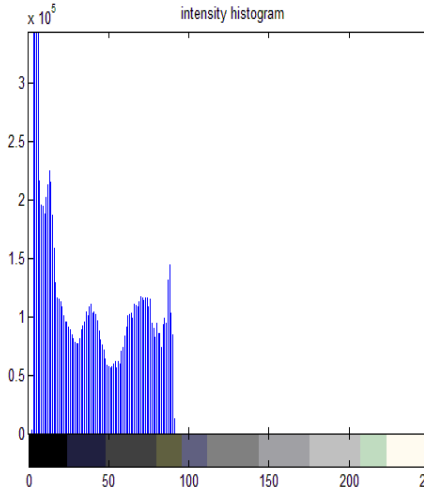


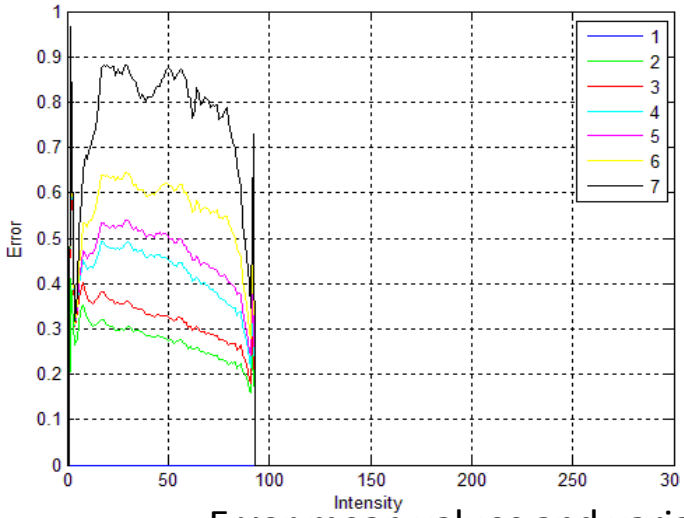


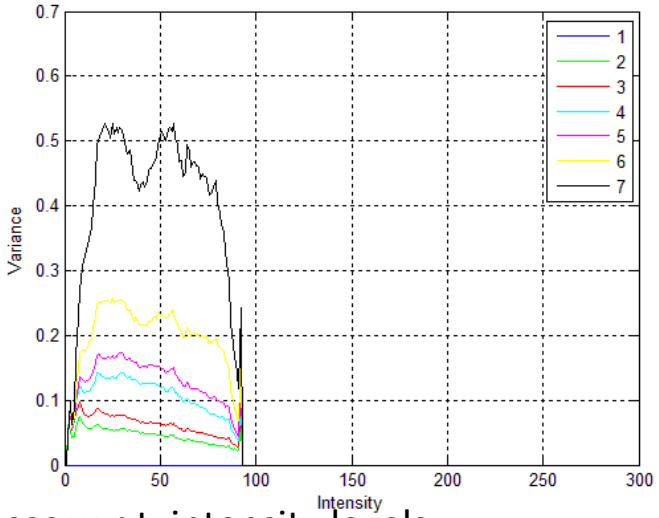
Error mean values and variances w.r.t. intensity gradients

Example 1: Error analysis in the HSI, La*b*, YCbCr spaces

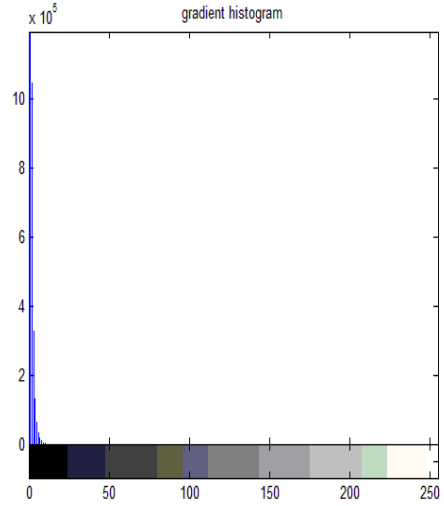
Color Photograph sample: L channel [0~100]

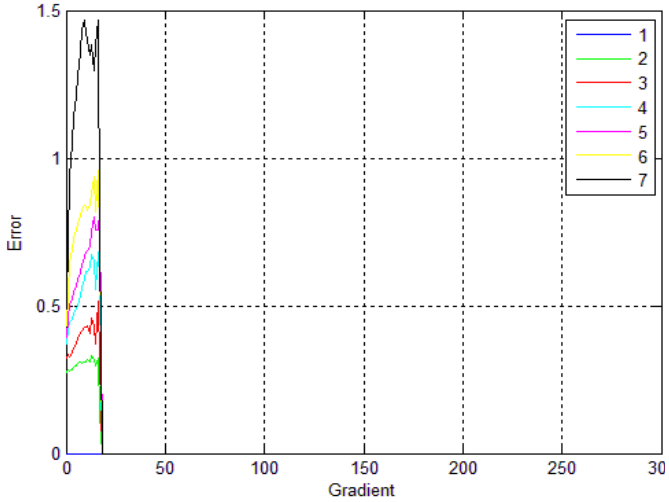


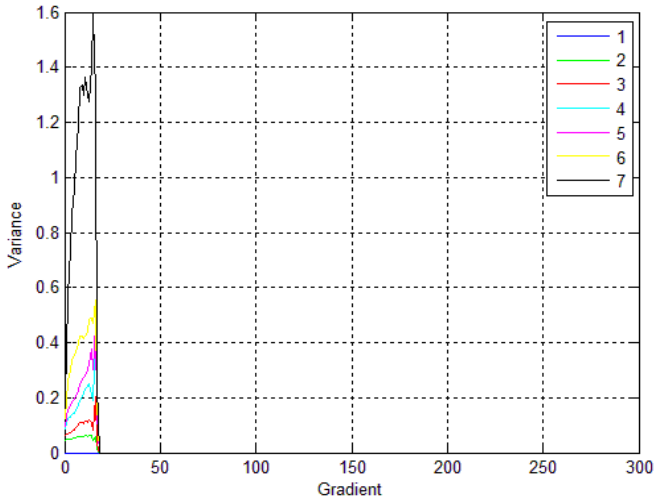




Error mean values and variances w.r.t. intensity levels



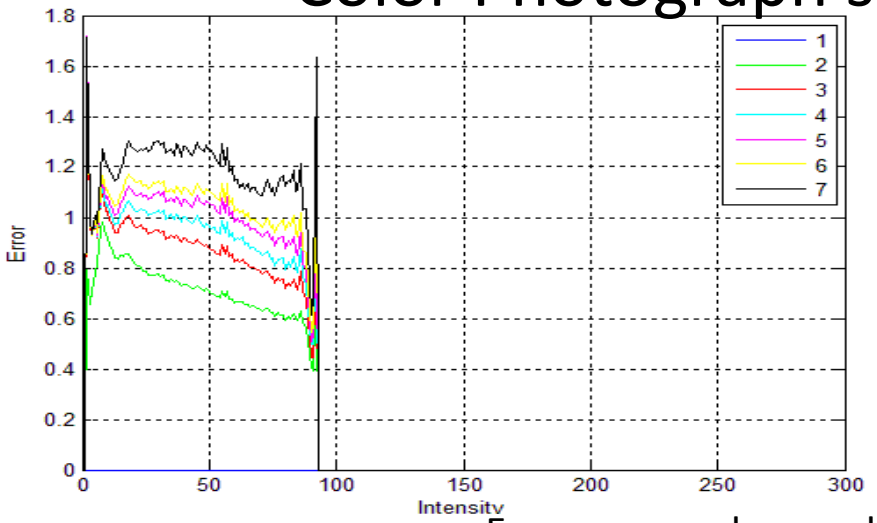


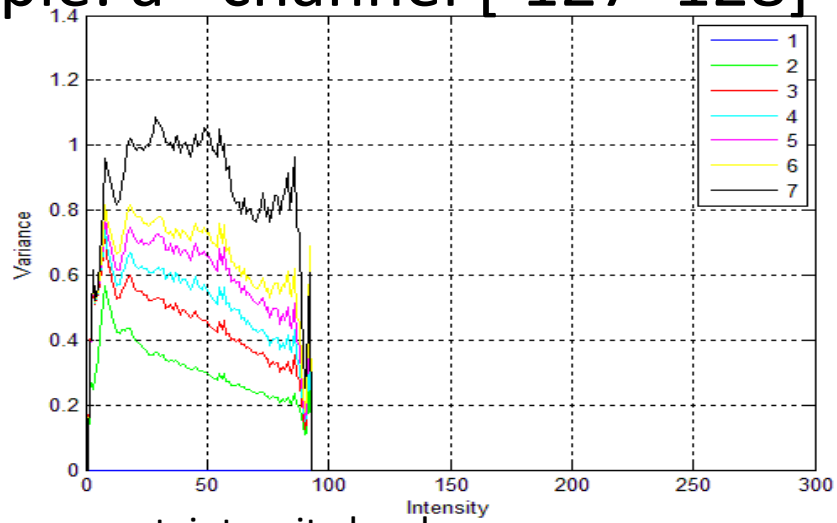


Error mean values and variances w.r.t. intensity gradients

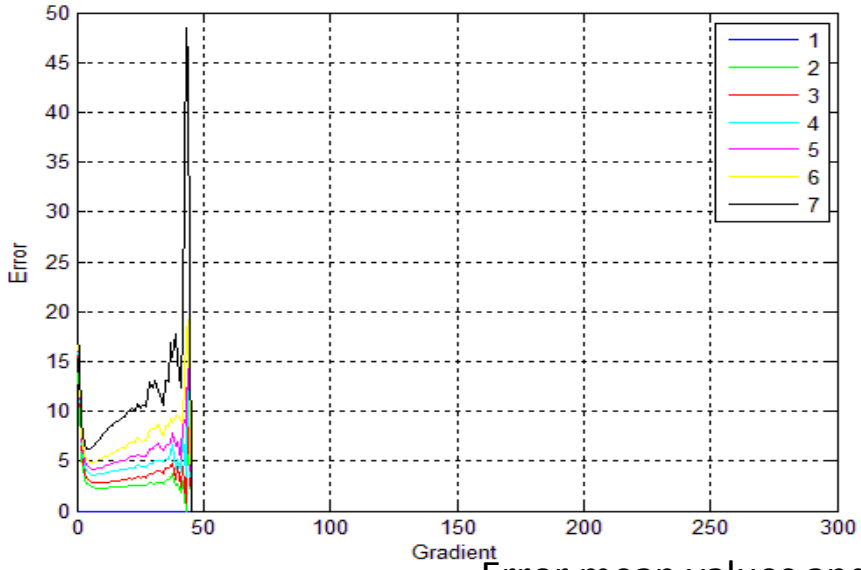
Example 1: Error analysis in the HSI, La*b*, YCbCr spaces

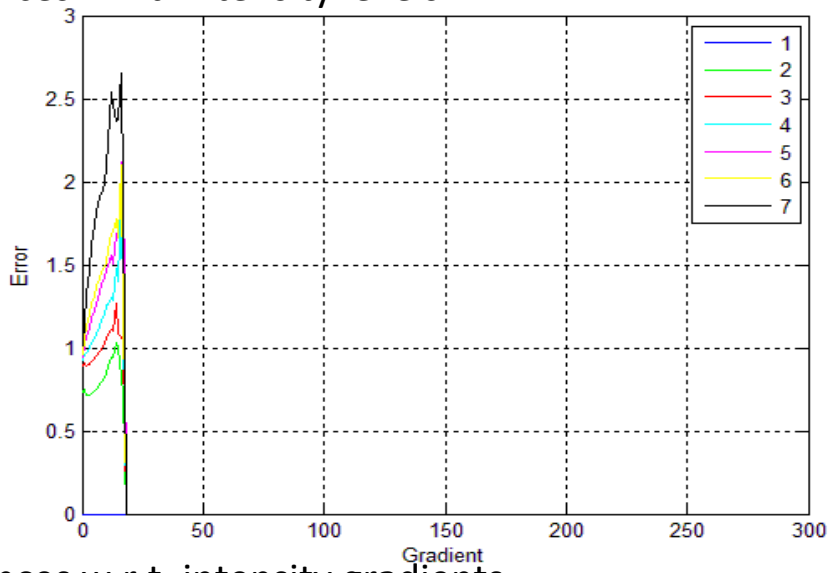
Color Photograph sample: a* channel [-127~128]





Error mean values and variances w.r.t. intensity levels

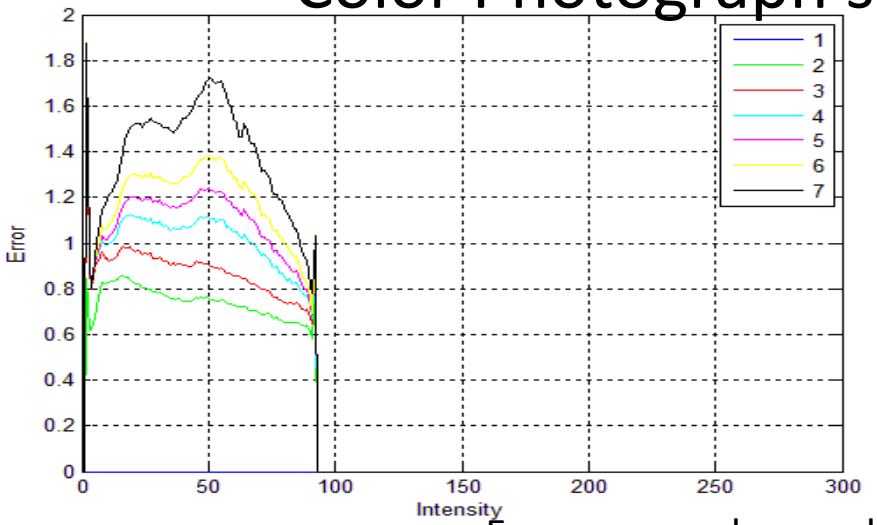


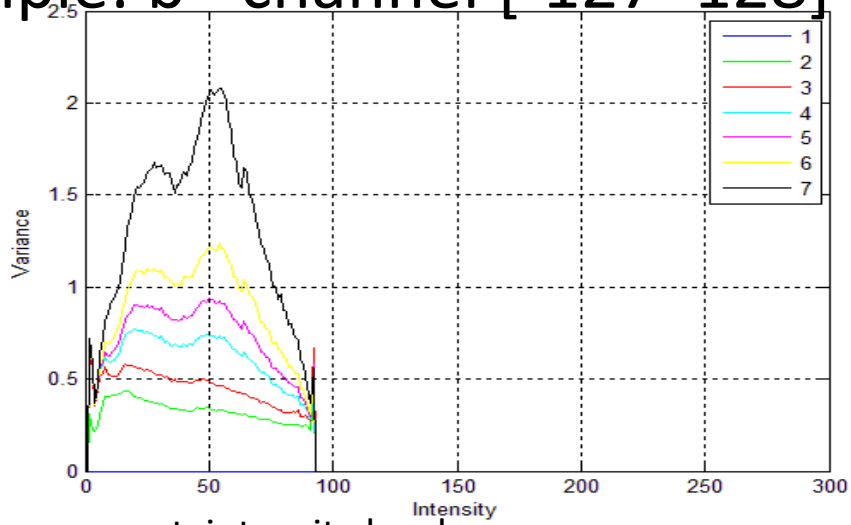


Error mean values and variances w.r.t. intensity gradients

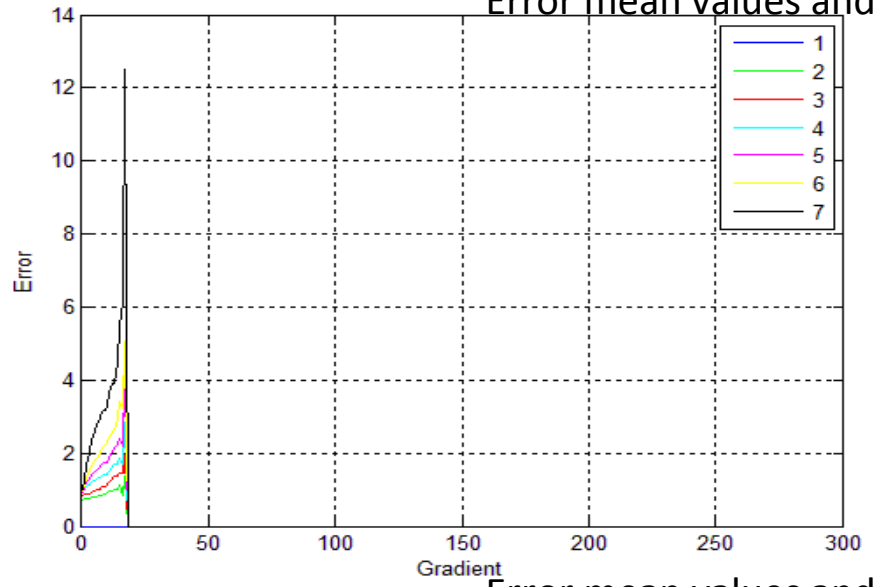
Example 1: Error analysis in the HSI, La*b*, YCbCr spaces

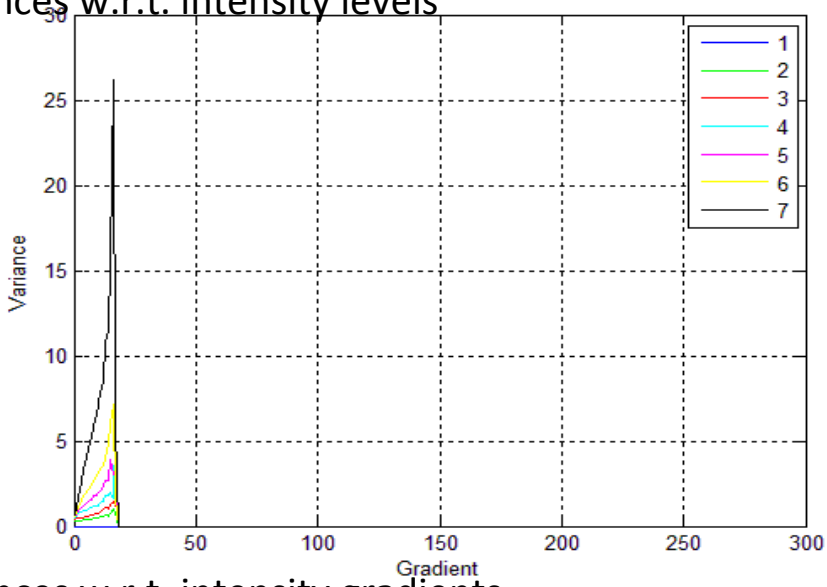
Color Photograph sample: b* channel [-127~128]





Error mean values and variances w.r.t. intensity levels

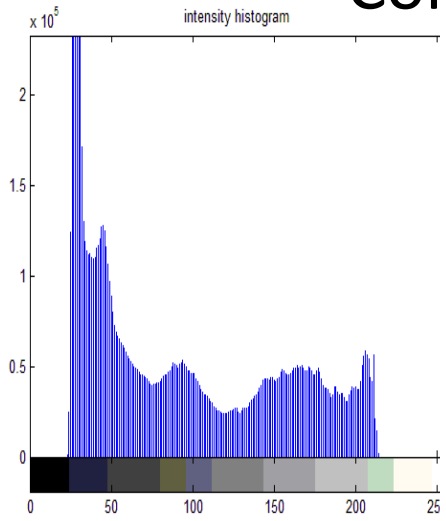


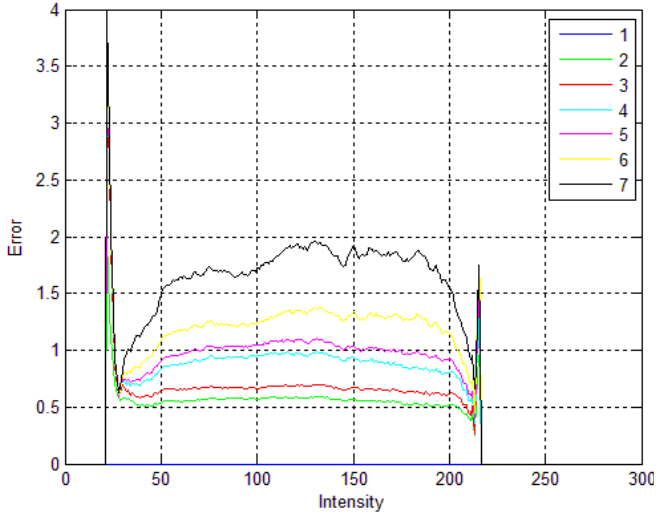


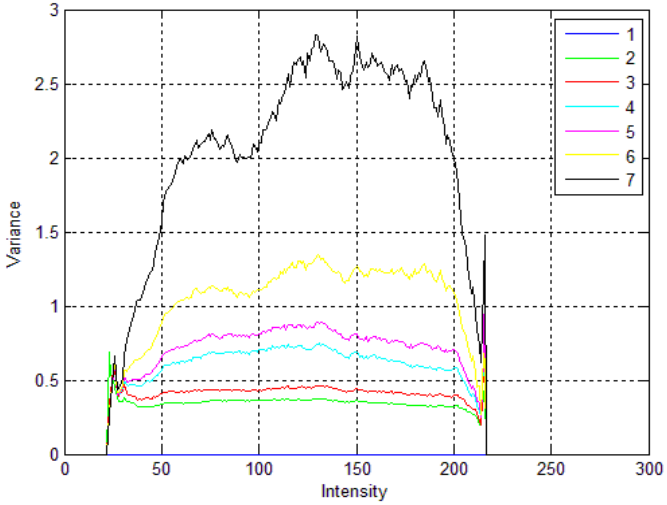
Error mean values and variances w.r.t. intensity gradients

Example 1: Error analysis in the HSI, La*b*, YCbCr spaces

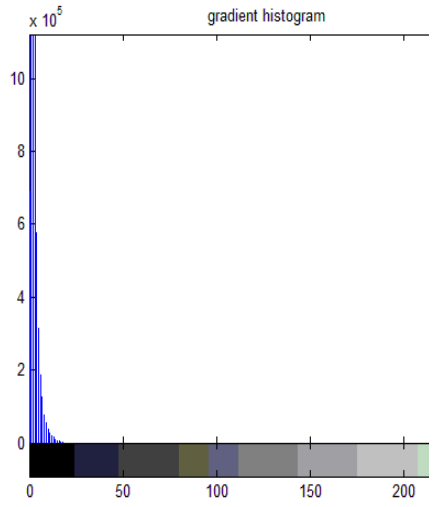
Color Photograph sample: Y channel [16~235]

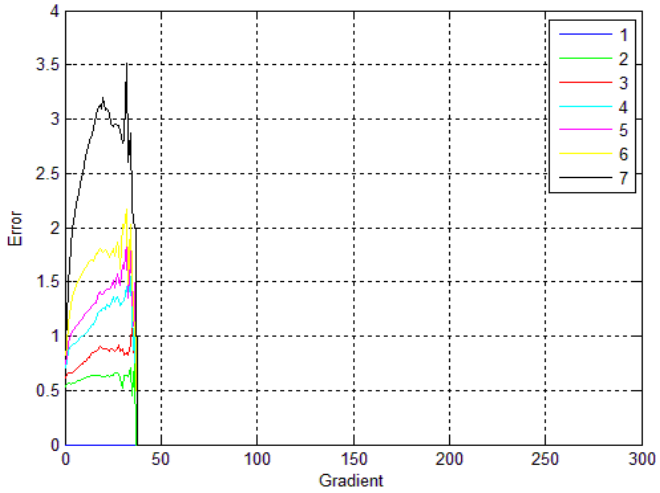


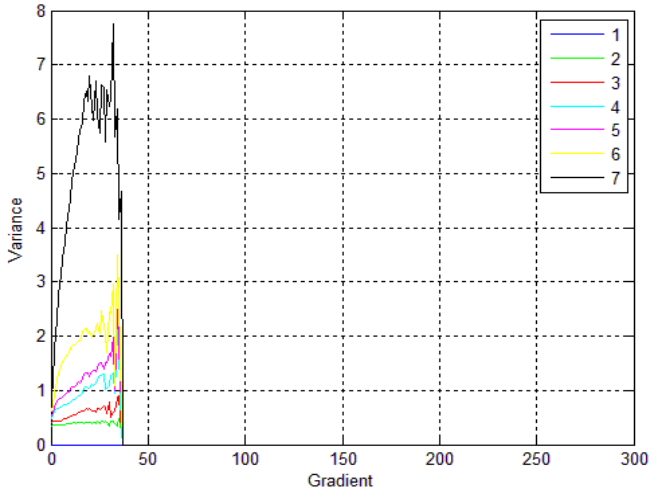




Error mean values and variances w.r.t. intensity levels



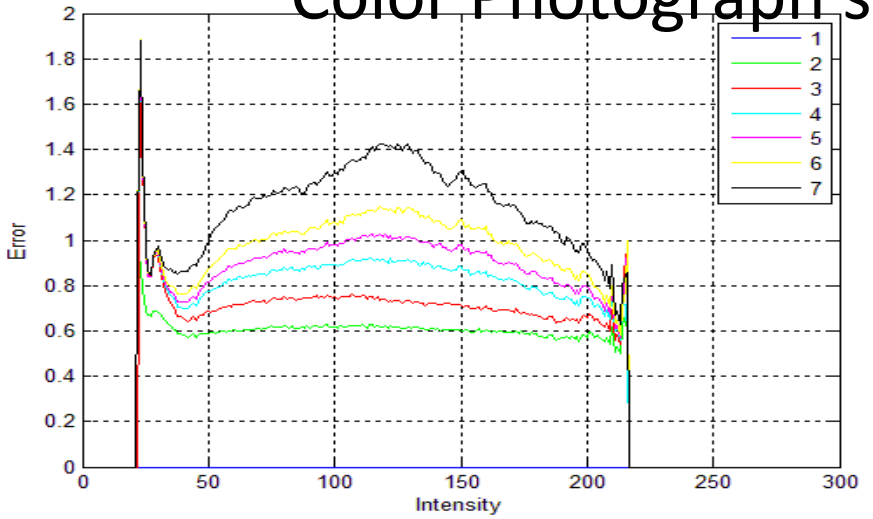


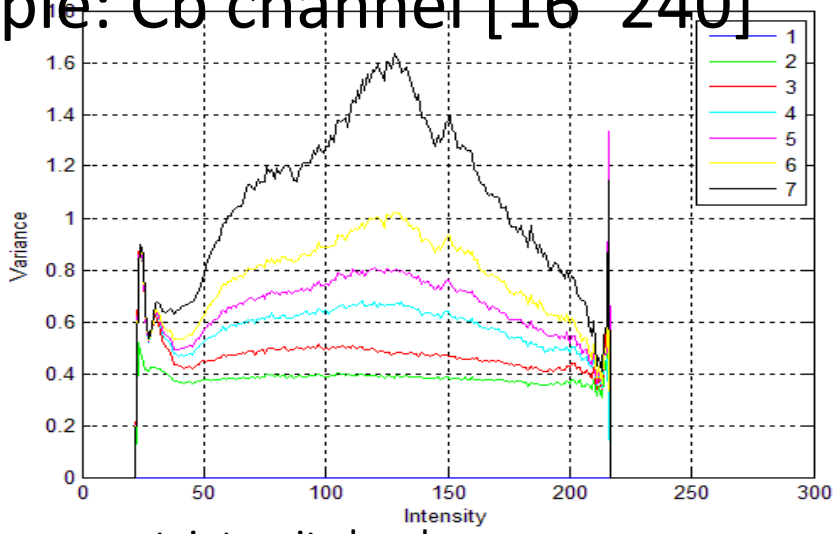


Error mean values and variances w.r.t. intensity gradients

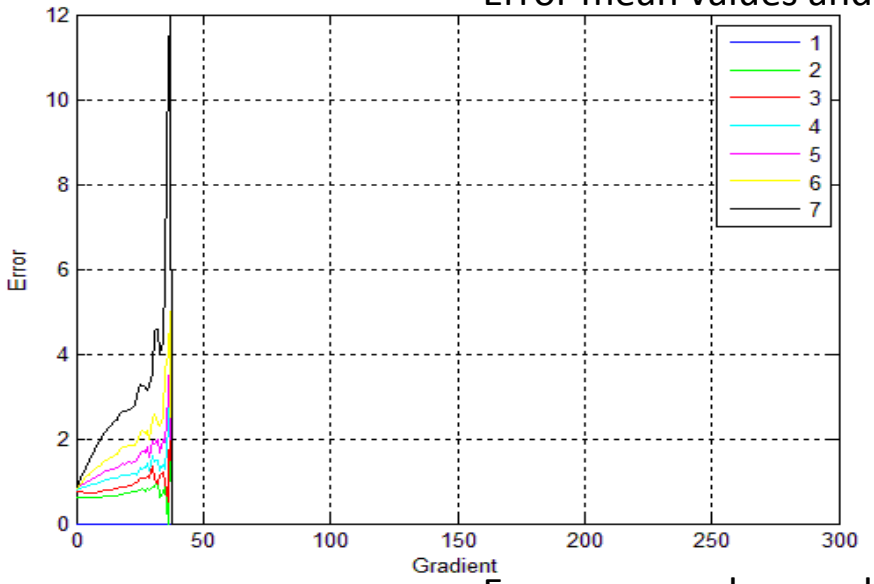
Example 1: Error analysis in the HSI, La*b*, YCbCr spaces

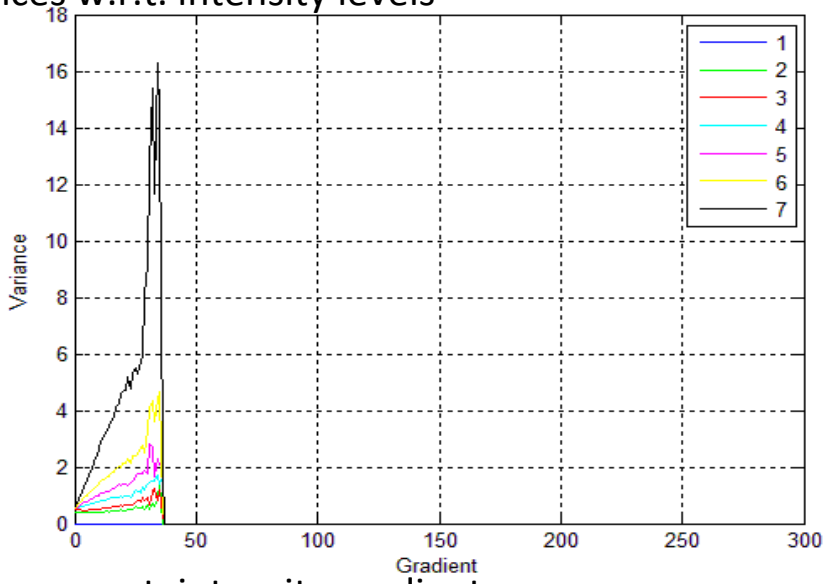
Color Photograph sample: Cb channel [16~240]





Error mean values and variances w.r.t. intensity levels

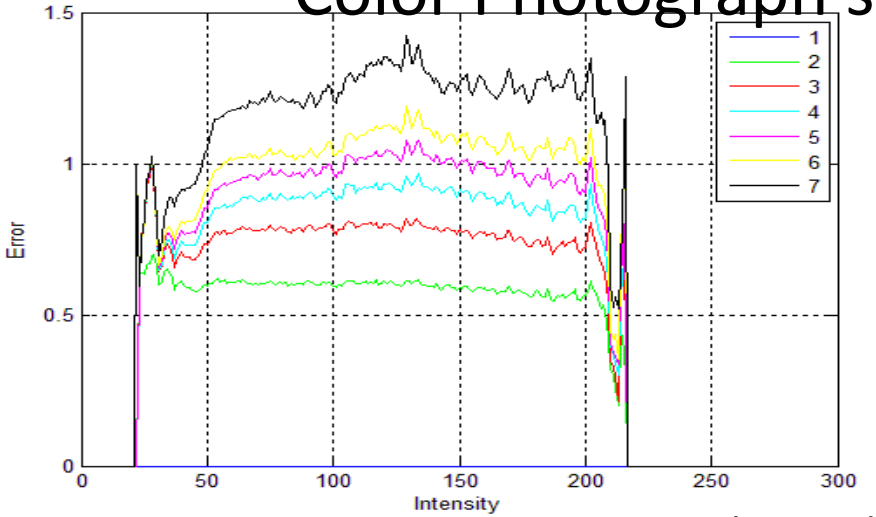


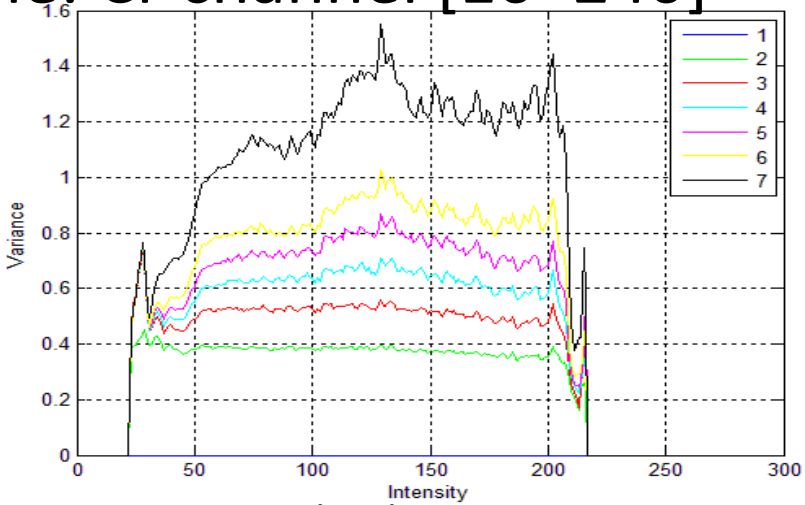


Error mean values and variances w.r.t. intensity gradients

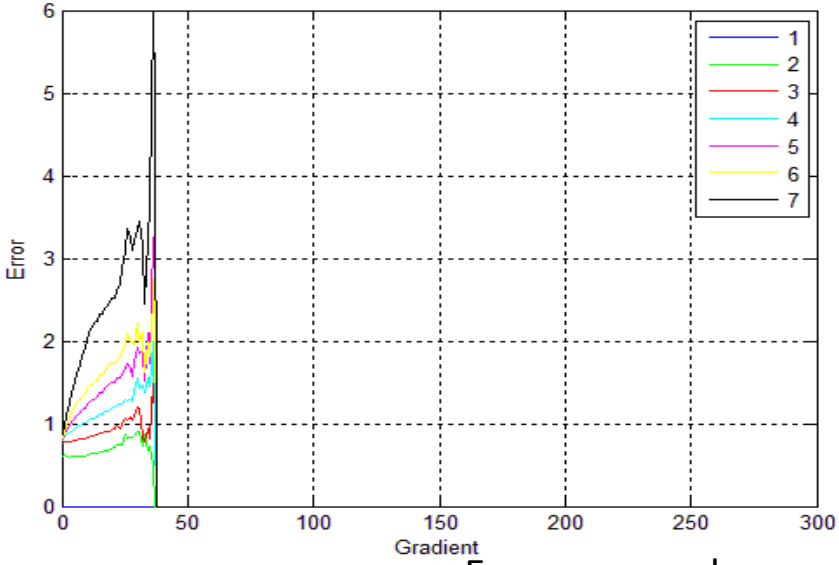
Example 1: Error analysis in the HSI, La*b*, YCbCr spaces

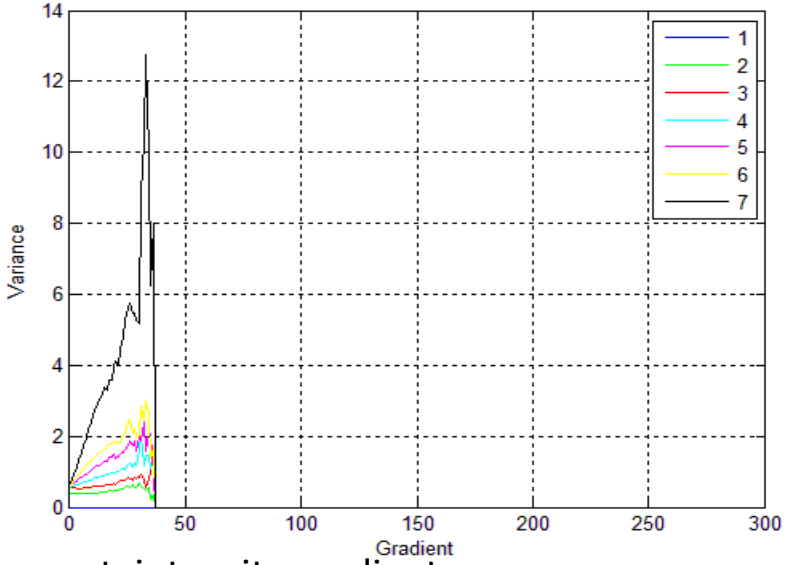
Color Photograph sample: Cr channel [16~240]





Error mean values and variances w.r.t. intensity levels

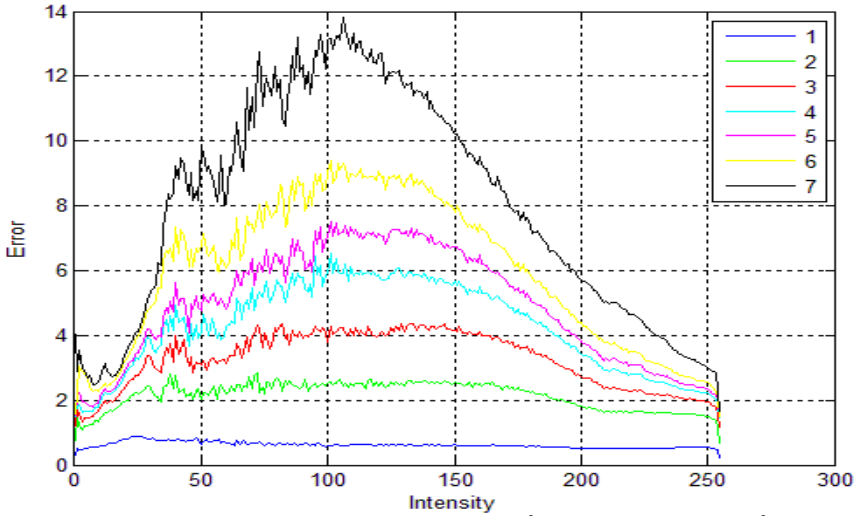


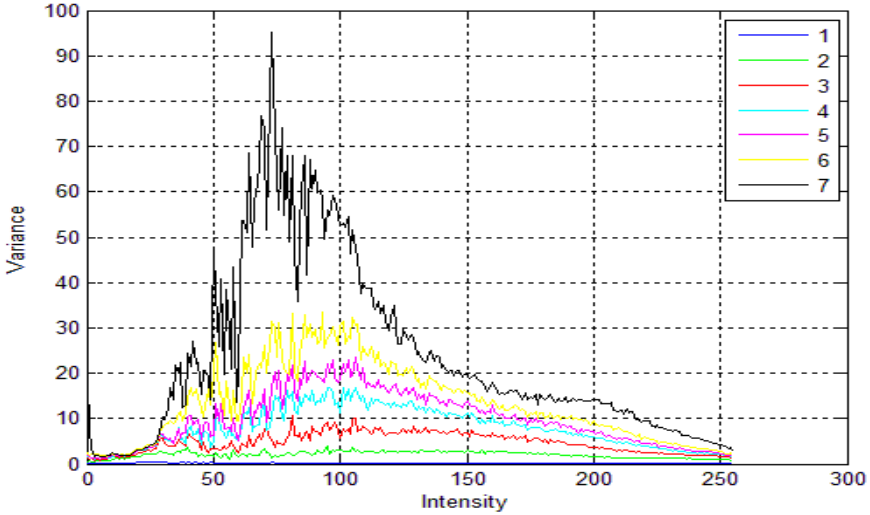


Error mean values and variances w.r.t. intensity gradients

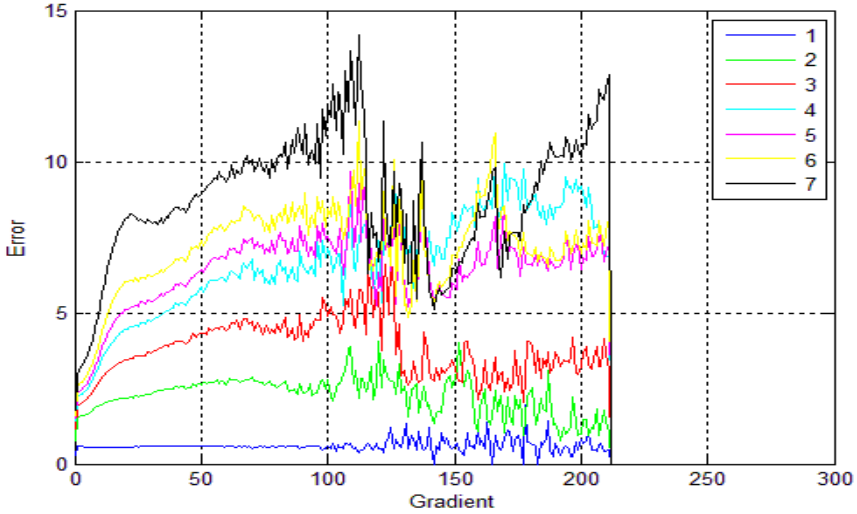
Example 2: Delta E (2000) Analysis

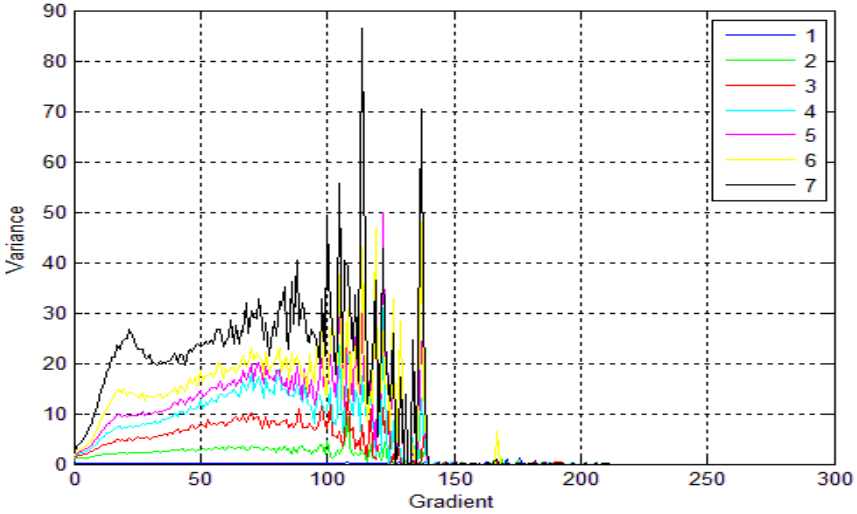
Copyright card sample #1





Delta E mean values and variances w.r.t. intensity levels

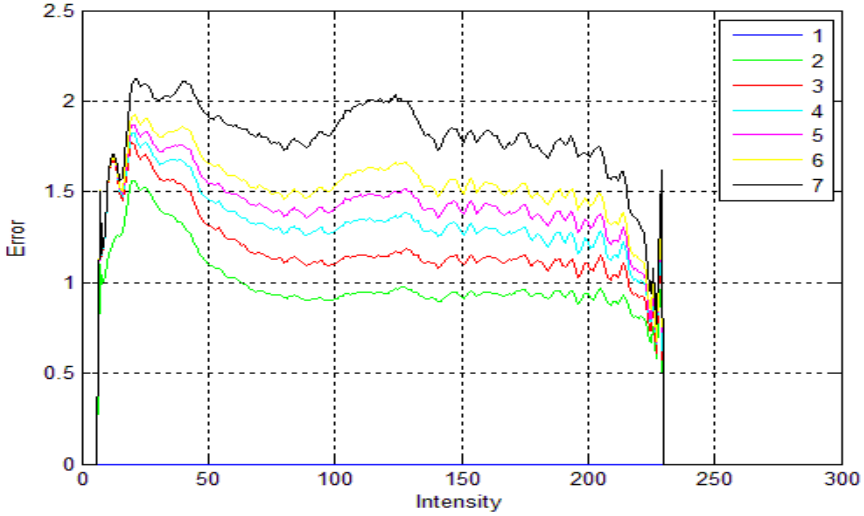


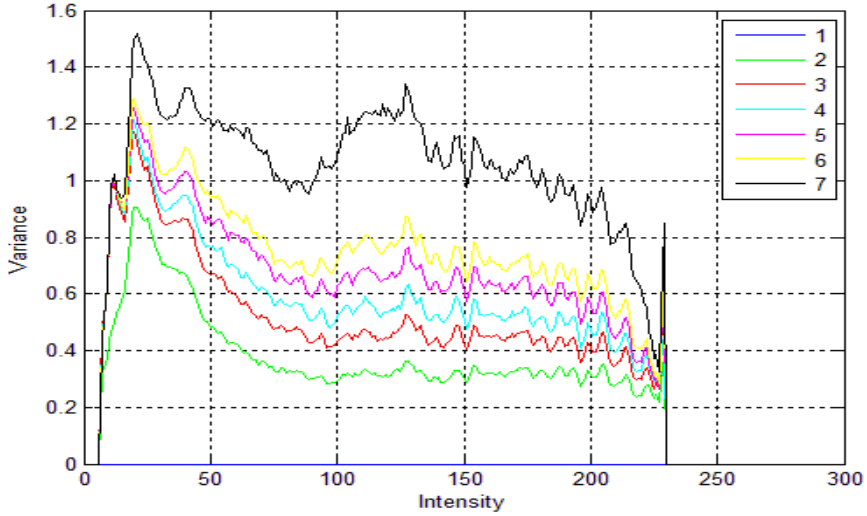


Delta E mean values and variances w.r.t. intensity gradients

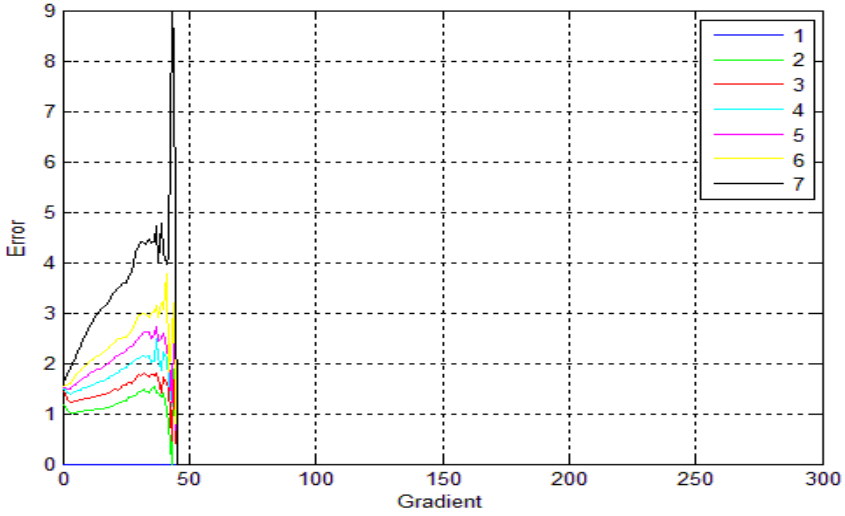
Example 2: Delta E (2000) Analysis

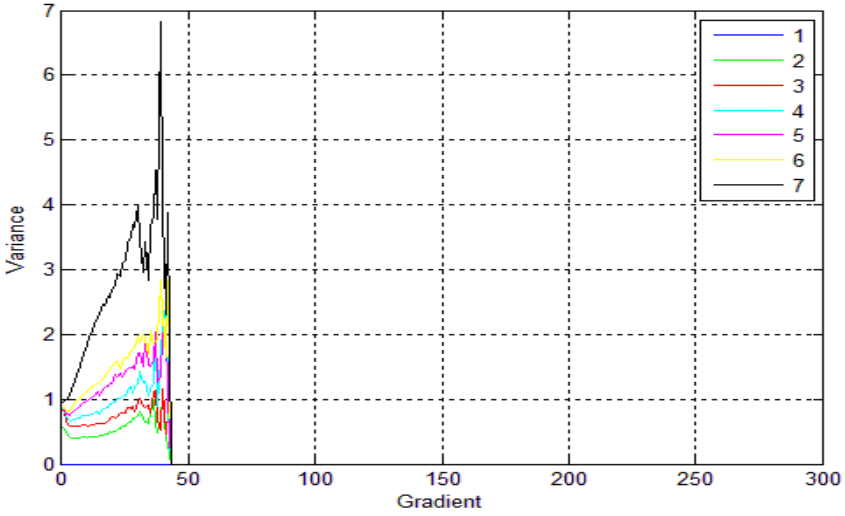
Color Photograph sample





Delta E mean values and variances w.r.t. intensity levels

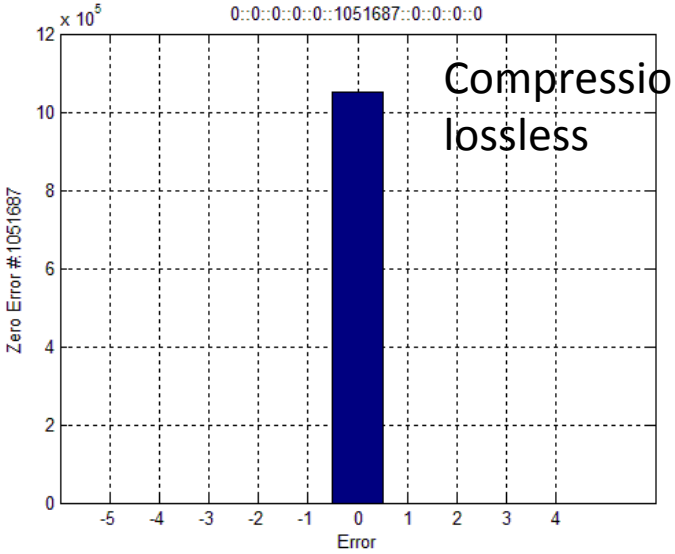




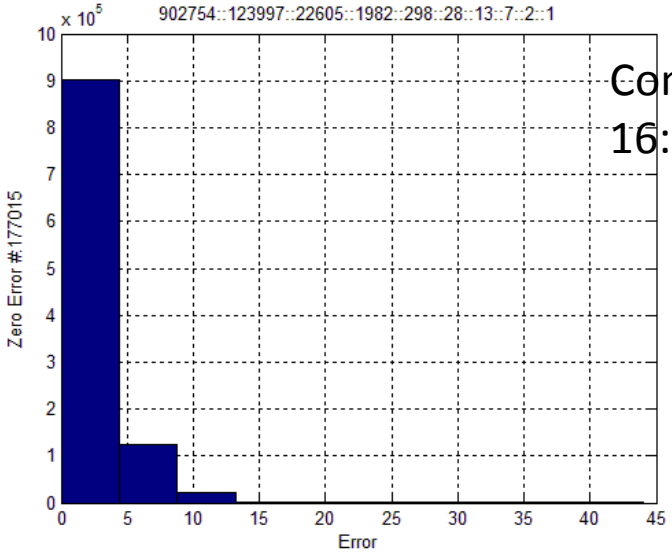
Delta E mean values and variances w.r.t. intensity gradients

Example 3: Error Histogram Analysis

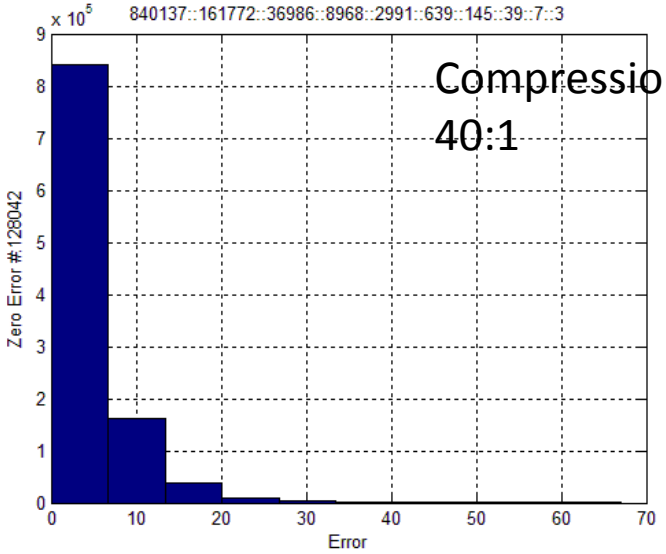
Gray photograph sample



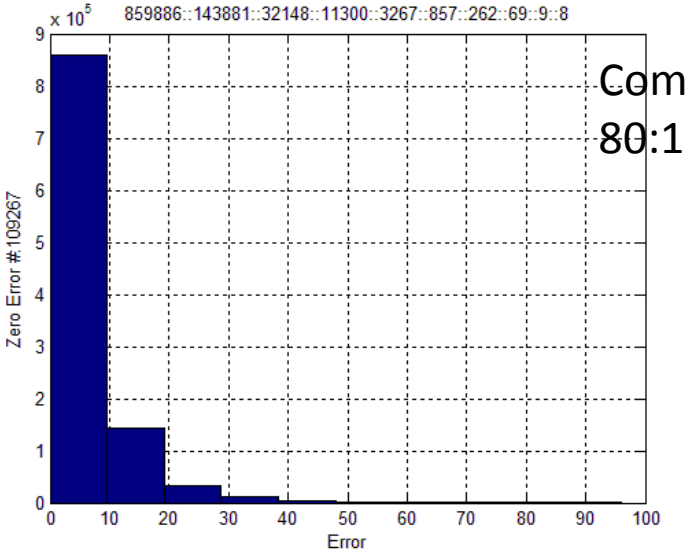
Compression 1
lossless



Compression 2
16:1



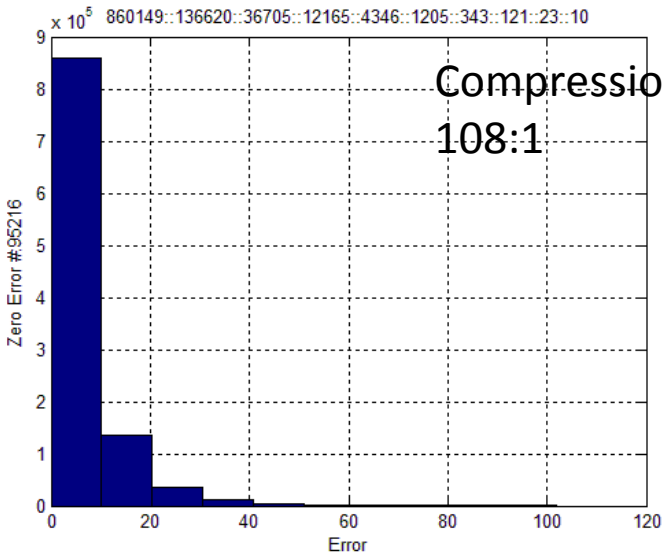
Compression 3
40:1



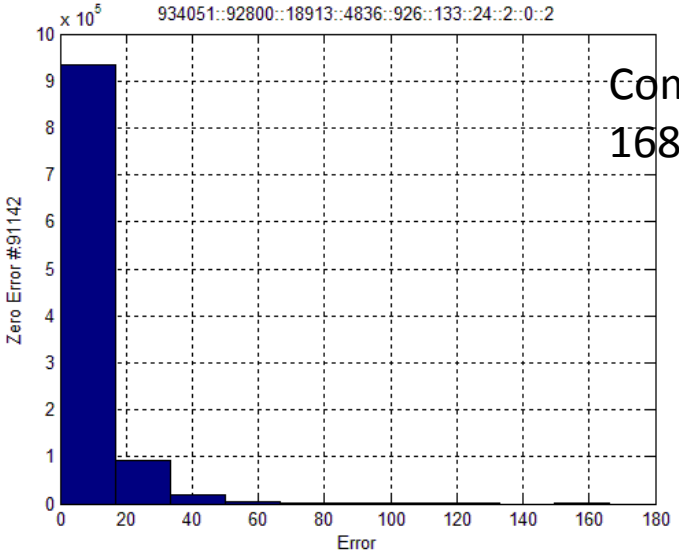
Compression 4
80:1

Example 3: Error Histogram Analysis

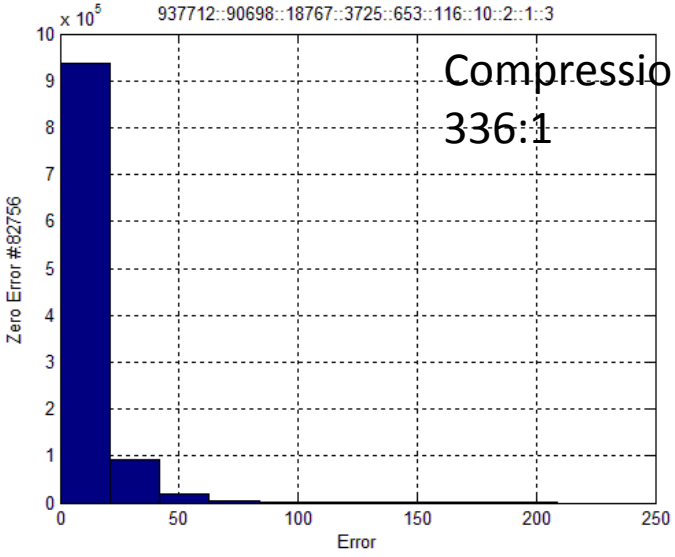
Gray photograph sample



Compression 5
108:1



Compression 6
168:1

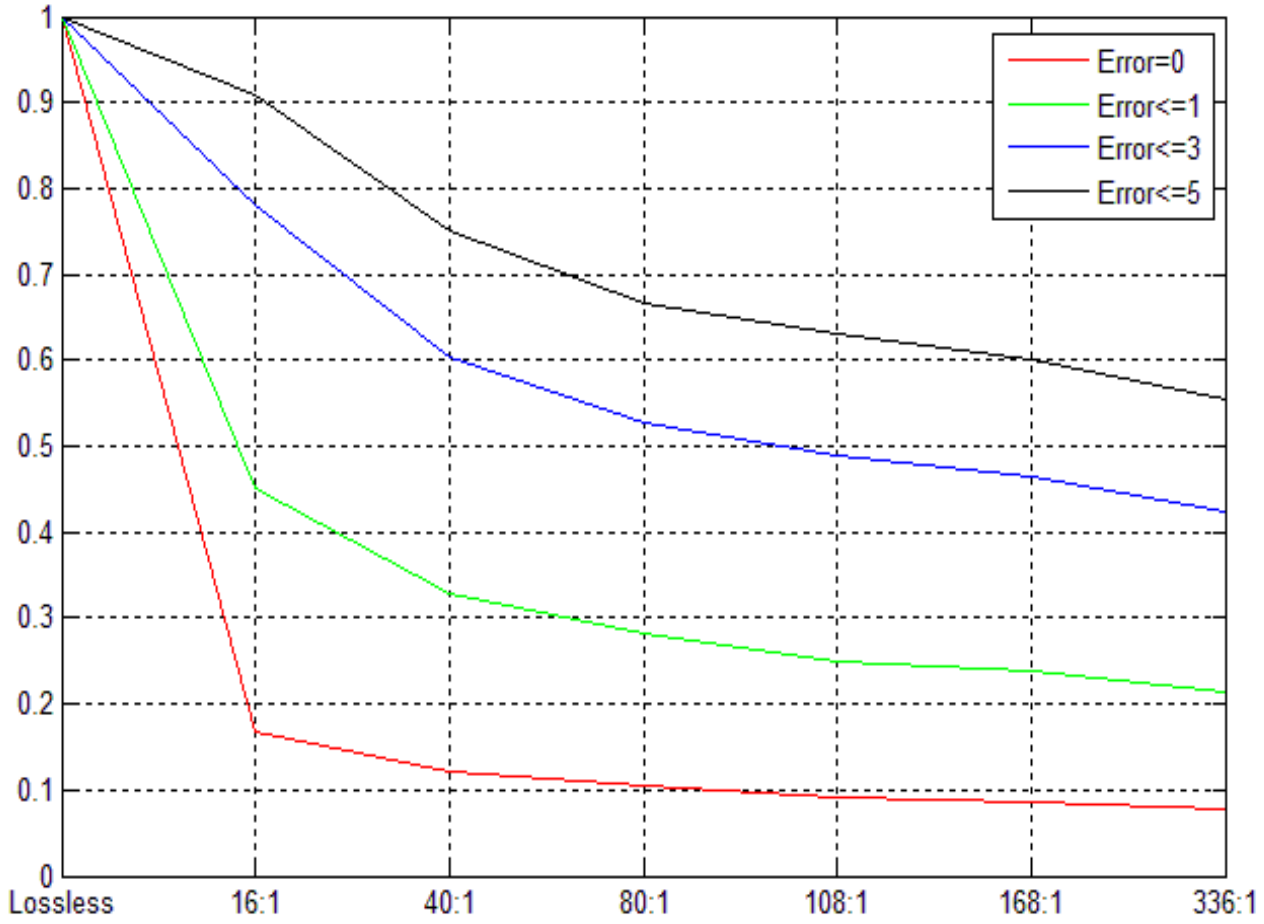


Compression 7
336:1

Example 3: Error Histogram Analysis

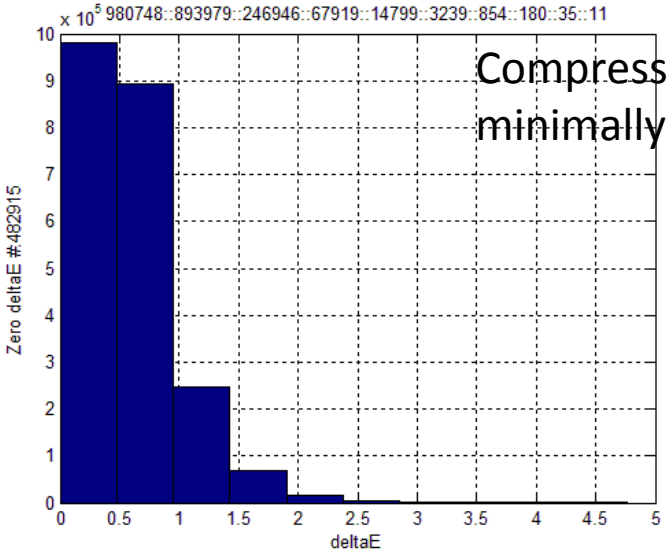
Gray photograph sample

Percentage of points with error = 0, <=1, 3, 5

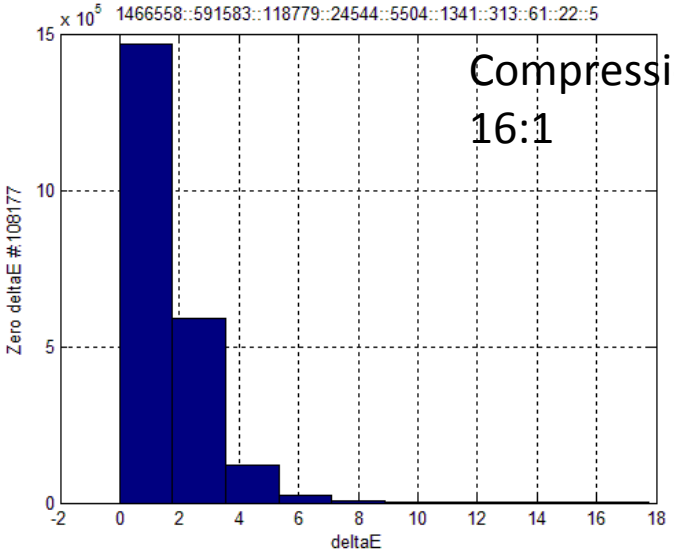


Example 4: Delta E (2000) Histograms

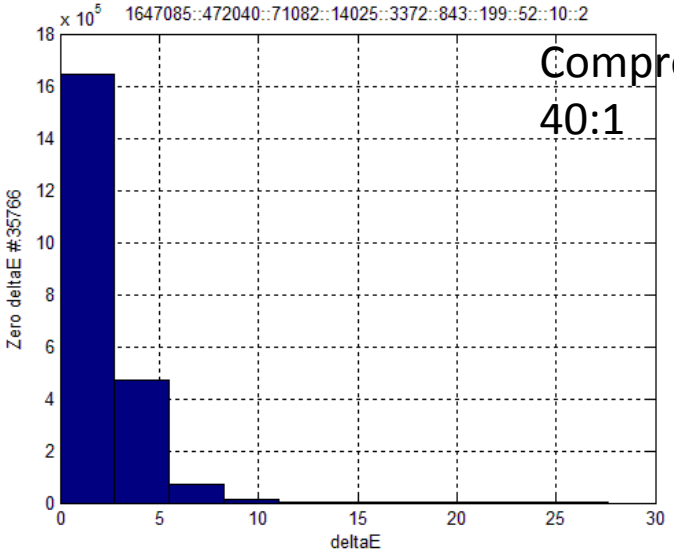
Copyright card sample #1



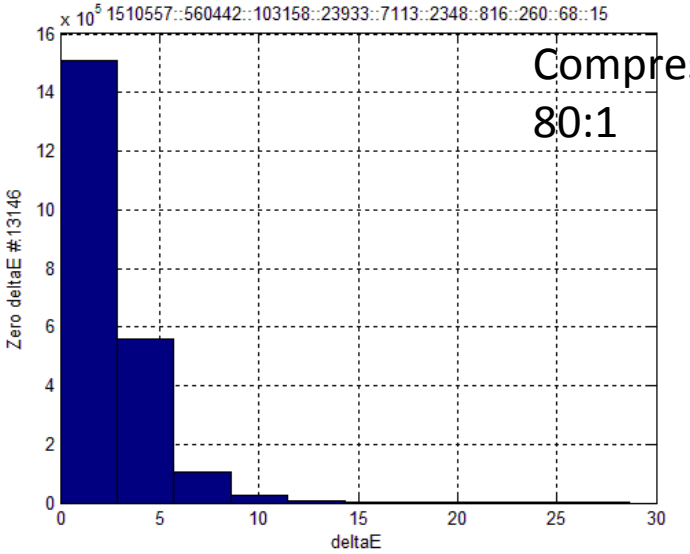
Compression 1
minimally lossy



Compression 2
16:1



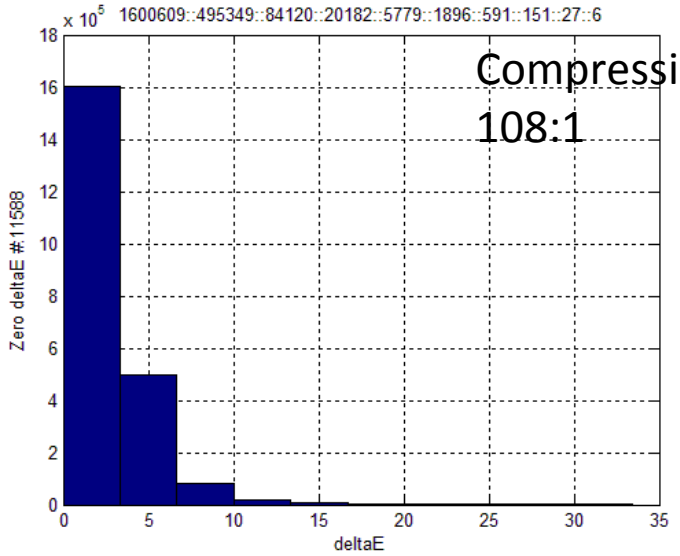
Compression 3
40:1



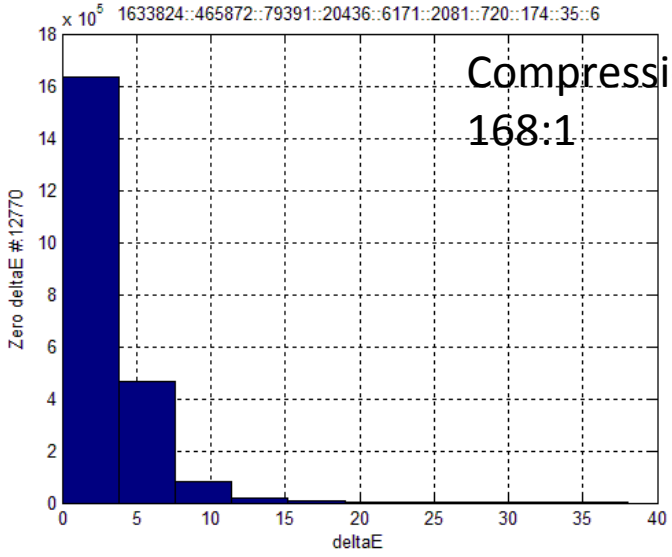
Compression 4
80:1

Example 4: Delta E (2000) Histograms

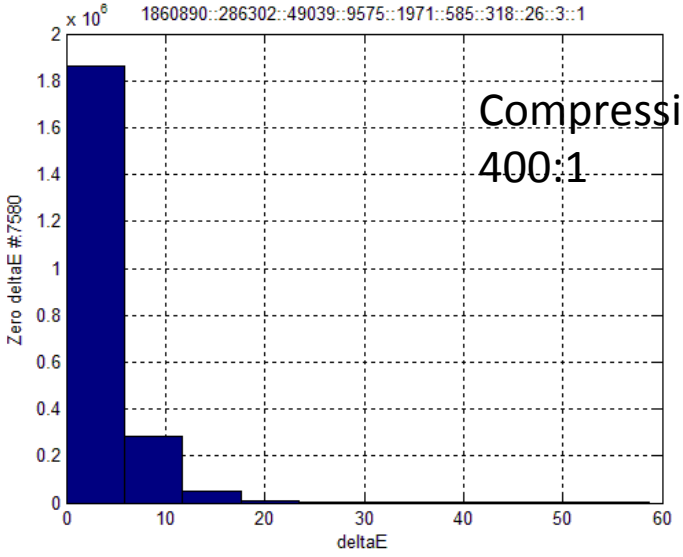
Copyright card sample #1



Compression 5
108:1



Compression 6
168:1

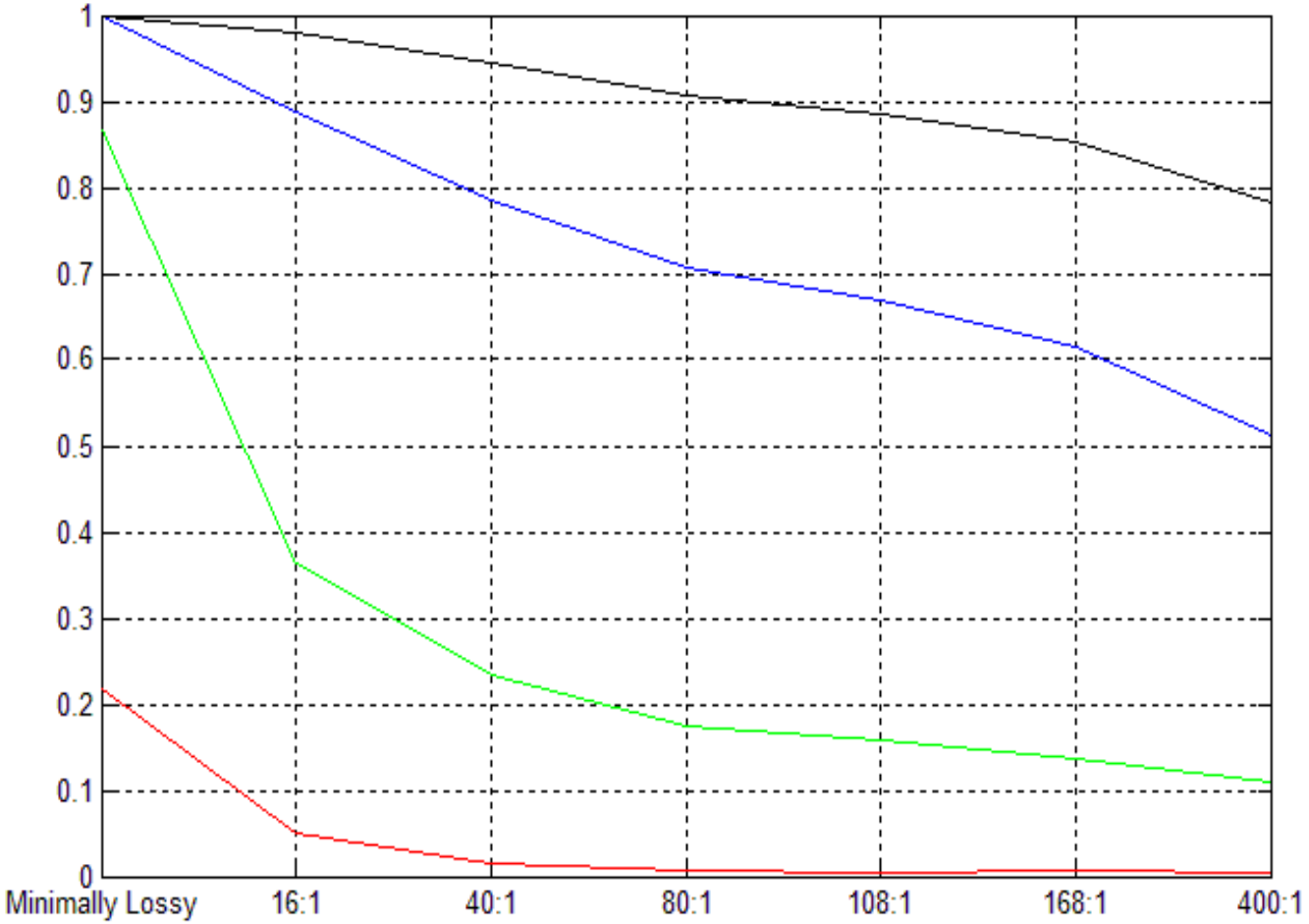


Compression 7
400:1

Example 4: Delta E (2000) Histograms

Copyright card sample #1

Percentage of points with delta E = 0, <=1, 3, 5



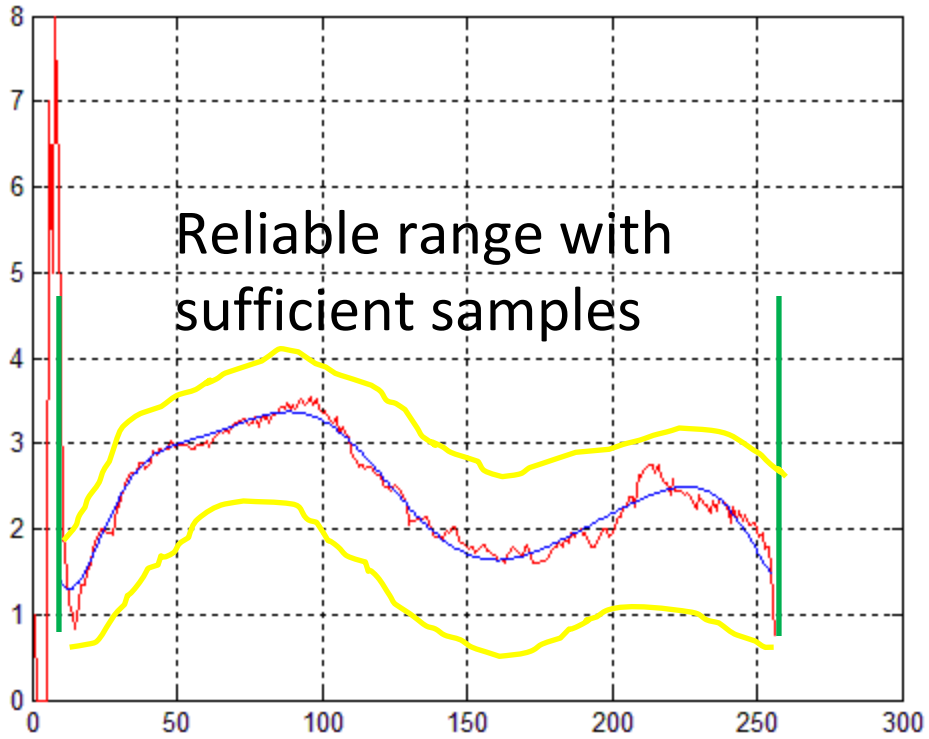
Observations

- Larger compression produces larger error
- Error (delta E) w.r.t. intensity:
  - Low compression levels produce “smoother” (i.e., smaller variance) curves for both error mean values and variances.
- Error (delta E) w.r.t. gradient:
  - Large gradients (i.e., edges) usually introduce large errors.
  - Error increases fast in a small range of gradient (small gradients), then becomes stable (flat curves).
- Small number of sample points (see the histograms) introduce large variance to the curves.

Future Work

- Identify the measurements which match human qualitative perception, this provides a more meaningful way to measure the acceptable error.

Possible solutions:
(1). the curvature mean value and variance of the fitted regression function for the nonlinear Gaussian model;
(2) low and up limits according to the variances (e.g., the curve of 40:1).

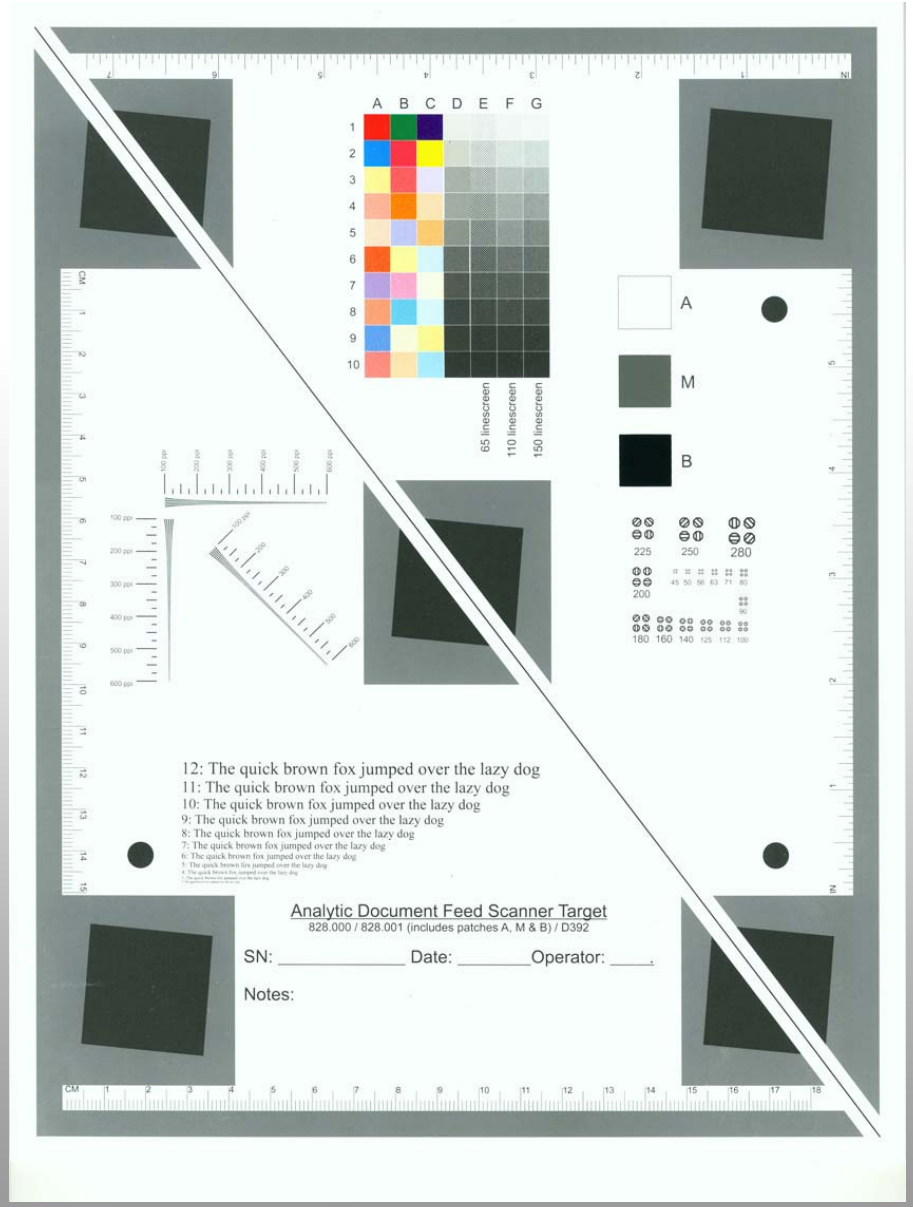


Future Work

- Repeat the experiments and analyses for a larger amount of images from different imaging technologies.
- Repeat the experiments and analyses for other image compression techniques, e.g., jpg, tif.
- 2D distribution modeling of the error, e.g., Markov random field.

DICE

1)Upgrade and New Targets


A B C D E F G
1
2
3
4
5
6
7
8
9
10
65 linescreen
110 linescreen
150 linescreen
A
M
B
225
250
280
200
180
160
140
12: The quick brown fox jumped over the lazy dog
11: The quick brown fox jumped over the lazy dog
10: The quick brown fox jumped over the lazy dog
9: The quick brown fox jumped over the lazy dog
8: The quick brown fox jumped over the lazy dog
7: The quick brown fox jumped over the lazy dog
6: The quick brown fox jumped over the lazy dog
Analytic Document Feed Scanner Target
828.000 / 828.001 (includes patches A, M & B) / D392
SN: _____ Date: _____Operator: ____.
Notes:

Documents and Wiki

a) Discuss Revision Priorities and Timeline

b) Additional Documents

Priorities and New Work

Meeting Notes and Advisory Board

Next Meeting-

Plan to meet once a quarter.